JMIR Public Health and Surveillance

Impact Factor (2022): 8.5
Volume 8 (2022), Issue 7   ISSN 2369-2960   Editor in Chief: Travis Sanchez, PhD, MPH

Contents

Original Papers

Viewpoint

Reviews

Corrigenda and Addenda

Short Paper

Original Paper

Estimating Cardiorespiratory Fitness Without Exercise Testing or Physical Activity Status in Healthy Adults: Regression Model Development and Validation

Robert Sloan[1], PhD; Marco Visentini-Scarzanella[1], PhD; Susumu Sawada[2], PhD; Xuemei Sui[3], MD, PhD; Jonathan Myers[4], PhD

[1]Department of Social and Behavioral Medicine, Kagoshima University Graduate Medical School, Kagoshima, Japan
[2]Faculty of Sport Sciences, Waseda University, Saitama, Japan
[3]Department of Exercise Science, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States
[4]Division of Cardiovascular Medicine, Veterans Affairs Palo Alto Health Care System, Stanford University, Palo Alto, CA, United States

Corresponding Author:
Robert Sloan, PhD
Department of Social and Behavioral Medicine
Kagoshima University Graduate Medical School
8-35-1 Sakuragaoka
Kagoshima
Japan
Phone: 81 99 275 5751
Email: rsloan@m.kufm.kagoshima-u.ac.jp

Abstract

Background: Low cardiorespiratory fitness (CRF) is an independent predictor of morbidity and mortality. Most health care settings use some type of electronic health record (EHR) system. However, many EHRs do not have CRF or physical activity data collected, thereby limiting the types of investigations and analyses that can be done.

Objective: This study aims to develop a nonexercise equation to estimate and classify CRF (in metabolic equivalent tasks) using variables commonly available in EHRs.

Methods: Participants were 42,676 healthy adults (female participants: n=9146, 21.4%) from the Aerobics Center Longitudinal Study examined from 1974 to 2005. The nonexercise estimated CRF was based on sex, age, measured BMI, measured resting heart rate, measured resting blood pressure, and smoking status. A maximal treadmill test measured CRF.

Results: After conducting nonlinear feature augmentation, separate linear regression models were used for male and female participants to calculate correlation and regression coefficients. Cross-classification of actual and estimated CRF was performed using low CRF categories (lowest quintile, lowest quartile, and lowest tertile). The multiple correlation coefficient (R) was 0.70 (mean deviation 1.33) for male participants and 0.65 (mean deviation 1.23) for female participants. The models explained 48.4% (SE estimate 1.70) and 41.9% (SE estimate 1.56) of the variance in CRF for male and female participants, respectively. Correct category classification for low CRF (lowest tertile) was found in 77.2% (n=25,885) of male participants and 74.9% (n=6,850) of female participants.

Conclusions: The regression models developed in this study provided useful estimation and classification of CRF in a large population of male and female participants. The models may provide a practical method for estimating CRF derived from EHRs for population health research.

(JMIR Public Health Surveill 2022;8(7):e34717) doi:10.2196/34717

KEYWORDS

nonexercise estimated cardiorespiratory fitness; public health; surveillance; epidemiology; electronic health record; EHR; fitness; cardiorespiratory; physical activity; regression model; nonexercise equation

Introduction

Background

The use of data from electronic health records (EHRs) beyond day-to-day medical management is rapidly emerging in the fields of digital health, public health, and epidemiology [1,2]. However, access to cardiorespiratory fitness (CRF), a valuable health metric, is limited. This limitation is primarily due to the medical service (cardiopulmonary stress test) being costly, time-consuming, and generally focused on cardiac patients [3-5]. CRF is a comprehensive measure of one’s functional capacity ($\text{mL O}_2 \cdot \text{kg}^{1} \cdot \text{min}^{1}$) driven by the combination of heart, lung, and muscle function [6]. It is an important marker of health status in the general adult population [3,6]. Further demonstrating the importance of CRF, the American Heart Association released a scientific statement proposing that CRF be considered a clinical vital sign. The scientific rationale behind the statement is driven by the voluminous evidence that demonstrates that low CRF is a strong independent predictor of adverse health outcomes (ie, all-cause mortality, cancer, stroke, heart disease, and diabetes incidence) [3].

Prior Work

To help increase accessibility to CRF data, researchers have developed an array of nonexercise estimated CRF (NEECRF) equations to estimate CRF [7-9]. NEECRF equations commonly include age, gender, resting heart rate, smoking status, BMI, and self-reported physical activity status (PAS) [8,10-12]. Studies have shown NEECRF to predict all-cause and cardiovascular disease mortality on par with measured CRF [13,14]. However, the assessment of PAS required to calculate NEECRF in patients is not typically conducted or documented in health care settings [15]. Therefore, using an NEECRF model without PAS (non-PAS) may be more feasible.

In a 2019 comprehensive NEECRF review, a few peer-reviewed non-PAS NEECRF equations were identified in adult populations reporting correlations and SE estimates [8]. Most of the equations were developed using only age, height, and weight combinations, and some used variables not commonly found in EHRs (eg, waist girth, predicted/ideal weight, or exercise mode).

Though correlations were moderate to high, the sample populations were too small to determine the classification accuracy of low CRF. Accurate classification of low CRF is essential for large-scale investigations [3,16]. Low CRF (lowest tertile) classification was recently investigated by Peterman et al [9], who found Baynard’s [7] simplified non-PAS NEECRF equation to have poor classification ability in a sizable (n=4871) adult population. Other researchers have also suggested that simple non-PAS NEECRF formulas have low validity, and nuanced approaches are warranted [16,17]. Investigators have also noted that valid non-PAS NEECRF models may have broad applications for public health, epidemiology, surveillance, practice, and research [3,8,9,16,18].

Goal of This Study

Because it is standard practice to assess and document resting heart rate, blood pressure, BMI, and smoking status during a typical clinic visit, the primary aim of this study was to develop new models for NEECRF that could potentially be used in large-scale population health investigations using variables commonly found in EHRs. To accomplish this, we compared a non-PAS NEECRF equation to clinically measured CRF and evaluated its ability to estimate and classify CRF.

Methods

Study Sample

The Aerobics Center Longitudinal Study (ACLS) is a prospective epidemiological investigation of participants that began in 1970 [19]. The original data set for this investigation included 43,257 healthy adults who voluntarily participated in a comprehensive preventive medical examination at the Cooper Clinic in Dallas, Texas between 1974 and 2005. At baseline, all participants were free of diabetes, heart disease, stroke, cancer, positive electrocardiograms, and completed a maximal graded exercise test. Each participant gave informed consent to join the longitudinal study. The research population demographic primarily consists of Caucasian college-educated adults of middle to high socioeconomic status with an average age at baseline of 43.5 (range 20-79) years.

Ethical Approval

The study was reviewed and approved annually by the Cooper Institute Institutional Review Board, and all participants provided written informed consent.

Measurements

Predictor variables were assessed during a preventive health examination that included objective measurements of age, BMI, resting heart rate, systolic blood pressure, diastolic blood pressure, and self-reported smoking status between 1974 and 2005 at the Cooper Clinic. Age was verified at the time of the examination. Height and weight were measured on a calibrated scale using US customary units and converted to metric scales for this investigation. BMI was calculated from measured height and weight as kg/m^2. Manual auscultation was used to measure resting blood pressure while seated. Resting heart rate was calculated using the R-R interval on an electrocardiogram while seated. CRF was expressed as absolute metabolic equivalent tasks (METs; $1 \text{ MET} = 3.5 \text{ mL O}_2 \cdot \text{kg}^{1} \cdot \text{min}^{1}$) based on the total duration of a symptom-limited maximal Balke graded exercise test [6]. Following American College for Sports Medicine Guidelines, patients were encouraged to give maximal effort, and the test end point was volitional exhaustion or termination by the physician for medical reasons [6]. METs were calculated based on the final treadmill speed and grade [6]. The Balke graded exercise test is highly correlated (r=0.94) with maximal graded cardiopulmonary exercise testing [20,21]. A standardized medical questionnaire was used to ascertain demographic information, lifestyle habits, and chronic disease status. More detailed information on the preventive health examination is available in prior ACLS publications [10,13].

Statistical Analysis

Data Exclusion

Data were first examined for outliers and skewness. We removed the outliers from the data set by removing participants with predictor variables (continuous) beyond the ±3 σ interval. To do this, we calculated the mean and SD of each predictor variable, excluding smoking (categorical). Any participant with at least one predictor variable above 3 SDs or below 3 SDs was flagged as an outlier. After removing outliers and incomplete entries (n=581), the data set comprised 33,530 male participants and 9146 female participants, 98.7% of the original 43,257 participants.

Regression

The main analysis was based on apparently healthy adults at baseline. Using a supervised machine learning technique, we conducted separate linear regression analyses for men and women to predict non-PAS NEECRF based on nonlinear augmentation of the predictor variables [1]. We also considered advanced machine learning models but did not find them advantageous. The male and female non-PAS NEECRF prediction equations were formulated to minimize the average mean squared error, where *N* is the number of samples in our data set:

The prediction equation used age, height (Ht), weight (Wt), BMI, resting heart rate (rHR), systolic blood pressure (SBP), diastolic blood pressure (DBP), and smoking. All variables were continuous except for smoking status (nonsmoker=0, current smoker=1). Data were standardized by subtracting the mean and dividing by the SD for each variable. Next, separate models were then trained for male and female participants. We augmented the original 8 variables with second order and interaction terms, and regressed them linearly to the dependent variable for training. In this way, the nonlinearity was transferred from the regressor to the independent variables, while the model's overall interpretability was maintained. The augmentation procedure added the following 28 second order and interaction terms: Wt^2, Wt × Ht, Wt × Age, Wt × rHR, Wt × SBP, Wt × DBP, Wt × BMI, Ht^2, Ht × Age, Ht × rHR, Ht × SBP, Ht × DBP, Ht × BMI, Age^2, Age × rHR, Age × SBP, Age × DBP, Age × BMI, rHR^2, Ht × SBP, Ht × DBP, rHR × BMI, SBP^2, SBP × DBP, SBP × BMI, DBP^2, and DBP × BMI (for a total of 36 variables). Because smoking status was a categorical variable, it was not used to create the additional variables. The augmented data set was input into an elastic net linear regressor and trained and evaluated via 10-fold cross-validation [22]. Optimal model hyperparameters were calculated for the male (α=.001, λ=1.0) and female (α=.004, λ=1.0) data sets through the cross-validation procedure. Pearson correlation coefficients were then calculated using the non-PAS NEECRF equations for the male and female data sets (shown in Multimedia Appendix 1). Lastly, for comparison, we cross-validated Baynard's [7] simplified non-PAS NEECRF equation (77.96 − 10.35 (sex; M=0, F=1) − 0.92 (BMI) − 0.32 (age)) [9].

Classification Accuracy

We cross-classified non-PAS NEECRF and CRF for three specified cut points (lowest quintile, lowest quartile, and lowest tertile) commonly used in epidemiological investigations to define low CRF [3]. Next, CRF distributions were then grouped by males and females where α served as the value of the α-th percentile using the calculation non-PAS NEECRF>α. After classification, we determined the non-PAS NEECRF accuracy, sensitivity, positive predictive value, and F_1 score. The reference standard was the measured CRF, and low CRF was defined as a positive test. All analyses were performed in scikit-learn version 0.22.2 (NumFOCUS).

Results

Descriptive statistics are provided in Table 1. Correlation coefficients between each independent variable and CRF are presented in Multimedia Appendix 1. The multiple Rs and mean deviations for non-PAS NEECRF (in METs) were high at 0.70 (mean deviation 1.33) for male participants and moderate at 0.65 (mean deviation 1.23) for female participants. The models explained 48.4% (SE estimate 1.70, 95% CI 0.05-3.97) of the variance in CRF for male participants and 41.9% (SE estimate 1.56, 95% CI 0.05-3.48) for female participants. Multimedia Appendix 1 provides a simple independent variable input Google Sheet for researchers and data scientists to easily calculate NEECRF. Table 2 provides the findings regarding the accuracy, positive predictive, and sensitivity values using the lowest quintile, quartile, and tertile to classify low CRF for male and female participants. While overall classification accuracy was meaningful for a nondiagnostic test across the three models, the optimal model was the lowest tertile. Combined male and female positive predictive value were 0.60, sensitivity 0.67, and F_1 score 0.63 [23,24]. The F_1 score is the best practice summary metric consisting of the harmonic mean of positive predictive value and sensitivity for classification (0=low, 1=high) [23]. Assuming a balanced data set (n=2529) by Peterman et al [16], we calculated the F_1 scores from their reported findings for equations applicable to EHRs [16]. We found F_1 scores ranging from 0.04 to 0.56. Based on a residual plot (Multimedia Appendix 1), we found the model was most accurate for CRF MET values in the 7.5 to 12.5 METs range but tended to underestimate MET values >12.5 METs and overestimate MET values <7.5 METs. Notably, this is a common finding in non-PAS NEECRF studies [3]. Similar to Peterman et al [9], we cross-validated Baynard's [7] non-PAS NEECRF equation with our reference CRF data set and found low positive correlations for male (*r*=0.49, mean deviation 1.60) and female (*r*=0.43, mean deviation 1.46) participants.

Table 1. Baseline characteristics of participants.

	All (N=42,676)	Male participants (n=33,530)	Female participants (n=9146)
Age (years), mean (SD)	44.1 (9.6)	44.1 (9.5)	44.1 (10.2)
BMI (kg/m^2), mean (SD)	25.8 (3.8)	26.5 (3.5)	23.3 (3.7)
Resting heart rate (bpm), mean (SD)	60.9 (10.5)	60.1 (10.4)	64.0 (10.1)
Systolic blood pressure (mmHg), mean (SD)	118.9 (13.4)	120.7 (12.7)	112.4 (13.9)
Diastolic blood pressure (mmHg), mean (SD)	80.0 (9.6)	81.1 (9.3)	76.0 (9.3)
Smoker, n (%)	6361 (14.9)	5569 (16.7)	792 (9.7)
Measured CRF[a] (mL/kg/min), mean (SD)	11.5 (2.5)	12.0 (2.4)	9.8 (2.0)

[a] CRF: cardiorespiratory fitness (maximal oxygen consumption).

Table 2. Predictive accuracy of nonexercise estimated CRF classification of lowest cardiovascular fitness against the reference CRF.

Group	Low CRF[a] (lowest quintile)			Low CRF (lowest quartile)			Low CRF (lowest tertile)		
	ACC[b] (%)	PPV[c] (%)	SEN[d] (%)	ACC (%)	PPV (%)	SEN (%)	ACC (%)	PPV (%)	SEN (%)
Male participants	81.5	44.7	69.2	79.5	50.9	67.4	77.2	59.9	66.7
Female participants	80.6	43.9	57.8	77.9	55.4	55.6	74.9	60.4	66.9

[a] CRF: cardiorespiratory fitness.
[b] ACC: accuracy.
[c] PPV: positive predictive value.
[d] SEN: sensitivity.

Discussion

Principal Findings

Using a greater combination of clinical measures commonly found in EHRs, this study compared non-PAS NEECRF with objectively measured CRF in the largest population to date [10,11,25-27]. Overall, our model may provide a more applicable method for estimating and classifying CRF than previous methods [8,9,16]. Moreover, because the vital signs and medical information used to calculate non-PAS NEECRF are routinely captured during health care visits, our approach places nominal demand on health care staff and patients for collecting data. From a public health perspective, a moderate positive predictive value is practical, given that non-PAS NEECRF is a nondiagnostic test that is no cost and easily accessible [24]. Likely, some individuals classified with low fitness may be at the lower end of the fit spectrum and benefit from health promotion [24]. From a clinical perspective and considering moderate sensitivity, we concur with previous investigators that, while estimation equations are applicable for epidemiological investigations, they should not replace clinical exercise testing for patient diagnosis and management [3,16]. Our findings show that the ACLS non-PAS NEECRF may provide a useful assessment of CRF to conduct population health research.

Comparison to Prior Work

Comparatively, PAS-based NEECRF models have demonstrated higher positive correlation values (0.71-0.93) along with a higher degree (~90%) of correct classification accuracy for low CRF than non-PAS NEECRF models [3]. Because physical activity is a key contributor to CRF, including a PAS variable in an NEECRF model improves accuracy [3,8,16]. However, a recent review of distinct EHRs across 20 countries found that only 18.8% of family practice clinics had structured PAS questionaries embedded within the EHR, with documented PAS in the EHR ranging from 10% to 86% [15]. Notably, no validated questionaries designed for PAS-based NEECRF calculations were used. Therefore, the ability to conduct large-scale studies aggregating existing EHR data across local, domestic, or international systems to predict CRF is unlikely. Conversely, our model may provide a global approach to aggregating EHR data across systems to predict CRF and conduct analyses.

In 2019, Wang et al [8] provided a comprehensive list of peer-reviewed non-PAS NEECRF models that used some combination of age, BMI, or gender to predict CRF [8]. Samples were from small populations, and moderate to high correlations were reported generally without SE estimates [8]. Notably, the validity and usefulness of these simplified types of equations have been called into question [8,9,16,17]. We found that the findings from such investigations may be limited because the studies only reported correlations and lacked sufficient sample sizes to calculate prediction values for low CRF. It is also important to note that high correlation values do not necessarily result in a more accurate classification of low CRF [9,16].

Recently, Peterman et al [16] determined the ability of 7 non-PAS NEECRF equations to accurately classify low NEECRF (tertile) compared to measured CRF in a demographically comparable cohort (n=2529) to ACLS. Only 3 of the 7 equations apply because of variables not commonly found in EHRs. On balance, the classification accuracy of low

CRF (tertile) appeared to be better for ACLS. In a separate investigation, Peterman et al [9] also tested Baynard's [7] simplified non-PAS NEECRF using the Ball State cohort (n=4871) to assess classification accuracy. The equation had an *r* of 0.76; however, there was poor accuracy for detecting individuals positive for low CRF (37%). We also cross-validated Baynard's [7] equation with our data set and found low correlations, thus did not attempt to determine accuracy [9]. Although the ACLS and Ball State cohorts are demographically similar, our finding is expected to some degree because the ACLS equation is specifically trained from the ACLS data set [9,16]. Moreover, these equations need to be tested in epidemiological investigations with EHR data to see how well they predict health outcomes.

Limitations

This study is not without limitations. Our study's primary limitation regarding correlation was that the measured reference CRF was conducted using a Balke graded maximal exercise test that estimates absolute METs. This testing strongly correlates with adults' maximal graded cardiopulmonary exercise testing and is routinely used for clinical and epidemiological purposes [3,20,21]. Though the ACLS data set provides the largest known healthy female data set for clinically measured CRF, a larger female sample size may have provided a slightly better predictive model [3,27]. Nonetheless, our analyses demonstrated reasonable correlation and classification. Our analyses are based on a large predominantly Caucasian cohort; it is unknown if the results generalize to other ethnic groups. Notably, the homogeneity of the ACLS cohort may have strengthened the internal validity of our results by limiting possible confounders. The main strength of this investigation is that it was conducted on the largest cohort to date with a larger number of objectively measured predictive variables to estimate non-PAS NEECRF.

Conclusions

The ACLS non-PAS NEECRF equation may provide a useful population health metric for CRF. More work should be conducted regarding diverse populations, the incidence of chronic conditions, and longitudinal repeated measures analyses toward improving public health and surveillance capability.

Acknowledgments

The authors are grateful to the Cooper Clinic physicians and technicians for collecting the baseline data and the Cooper Institute staff for data entry and management. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Japan Society for the Promotion of Science.

This research was supported by the Japan Society for the Promotion of Science KAKENHI Grant 19K19437.

Authors' Contributions

RS conceptualized the study. RS and MVS designed the methodology. MVS used the software. RS and MVS validated the study. MVS conducted the formal analysis. RS conducted the investigation and obtained the resources. XS curated the data. RS and MVS wrote and prepared the original draft. RS, MVS, JM, XS, and SS reviewed and edited the manuscript. RS supervised the study and acquired the funding. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[PDF File (Adobe PDF File), 123 KB - publichealth_v8i7e34717_app1.pdf]

References

1. Wiemken TL, Kelley RR. Machine learning in epidemiology and health outcomes research. Annu Rev Public Health 2020 Apr 02;41:21-36. [doi: 10.1146/annurev-publhealth-040119-094437] [Medline: 31577910]
2. Kim J. Big data, health informatics, and the future of cardiovascular medicine. J Am Coll Cardiol 2017 Feb 21;69(7):899-902 [FREE Full text] [doi: 10.1016/j.jacc.2017.01.006] [Medline: 28209228]
3. Ross R, Blair SN, Arena R, Church TS, Després JP, Franklin BA, American Heart Association Physical Activity Committee of the Council on Lifestyle and Cardiometabolic Health, Council on Clinical Cardiology, Council on Epidemiology and Prevention, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Stroke Council. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. Circulation 2016 Dec 13;134(24):e653-e699. [doi: 10.1161/CIR.0000000000000461] [Medline: 27881567]
4. Sakr S, Elshawi R, Ahmed A, Qureshi WT, Brawner C, Keteyian S, et al. Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford ExercIse Testing (FIT) Project. PLoS One 2018;13(4):e0195344 [FREE Full text] [doi: 10.1371/journal.pone.0195344] [Medline: 29668729]

5. Al-Mallah MH, Keteyian SJ, Brawner CA, Whelton S, Blaha MJ. Rationale and design of the Henry Ford Exercise Testing Project (the FIT project). Clin Cardiol 2014 Aug;37(8):456-461. [doi: 10.1002/clc.22302] [Medline: 25138770]
6. American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. 10th ed. Philadelphia, PA: Wolters Kluwer Health; 2017.
7. Baynard T, Arena RA, Myers J, Kaminsky LA. The role of body habitus in predicting cardiorespiratory fitness: The FRIEND Registry. Int J Sports Med 2016 Oct;37(11):863-869. [doi: 10.1055/s-0042-110572] [Medline: 27490112]
8. Wang Y, Chen S, Lavie CJ, Zhang J, Sui X. An overview of non-exercise estimated cardiorespiratory fitness: estimation equations, cross-validation and application. J Sci Sport Exerc 2019 Jun 10;1(1):38-53. [doi: 10.1007/s42978-019-0003-x]
9. Peterman JE, Harber MP, Imboden MT, Whaley MH, Fleenor BS, Myers J, et al. Accuracy of exercise-based equations for estimating cardiorespiratory fitness. Med Sci Sports Exerc 2021 Jan;53(1):74-82. [doi: 10.1249/MSS.0000000000002435] [Medline: 32694370]
10. Jackson AS, Sui X, O'Connor DP, Church TS, Lee D, Artero EG, et al. Longitudinal cardiorespiratory fitness algorithms for clinical settings. Am J Prev Med 2012 Nov;43(5):512-519. [doi: 10.1016/j.amepre.2012.06.032] [Medline: 23079174]
11. Jackson AS, Wier LT, Ayers GW, Beard EF, Stuteville JE, Blair SN. Changes in aerobic power of women, ages 20-64 yr. Med Sci Sports Exerc 1996 Jul;28(7):884-891. [doi: 10.1097/00005768-199607000-00016] [Medline: 8832543]
12. Jurca R, Jackson AS, LaMonte MJ, Morrow JR, Blair SN, Wareham NJ, et al. Assessing cardiorespiratory fitness without performing exercise testing. Am J Prev Med 2005 Oct;29(3):185-193. [doi: 10.1016/j.amepre.2005.06.004] [Medline: 16168867]
13. Artero EG, Jackson AS, Sui X, Lee D, O'Connor DP, Lavie CJ, et al. Longitudinal algorithms to estimate cardiorespiratory fitness: associations with nonfatal cardiovascular disease and disease-specific mortality. J Am Coll Cardiol 2014 Jun 03;63(21):2289-2296 [FREE Full text] [doi: 10.1016/j.jacc.2014.03.008] [Medline: 24703924]
14. Stamatakis E, Hamer M, O'Donovan G, Batty GD, Kivimaki M. A non-exercise testing method for estimating cardiorespiratory fitness: associations with all-cause and cardiovascular mortality in a pooled analysis of eight population-based cohorts. Eur Heart J 2013 Mar;34(10):750-758. [doi: 10.1093/eurheartj/ehs097] [Medline: 22555215]
15. Lindeman C, McCurdy A, Lamboglia CG, Wohlers B, Pham ANQ, Sivak A, et al. The extent to which family physicians record their patients' exercise in medical records: a scoping review. BMJ Open 2020 Feb 12;10(2):e034542 [FREE Full text] [doi: 10.1136/bmjopen-2019-034542] [Medline: 32054628]
16. Peterman JE, Whaley MH, Harber MP, Fleenor BS, Imboden MT, Myers J, et al. Comparison of non-exercise cardiorespiratory fitness prediction equations in apparently healthy adults. Eur J Prev Cardiol 2021 Apr 10;28(2):142–148. [doi: 10.1177/2047487319881242] [Medline: 33838037]
17. Malek MH, Berger DE, Housh TJ, Coburn JW, Beck TW. Validity of VO2max equations for aerobically trained males and females. Med Sci Sports Exerc 2004 Aug;36(8):1427-1432. [doi: 10.1249/01.mss.0000135795.60449.ce] [Medline: 15292753]
18. Radford NB, DeFina LF, Leonard D, Barlow CE, Willis BL, Gibbons LW, et al. Cardiorespiratory fitness, coronary artery calcium, and cardiovascular disease events in a cohort of generally healthy middle-age men: results from the cooper center longitudinal study. Circulation 2018 May 01;137(18):1888-1895. [doi: 10.1161/CIRCULATIONAHA.117.032708] [Medline: 29343464]
19. Blair SN, Kohl HW, Paffenbarger RS, Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality. A prospective study of healthy men and women. JAMA 1989 Nov 03;262(17):2395-2401. [doi: 10.1001/jama.262.17.2395] [Medline: 2795824]
20. Balke B, Ware RE. An experimental study of physical fitness of Air Force personnel. U S Armed Forces Med J 1959 Jun;10(6):675-688. [Medline: 13659732]
21. Pollock ML, Foster C, Schmidt D, Hellman C, Linnerud AC, Ward A. Comparative analysis of physiologic responses to three different maximal graded exercise test protocols in healthy women. Am Heart J 1982 Mar;103(3):363-373. [doi: 10.1016/0002-8703(82)90275-7] [Medline: 7064770]
22. Zou H, Hastie T. Regularization and variable selection via the elastic net. J Royal Stat Soc Ser B Stat Methodology 2005 Apr;67(2):301-320. [doi: 10.1111/j.1467-9868.2005.00503.x]
23. Xiang L, Deng K, Mei Q, Gao Z, Yang T, Wang A, et al. Population and age-based cardiorespiratory fitness level investigation and automatic prediction. Front Cardiovasc Med 2021;8:758589. [doi: 10.3389/fcvm.2021.758589] [Medline: 35071342]
24. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017;5:307. [doi: 10.3389/fpubh.2017.00307] [Medline: 29209603]
25. Zeiher J, Ombrellaro KJ, Perumal N, Keil T, Mensink GBM, Finger JD. Correlates and determinants of cardiorespiratory fitness in adults: a systematic review. Sports Med Open 2019 Sep 03;5(1):39. [doi: 10.1186/s40798-019-0211-2] [Medline: 31482208]
26. Ahmadian HR, Sclafani JJ, Emmons EE, Morris MJ, Leclerc KM, Slim AM. Comparison of predicted exercise capacity equations and the effect of actual versus ideal body weight among subjects undergoing cardiopulmonary exercise testing. Cardiol Res Pract 2013;2013:940170. [doi: 10.1155/2013/940170] [Medline: 23653881]
27. Rodgers JL. Degrees of freedom at the start of the second 100 years: a pedagogical treatise. Adv Methods Practices Psychological Sci 2019 Dec 05;2(4):396-405. [doi: 10.1177/2515245919882050]

Abbreviations

ACLS: Aerobics Center Longitudinal Study
CRF: cardiorespiratory fitness
DBP: diastolic blood pressure
EHR: electronic health record
Ht: height
MET: metabolic equivalent task
NEECRF: nonexercise estimated cardiorespiratory fitness
PAS: physical activity status
rHR: resting heart rate
SBP: systolic blood pressure
Wt: weight

Edited by H Bradley; submitted 04.11.21; peer-reviewed by L Larsen, C Lavie, J Peterman; comments to author 24.02.22; revised version received 14.03.22; accepted 27.05.22; published 06.07.22.

Please cite as:
Sloan R, Visentini-Scarzanella M, Sawada S, Sui X, Myers J
Estimating Cardiorespiratory Fitness Without Exercise Testing or Physical Activity Status in Healthy Adults: Regression Model Development and Validation
JMIR Public Health Surveill 2022;8(7):e34717
URL: https://publichealth.jmir.org/2022/7/e34717
doi:10.2196/34717
PMID:35793133

Viewpoint

Use of Facial Morphology to Determine Nutritional Status in Older Adults: Opportunities and Challenges

Wesley Tay[1], BND; Rina Quek[1], BSc, MSc; Bhupinder Kaur[1], BSc, MSc; Joseph Lim[1], BSc; Christiani Jeyakumar Henry[1,2], BSc, MSc, PhD

[1]Clinical Nutrition Research Centre, Singapore Institute of Food and Biotechnology Innovation, Agency for Science, Technology and Research, Singapore, Singapore

[2]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
Christiani Jeyakumar Henry, BSc, MSc, PhD
Clinical Nutrition Research Centre
Singapore Institute of Food and Biotechnology Innovation
Agency for Science, Technology and Research
#07-02 MD6 Centre for Translational Medicine
14 Medical Drive
Singapore, 117599
Singapore
Phone: 65 64070793
Email: jeya_henry@sifbi.a-star.edu.sg

Abstract

Undiagnosed malnutrition is a significant problem in high-income countries, which can reduce the quality of life of many individuals, particularly of older adults. Moreover, it can also inflate the costs of existing health care systems because of the many metabolic complications that it can cause. The current methods for assessing malnutrition can be cumbersome. A trained practitioner must be present to conduct an assessment, or patients must travel to facilities with specialized equipment to obtain their measurements. Therefore, digital health care is a possible way of closing this gap as it is rapidly gaining traction as a scalable means of improving efficiency in the health care system. It allows for the remote monitoring of nutritional status without requiring the physical presence of practitioners or the use of advanced medical equipment. As such, there is an increasing interest in expanding the range of digital applications to facilitate remote monitoring and management of health issues. In this study, we discuss the feasibility of a novel digital remote method for diagnosing malnutrition using facial morphometrics. Many malnutrition screening assessments include subjective assessments of the head and the face. Facial appearance is often used by clinicians as the first point of qualitative indication of health status. Hence, there may be merit in quantifying these subtle but observable changes using facial morphometrics. Modern advancements in artificial intelligence, data science, sensors, and computing technologies allow facial features to be accurately digitized, which could potentially allow these previously intuitive assessments to be quantified. This study aims to stimulate further discussion and discourse on how this emerging technology can be used to provide real-time access to nutritional status. The use of facial morphometrics extends the use of currently available technology and may provide a scalable, easily deployable solution for nutritional status to be monitored in real time. This will enable clinicians and dietitians to keep track of patients remotely and provide the necessary intervention measures as required, as well as providing health care institutions and policy makers with essential information that can be used to inform and enable targeted public health approaches within affected populations.

(JMIR Public Health Surveill 2022;8(7):e33478) doi:10.2196/33478

KEYWORDS

malnutrition; facial recognition; facial morphology; telemonitoring; 3D scans; digital health; digital nutrition; public health nutrition; mobile phone

Introduction

Background

Malnutrition in older adults in many high-income countries is an unrecognized condition that is increasing in prevalence as the population ages [1]. The impact of malnutrition on morbidity and mortality in older adults is a problem that necessitates urgent attention. The aging process brings about biological, physiological, social, and psychological changes, coupled with a higher morbidity prevalence, which can negatively affect nutritional status [2].

Malnutrition in older adults, also coined as *anorexia of aging*, has severe implications, resulting in a significantly poor quality of life [3,4]. The loss of swallowing function, polypharmacy, cachexia, and lack of social support can exacerbate malnutrition [5,6]. Malnutrition occurs because of insufficient food intake, the inability of the body to absorb nutrients, or because of the presence of catabolic diseases such as chronic obstructive pulmonary disease or cancer. This can cause a reduction in fat and muscle mass, further leading to impaired physical and cognitive function [7]. Malnutrition can occur gradually over significant periods without being detected, as the symptoms may not be critical to warrant immediate medical attention. However, the severity of malnutrition may be significantly elevated if it is associated with a disease or inflammation. In geriatric patients with diseases, it is evident that a reduction in food intake causes increased muscle breakdown because of increased resting energy expenditure [8].

Malnutrition has been found to lead to sarcopenia, and sarcopenia may aggravate malnutrition, leading to the aggravation of morbidity and mortality [9,10]. Sarcopenia is defined as an age-related loss of muscle mass, where muscle mass is lost involuntarily over time, leading to a decrease in muscle protein synthesis, strength, and function [11]. In a recent review, both malnutrition and sarcopenia were examined concomitantly, and the findings for the management and interventions of these 2 conditions were found to overlap considerably [12]. Hence, the newly proposed clinical syndrome, called malnutrition-sarcopenia syndrome, should be assessed simultaneously to improve health outcomes [13]. Older adults have an increased risk of malnutrition compared with other adult populations, and it has been estimated that 2% to 16% of community-dwelling older adults are deficient in their protein and energy needs [14]. The effects of poor skeletal muscle mass (SMM) can be detrimental as they reduce the prognosis of many diseases, including diabetes, heart disease, chronic kidney disease, sepsis, and cancer [15-20]. In a community-based, cross-sectional study conducted on 722 stroke- and dementia-free participants aged 50 to 75 years, an increase in SMM was observed to protect against ischemic stroke, especially in men [21]. In another study in which the prevalence of sarcopenia in 414 patients with diabetes and 396 control participants was examined, type 2 diabetes was independently associated with sarcopenia, with a 15.7% prevalence in patients with diabetes compared with 6.9% in the control group [17].

Apart from physiological and metabolic implications, malnutrition and sarcopenia in older adults have been widely established to have financial implications and are burdens to health care resources [22]. Patients who are malnourished are associated with more frequent hospitalizations, increased length of hospital stays, and increased readmissions and hospitalization costs [23]. Approximately 40% of patients are malnourished upon hospital admission, and depending on the severity, they are faced with approximately 31% to 38% increase in hospital costs [24].

However, it is difficult to detect subclinical malnutrition because of the absence of clear indicators. This could lead to a lack of identifying poor nutritional practices among older adults, as malnutrition does not often present with immediate problems. Therefore, better screening options for the diagnosis of subclinical malnutrition must be recognized and identified [25,26]. This could potentially ease the burden on the health care system, reduce the cost of resources, improve the quality of patients' lives, and extend the healthy living years of older adults.

It is imperative that malnutrition and the prognostic value of treatment options to manage malnutrition be identified. The face is often used by health practitioners to make qualitative assessments of health and nutritional status. Modern advancements in sensor and computing technologies enable features in the face to be accurately digitized and could potentially allow these previously intuitive assessments to be quantified. In this paper, we discuss the feasibility of a novel digital remote method for diagnosing malnutrition using facial morphometrics.

Current Methods of Diagnosing Malnutrition

Until 2016, there were no universally accepted screening tools or criteria used as the gold standard to identify malnutrition risk [27]. The Global Leadership Initiative on Malnutrition assembled a set of criteria for identifying and diagnosing malnutrition in older adults in hospitals [27]. This initiative comprises a 2-step method in which a validated screening tool is used to identify patients who are at risk of malnutrition. This is followed by an assessment to diagnose and grade the severity of malnutrition. The Global Leadership Initiative on Malnutrition grouped the diagnosis of malnutrition into two distinct categorical criteria: phenotypic (which included weight loss, low BMI, and reduced SMM) and the other etiologic criteria comprising reduced food intake, disease burden, or inflammation [27].

One of the most widely used traditional methods for assessing malnutrition is the subjective global assessment tool. It is an integrated tool that assesses nutritional status based on patients' past records and physical examinations, and subsequently classifies patients into three categories: well-nourished, moderately malnourished, or severely malnourished [28]. Anthropometric measurements such as midarm circumference and triceps skinfold thickness have been used as basic methods of correlating the percentage of body fat and lean muscle mass [29]. In addition to the subjective global assessment, bioelectrical impedance analysis and ultrasound techniques have been recognized as accurate methods for measuring lean body tissue, which is a key indicator of nutritional status [30]. Other established methods used to assess SMM include scans using computed tomography (CT), dual-energy x-ray absorptiometry, and magnetic resonance imaging (MRI), which are costly and time consuming [31,32]. Analytic morphomics is an emerging field that uses cross-sectional images to provide a global assessment of a patient beyond the localized specific pathology of interest (Englesbe et al [33]). Through the application of CT-based analytic morphomics, Lee et al [31] developed the Morphomic Malnutrition Score as a means of providing standardization and objective measurements for conventional diagnostics. This method provides an in-depth assessment to distinguish between healthy individuals and individuals who are severely malnourished. Although these methods are highly accurate, conducting them can be cumbersome and resource intensive, as they require a trained practitioner to be present to conduct the assessment, and patients need to travel to specialized facilities to obtain their measurements. In cases where older adults do not frequent clinics or hospitals, malnutrition and sarcopenia could be left undetected for extended periods or until the onset of a disease, leading to several negative implications. If malnutrition monitoring and assessments could be delivered or enhanced through the use of readily accessible and cost-effective imaging technologies, this could greatly reduce the burden on rapidly inflating health care costs. More importantly, the implementation of such methods would significantly improve access to and quality of care for older adult patients who are malnourished.

Telemonitoring and Digital Health Care Systems

Overview

Digital health care is one such approach that allows for malnutrition and nutritional status to be monitored remotely without requiring the physical presence of practitioners or advanced equipment. As we enter the Fourth Industrial Revolution, the widespread availability and improvements in data science and infocommunication (integration of information and telecommunication technology sectors) technologies have rapidly expanded the vast range of available services and applications in the health care sector [34-39]. Digital health care, in its many forms, is rapidly gaining traction as a scalable means of improving efficiency in the health care system, allowing health care institutions to extend their reach and increase the number of touchpoints they have with patients. Increased levels of connectivity offered by these technologies are changing the ways in which the health care system is structured and administered.

Telemonitoring is the real-time monitoring of patients using mobile technologies to conduct routine medical tests and communicate results to health care workers for evaluation [38]. This capitalizes on the recent progress in Internet of Medical Things devices and widespread internet access, which can range from data gathered from a single device to an entire network of personal digital devices, connected medical devices, implants, and other sensors [38]. The integration of these sensors and collection of accurate real-time data can support clinical decisions by rapidly communicating changes in physiological or biochemical states, improving the overall quality of health care services and leading to more effective management of chronic diseases [34,37,38,40-46]. Digital applications allow conditions to be managed at home or within local communities, offering patients greater levels of independence and autonomy, reducing travel time, and ensuring better equity for patients in rural areas [34,38]. As such, many studies have shown high levels of satisfaction, empowerment, and reassurance expressed by both providers and end users of telemonitoring solutions [34,37,45,47].

Telemonitoring is also considered an efficient use of health care resources [34,37,38,40-44,46,48]. Many conventional tests that require the presence of a health care worker can now be conducted automatically with the use of specialized, portable equipment, reducing the number of face-to-face interactions required of health care workers [34,38]. Patients are also not required to be physically present at health care institutions. The reduction in traffic and congestion allows for the prioritization of more serious cases that require immediate care, a benefit most notably realized during the COVID-19 pandemic [34,49,50]. The improved management of chronic diseases also reduces the number of hospitalization events and the length of hospital stays for many conditions and, consequently, reduces overall health care expenditure.

Given the many benefits that telemonitoring has to offer, it is critical that new applications are developed to broaden the scope of conditions to which telemonitoring can be applied. In both high- and low-income nations, smartphones have been used as platforms for several aspects of life, such as identity verification and personal banking, and are increasingly being adopted for use in the capture, management, and transmission of personal health data [51-54]. These smart devices are outfitted with sensors such as GPS trackers, accelerometers, cameras, and depth sensors and are progressively updated each year with newer and more powerful components. The incorporation of these advanced sensors may offer a scalable solution for population-wide telemonitoring of nutritional status and allow early detection of malnutrition in the community.

Facial Morphometrics and Malnutrition Screening

A relatively unexplored method that may offer predictive capabilities in the area of malnutrition screening is the application of facial morphometrics. Facial morphometrics is the study of the contours and structures of faces using geometric mapping tools that allow landmarks and features to be identified [55]. Although facial changes during weight loss and undernutrition have been reported since the time of Leonardo Da Vinci, their implications and ability to predict nutritional status have been recognized only recently [56-62].

The newly integrated depth sensors in personal smart devices may present a novel solution that may be beneficial for the remote monitoring of malnutrition. These depth sensors capture 3D depth data of objects, enabling the use of applications such as augmented reality–based applications, as well as the generation of 3D point clouds for 3D modeling [63]. Most recently, a notable application that highlights the current level of sophistication and accuracy of depth sensors is Apple's 3D face recognition–based encryption system, known as Face ID [64]. The degree of security offered by the Face ID system can be attributed to the level of 3D detail that the True Depth camera can capture. A total of 30,000 infrared dots are projected onto the target object and captured by an infrared-sensitive camera, allowing fine details such as contours and recesses to be detected by the camera [65]. This granularity has enabled its use in a wide range of applications, ranging from the analysis of human facial emotions to the customization of prosthetics and medical equipment requiring precise fit [63,66-71].

With depth cameras becoming the mainstay in smartphones beyond the iPhone, there may be potential for developing telemonitoring applications that involve the extraction and correlation of facial morphological features with health indices. Typical facial landmarks include the eyes, nose, chin, cheekbones, and the overall shape of the face. The extraction and analysis of features such as geodesic distances or ratios between these landmarks can serve as informative discriminatory tools in the quantification of shape and variation [72].

These quantified features could serve as proxy markers for the nutritional status of the human body. This has been demonstrated in several studies, which have observed correlations between superficial tissues that affect the overall morphology of the face such as the zygomatic fat pads and masseter and temporalis muscles with nutritional status and whole-body SMM [73-83]. The quantification of these tissues involves the use of techniques such as ultrasound, MRI, and CT scans. However, these methods can be expensive, time consuming, and stressful for patients. Therefore, the application of facial morphometrics using widely available depth sensors could provide a more accessible and less invasive solution for larger populations.

The use of facial morphometrics in the field of health sciences is not new. However, most of the research has primarily been conducted in the fields of aesthetics, orthodontics, and prosthetics. There have been attempts to explore the correlation between facial features and nutrition-focused indices such as BMI and waist-to-hip ratio [84-91]. Earlier studies used 2D images and relied on planar distances or 2D angularity to identify features that account for BMI variations [87,89,92-99]. Recent studies have explored the use of 3D meshes for more detailed geometric modeling of faces using tools such as 3D scanners or stereophotogrammetry [86,100-102]. However, these technologies require a specialized setup, and individuals must be physically present for the scans to be conducted, thus limiting the number of individuals that can be recruited. The widespread availability of personal smartphones integrated with accurate depth cameras presents an opportunity for more scalable research in this area, allowing the acquisition of larger data sets within a shorter amount of time.

With these capabilities in mind, this paper further discusses the relevance of certain facial features that have the potential to provide diagnostic value in identifying changes in SMM and overall nutritional status.

Facial Features With Clinical Significance

Overview

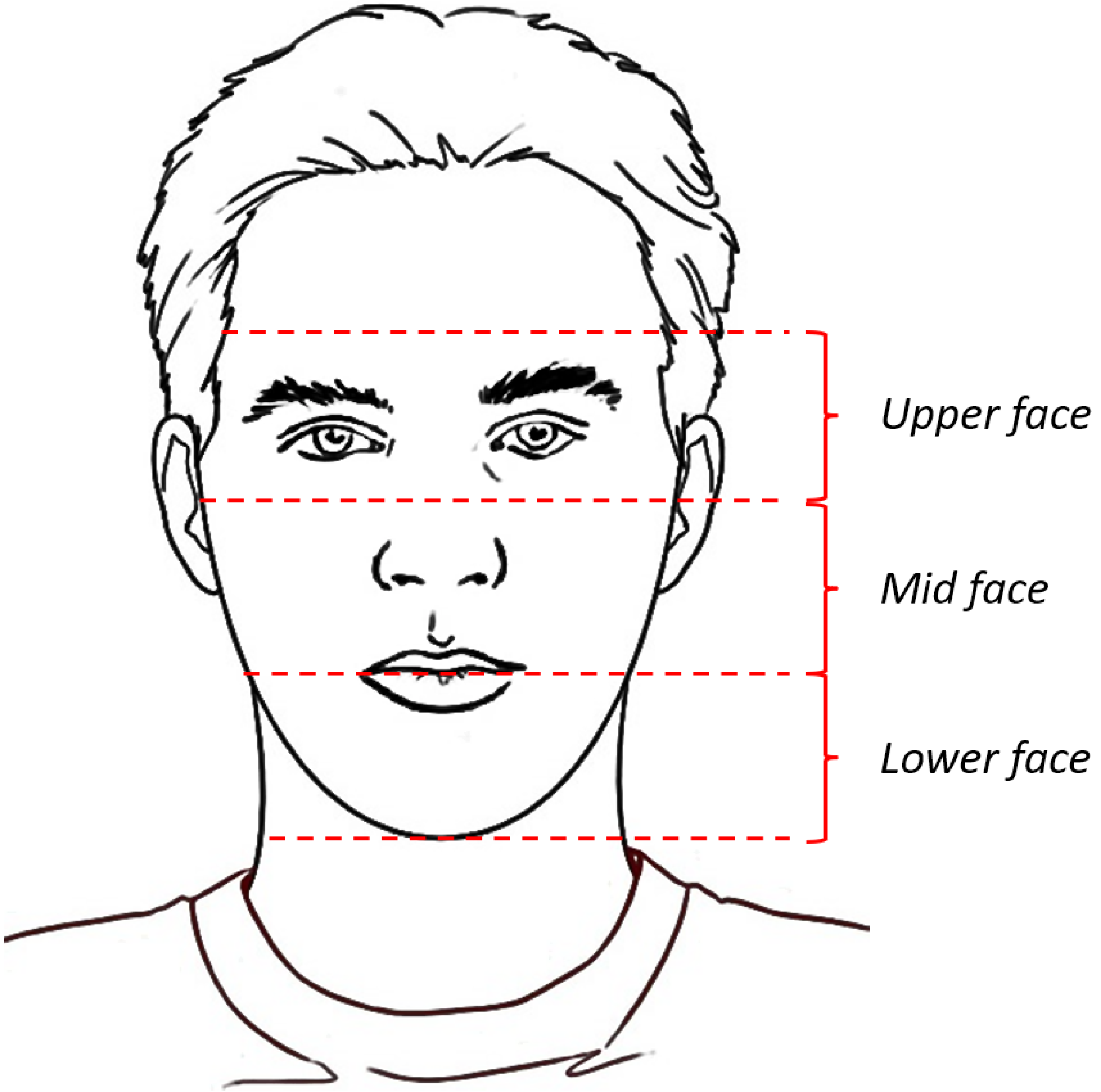
With imaging tools capable of deriving facial morphology accurately, both direct and indirect tissue losses can potentially be captured as additional features for analysis. Direct measurements of tissue loss can include the quantification of protrusions and depressions that form from tissue atrophy, whereas indirect measurements include features that become more prominent when there is a loss of fat or muscle tissue, such as certain bone processes and ridges, or the angulation of the jawline [73,100,103].

The degree of bone prominence is important in determining nutritional status [104,105]. Physical examinations of malnutrition typically observe the prominence of bone structures in the clavicle, scapular, acromion, and patellar regions to determine the degree of muscle and subcutaneous fat loss [104]. Certain superficial structures that have a lesser degree of covering tissue, such as the zygomatic process or mandible, are immediately distinguishable. Other bone structures such as those in the orbit, zygomatic bone, and other anatomical details of the mandible or jaw may become more prominent in the case of malnourishment, following the loss of muscle and fat tissue [105].

Segmentation of the Face

To discuss potential features, we segment the face into three regions: upper face, midface, and lower face (Figure 1).

Figure 1. Illustration of face segments of interest.



Upper Face

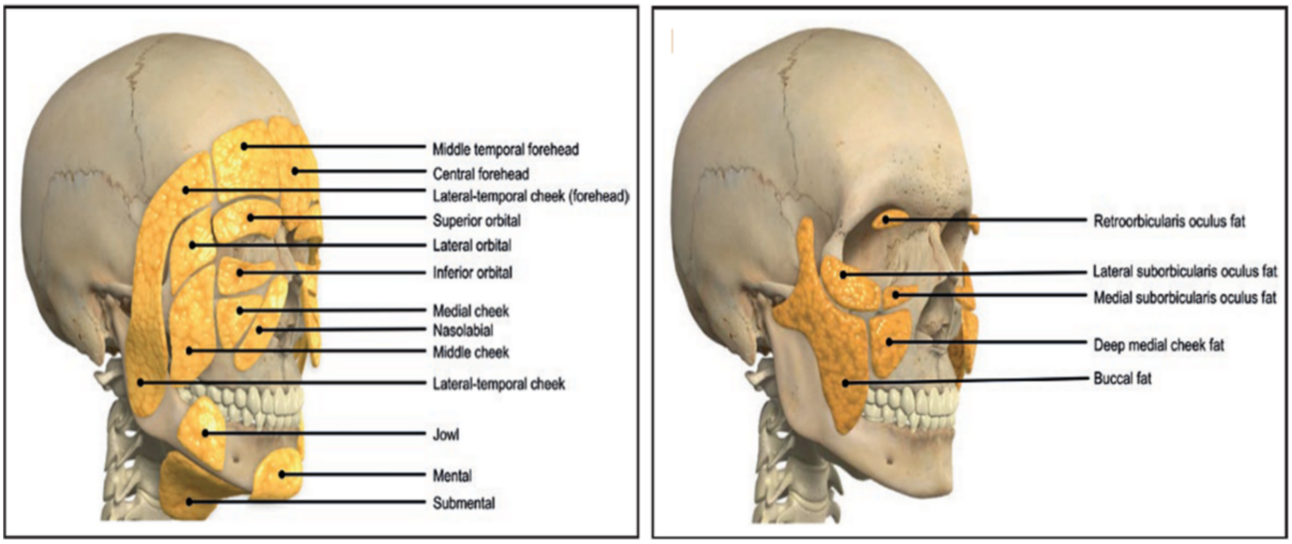
The upper face comprises the regions between both temples, including the eyes and the surrounding periorbital region. The primary superficial tissues that contribute to craniofacial morphology in this region include the temporalis muscle, periorbital fat pads, and bones that make up the orbital rim [59,106,107]. These features are significantly affected by malnutrition and loss of fat or skeletal muscle in the region [104,108].

The temporalis muscle, as well as the superficial and deep temporal fat pads, contribute to most of the volume in the temple regions [109]. As it is easily accessible and palpable, the temporalis muscle is frequently used as an indicator of malnutrition in physical examinations [104]. The temporalis muscle is a fan-shaped muscle situated beneath the temples, which runs from the side of the skull down to the back of the lower jaw, forming a convexly shaped area that blends in with the lateral orbital rim [107]. On several occasions, this muscle has been shown to be an effective surrogate marker for sarcopenia and SMM in the rest of the body. Many studies have reported correlations between the temporalis muscle and other indicators of SMM, such as the psoas muscle, lumbar skeletal muscle cross-sectional area, and patient grip strength [75,76,78,82,83,110-112].

In the case of atrophy of the underlying fat pads and the temporalis muscle, this region could appear sunken, thus increasing the prominence of the temporal orbital rim [106]. The frontal process of the zygomatic bone and the superior border of the zygomatic arch become more prominent and can result in sharp angulation along the orbital rim [109]. These features cast a shadow that accents the temporal hollows, resulting in a gaunt appearance [106].

The morphology of the periorbital region also changes significantly in the case of tissue loss and is largely determined by the volume of the periorbital fat pads. These comprise the temporal fat pad, the preaponeurotic fat pad, the nasal fat pads, and the central fat pad [113] (Figure 2). In cases of lipoatrophy, the loss of volume in these fat pads has been observed to accentuate the superior, medial, and lateral edges of the orbital rim, contributing to a sunken, skeletonized appearance [109]. The increasing hollowness of the orbit coupled with skin laxity can also contribute to the development of infraorbital dark circles [114].

Figure 2. Overview of facial fat pads. Superficial fat compartments in the midface include the nasolabial fat pad, medial cheek, buccal fat pad, and deep medial cheek fat (Swift et al [115]).



Certain confounders may arise because of the natural changes around the eye regions, which may complicate the identification of malnutrition. Age contributes to a decrease in brow projection from the superior orbital rim [106]. Youthful brows have a more prominent anterior projection because of the underlying support from the retro-orbicularis oculi fat. During aging, subsequent remodeling and thinning of the superomedial and inferolateral orbital rim can result in changes to the morphology and decreased projection of the brows from the superior orbital rim [106]. This makes the superciliary arch more prominent and visible in older adults. Therefore, it is important to carefully distinguish this from excess atrophy of fat and muscle tissue.

It has been well established that the morphology around the periorbital fat pads is affected by many factors because of aging [59,116,117]. The tissues around the eyes, such as the muscles that support the eyelids, tend to weaken. Fat pads that support the eye may migrate lower to take the form of bags, and a reduction in fat and underlying muscle fibers can also give the eyelids a sunken and hollow appearance. The loss of skin elasticity and development of eyelid folds or ptosis can lead to additional contours forming around the eye [59]. Future research will need to account for these variabilities that may exist because of age, gender, and ethnicity to differentiate between malnutrition and the natural process of aging [72].

Midface

The midface comprises the regions bound by the cheekbones as the upper boundary and the lips as the lower boundary. Key tissue groups that affect the morphology of this region include the masseter muscle, the various zygomatic-malar fat pads, and the zygomatic bone [106,107].

The masseter muscle is a masticatory muscle located over the zygomatic bone, running from the temporal bone to the lower jaw [107]. Given the prominence and superficiality of the masseter muscle on the side of the face, it contributes to the facial contour of the cheek. The thickness of the masseter muscle significantly affects the overall morphology of the face [107,118-128]. The morphological effects of masseter muscle thickness have been well documented in the field of aesthetic medicine, where prescriptive volume loss through the use of botulinum toxin type A demonstrates noticeable changes in the shape of the face [118]. Given the superficiality of the muscle, Murakami et al [77] developed a tool to estimate the mass of the masseter muscle using surface measurements obtained with a vernier caliper. These were paired with thickness measurements derived from ultrasound images to determine the overall mass of the masseter muscle. This was shown to be an accurate method of assessment when compared with the results measured using MRI [77].

Similar to the usability of the temporalis muscle as a surrogate marker for SMM, the masseter muscle has also been found to be correlated with whole-body SMM and the psoas, limb, and trunk muscles [74,77,79,80,126,129]. Relationships have been established between the masseter muscle and aspects of physical fitness in older adults [130]. In some studies, the thickness of the masseter muscle was found to be a better predictor of mortality and sarcopenia than the psoas muscle, which is a commonly referenced muscle in the diagnosis of sarcopenia [78-81,110,131,132].

Given the masseter muscle’s key function as one of the primary muscles of mastication, its strength has also been found to be correlated with the chewing ability [74,130,133]. Research has shown that the development or attrition of the masticatory muscles is related to chewing activity and is indirectly linked to dietary habits [74,77,122,123,130]. In patients who experience loss of appetite or decreased food intake because of medical, social, or psychological reasons, the masticatory muscles have been found to rescind in size [77,81,134]. This will likely affect the external morphology of the cheeks and jaw and could potentially indicate a risk of malnutrition.

There are several fat compartments in the zygomatic-malar region of the face. The deep midface compartments include the medial and lateral suborbicularis oculi fat, the medial portion of the buccal fat pad, and the deep medial cheek fat. Superficial compartments include the nasolabial and middle cheek superficial fat compartments (Figure 2) [107,116]. The fat pads in this region are typically considered in malnutrition-specific physical examinations. In these examinations, practitioners typically look for hollow, sunken, and narrow faces as these can be an indication of malnutrition [104,108]. In extreme scenarios of tissue loss such as HIV-specific lipoatrophy, exacerbated atrophy in these fat pads can manifest as significant concavities in the buccal and malar areas [105] (Figure 3). This greatly accentuates the zygomatic bone and can form many depressions in the midfacial region.

Figure 3. Visible atrophy of adipose tissue and protruding facial bone structures in both cheeks because of antiretroviral therapy for HIV, indicated by red arrows (Szczerkowska-Dobosz et al [105]).

Subcutaneous buccal fat pads are also a significant area of aging-associated fat loss [59]. The loss of overall volume, decreased tissue support for facial features, and a reduction in skin elasticity can lead to the hollowing of the cheeks. These factors can affect the morphology of the face and, as such, will need to be accounted for when identifying features correlated with malnutrition.

Lower Face

The lower face comprises the mandible and the surrounding tissues of the lower cheeks, chin, and neck. These structures contribute significantly to the overall shape of the face and are often easily perceptible, except in cases of excess facial hair. As such, the lower face is a key point of reference in many facial morphological studies and BMI prediction algorithms [89,95,98,135,136].

The lower mandible contributes to the morphology of the lower face and accentuates the jawline. Aesthetic surgeons refer to a youthful jawline as one characterized by a straight line from the chin to the mandibular angle [137,138]. Excess subcutaneous fat in the region may cause certain skinfolds and disrupt this angular definition of the jawline, which can lead to the lower face becoming more squarish than oval [84,89]. Conversely, malnutrition and periods of unintentional weight loss can lead to atrophy of subcutaneous fat and the underlying muscle tissues, resulting in a more pronounced jawline.

The jawline can be a potentially important feature in predicting malnutrition or sarcopenia. In a study exploring the relationship between face shape and waist-to-hip ratio, Mayer et al [89] showed that a 2D face shape had a correlation coefficient of 0.63 (Figure 4). Participants with a lower BMI were likely to have jawlines that were sharper and more angular, whereas the jawlines of participants with a higher BMI were wider and rounder. In the Coetzee et al [84] study identifying quantifiable visual cues that contribute to weight, 2 out of the 3 primary cues that showed predictive potential to determine BMI from facial images were the perimeter-to-area ratio and the cheek-to-jaw width ratio (Figure 5). These features are largely dependent on the amount of tissue in the lower facial regions, thus indicating the importance of identifying the jawline as a differential landmark.

Figure 4. Changes in BMI (kg/m2) as reflected in face shape (Mayer et al [89]).

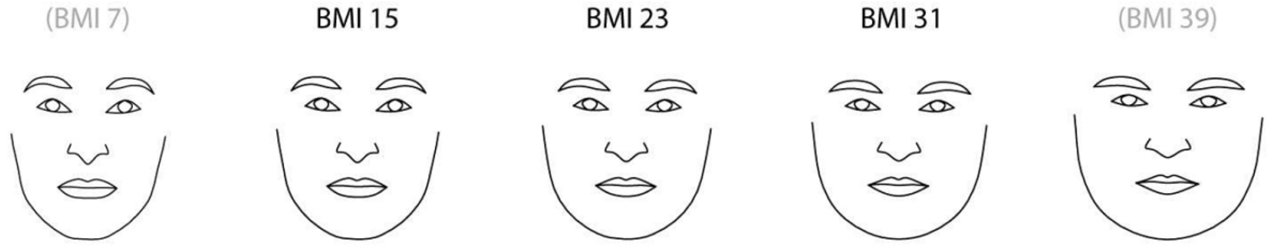
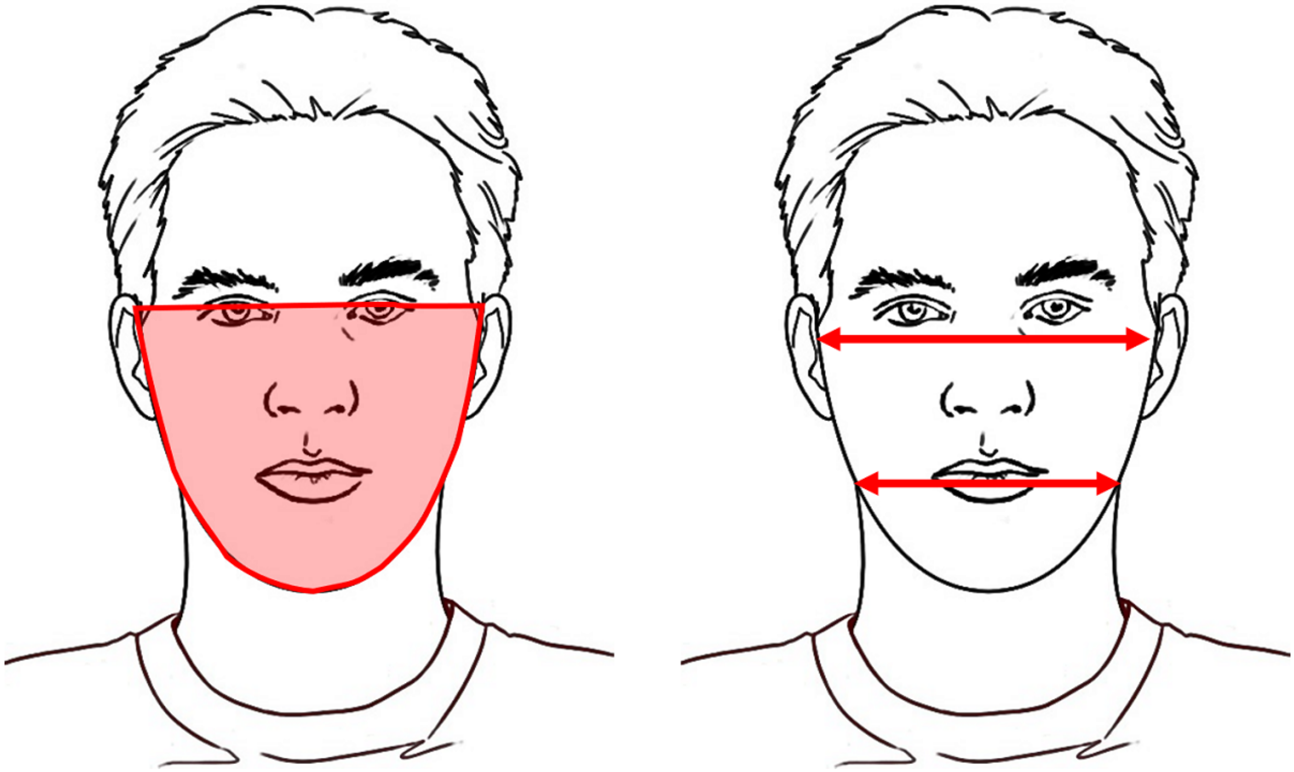


Figure 5. Perimeter-to-area ratio (left) and cheek-to-jaw width ratio (right; adapted from Coetzee et al [84]).



Although subcutaneous fat affects the overall shape of the jawline, aging is another important factor that changes the overall morphology of this facial feature [106,135,138-140]. A strong mandibular border with prominent jawline definitions is typically a sign of a youthful face, whereas aging contributes to squaring of the face and a loss of jawline definition. This is because of the descent of the hyoid bone and larynx, leading to a blunting of the cervicomental angle. Together with ptosis of the unsupported skin, chin pad, and facial portion of the platysma muscle, it results in the development of ptotic jowls [106,135,141]. This can potentially complicate the analysis of mandibular morphology and can affect its usability in malnutrition assessment, especially in older adults. Other confounding factors that may mask the contours of the chin and mandible regions include facial hair or aesthetic fillers.

Challenges

Facial morphology technology provides several exciting opportunities as a diagnostic tool for malnutrition and body composition. However, there remain important challenges that need to be overcome.

Developing a comprehensive database that accounts for potential confounders and isolating the degree of correlation between various facial morphological features and overall SMM and nutritional status is necessary. Many factors can influence facial variations outside of weight and SMM, such as gender, ethnicity, age, and medical conditions [72,142,143]. The rate and degree of facial development throughout the course of growth spurts and aging may also vary depending on ethnicity, which could further complicate the effects of either parameter [144]. The development of facial morphology databases will need to account for these factors such that specific features most relevant to malnutrition can be isolated.

Facial morphology holds potential as a diagnostic tool in the context of malnutrition on the premise that SMM is correlated with facial tissues. However, this correlation may not always be consistent, particularly for complex medical conditions. Certain diseases such as critical illness myopathy have been noted to affect muscles in the body differently, resulting in significant muscle loss in the limbs and trunk while sparing the masseter and other craniofacial muscles [142]. Controlling for these potential confounders will become an important consideration if facial morphology is to be used as a population-wide diagnostic tool.

There can also be disproportionate tissue loss in specific parts of the face, which may confound the degree of importance of the respective facial regions. For example, patients undergoing highly active antiretroviral therapy for HIV infections or patients with connective tissue diseases such as panniculitis are known to develop asymmetrical and dynamic facial lipoatrophy [105]. This form of lipoatrophy can manifest in different ways in age- or lifestyle-related malnutrition, and care should be taken to distinguish between the two.

The use of facial morphology as a diagnostic tool may also require additional considerations because of aesthetic variables. The presence of facial hair, such as beards and mustaches, may impair the ability of features derived from the lower half of the face to be used in the analysis. Cosmetic or reconstructive surgery used for the concealment of facial anomalies or to improve one’s appearance is also becoming relatively common in many countries [145]. These can have a significant impact on the morphology of the face and are important factors to consider in these demographics.

Certain ethical and societal considerations may need to be reviewed to increase the adoption of facial morphometrics as a tool in public health. The face is an important and sensitive identifiable feature, and any application incorporating digital 3D scans of the face will need to ensure that patient privacy and data security are not compromised. The development of digital health applications using personal smartphones may also widen the digital divide and potentially alienate certain segments of the population [146]. Populations of lower socioeconomic status may not have access to these devices, and older adult patients may have difficulties navigating certain aspects of newer technologies [35]. These populations are the most vulnerable to malnutrition and its associated complications. To ensure that these segments of the population who may not be able to afford or be familiar with digital technologies are not excluded, there is a need to consider both affordability as a potential barrier to entry, as well as the ergonomics and accessibility of the user experience.

Discussion

Conclusions

Undiagnosed malnutrition is becoming a significant problem that can reduce the quality of life of many individuals and inflate the costs of existing health care systems because of the many metabolic complications it can cause. Many high-income countries are faced with a rapidly aging population that may lead to poor food intake, disease etiology, and improper access to food, which may give rise to an increase in the prevalence of malnutrition. Given the need for specialized equipment and trained practitioners for the diagnosis of malnutrition, the development of simpler and more scalable methods for assessing malnutrition has been a long-standing challenge for public health.

Recent advancements in artificial intelligence, computing, data science, and infocommunication technologies present a wealth of opportunities that will shape and revolutionize the landscape of health care. This paper aims to stimulate further discussion and discourse on how adopting this emerging technology could provide real-time access to nutritional status. We explored the feasibility of a novel, noninvasive approach to tackling malnutrition in the community. The use of facial morphometrics extends the use of currently available technology, providing a scalable and easily deployable method for the real-time monitoring of nutritional status. Clinicians and dietitians will be able to keep track of patients remotely and provide the necessary intervention measures as required. Finally, it will furnish health care institutions and policy makers with essential information that can facilitate necessary interventions targeted at affected populations.

Acknowledgments

The study was supported by the Biomedical Science Institute Industry Alignment Fund and the Food Structure Engineering for Nutrition and Health (H17/01/a0/A11 and H18/01/a0/B11).

Authors' Contributions

This study was conceptualized by WT, RQ, and CJH. The original draft was prepared by WT and RQ. Reviewing and editing was done by WT, RQ, JL, BK, and CJH. All the authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Visvanathan R. Under-nutrition in older people: a serious and growing global problem!. J Postgrad Med 2003;49(4):352-360 [FREE Full text] [Medline: 14699239]
2. Ahmed T, Haboubi N. Assessment and management of nutrition in older people and its importance to health. Clin Interv Aging 2010 Aug 09;5:207-216 [FREE Full text] [doi: 10.2147/cia.s9664] [Medline: 20711440]
3. Rasheed S, Woods RT. Malnutrition and quality of life in older people: a systematic review and meta-analysis. Ageing Res Rev 2013 Mar;12(2):561-566. [doi: 10.1016/j.arr.2012.11.003] [Medline: 23228882]
4. Wysokiński A, Sobów T, Kłoszewska I, Kostka T. Mechanisms of the anorexia of aging-a review. Age (Dordr) 2015 Aug;37(4):9821 [FREE Full text] [doi: 10.1007/s11357-015-9821-x] [Medline: 26232135]
5. Sampson G. Weight loss and malnutrition in the elderly--the shared role of GPs and APDs. Aust Fam Physician 2009 Jul;38(7):507-510 [FREE Full text] [Medline: 19575069]
6. Nagaratnamon N. Malnutrition and malabsorption in the elderly. In: Advanced Age Geriatric Care. Cham: Springer; 2019.
7. Cederholm T, Barazzoni R, Austin P, Ballmer P, Biolo G, Bischoff SC, et al. ESPEN guidelines on definitions and terminology of clinical nutrition. Clin Nutr 2017 Feb;36(1):49-64. [doi: 10.1016/j.clnu.2016.09.004] [Medline: 27642056]
8. Jensen GL, Mirtallo J, Compher C, Dhaliwal R, Forbes A, Grijalba RF, International Consensus Guideline Committee. Adult starvation and disease-related malnutrition: a proposal for etiology-based diagnosis in the clinical practice setting from the International Consensus Guideline Committee. Clin Nutr 2010 Apr;29(2):151-153. [doi: 10.1016/j.clnu.2009.11.010] [Medline: 20071059]
9. Mithal A, Bonjour J, Boonen S, Burckhardt P, Degens H, El Hajj Fuleihan G, IOF CSA Nutrition Working Group. Impact of nutrition on muscle mass, strength, and performance in older adults. Osteoporos Int 2013 May;24(5):1555-1566. [doi: 10.1007/s00198-012-2236-y] [Medline: 23247327]
10. Wakabayashi H, Matsushima M, Uwano R, Watanabe N, Oritsu H, Shimizu Y. Skeletal muscle mass is associated with severe dysphagia in cancer patients. J Cachexia Sarcopenia Muscle 2015 Dec;6(4):351-357 [FREE Full text] [doi: 10.1002/jcsm.12052] [Medline: 26673551]
11. Cruz-Jentoft AJ, Sayer AA. Sarcopenia. Lancet 2019 Jun 29;393(10191):2636-2646. [doi: 10.1016/S0140-6736(19)31138-9] [Medline: 31171417]
12. Juby AG, Mager DR. A review of nutrition screening tools used to assess the malnutrition-sarcopenia syndrome (MSS) in the older adult. Clin Nutr ESPEN 2019 Aug;32:8-15. [doi: 10.1016/j.clnesp.2019.04.003] [Medline: 31221295]
13. Vandewoude MF, Alish CJ, Sauer AC, Hegazi RA. Malnutrition-sarcopenia syndrome: is this the future of nutrition screening and assessment for older adults? J Aging Res 2012;2012:651570 [FREE Full text] [doi: 10.1155/2012/651570] [Medline: 23024863]
14. Amarya S, Singh K, Sabharwal M. Changes during aging and their association with malnutrition. J Clin Gerontol Geriatrics 2015 Sep;6(3):78-84. [doi: 10.1016/j.jcgg.2015.05.003]
15. Dodson S, Baracos VE, Jatoi A, Evans WJ, Cella D, Dalton JT, et al. Muscle wasting in cancer cachexia: clinical implications, diagnosis, and emerging treatment strategies. Annu Rev Med 2011;62:265-279. [doi: 10.1146/annurev-med-061509-131248] [Medline: 20731602]
16. Kalyani RR, Corriere M, Ferrucci L. Age-related and disease-related muscle loss: the effect of diabetes, obesity, and other diseases. Lancet Diabetes Endocrinol 2014 Oct;2(10):819-829 [FREE Full text] [doi: 10.1016/S2213-8587(14)70034-8] [Medline: 24731660]
17. Kim TN, Park MS, Yang SJ, Yoo HJ, Kang HJ, Song W, et al. Prevalence and determinant factors of sarcopenia in patients with type 2 diabetes: the Korean Sarcopenic Obesity Study (KSOS). Diabetes Care 2010 Jul;33(7):1497-1499 [FREE Full text] [doi: 10.2337/dc09-2310] [Medline: 20413515]
18. Shibahashi K, Sugiyama K, Kashiura M, Hamabe Y. Decreasing skeletal muscle as a risk factor for mortality in elderly patients with sepsis: a retrospective cohort study. J Intensive Care 2017;5:8 [FREE Full text] [doi: 10.1186/s40560-016-0205-9] [Medline: 28096999]
19. Tsuchida K, Fujihara Y, Hiroki J, Hakamata T, Sakai R, Nishida K, et al. Significance of sarcopenia evaluation in acute decompensated heart failure. Int Heart J 2018 Jan 27;59(1):143-148 [FREE Full text] [doi: 10.1536/ihj.17-057] [Medline: 29332917]
20. Wang XH, Mitch WE. Mechanisms of muscle wasting in chronic kidney disease. Nat Rev Nephrol 2014 Sep;10(9):504-516 [FREE Full text] [doi: 10.1038/nrneph.2014.112] [Medline: 24981816]

21. Minn Y, Suk S. Higher skeletal muscle mass may protect against ischemic stroke in community-dwelling adults without stroke and dementia: the PRESENT project. BMC Geriatr 2017 Feb 03;17(1):45 [FREE Full text] [doi: 10.1186/s12877-017-0433-4] [Medline: 28158989]
22. Sousa AS, Guerra RS, Fonseca I, Pichel F, Ferreira S, Amaral TF. Financial impact of sarcopenia on hospitalization costs. Eur J Clin Nutr 2016 Sep;70(9):1046-1051. [doi: 10.1038/ejcn.2016.73] [Medline: 27167668]
23. Rodríguez-Sánchez B, Sulo S, Carnicero JA, Rueda R, Rodríguez-Mañas L. Malnutrition prevalence and burden on healthcare resource use among Spanish community-living older adults: results of a longitudinal analysis. Clinicoecon Outcomes Res 2020;12:355-367 [FREE Full text] [doi: 10.2147/CEOR.S256671] [Medline: 32765021]
24. Curtis LJ, Bernier P, Jeejeebhoy K, Allard J, Duerksen D, Gramlich L, et al. Costs of hospital malnutrition. Clin Nutr 2017 Oct;36(5):1391-1396. [doi: 10.1016/j.clnu.2016.09.009] [Medline: 27765524]
25. Cowan DT, Roberts JD, Fitzpatrick JM, While AE, Baldwin J. Nutritional status of older people in long term care settings: current status and future directions. Int J Nurs Stud 2004 Mar;41(3):225-237. [doi: 10.1016/S0020-7489(03)00131-7] [Medline: 14967179]
26. Rauscher C. Malnutrition among the elderly. Can Fam Physician 1993 Jun;39:1395-1403 [FREE Full text] [Medline: 8324408]
27. Cederholm T, Jensen GL, Correia MI, Gonzalez MC, Fukushima R, Higashiguchi T, GLIM Core Leadership Committee, GLIM Working Group. GLIM criteria for the diagnosis of malnutrition - a consensus report from the global clinical nutrition community. J Cachexia Sarcopenia Muscle 2019 Feb;10(1):207-217 [FREE Full text] [doi: 10.1002/jcsm.12383] [Medline: 30920778]
28. Detsky AS, McLaughlin JR, Baker JP, Johnston N, Whittaker S, Mendelson RA, et al. What is subjective global assessment of nutritional status? JPEN J Parenter Enteral Nutr 1987;11(1):8-13. [doi: 10.1177/014860718701100108] [Medline: 3820522]
29. Buchard B, Boirie Y, Cassagnes L, Lamblin G, Coilly A, Abergel A. Assessment of malnutrition, sarcopenia and frailty in patients with cirrhosis: which tools should we use in clinical practice? Nutrients 2020 Jan 09;12(1):186 [FREE Full text] [doi: 10.3390/nu12010186] [Medline: 31936597]
30. Earthman CP. Body composition tools for assessment of adult malnutrition at the bedside: a tutorial on research considerations and clinical applications. JPEN J Parenter Enteral Nutr 2015 Sep;39(7):787-822. [doi: 10.1177/0148607115595227] [Medline: 26287016]
31. Lee C, Raymond E, Derstine BA, Glazer JM, Goulson R, Rajasekaran A, et al. Morphomic malnutrition score: a standardized screening tool for severe malnutrition in adults. JPEN J Parenter Enteral Nutr 2018 Nov;42(8):1263-1271. [doi: 10.1002/jpen.1175] [Medline: 29786877]
32. Mullie L, Afilalo J. CoreSlicer: a web toolkit for analytic morphomics. BMC Med Imaging 2019 Feb 11;19(1):15 [FREE Full text] [doi: 10.1186/s12880-019-0316-6] [Medline: 30744586]
33. Englesbe MJ, Lee JS, He K, Fan L, Schaubel DE, Sheetz KH, et al. Analytic morphomics, core muscle size, and surgical outcomes. Ann Surg 2012 Aug;256(2):255-261. [doi: 10.1097/SLA.0b013e31826028b1] [Medline: 22791101]
34. Bringing health care to the patient: an overview of the use of telemedicine in OECD countries. OECD. URL: https://www.oecd-ilibrary.org/social-issues-migration-health/bringing-health-care-to-the-patient_8e56ede7-en [accessed 2020-12-29]
35. Huxhold O, Hees E, Webster NJ. Towards bridging the grey digital divide: changes in internet access and its predictors from 2002 to 2014 in Germany. Eur J Ageing 2020 Sep;17(3):271-280 [FREE Full text] [doi: 10.1007/s10433-020-00552-z] [Medline: 32904732]
36. Lee J, Lim J. The prospect of the fourth industrial revolution and home healthcare in super-aged society. Ann Geriatr Med Res 2017 Sep 30;21(3):95-100. [doi: 10.4235/agmr.2017.21.3.95]
37. Salehi S, Olyaeemanesh A, Mobinizadeh M, Nasli-Esfahani E, Riazi H. Assessment of remote patient monitoring (RPM) systems for patients with type 2 diabetes: a systematic review and meta-analysis. J Diabetes Metab Disord 2020 Jun;19(1):115-127 [FREE Full text] [doi: 10.1007/s40200-019-00482-3] [Medline: 32550161]
38. Goy A, Nishtar S, Dzau V, Balatbat C, Diabo R. Health and healthcare in the Fourth Industrial Revolution: Global Future Council on the Future of Health and Healthcare 2016-2018. World Economic Forum. 2019. URL: https://apo.org.au/node/235506 [accessed 2020-12-21]
39. Sallai G. Defining infocommunications and related terms. Acta Polytechnica Hungarica 2012;9(6):5-15.
40. Alves DS, Times VC, Da Silva EM, Melo PS, Novaes MD. Advances in obstetric telemonitoring: a systematic review. Int J Med Inform 2020 Feb;134:104004. [doi: 10.1016/j.ijmedinf.2019.104004] [Medline: 31816495]
41. Ding H, Chen SH, Edwards I, Jayasena R, Doecke J, Layland J, et al. Effects of different telemonitoring strategies on chronic heart failure care: systematic review and subgroup meta-analysis. J Med Internet Res 2020 Nov 13;22(11):e20032 [FREE Full text] [doi: 10.2196/20032] [Medline: 33185554]
42. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]

43. Kim Y, Park J, Lee B, Jung C, Park D. Comparative effectiveness of telemonitoring versus usual care for type 2 diabetes: a systematic review and meta-analysis. J Telemed Telecare 2019 Dec;25(10):587-601. [doi: 10.1177/1357633X18782599] [Medline: 30012042]
44. Lanssens D, Vandenberk T, Thijs IM, Grieten L, Gyselaers W. Effectiveness of telemonitoring in obstetrics: scoping review. J Med Internet Res 2017 Sep 27;19(9):e327 [FREE Full text] [doi: 10.2196/jmir.7266] [Medline: 28954715]
45. Lee PA, Greenfield G, Pappas Y. The impact of telehealth remote patient monitoring on glycemic control in type 2 diabetes: a systematic review and meta-analysis of systematic reviews of randomised controlled trials. BMC Health Serv Res 2018 Jun 26;18(1):495 [FREE Full text] [doi: 10.1186/s12913-018-3274-8] [Medline: 29940936]
46. Yun JE, Park J, Park H, Lee H, Park D. Comparative effectiveness of telemonitoring versus usual care for heart failure: a systematic review and meta-analysis. J Card Fail 2018 Jan;24(1):19-28. [doi: 10.1016/j.cardfail.2017.09.006] [Medline: 28939459]
47. Li X, Xie Y, Zhao H, Zhang H, Yu X, Li J. Telemonitoring interventions in COPD patients: overview of systematic reviews. Biomed Res Int 2020;2020:5040521 [FREE Full text] [doi: 10.1155/2020/5040521] [Medline: 32016115]
48. OECD. Tackling Wasteful Spending on Health. Washington, DC: OECD; 2017.
49. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
50. Portnoy J, Waller M, Elliott T. Telemedicine in the era of COVID-19. J Allergy Clin Immunol Pract 2020 May;8(5):1489-1491 [FREE Full text] [doi: 10.1016/j.jaip.2020.03.008] [Medline: 32220575]
51. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
52. Jayaraman PP, Forkan AR, Morshed A, Haghighi PD, Kang Y. Healthcare 4.0: a review of frontiers in digital health. WIREs Data Mining Knowl Discov 2019 Dec 25;10(2). [doi: 10.1002/widm.1350]
53. Wisniewski H, Torous J. Digital navigators to implement smartphone and digital tools in care. Acta Psychiatr Scand 2020 Apr;141(4):350-355 [FREE Full text] [doi: 10.1111/acps.13149] [Medline: 31930477]
54. Anowar F, Ashraf M, Islam A, Ahmed E, Chowdhury A. A review on diabetes self-management applications for android smartphones: perspective of developing countries. In: Proceedings of the 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). 2020 Presented at: 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA); Jun 26-28, 2020; Ankara, Turkey. [doi: 10.1109/hora49412.2020.9152932]
55. Finka LR, Luna SP, Brondani JT, Tzimiropoulos Y, McDonagh J, Farnworth MJ, et al. Geometric morphometrics for the study of facial expressions in non-human animals, using the domestic cat as an exemplar. Sci Rep 2019 Jul 08;9(1):9883 [FREE Full text] [doi: 10.1038/s41598-019-46330-5] [Medline: 31285531]
56. Coetzee V, Perrett DI, Stephen ID. Facial adiposity: a cue to health? Perception 2009;38(11):1700-1711. [doi: 10.1068/p6423] [Medline: 20120267]
57. Foo YZ, Simmons LW, Rhodes G. Predictors of facial attractiveness and health in humans. Sci Report 2017 Feb 03;7:39731 [FREE Full text] [doi: 10.1038/srep39731] [Medline: 28155897]
58. Henderson AJ, Holzleitner IJ, Talamas SN, Perrett DI. Perception of health from facial cues. Philos Trans R Soc Lond B Biol Sci 2016 May 05;371(1693):20150380 [FREE Full text] [doi: 10.1098/rstb.2015.0380] [Medline: 27069057]
59. Kaur M, Garg RK, Singla S. Analysis of facial soft tissue changes with aging and their effects on facial morphology: a forensic perspective. Egyptian J Forensic Sci 2015 Jun;5(2):46-56. [doi: 10.1016/j.ejfs.2014.07.006]
60. Re DE, Rule NO. About face: new directions for the physician's general survey. Curr Dir Psychol Sci 2016 Feb 10;25(1):65-69. [doi: 10.1177/0963721415620144]
61. Seitz MP, Minhas S, Khouzam A, Khouzam N, Harper Y. The face is the mirror of the soul. The cardiovascular physical exam is not yet dead!. Curr Probl Cardiol 2021 Mar;46(3):100644. [doi: 10.1016/j.cpcardiol.2020.100644] [Medline: 32600656]
62. Stephen ID, Hiew V, Coetzee V, Tiddeman BP, Perrett DI. Facial shape analysis identifies valid cues to aspects of physiological health in Caucasian, Asian, and African populations. Front Psychol 2017;8:1883 [FREE Full text] [doi: 10.3389/fpsyg.2017.01883] [Medline: 29163270]
63. Breitbarth A, Schardt T, Kind C, Brinkmann J, Dittrich P, Notni G. Measurement accuracy and dependence on external influences of the iPhone X TrueDepth sensor. In: Proceedings of the Volume 11144, Photonics and Education in Measurement Science 2019. 2019 Presented at: Proceedings Volume 11144, Photonics and Education in Measurement Science 2019; 2019; Jena, Germany. [doi: 10.1117/12.2530544]
64. Face ID and your Apps. Apple. URL: https://developer.apple.com/videos/play/tutorials/face-id-and-your-apps/ [accessed 2022-01-20]
65. Face tracking with ARKit. Apple Developer Documentation. URL: https://developer.apple.com/videos/play/tech-talks/601/ [accessed 2022-01-20]
66. LeCompte MC, Chung SA, McKee MM, Marshall TG, Frizzell B, Parker M, et al. Simple and rapid creation of customized 3-dimensional printed bolus using iPhone X true depth camera. Pract Radiat Oncol 2019;9(4):e417-e421. [doi: 10.1016/j.prro.2019.03.005] [Medline: 30926481]

67. Park H, Lee J, Ko J. Achieving real-time sign language translation using a smartphone's true depth images. In: Proceedings of the 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS). 2020 Presented at: 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS); Jan 07-11, 2020; Bengaluru, India. [doi: 10.1109/comsnets48256.2020.9027420]
68. Scherr S, Kammler C, Elberzhager F. Detecting user emotions with the true-depth camera to support mobile app quality assurance. In: Proceedings of the 2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). 2019 Presented at: 2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA); Aug 28-30, 2019; Kallithea, Greece. [doi: 10.1109/seaa.2019.00034]
69. Scherr S, Mennig P, Kammler C, Elberzhager F. On the road to enriching the app improvement process with emotions. In: Proceedings of the 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW). 2019 Presented at: 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW); Sep 23-27, 2019; Jeju, Korea (South). [doi: 10.1109/rew.2019.00018]
70. Sini F, Bruno G, Chiabert P, Segonds F. A lean quality control approach for additive manufacturing. In: Product Lifecycle Management Enabling Smart X. Cham: Springer; 2020.
71. Alfaro-Santafé J, Gómez-Bernal A, Lanuza-Cerzócimo C, Alfaro-Santafé JV, Pérez-Morcillo A, Almenar-Arasanz AJ. Three-axis measurements with a novel system for 3D plantar foot scanning: iPhone X. Footwear Sci 2020 Mar 05;12(2):123-131. [doi: 10.1080/19424280.2020.1734867]
72. Agbolade O, Nazri A, Yaakob R, Ghani AA, Cheah YK. Morphometric approach to 3D soft-tissue craniofacial analysis and classification of ethnicity, sex, and age. PLoS One 2020;15(4):e0228402 [FREE Full text] [doi: 10.1371/journal.pone.0228402] [Medline: 32271782]
73. Al-Taee R, Al-Saedi AI, Nahidh M. Does body mass index affect soft tissue facial anthropometric measurements? Int J Morphol 2021 Apr;39(2):520-526. [doi: 10.4067/s0717-95022021000200520]
74. Hwang Y, Lee YH, Cho DH, Kim M, Lee D, Cho HJ. Applicability of the masseter muscle as a nutritional biomarker. Medicine (Baltimore) 2020 Feb;99(6):e19069 [FREE Full text] [doi: 10.1097/MD.0000000000019069] [Medline: 32028430]
75. Katsuki M, Suzuki Y, Kunitoki K, Sato Y, Sasaki K, Mashiyama S, et al. Temporal muscle as an indicator of sarcopenia is independently associated with hunt and Kosnik grade on admission and the modified Rankin scale score at 6 months of patients with subarachnoid hemorrhage treated by endovascular coiling. World Neurosurg 2020 May;137:e526-e534. [doi: 10.1016/j.wneu.2020.02.033] [Medline: 32061954]
76. Leitner J, Pelster S, Schöpf V, Berghoff AS, Woitek R, Asenbaum U, et al. High correlation of temporal muscle thickness with lumbar skeletal muscle cross-sectional area in patients with brain metastases. PLoS One 2018;13(11):e0207849 [FREE Full text] [doi: 10.1371/journal.pone.0207849] [Medline: 30496307]
77. Murakami M, Iijima K, Watanabe Y, Tanaka T, Iwasa Y, Edahiro A, et al. Development of a simple method to measure masseter muscle mass. Gerodontology 2020 Dec;37(4):383-388. [doi: 10.1111/ger.12487] [Medline: 32662134]
78. Ranganathan K, Terjimanian M, Lisiecki J, Rinkinen J, Mukkamala A, Brownley C, et al. Temporalis muscle morphomics: the psoas of the craniofacial skeleton. J Surg Res 2014 Jan;186(1):246-252. [doi: 10.1016/j.jss.2013.07.059] [Medline: 24079810]
79. Umeki K, Watanabe Y, Hirano H, Edahiro A, Ohara Y, Yoshida H, et al. The relationship between masseter muscle thickness and appendicular skeletal muscle mass in Japanese community-dwelling elders: a cross-sectional study. Arch Gerontol Geriatr 2018;78:18-22. [doi: 10.1016/j.archger.2018.05.014] [Medline: 29883805]
80. Wallace JD, Calvo RY, Lewis PR, Brill JB, Shackford SR, Sise MJ, et al. Sarcopenia as a predictor of mortality in elderly blunt trauma patients: comparing the masseter to the psoas using computed tomography. J Trauma Acute Care Surg 2017 Jan;82(1):65-72. [doi: 10.1097/TA.0000000000001297] [Medline: 27820555]
81. Yamaguchi K, Tohara H, Hara K, Nakane A, Yoshimi K, Nakagawa K, et al. Factors associated with masseter muscle quality assessed from ultrasonography in community-dwelling elderly individuals: a cross-sectional study. Arch Gerontol Geriatr 2019;82:128-132. [doi: 10.1016/j.archger.2019.02.003] [Medline: 30780049]
82. yeşil çınkır H. The prognostic effects of temporal muscle thickness and inflammatory-nutritional parameters on survival in lung cancer patients with brain metastasis. TJ Oncol 2020:119-126. [doi: 10.5505/tjo.2019.2030]
83. Steindl A, Leitner J, Schwarz M, Nenning K, Asenbaum U, Mayer S, et al. Sarcopenia in neurological patients: standard values for temporal muscle thickness and muscle strength evaluation. J Clin Med 2020 Apr 28;9(5):1272 [FREE Full text] [doi: 10.3390/jcm9051272] [Medline: 32354003]
84. Coetzee V, Chen J, Perrett DI, Stephen ID. Deciphering faces: quantifiable visual cues to weight. Perception 2010;39(1):51-61. [doi: 10.1068/p6560] [Medline: 20301846]
85. de Jager S, Coetzee N, Coetzee V. Facial adiposity, attractiveness, and health: a review. Front Psychol 2018;9:2562 [FREE Full text] [doi: 10.3389/fpsyg.2018.02562] [Medline: 30622491]
86. Erten O, Yılmaz BN. Three-dimensional imaging in orthodontics. Turk J Orthod 2018 Sep;31(3):86-94 [FREE Full text] [doi: 10.5152/TurkJOrthod.2018.17041] [Medline: 30206567]
87. Kocabey E, Camurcu M, Ofli F, Aytar Y, Marin J, Torralba A, et al. Face-to-BMI: using computer vision to infer body mass index on social media. In: Proceedings of the ICWSM'17. 2017 Presented at: ICWSM'17; 2017; Québec, Canada.

88. Lee BJ, Kim JY. Predicting visceral obesity based on facial characteristics. BMC Complement Altern Med 2014 Jul 16;14(1):248 [FREE Full text] [doi: 10.1186/1472-6882-14-248] [Medline: 25030087]
89. Mayer C, Windhager S, Schaefer K, Mitteroecker P. BMI and WHR are reflected in female facial shape and texture: a geometric morphometric image analysis. PLoS ONE 2017 Jan 4;12(1):e0169336 [FREE Full text] [doi: 10.1371/journal.pone.0169336] [Medline: 28052103]
90. Ulrich L, Dugelay J, Vezzetti E, Moos S, Marcolin F. Perspective morphometric criteria for facial beauty and proportion assessment. Applied Sci 2019 Dec 18;10(1):8. [doi: 10.3390/app10010008]
91. Coetzee V, Re D, Perrett DI, Tiddeman BP, Xiao D. Judging the health and attractiveness of female faces: is the most attractive level of facial adiposity also considered the healthiest? Body Image 2011 Mar;8(2):190-193. [doi: 10.1016/j.bodyim.2010.12.003] [Medline: 21354874]
92. Barr M, Guo G, Colby S, Olfert M. Detecting body mass index from a facial photograph in lifestyle intervention. Technologies 2018 Aug 31;6(3):83. [doi: 10.3390/technologies6030083]
93. Haritosh A, Gupta A, Chahal E, Misra A, Chandra S. A novel method to estimate Height, Weight and Body Mass Index from face images. In: Proceedings of the 2019 Twelfth International Conference on Contemporary Computing (IC3). 2019 Presented at: 2019 Twelfth International Conference on Contemporary Computing (IC3); Aug 08-10, 2019; Noida, India. [doi: 10.1109/ic3.2019.8844872]
94. Jiang M, Shang Y, Guo G. On visual BMI analysis from facial images. Image Vision Comput 2019 Sep;89:183-196. [doi: 10.1016/j.imavis.2019.07.003]
95. Lee BJ, Do J, Kim JY. A classification method of normal and overweight females based on facial features for automated medical applications. J Biomed Biotechnol 2012;2012:834578 [FREE Full text] [doi: 10.1155/2012/834578] [Medline: 22919277]
96. Othman SA, Majawit LP, Wan Hassan WN, Wey MC, Mohd Razi R. Anthropometric study of three-dimensional facial morphology in Malay adults. PLoS One 2016;11(10):e0164180 [FREE Full text] [doi: 10.1371/journal.pone.0164180] [Medline: 27706220]
97. Pham DD, Do J, Ku B, Lee HJ, Kim H, Kim JY. Body mass index and facial cues in sasang typology for young and elderly persons. Evid Based Complement Alternat Med 2011;2011:749209 [FREE Full text] [doi: 10.1155/2011/749209] [Medline: 21423637]
98. Wen L, Guo G. A computational approach to body mass index prediction from face images. Image Vision Comput 2013 May;31(5):392-400. [doi: 10.1016/j.imavis.2013.03.001]
99. Wolffhechel K, Hahn AC, Jarmer H, Fisher CI, Jones BC, DeBruine LM. Testing the utility of a data-driven approach for assessing BMI from face images. PLoS One 2015;10(10):e0140347 [FREE Full text] [doi: 10.1371/journal.pone.0140347] [Medline: 26460526]
100. Pascali MA, Giorgi D, Bastiani L, Buzzigoli E, Henriquez P, Matuszewski BJ, et al. Face morphology: can it tell us something about body weight and fat? Comput Biol Med 2016 Sep 01;76:238-249. [doi: 10.1016/j.compbiomed.2016.06.006] [Medline: 27504744]
101. Peters F, Kroh A, Neumann UP, Ulmer FT, Möhlhenrich SC, Bock A, et al. Morphological changes of the human face after massive weight-loss due to bariatric surgery. J Craniomaxillofac Surg 2020 Jul;48(7):694-699. [doi: 10.1016/j.jcms.2020.05.010] [Medline: 32513431]
102. Sforza C, Dellavia C, De Menezes M, Rosati R, Ferrario V. Three-dimensional facial morphometry: from anthropometry to digital morphology. In: Handbook of Anthropometry. New York: Springer; 2012.
103. Giorgi D, Pascali M, Raccichini G, Colantonio S, Salvetti O. Morphological analysis of 3D faces for weight gain assessment. In: Proceedings of the 2015 Eurographics Workshop on 3D Object Retrieval. 2015 Presented at: 3DOR '15: Proceedings of the 2015 Eurographics Workshop on 3D Object Retrieval; May 2 - 3, 2015; Zurich Switzerland.
104. Fischer M, JeVenn A, Hipskind P. Evaluation of muscle and fat loss as diagnostic criteria for malnutrition. Nutr Clin Pract 2015 Apr;30(2):239-248. [doi: 10.1177/0884533615573053] [Medline: 25753808]
105. Szczerkowska-Dobosz A, Olszewska B, Lemańska M, Purzycka-Bohdan D, Nowicki R. Acquired facial lipoatrophy: pathogenesis and therapeutic options. Postepy Dermatol Alergol 2015 Apr;32(2):127-133 [FREE Full text] [doi: 10.5114/pdia.2014.40971] [Medline: 26015783]
106. Ko AC, Korn BS, Kikkawa DO. The aging face. Surv Ophthalmol 2017;62(2):190-202. [doi: 10.1016/j.survophthal.2016.09.002] [Medline: 27693312]
107. von Arx T, Nakashima MJ, Lozanoff S. The face – a musculoskeletal perspective. A literature review. Swiss Dent J 2018 Sep 10;128(9):678-688 [FREE Full text] [Medline: 30056693]
108. White JV, Guenter P, Jensen G, Malone A, Schofield M. Consensus statement: academy of nutrition and dietetics and American society for parenteral and enteral nutrition. JPEN J Parenter Enteral Nutr 2012 Apr 24;36(3):275-283. [doi: 10.1177/0148607112440285] [Medline: 22535923]
109. Coleman S, Saboeiro A, Sengelmann R. A comparison of lipoatrophy and aging: volume deficits in the face. Aesthetic Plast Surg 2009 Jan;33(1):14-21. [doi: 10.1007/s00266-008-9258-z] [Medline: 18987910]
110. Bentov I, Kaplan SJ, Pham TN, Reed MJ. Frailty assessment: from clinical to radiological tools. Br J Anaesth 2019 Jul;123(1):37-50 [FREE Full text] [doi: 10.1016/j.bja.2019.03.034] [Medline: 31056240]

111. Chung SJ, Kim YJ, Yoo HS, Jung JH, Baik K, Lee HS, et al. Temporalis muscle thickness as an indicator of sarcopenia is associated with long-term motor outcomes in Parkinson's disease. J Gerontol A Biol Sci Med Sci 2021 Nov 15;76(12):2242-2248. [doi: 10.1093/gerona/glab082] [Medline: 33754634]
112. Liu F, Xing D, Zha Y, Wang L, Dong W, Li L, et al. Predictive value of temporal muscle thickness measurements on cranial magnetic resonance images in the prognosis of patients with primary glioblastoma. Front Neurol 2020;11:523292 [FREE Full text] [doi: 10.3389/fneur.2020.523292] [Medline: 33304303]
113. Jindal K, Sarcia M, Codner MA. Functional considerations in aesthetic eyelid surgery. Plast Reconstr Surg 2014 Dec;134(6):1154-1170. [doi: 10.1097/PRS.0000000000000748] [Medline: 25415085]
114. Roh MR, Chung KY. Infraorbital dark circles: definition, causes, and treatment options. Dermatol Surg 2009 Aug;35(8):1163-1171. [doi: 10.1111/j.1524-4725.2009.01213.x] [Medline: 19469797]
115. Swift A, Liew S, Weinkle S, Garcia JK, Silberberg MB. The facial aging process from the "Inside out". Aesthet Surg J 2021 Sep 14;41(10):1107-1119 [FREE Full text] [doi: 10.1093/asj/sjaa339] [Medline: 33325497]
116. Wan D, Amirlak B, Rohrich R, Davis K. The clinical importance of the fat compartments in midfacial aging. Plast Reconstr Surg Glob Open 2013 Dec;1(9):e92 [FREE Full text] [doi: 10.1097/GOX.0000000000000035] [Medline: 25289286]
117. Lambros V. Observations on periorbital and midface aging. Plast Reconstr Surg 2007 Oct;120(5):1367-1376. [doi: 10.1097/01.prs.0000279348.09156.c3] [Medline: 17898614]
118. Almukhtar RM, Fabi SG. The masseter muscle and its role in facial contouring, aging, and quality of life: a literature review. Plast Reconstr Surg 2019 Jan;143(1):39e-348. [doi: 10.1097/PRS.0000000000005083] [Medline: 30303926]
119. Castelo PM, Bonjardim LR, Pereira LJ, Gavião MB. Facial dimensions, bite force and masticatory muscle thickness in preschool children with functional posterior crossbite. Braz Oral Res 2008;22(1):48-54 [FREE Full text] [doi: 10.1590/s1806-83242008000100009] [Medline: 18425245]
120. Eckardt L, Harzer W. Facial structure and functional findings in patients with progressive muscular dystrophy (Duchenne). Am J Orthod Dentofacial Orthop 1996 Aug;110(2):185-190. [doi: 10.1016/s0889-5406(96)70107-5] [Medline: 8760845]
121. Houston K, Buschang PH, Iannaccone ST, Seale NS. Craniofacial morphology of spinal muscular atrophy. Pediatr Res 1994 Aug;36(2):265-269. [doi: 10.1203/00006450-199408000-00020] [Medline: 7970943]
122. Kiliaridis S. Masticatory muscle influence on craniofacial growth. Acta Odontol Scand 1995 Jun;53(3):196-202. [doi: 10.3109/00016359509005972] [Medline: 7572097]
123. Kubota M, Nakano H, Sanjo I, Satoh K, Sanjo T, Kamegai T, et al. Maxillofacial morphology and masseter muscle thickness in adults. Eur J Orthod 1998 Oct;20(5):535-542. [doi: 10.1093/ejo/20.5.535] [Medline: 9825556]
124. Pereira LJ, Gavião MB, Bonjardim LR, Castelo PM, van der Bilt A. Muscle thickness, bite force, and craniofacial dimensions in adolescents with signs and symptoms of temporomandibular dysfunction. Eur J Orthod 2007 Feb;29(1):72-78. [doi: 10.1093/ejo/cjl055] [Medline: 17290018]
125. Raadsheer MC, Kiliaridis S, Van Eijden TM, Van Ginkel FC, Prahl-Andersen B. Masseter muscle thickness in growing individuals and its relation to facial morphology. Arch Oral Biol 1996 Apr;41(4):323-332. [doi: 10.1016/0003-9969(95)00136-0] [Medline: 8771323]
126. Reis Durão AP, Morosolli A, Brown J, Jacobs R. Masseter muscle measurement performed by ultrasound: a systematic review. Dentomaxillofac Radiol 2017 Aug;46(6):20170052 [FREE Full text] [doi: 10.1259/dmfr.20170052] [Medline: 28467130]
127. Satiroğlu F, Arun T, Işik F. Comparative data on facial morphology and muscle thickness using ultrasonography. Eur J Orthod 2005 Dec;27(6):562-567. [doi: 10.1093/ejo/cji052] [Medline: 16135538]
128. Soyoye OA, Otuyemi OD, Kolawole KA, Ayoola OO. Relationship between masseter muscle thickness and maxillofacial morphology in pre-orthodontic treatment patients. Int Orthod 2018 Dec;16(4):698-711. [doi: 10.1016/j.ortho.2018.09.015] [Medline: 30348502]
129. Yamaguchi K, Tohara H, Hara K, Nakane A, Kajisa E, Yoshimi K, et al. Relationship of aging, skeletal muscle mass, and tooth loss with masseter muscle thickness. BMC Geriatr 2018 Mar 08;18(1):67 [FREE Full text] [doi: 10.1186/s12877-018-0753-z] [Medline: 29519234]
130. Gaszynska E, Godala M, Szatko F, Gaszynski T. Masseter muscle tension, chewing ability, and selected parameters of physical fitness in elderly care home residents in Lodz, Poland. Clin Interv Aging 2014;9:1197-1203 [FREE Full text] [doi: 10.2147/CIA.S66672] [Medline: 25092969]
131. Tanabe C, Reed MJ, Pham TN, Penn K, Bentov I, Kaplan SJ. Association of brain atrophy and masseter sarcopenia with 1-year mortality in older trauma patients. JAMA Surg 2019 Aug 01;154(8):716-723 [FREE Full text] [doi: 10.1001/jamasurg.2019.0988] [Medline: 31066880]
132. Uhlich R, Hu P. Sarcopenia diagnosed using masseter muscle area predictive of early mortality following severe traumatic brain injury. Neural Regen Res 2018 Dec;13(12):2089-2090 [FREE Full text] [doi: 10.4103/1673-5374.241451] [Medline: 30323130]
133. Alabdullah M, Saltaji H, Abou-Hamed H, Youssef M. Association between facial growth pattern and facial muscle activity: a prospective cross-sectional study. Int Orthod 2015 Jun;13(2):181-194. [doi: 10.1016/j.ortho.2015.03.011] [Medline: 25986702]

134. Schimmel M, Katsoulis J, Genton L, Müller F. Masticatory function and nutrition in old age. Swiss Dent J 2015;125(4):449-454 [FREE Full text] [Medline: 26169366]
135. Coleman SR, Grover R. The anatomy of the aging face: volume loss and changes in 3-dimensional topography. Aesthet Surg J 2006;26(1S):S4-S9. [doi: 10.1016/j.asj.2005.09.012] [Medline: 19338976]
136. Ferrario VF, Dellavia C, Tartaglia GM, Turci M, Sforza C. Soft tissue facial morphology in obese adolescents: a three-dimensional noninvasive assessment. Angle Orthod 2004 Feb;74(1):37-42 [FREE Full text] [doi: 10.1043/0003-3219(2004)074<0037:STFMIO>2.0.CO;2] [Medline: 15038489]
137. Fedok FG. The rejuvenation of the aged central lower face: a contemporary perspective. Facial Plast Surg 2019 Apr;35(2):121-128. [doi: 10.1055/s-0039-1683857] [Medline: 30943555]
138. Yutskovskaya YA, Sergeeva AD, Kislitsyna A, Landau M. Contouring of lower face and chin in consideration of facial morphotypes and shapes- is it a more accurate approach? Madridge J Dermatol Res 2017 Dec 29;2(1):26-31. [doi: 10.18689/mjdr-1000107]
139. Donofrio LM. Fat distribution: a morphologic study of the aging face. Dermatol Surg 2000 Dec;26(12):1107-1112. [Medline: 11134986]
140. Wysong A, Joseph T, Kim D, Tang JY, Gladstone HB. Quantifying soft tissue loss in facial aging: a study in women using magnetic resonance imaging. Dermatol Surg 2013 Dec;39(12):1895-1902. [doi: 10.1111/dsu.12362] [Medline: 24238002]
141. Shridharani SM. Novel surface anatomic landmarks of the Jowl to guide treatment with ATX-101. Plast Reconstr Surg Glob Open 2019 Oct;7(10):e2459 [FREE Full text] [doi: 10.1097/GOX.0000000000002459] [Medline: 31772888]
142. Hebert S, Willoughby C, Andrade F, McLoon L. Masticatory muscle response to neuromuscular diseases and specific pathologies. In: Craniofacial Muscles. New York, NY: Springer; 2012.
143. Wen YF, Wong HM, Lin R, Yin G, McGrath C. Inter-ethnic/racial facial variations: a systematic review and Bayesian meta-analysis of photogrammetric studies. PLoS One 2015;10(8):e0134525 [FREE Full text] [doi: 10.1371/journal.pone.0134525] [Medline: 26247212]
144. Kau CH, Wang J, Davis M. A cross-sectional study to understand 3d facial differences in a population of African Americans and Caucasians. Eur J Dent 2019 Oct;13(4):485-496 [FREE Full text] [doi: 10.1055/s-0039-3400551] [Medline: 31891965]
145. Bouguila J, Khochtali H. Facial plastic surgery and face recognition algorithms: interaction and challenges. A scoping review and future directions. J Stomatol Oral Maxillofac Surg 2020 Dec;121(6):696-703. [doi: 10.1016/j.jormas.2020.06.007] [Medline: 32574869]
146. Hong YA, Zhou Z, Fang Y, Shi L. The digital divide and health disparities in China: evidence from a national survey and policy implications. J Med Internet Res 2017 Sep 11;19(9):e317 [FREE Full text] [doi: 10.2196/jmir.7786] [Medline: 28893724]

Abbreviations

CT: computed tomography
MRI: magnetic resonance imaging
SMM: skeletal muscle mass

Edited by Y Khader; submitted 16.09.21; peer-reviewed by S Soofi, YL Leung; accepted 03.03.22; published 18.07.22.

Please cite as:
Tay W, Quek R, Kaur B, Lim J, Henry CJ
Use of Facial Morphology to Determine Nutritional Status in Older Adults: Opportunities and Challenges
JMIR Public Health Surveill 2022;8(7):e33478
URL: https://publichealth.jmir.org/2022/7/e33478
doi:10.2196/33478
PMID:35849429

Review

Assessing the Implementation of Digital Innovations in Response to the COVID-19 Pandemic to Address Key Public Health Functions: Scoping Review of Academic and Nonacademic Literature

Joseph Francombe[1], BA, MPhil, PhD; Gemma-Claire Ali[1], MA, MSc; Emily Ryen Gloinson[1], BA, MA; Carolina Feijao[1], BSc, MSc, PhD; Katherine I Morley[1], BA, BSc, PhD; Salil Gunashekar[1], BSc, MSc, PhD; Helena de Carvalho Gomes[2], MPH, MD

[1]RAND Europe, Cambridge, United Kingdom

[2]European Centre for Disease Prevention and Control, Stockholm, Sweden

Corresponding Author:
Helena de Carvalho Gomes, MPH, MD
European Centre for Disease Prevention and Control
Gustav III:s Boulevard 40
Solna
Stockholm, 169 73
Sweden
Phone: 46 858601000
Email: helena.gomes@ecdc.europa.eu

Abstract

Background: Digital technologies have been central to efforts to respond to the COVID-19 pandemic. In this context, a range of literature has reported on developments regarding the implementation of new digital technologies for COVID-19–related surveillance, prevention, and control.

Objective: In this study, scoping reviews of academic and nonacademic literature were undertaken to obtain an overview of the evidence regarding digital innovations implemented to address key public health functions in the context of the COVID-19 pandemic. This study aimed to expand on the work of existing reviews by drawing on additional data sources (including nonacademic sources) by considering literature published over a longer time frame and analyzing data in terms of the number of unique *digital innovations*.

Methods: We conducted a scoping review of the academic literature published between January 1, 2020, and September 15, 2020, supplemented by a further scoping review of selected nonacademic literature published between January 1, 2020, and October 13, 2020. Both reviews followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach.

Results: A total of 226 academic articles and 406 nonacademic articles were included. The included articles provided evidence of 561 (academic literature) and 497 (nonacademic literature) unique digital innovations. The most common implementation settings for digital innovations were the United States, China, India, and the United Kingdom. Technologies most commonly used by digital innovations were those belonging to the high-level technology group of *integrated and ubiquitous fixed and mobile networks*. The key public health functions most commonly addressed by digital innovations were communication and collaboration and surveillance and monitoring.

Conclusions: Digital innovations implemented in response to the COVID-19 pandemic have been wide ranging in terms of their implementation settings, the digital technologies used, and the public health functions addressed. However, evidence gathered through this study also points to a range of barriers that have affected the successful implementation of digital technologies for public health functions. It is also evident that many digital innovations implemented in response to the COVID-19 pandemic are yet to be formally evaluated or assessed.

(JMIR Public Health Surveill 2022;8(7):e34605) doi:10.2196/34605

KEYWORDS

digital technologies; COVID-19; key public health functions; scoping review; digital health; pandemic; surveillance; mobile phone

Introduction

Background

Digital technologies, such as artificial intelligence (AI), robotics, and wearables, have been widely used in worldwide efforts to respond to the COVID-19 pandemic. In this context, a range of studies has reported on developments regarding the implementation of new digital technologies for COVID-19–related surveillance, prevention, and control. To consolidate this literature, several reviews have been undertaken [1-4]. Broadly, the aim of these reviews has been to describe the characteristics of digital technologies that have been reported on within the early scientific literature. Golinelli et al [2] searched MEDLINE and medRxiv to identify the relevant literature on the use of digital technologies in health care during the COVID-19 pandemic. The included papers were then analyzed in terms of article characteristics and the type of technology and *patient needs addressed*. A review conducted by Budd et al [1] provided a qualitative overview of the breadth of digital innovations introduced as part of the global public health response to the COVID-19 pandemic, the types of public health activities they addressed, and the key potential barriers to their implementation. Vargo et al [4] further reviewed digital technology use during the COVID-19 pandemic based on searches of 4 databases: Web of Science, Scopus, PubMed, and Google Scholar. The review synthesized the evidence from included papers in relation to 4 key areas of *technologies*, *users*, *activities*, and *effects* within the spheres of *health care*, *education*, *work*, and *daily life* [4]. More recently, Mbunge et al [3] undertook a critical review of emerging technologies for tackling the COVID-19 pandemic, focusing on prevention, surveillance, and containment, based on searches of the following sources: Google Scholar, Scopus, ScienceDirect, PubMed, IEEE Xplore Digital Library, ACM Digital Library, Wiley Library, and SpringerLink [3]. Although providing valuable overviews of the digital response to the COVID-19 pandemic, existing reviews have also been limited by a focus on academic sources, thereby potentially missing developments reported in wider nonacademic literature while also tending to focus on the early period of the pandemic.

Study Aims

In this study, we present the findings of a further scoping review on the implementation of digital technologies in response to the COVID-19 pandemic. The scoping review expands on the work of existing reviews by drawing on additional data sources while also considering literature published over a longer time frame (ie, January to September 2020). Although focusing on academic literature, the scoping review also goes beyond existing reviews by presenting evidence from a complementary review of nonacademic sources, including web-based technology-related news sources and news feeds (covering news articles, press releases, and blogs). The incorporation of wider nonacademic sources into this review allows for the consideration of technological developments in the private or public sector, which are not necessarily oriented toward research publications, thus helping to capture more up-to-date information on the implementation of digital technologies in response to the COVID-19 pandemic.

This scoping review goes beyond existing reviews by using the concept of *digital innovations*. By digital innovations, we refer to the application ≥1 digital technology to address COVID-19–related key public health functions within a single application in a specific context. An example of a digital innovation captured by this scoping review is Austria’s contact-tracing app *Stopp Corona*. The app combines 2 digital technologies of interest in this review—smartphone apps and Bluetooth—into a single digital innovation [5,6]. By analyzing data regarding the *number of implemented digital innovations* and their characteristics, this study goes another step beyond existing reviews, all of which have analyzed digital technology trends by considering the *number of papers* reporting on different technology types and functions [1-4]

The specific research questions addressed by this scoping review were as follows:

1. What are the main characteristics of the literature reporting on digital innovations used in the context of the COVID-19 pandemic?
2. What has been the geographical setting of the digital innovations implemented in response to the COVID-19 pandemic?
3. What types of implemented digital innovations have been discussed in the academic and nonacademic literature in relation to COVID-19–related surveillance, prevention, and control?
4. Which key public health functions have been addressed by the digital innovations implemented in response to the COVID-19 pandemic?

Methods

Overview

The scoping review followed the approach specified in PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [7]. A completed PRISMA-ScR checklist for the review is presented in Multimedia Appendix 1. A study protocol was developed presenting key elements of the proposed approach and methods to be used. The approach comprised 2 parallel methodological approaches—one for the review of academic literature and the other for the review of nonacademic literature.

Search Strategy

For the academic literature search, we developed and ran a search strategy in 2 bibliographic databases—EMBASE and Scopus—using the same strategy for both databases. The search terms used in this study are presented in Multimedia Appendix 2. The search was limited to articles published between January 1 and September 15, 2020 (the date of the search), and included English-language and non–English-language articles. The search strategy drew on a strategy developed for a previous scoping review on digital technologies for infectious disease surveillance, prevention, and control, which was peer-reviewed using the Peer-Review of Electronic Search Strategies approach [8]. In addition to the database searches, we also conducted targeted searches using Google Scholar to identify a small number of additional academic articles where the results of the database search strategies revealed evidence gaps.

For the nonacademic literature search, we ran a search strategy using the news aggregation software Feedly [9]. To conduct the Feedly search, we identified relevant information sources covering digital technological innovation and health innovation based on expert consultation and internal piloting. The search strategy applied to these information sources was broadly aligned with that used for the academic literature search. However, because of the limitations of the Feedly search function, the search string used was shorter and more generic than that used for the academic search. The used search terms are presented in Multimedia Appendix 2. The Feedly search traced backward to capture articles published between January 1 and September 15, 2020 (the date of the initial search), but also captured articles published over a period of 4 weeks from the initial search date (ie, between September 15 and October 13, 2020). To manage the scope of the study, the nonacademic literature search was limited to English-language sources only. As with the academic literature search, we also conducted targeted searches to identify a small number of additional nonacademic articles. In addition to a systematic Google search, targeted searching of nonacademic literature included scraping selected websites to identify relevant articles. The selection of websites focused on addressing gaps in the evidence produced by the Feedly-based searches, specifically the limited number of articles reporting on developments within the European Union or European Economic Area (EU/EEA) region compared with those reporting on technological developments outside the EU or EEA.

Study Selection

Articles captured by both searches were screened against defined inclusion and exclusion criteria to determine their eligibility for the study. Multimedia Appendix 3 presents the used inclusion and exclusion criteria. In the review of academic literature, non-English articles were included in the study selection but only if an English-language abstract or summary was available. Screening was undertaken by 2 study teams, each comprising 2 researchers—one study team for the review of academic literature and one study team for the review of nonacademic literature. Before commencing the study selection, both study teams engaged in pilot screening exercises for 100 articles to ensure consistency in the application of the eligibility criteria. The 2 reviewers discussed the areas of uncertainty or disagreement until full agreement on inclusion or exclusion was reached. To further ensure consistency across the study teams, we held weekly cross-project meetings. During these meetings, any articles for which a reviewer was unsure were marked and discussed with the other study team to determine inclusion or exclusion. A shared log of the inclusion and exclusion decisions was maintained across the 2 study teams.

Data Extraction

We extracted data from the included articles using Microsoft Excel extraction templates: one template for the review of academic literature and one for the review of nonacademic literature. Both extraction templates included columns to capture information relating to the core research questions regarding the types and nature of implemented digital innovations, as well as broader information regarding the article type and identified barriers to implementing digital innovations in the discussed countries and regions. Where possible, drop-down menus were used to limit the range of responses that could be submitted, thereby facilitating data filtering and analysis. To ensure a consistent extraction approach, the 2 project teams conducted pilot extraction exercises using a small number of articles. The used extraction templates and drop-down menus are presented in Multimedia Appendix 4.

Data Analysis

To analyze the extracted data, we used the software package R. Descriptive quantitative analysis focused on statistical and graphical summaries for each column of data captured using drop-down menus in the extraction template, together with relevant cross-analyses. Data extracted on barriers to the implementation of digital innovations were analyzed qualitatively.

Key Study Variables

High-Level Technology Groups and Specific Digital Technologies

In extracting data on digital innovations, we coded data on the specific digital technologies that have been used within these innovations, as well as the technology groups to which these technologies belong. The used coding approach drew on an earlier scoping review of the use of digital technologies for the prevention, surveillance, and control of infectious diseases. In this study, specific technologies identified in the literature were clustered into high-level technology groups of similar or conceptually related digital technologies [10]. For this study, definitions for each specific digital technology and each high-level technology group were established using the European Commission’s Digital Single Market glossary, supplemented, where necessary, by definitions from relevant academic literature [11]. The coding approach is presented in Textbox 1.

Textbox 1. Coding of specific digital technologies into high-level technology groups.

High-level technology group and specific digital technology

- Advanced manufacturing technologies
  - 3D printing
- Autonomous devices and systems
  - Drones
  - Robotics
- Blockchain or distributed ledger technology
  - Blockchain or distributed ledger technology
- Cloud computing or cloud-based networks
  - Cloud computing or cloud-based networks
- Cognitive technologies
  - Artificial intelligence
  - Expert systems
  - Machine learning
  - Natural language processing
  - Facial recognition
  - (Artificial) neural networks
- Crowdsourcing platforms
  - Crowdsourcing
- Data analytics (including big data)
  - Data mining
  - Data analytics
  - Big data
  - Health informatics
  - Parallel computing
  - Social media and mobile data analysis
- eHealth
  - Digital health, eHealth, and mobile health
  - Electronic health records
  - Telemedicine
- Imaging and sensing technologies (including Geographic Information System)
  - Geographic Information System
  - Image processing
  - Infrared sensing
  - Satellite communication or imaging (including earth observation and remote sensing)
- Immersive technologies
  - Virtual or augmented reality
- Integrated and ubiquitous fixed and mobile networks
  - Cellular networks

  - SMS text message communications
  - Smartphone apps
  - Bluetooth
  - Smartphones and tablet computing devices

- Web-based tools and platforms
  - Web-based learning platforms
  - Web-based self-assessment tools
  - Information management tools

- Social media
  - Microblogging platforms
  - Blogging platforms
  - Instant messaging platforms
  - Networking platforms
  - Photograph or video-sharing platforms

- Internet of Things
  - Internet of Things
  - Wireless sensor networks
  - Biosensors

- Nanotechnology and microsystems
  - Digital DNA, RNA or protein analysis
  - Lab-on-chip
  - Nanotechnology

- Quantum computing
  - Quantum computing

- Simulation
  - Mathematical models or simulations

- Wearables (including ingestibles)
  - Wearables (including smart fabrics and ingestibles)

Key Public Health Functions

We also coded each digital innovation as fulfilling ≥1 of the following seven key public health functions: (1) screening and diagnostics, (2) surveillance and monitoring, (3) contact tracing, (4) forecasting, (5) signal or outbreak detection and validation, (6) pandemic response, and (7) communication and collaboration. The use of these public health key functions followed a mapping of the European Centre for Disease Prevention and Control's priorities against the 10 essential public health operations of the World Health Organization's Regional Office for Europe [12] and the US Center for Disease Control's 10 essential public health services [13]. On the basis of this mapping exercise, the public health key functions used in this study were also refined to suit the COVID-19 context. For example, to better reflect the diverse range of activities undertaken to ensure safe access to and management of essential resources during the COVID-19 pandemic, including at the population level, *pandemic response* was included as a key public health function. Similarly, to reflect its centrality in response to the pandemic, contact tracing was also included as a distinct key public health function. Textbox 2 presents the high-level definitions of these key public health functions for the purposes of this study. The key public health functions used in this study were neither exhaustive nor definitive. For example, the used functions do not cover the application of emerging digital technologies to the development of treatments or vaccines. Other studies may adopt alternative approaches to identifying and defining key public health functions.

As with the classification of technologies, our data extraction template included columns to record instances in which a digital innovation addressed >1 key public health function. For example, digital innovations performing surveillance and monitoring functions and signal or outbreak detection and validation were coded with both these public health functions. The coding of key public health functions was based on data extracted from the academic or nonacademic sources being reviewed. The emphasis within the article guided the assessment of how the codes were applied. The collation of data from multiple sources on the same innovations allowed us to capture where digital innovations addressed >2 key public health functions.

Textbox 2. Key public health functions and definitions.

Screening and diagnostics

- Identifying (including self-identifying) COVID-19 symptoms and the presence of SARS-CoV-2 in individuals

Surveillance and monitoring

- Systematic collection and analysis of relevant data such as SARS-CoV-2 infection rates and excess deaths along with ongoing monitoring of COVID-19 symptoms or adherence to COVID-19 restrictions at the individual and population levels

Contact tracing

- Identifying and alerting people who have been in contact with someone diagnosed with COVID-19 and who are therefore at high risk of having been exposed to SARS-CoV-2, so that they can take appropriate and sometimes mandated action (eg, self-isolating)

Forecasting

- Predicting COVID-19 infections or health outcomes at the individual and population levels

Signal or outbreak detection and validation

- Detecting and validating outbreaks of COVID-19

Pandemic response

- Responses to the pandemic that have helped widen safe access to and management of essential resources required by individuals and populations for COVID-19 prevention and response

Communication and collaboration

- Communication involves informing, educating, and empowering individuals and populations about COVID-19, and collaboration refers to working together across disciplines or sectors to share knowledge and improve the collective COVID-19 response

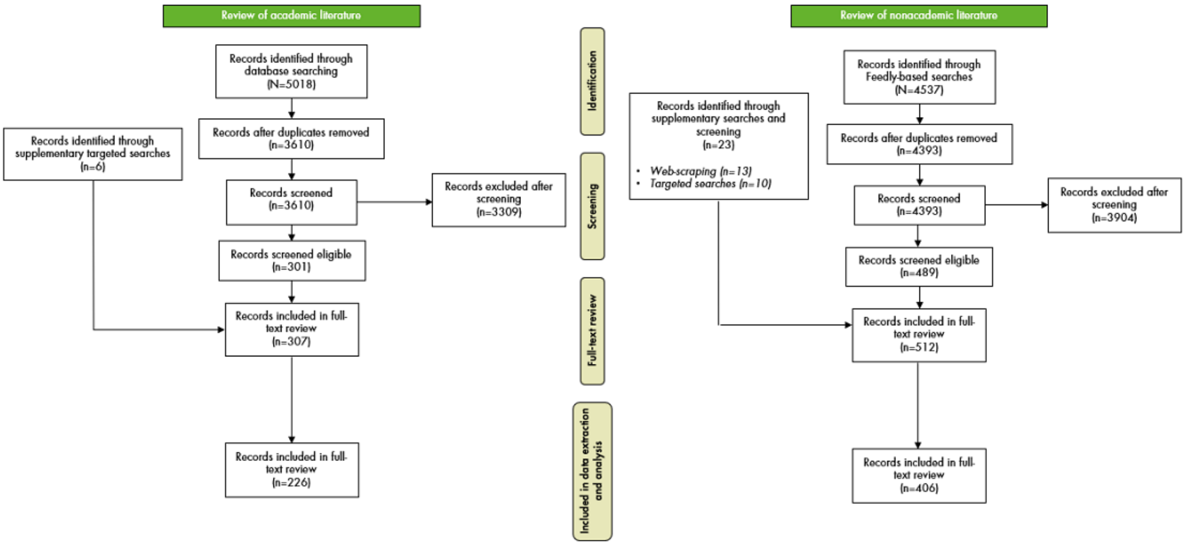
Results

Search Results

The search of academic databases returned a total of 5018 articles, of which 1408 (28.06%) were duplicates. Title and abstract screening of the remaining 3610 articles resulted in 3309 (91.66%) articles being excluded, with 301 (8.34%) deemed eligible for full-text review. Through targeted searches, we also included an additional 6 articles. Of these 307 articles, 81 (26.4%) were excluded during data extraction and analysis, resulting in 226 (73.6%) unique articles being included in the review of the academic literature.

For the review of nonacademic literature, the Feedly-based literature search returned a total of 4537 articles, of which 144 (3.17%) were duplicates. Title screening of the remaining 4393 articles resulted in 3904 (88.87%) articles being excluded, with 489 (11.13%) included for full-text review. We also included an additional 23 articles through targeted searching (n=10, 43% articles) and web scraping (n=13, 57% articles). Of these 512 articles, 107 (20.9%) were excluded during the data extraction and analysis. This resulted in 79.2% (406/512) unique nonacademic articles being included. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagrams for the 2 scoping reviews are presented in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagrams for the review of academic literature (left) and nonacademic literature (right; review time frame for academic literature: January 1, 2020, to September 15, 2020; review time frame for nonacademic literature: January 1, 2020, to October 13, 2020).



Characteristics of the Included Literature

All articles included in the review of academic literature fell into one of the following 3 article types: brief journal comment, editorial, letter, or opinion (98/226, 43.4% of articles), research articles (86/226, 38.1% articles), and detailed journal perspective, policy review, or practice reviews (42/226, 18.6% of articles). The 86 articles categorized as research articles were further analyzed in terms of their study type, with the following results: 67 (78%) primary studies, 2 (2%) mathematical models or simulations, 2 (2%) scoping reviews, and 14 (16%) other literature reviews. The study type of one of the research articles was unclear. All articles included in the review of nonacademic literature were news articles, blog posts, or press releases.

The publication dates of the included articles by month are shown in Table 1. For the review of academic literature, most included articles (203/226, 89.8% of articles) were published in April, May, June, July, and August, with June being the month in which the highest number of included articles (46/226, 20.4% articles) were published. For the review of nonacademic literature, a higher number of included articles (250/406, 61.6%) were published in March, April, and May. For comparison, 34% (138/406) of articles were published in June, July, August, and September. Across both reviews, a small number of articles were published in February, September, and October.

For articles identified by the review of academic literature, we analyzed the geographical location of key contributors' (first, last, and corresponding authors') organizations. The country with the highest number of key contributor organizational affiliations was the United States (64/226, 28.3% of articles). Other countries with a high number of key contributor organizational affiliations included China (23/226, 10.2% of articles), the United Kingdom (20/226, 8.8% of articles), and India (13/226, 5.8% of articles). Table 2 presents the 10 countries with the highest number of key contributor organizational affiliations.

Table 1. Number of articles by publication date (by month) in the review of academic literature and nonacademic literature (review time frame for academic literature: January 1, 2020, to September 15, 2020; review time frame for nonacademic literature: January 1, 2020, to October 13, 2020)[a].

Publication date	Included articles, n (%)
Review of academic literature (N=226)	
January	0 (0)
February	3 (1.3)
March	10 (4.4)
April	38 (16.8)
May	39 (17.3)
June	46 (20.4)
July	39 (17.3)
August	41 (18.1)
September	10 (4.4)
Review of nonacademic literature (N=406)[b]	
January	0 (0)
February	6 (1.5)
March	66 (16.3)
April	108 (26.7)
May	76 (18.8)
June	36 (8.9)
July	31 (7.7)
August	33 (8.1)
September	38 (9.4)
October	11 (2.7)

[a] Due to rounding, the percentages for the nonacademic review do not add up to 100.

[b] For one article within the review of nonacademic literature, no publication date was recorded. Although 406 articles were reviewed, the articles listed against months in the table therefore add up to 405.

Table 2. Top 10 countries in which key contributors' organizations were based in the review of academic literature (review time frame: January 1, 2020, to September 15, 2020; N=226).

Country of key contributors' organization	Included articles, n (%)
United States	64 (28.3)
China (mainland)	23 (10.2)
United Kingdom	20 (8.8)
India	13 (5.8)
Australia	8 (3.5)
France	8 (3.5)
Taiwan	8 (3.5)
Singapore	7 (3.1)
South Korea	7 (3.1)
Spain	6 (2.7)

Number of Digital Innovations Implemented in Response to the COVID-19 Pandemic

Through our review of the academic literature, we identified 561 instances of the implementation of digital innovations to tackle the COVID-19 pandemic. Our review of the nonacademic literature identified 497 digital innovations. The 2 reviews were conducted independently. As such, the digital innovations identified by the review of nonacademic literature are not necessarily presented as unique from those identified by the review of academic literature. Although there is likely to be some crossover between the innovations captured by the 2 reviews, it is also the case that the more experimental review of nonacademic literature has captured some innovations, in particular those developed by private companies, that have not been captured within the review of academic literature, particularly as developments occurred rapidly during the first few months of the pandemic in 2020.

Geographic Context of Digital Innovations Implemented in Response to the COVID-19 Pandemic

The identified digital innovations were analyzed by the geographical context in which they were implemented at both the regional and country levels. In the academic literature, 66.5% (373/561) of digital innovations were implemented in non-EU/EEA countries, with 21.4% (120/561) of digital innovations being implemented within EU/EEA countries. Approximately 12.1% (68/561) of digital innovations identified in our review were implemented worldwide. The countries with the highest number of implemented digital innovations, according to academic literature, were the United States of America (107/561, 19.1% of digital innovations), China (71/561, 12.7% of digital innovations), and India (28/561, 5% of digital innovations). The most common EU or EEA country implementation settings were France (18/561, 3.2% of digital innovations), Spain (18/561, 3.2% of digital innovations), and Italy (12/561, 2.1% of digital innovations; see Table 3).

Table 3. Top 10 countries in which the highest number of digital innovations have been implemented in the review of academic literature and nonacademic literature (review time frame for academic literature: January 1, 2020, to September 15, 2020; review time frame for nonacademic literature: January 1, 2020, to October 13, 2020).

Implementation setting	Digital innovations, n (%)
Review of academic literature (N=561)	
United States	107 (19.1)
China	71 (12.7)
India	28 (5.0)
United Kingdom	25 (4.5)
France	18 (3.2)
Spain	18 (3.2)
South Korea	16 (2.9)
Singapore	13 (2.3)
Canada	12 (2.1)
Italy	12 (2.1)
Review of nonacademic literature (N=497)	
United States	141 (28.4)
United Kingdom	60 (12.1)
China	38 (7.6)
Italy	13 (2.6)
Spain	11 (2.2)
Germany	10 (2)
Singapore	10 (2)
France	8 (1.6)
India	8 (1.6)
Australia, Israel, and The Netherlands	7 (1.4)

Evidence from the nonacademic literature supports this overarching picture, with most of the identified digital innovations implemented in non-EU or EEA countries (320/497, 64.4% of digital innovations) and a smaller number of digital innovations implemented within the EU or EEA (89/497, 17.9% of digital innovations). According to the nonacademic literature, the most common implementation settings were the United States (141/497, 28.4% of digital innovations), the United Kingdom (60/497, 12.1% of digital innovations), and China (38/497, 7.6% of digital innovations), with Italy (13/497, 2.6% of digital innovations), Spain (11/497, 2.2% of digital innovations), Germany (10/497, 2% of innovations), and France (8/497, 1.6% of digital innovations) being the most common EU or EEA country implementation settings (Table 3).

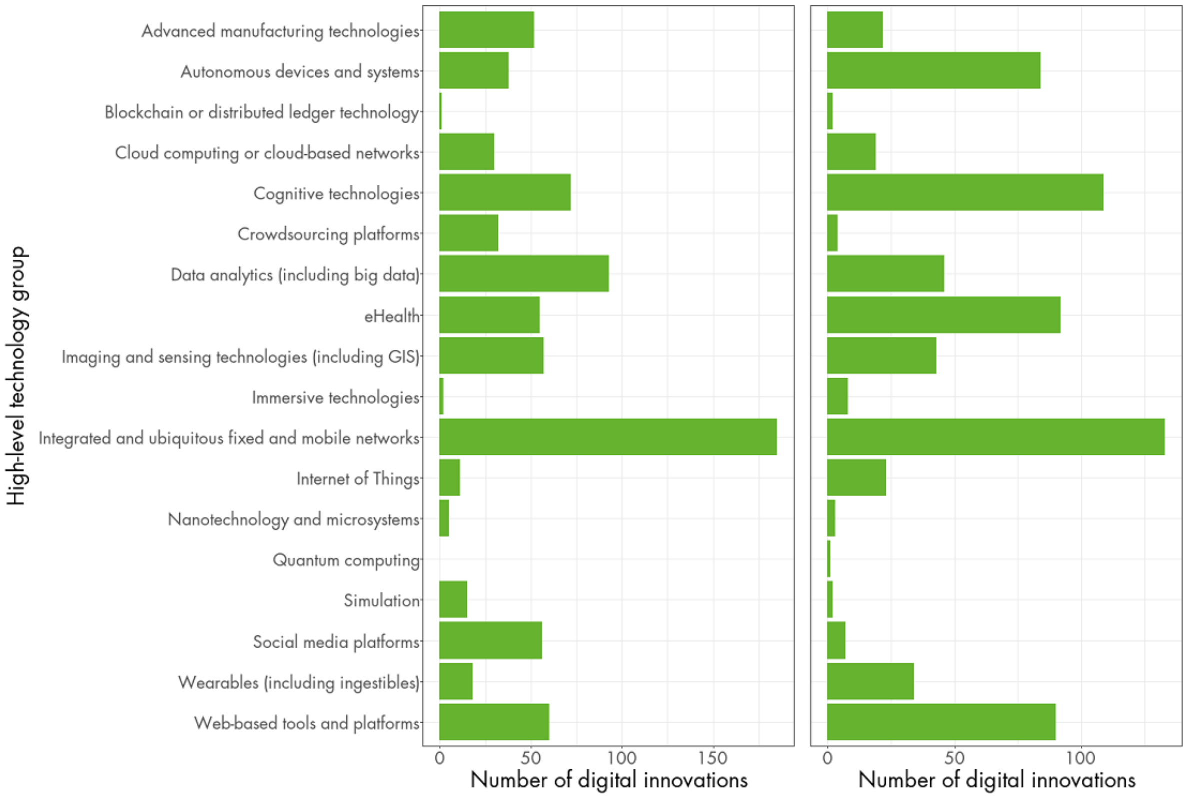
Types of Digital Innovations Implemented in Response to the COVID-19 Pandemic

The digital innovations were also analyzed in terms of the types of digital technology they incorporated. As explained previously, we analyzed innovations in terms of the specific digital technologies they incorporated and the high-level technology groups to which these specific technologies belonged.

In the academic literature, the most commonly implemented high-level technology group was *integrated and ubiquitous fixed and mobile networks* (185/561, 33% of digital innovations incorporated at least one digital technology falling under this high-level group; Figure 2, left frame). Other high-level technology groups with a high number of digital innovations were *data analytics* (93/561, 16.6% of digital innovations), *cognitive technologies* (72/561, 12.8% of digital innovations), and *web-based tools and platforms* (60/561, 10.7% of digital innovations). The most commonly used specific digital technology was smartphone apps (144/561, 25.7% of all digital innovations and 144/185, 77.8% of digital innovations within the *integrated and ubiquitous fixed and mobile networks* high-level technology group).

In the nonacademic literature, *integrated and ubiquitous fixed and mobile networks* were also the most commonly implemented high-level technology group (133/497, 26.8% of digital innovations; Figure 2, right frame). Other high-level technology groups with a high number of digital innovations were *cognitive technologies* (109/497, 21.9% of digital innovations), *eHealth* (92/497, 18.5% of digital innovations), and *web-based tools and platforms* (90/497, 18.1% of digital innovations). The most commonly included specific digital technology, according to the nonacademic literature, was AI (90/497, 18.1% of all digital innovations and 90/109, 82.6% of digital innovations within the *cognitive technologies* high-level technology group).

Figure 2. Number of digital innovations using each high-level technology group in the review of academic literature (left) and nonacademic literature (right; review time frame for academic literature: January 1, 2020, to September 15, 2020; review time frame for nonacademic literature: January 1, 2020, to October 13, 2020). GIS: Geographic Information Systems.



Key Public Health Functions Addressed by Digital Innovations Implemented in Response to the COVID-19 Pandemic

The key public health function addressed by the highest number of digital innovations in the academic literature was communication and collaboration (264/561, 47.1% of digital innovations addressed this public health function; Table 4). Other key public health functions addressed by a large number of digital innovations were surveillance and monitoring (199/561, 35.5% of digital innovations), pandemic response (126/561, 22.5% of digital innovations), and screening and diagnostics (103/561, 18.4% of digital innovations).

Table 4. Number of digital innovations addressing each key public health function in the review of academic literature and nonacademic literature (review time frame for academic literature: January 1, 2020, to September 15, 2020; review time frame for nonacademic literature: January 1, 2020, to October 13, 2020)[a].

Key public health function	Digital innovations, n (%)
Review of academic literature (N=561)	
Communication and collaboration	264 (47.1)
Surveillance and monitoring	199 (35.5)
Pandemic response	126 (22.5)
Screening and diagnostics	103 (18.4)
Contact tracing	77 (13.7)
Forecasting	30 (5.3)
Signal or outbreak detection and validation	8 (1.4)
Review of nonacademic literature (N=497)	
Surveillance and monitoring	197 (39.6)
Pandemic response	169 (34.0)
Screening and diagnostics	167 (33.6)
Communication and collaboration	130 (26.2)
Contact tracing	52 (10.5)
Forecasting	27 (5.4)
Signal or outbreak detection and validation	6 (1.2)

[a] For both the academic and nonacademic review, the number of digital innovations add up to more than the overall sample size (N) and the percentages add up to more than 100. This is because each digital innovation could be assigned more than one key public health function in our review.

The public health function addressed most commonly by digital innovations in the nonacademic literature was surveillance and monitoring (197/497, 39.6% of digital innovations), followed by pandemic response (169/497, 34% of digital innovations) and screening and diagnostics (167/497, 33.6% of digital innovations; Table 4). Compared with the academic literature, the function of communication and collaboration was addressed by a smaller number of innovations (130/497, 26.2% of digital innovations).

Cross-analysis

For each country in which digital innovations were implemented, we cross-analyzed the number of high-level technology groups with which implemented innovations were associated. According to the academic literature, China had implemented digital innovations across the largest number of high-level technology groups (digital innovations were implemented with technologies from 15/17, 88% of the high-level technology groups). This was followed by the United States (14/17, 82% of high-level technology groups) and the United Kingdom (13/17, 76% of high-level technology groups). The EU or EEA countries with digital innovations covering the highest number of technology groups were France (11/17, 65% of high-level technology groups) and Italy (9/17, 53% of high-level technology groups). In the nonacademic literature review, the countries implementing technologies covering the largest number of high-level technology groups were the United States (15/17, 88% of high-level technology groups), the United Kingdom (12/17, 71% of high-level technology groups), China (10/17, 59% of high-level technology groups), and Italy (10/17, 59% of high-level technology groups). Tables presenting a further analysis of the implementation setting and high-level technology groups are presented in Multimedia Appendix 5.

For each key public health function, we cross-analyzed the high-level technology groups with which digital innovations were most associated. For communication and collaboration (the public health function addressed most frequently by digital innovations in the academic literature), most digital innovations incorporated technologies within the following high-level technology groups: *integrated and ubiquitous fixed and mobile networks* (88/561, 15.7% of digital innovations), *social media platforms* (56/561, 10% of digital innovations) *data analytics(including big data*) (42/561, 7.5% of digital innovations), and *web-based tools and platforms* (38/561, 6.8% of digital innovations). For surveillance and monitoring (the public health function addressed most frequently by digital innovations in the nonacademic literature), most digital innovations incorporated technologies within the following high-level technology groups: *integrated and ubiquitous fixed and mobile networks* (60/497, 12.1% of digital innovations), *web-based tools and platforms* (44/497, 8.9% of digital innovations), and *wearables (including ingestibles* (28/497, 5.6% of digital innovations). Tables presenting a further analysis of key public health functions and high-level technology groups are presented in Multimedia Appendix 5.

Discussion

Summary of Key Findings

Following the COVID-19 pandemic, actors worldwide turned to digital technologies to assist the public health response. This study suggests that the most common implementation settings for digital innovations implemented to tackle the COVID-19 pandemic were the United States, the United Kingdom, China, and India. Meanwhile, within the EU/EEA region, Italy, Spain, Germany, and France were the most common implementation settings. The study suggested that a high number of digital innovations implemented in response to the COVID-19 pandemic used technologies within the *integrated and ubiquitous fixed and mobile network* technology group, including cellular networks, smartphone and tablet computing devices, smartphone apps, and Bluetooth. Smartphone apps have been the specific technology most used by digital innovations in response to the COVID-19 pandemic, with uses ranging from contact-tracing apps [5,6,14-23] and self-assessment apps [24] to apps supporting population surveillance and monitoring of regulation compliance [25]. The study also found that *data analytics (including big data)* and *cognitive technologies*, the latter including AI and machine learning, have also been incorporated into many COVID-19–related digital innovations. According to the results of this study, communication and collaboration and surveillance and monitoring have been the public health functions most commonly addressed by digital innovations implemented in response to the COVID-19 pandemic. Both these functions have been addressed by a wide range of technologies, covering nearly all high-level technology groups used in this study. Other functions addressed by a large number of innovations were screening and diagnostics and pandemic response.

Comparison With Other Studies

To the best of our knowledge, this study is the first to provide a broad overview of the geographical context in which digital technologies have been implemented in response to the COVID-19 pandemic. Mbunge et al [3] reviewed the evidence regarding leading countries in the application of AI models for COVID-19, finding that China was the country with the highest frequency in this respect, but did not review the geographic distribution of wider forms of digital innovation [3]. Although most digital innovations identified in our study were implemented in non-EU/EEA countries, it is also worth noting that in this study, EU/EEA countries featured more prominently as implementation settings when than an earlier scoping review that examined digital technologies implemented for infectious disease surveillance, prevention, and control more broadly [10].

This study's findings on the types of digital technologies used by digital innovations have commonality with the results of other reviews on digital technologies and public health in response to the COVID-19 pandemic. In their review, for example, Vargo et al [4] found that the types of technological hardware most reported in relation to the health care sector were computerized tomography machines (in most cases discussed in combination with AI-based learning approaches) and mobile devices, with computers or mobile apps being among the most prominent forms of software used. Similarly, a review of digital technologies in health care conducted by Golinelli et al [2] found that many studies reported on the use of AI tools, big data analytics, mobile apps, and mobile tracing. More broadly, the wider array of digital technologies identified in this scoping review aligns with the digital technologies identified in other reviews. For example, in their review of digital technologies for COVID-19 prevention, surveillance, and containment, Mbunge et al [3] identified the following emerging technologies to be relevant in tackling COVID-19: AI, social media platforms, Internet of Medical Things, virtual or augmented reality, blockchain, additive manufacturing, 5G cellular technology and smart applications, geographic information systems, big data, and autonomous robots.

The findings of this study also illustrate certain differences from other reviews. For example, in the study by Vargo et al [4], video-based communication platforms were found to be a commonly reported on technological software. Meanwhile, Golinelli et al [2] reported a high number of articles reporting on telehealth or telemedicine. The difference between the high reportage of such technologies in other reviews and the relatively lower numbers found in this review may perhaps be explained by the fact that this study included only telemedicine-based innovations when the innovations had been implemented specifically to tackle the COVID-19 pandemic.

The study's finding that digital technologies introduced in response to the COVID-19 pandemic are principally oriented toward 4 public health functions—communication and collaboration, surveillance and monitoring, screening and diagnostics, and pandemic response—is also in line with the findings of other reviews [1,2]. For example, Golinelli et al [2] identified the following 4 key *patient needs* that were addressed by technologies cited within the early scientific literature: *diagnosis*, *surveillance*, *prevention*, and *treatment*. Meanwhile, the review conducted by Budd et al [1] identified 4 overarching public health functions performed by technologies: *digital epidemiological surveillance*, *rapid case identification*, *interrupting community transmission*, and *public communication*. Each of these taxonomies broadly mirrors the functions most commonly addressed in this review, with the exception that, unlike the study by Golinelli et al [2], this study did not consider technologies related to COVID-19 treatment or therapeutics. In the sense that they emphasize the communicative or collaborative function of many COVID-19 digital innovations, the findings of this study are more closely aligned with the review conducted by Budd et al [1]. This emphasis on communication provides support for the broader literature on the role of social media in public health communication during the pandemic [26]. The literature has highlighted the role of social media in facilitating forms of communication such as scientific exchange and the transmission of information from formal public health agencies and other bodies, as well as the potential for such platforms to act as vectors for the spread of misinformation [1,26].

In conducting this scoping review, we encountered evidence of a range of barriers to the successful implementation of digital innovation, including potential risks. In addition to the limitations of the technologies themselves, these include investment and financial barriers, infrastructural barriers (including a lack of required physical and network infrastructure), human resource barriers, data availability and quality barriers, social barriers (including low uptake and low access to technologies), ethical barriers (including privacy concerns and risks of increased socioeconomic inequality), security and safety barriers, and legal or regulatory barriers. In their review, Budd et al [1] identified similar legal, ethical, and privacy barriers, as well as organizational and workforce barriers, to the implementation of technologies for the COVID-19 pandemic. The extent to which these factors present an obstacle to the implementation of technologies depends on the specific contexts (eg, geographical, cultural, political, and economic) within which technologies are developed and implemented. Therefore, the literature suggests that an effective rollout of technologies will require interventions tailored to the specific characteristics of target regions, recognizing both barriers and enablers that may exist [27]. For instance, in regions without the necessary infrastructure to support cellular and data coverage, automated applications that do not require continuous network access may be more appropriate than other applications [27].

Although this study presents evidence regarding the technologies used by digital innovations, the public health functions addressed, and barriers to implementation, it has not systematically examined the performance of individual technologies or the extent to which technologies have been evaluated or comparatively assessed (see the *Limitations* section). The wider literature provides examples of evaluative studies in specific contexts, including statistical evaluations of diagnostic accuracy [28], epidemic modeling [29,30], and qualitative evaluations [31], which demonstrate that the performance of digital technologies implemented in response to the COVID-19 pandemic can vary significantly, depending not only on endogenous technological factors but also on broader exogenous factors, including legal, infrastructural, and social issues [28,29,30,31]. It is also evident that a large number of the technologies introduced in response to the COVID-19 pandemic have not yet been formally evaluated or assessed.

Therefore, the rapid proliferation of digital public health technologies in response to the COVID-19 pandemic has underscored the need for further studies to evaluate the performance of emerging digital technologies, as well as rigorous oversight mechanisms [1,3]. Such approaches should help not only to verify the performance of new technologies but also to identify the underpinning barriers that stand in the way of those technologies realizing their potential. At the same time, oversight mechanisms should also help to strike a balance between the opportunities presented by new innovations and potential risks, such as ethical and privacy risks, that they may pose [32].

Limitations

This study has sought to provide a broad characterization of the evidence regarding the implementation of digital innovations to tackle the COVID-19 pandemic. This study followed a systematic approach in line with the PRISMA checklist for scoping reviews. Drawing on 2 scoping reviews—a review of academic literature, supported by a supplementary, experimental review of nonacademic literature—the review considers evidence from a wide range of sources, from peer-reviewed publications to news articles, press releases, and blogs. Although the methodological approach is well suited to the objectives of the study, it is also subject to several limitations.

The first limitation of the study relates to the scope of information sources. For the review of academic literature, we relied on 2 databases, EMBASE and Scopus, supplemented by structured targeted searches using Google Scholar. Similarly, our review of nonacademic literature was also limited in that it only considered articles published by a selected set of information sources. For example, in focusing on information sources available within Feedly, the nonacademic search strategy did not include national public health institute websites (although our search strategy identified several digital innovations developed and implemented by public health institutes). We cannot rule out the possibility that running the searches in additional databases, including those used by other reviews of digital technology use for the COVID-19 pandemic, might have led us to identify further examples of digital innovations implemented in response to the COVID-19 pandemic.

We adopted broad inclusion criteria during study selection to maximize the scope of the included evidence. However, it was also necessary to limit the review's scope to keep it manageable within the resources available for this study. Our decision to focus on *implemented* digital innovations, thereby excluding innovations still at the conceptual stage, was an example of this. Another was our focus on specific key public health functions, meaning that innovations oriented toward other functions were excluded. In both cases, such decisions led to the inevitable exclusion of digital innovations developed in response to the COVID-19 pandemic.

Another limitation of this scoping review was related to the categories used to code technologies and key public health functions. To extract data, we used drop-down menus to classify digital innovations by technology (specific digital technology and high-level technology groups) and by key public health function. The categories used for the drop-down menus (described earlier as key study variables) were carefully selected after several discussions between team members and drew on earlier studies [10]. Although these categories helped classify and organize the data for the purposes of quantitative analysis, inevitably, there is also an element of subjectivity in the application of these categories to digital innovations. It is also not claimed that these categories are definitive or exhaustive in any way. They represent only one approach to classifying implemented technologies and the role they have performed in supporting public health efforts.

We also faced some technical limitations in analyzing data on the types of nonacademic sources included in the review. Initially, the sources were categorized as news articles, blog posts, or press releases. However, as most sources were news articles, it was decided to merge these 3 categories into one during the analysis stage. The decision also reflected the challenges faced during the export of the included articles into a Microsoft Excel file (a measure taken to mitigate the potential risk of URL changes). With the article content exported to Microsoft Excel, it was not always possible to determine the original format (eg, whether a source was an original news article or an article based on a press release).

Finally, while reviewing evidence on the technologies used by digital innovations, the public health functions addressed, and the key barriers to implementation, a systematic evaluation of the performance of individual technologies and innovations is beyond the scope of this study. The incorporation of an evaluative aspect into the study was not feasible because of the limited amount of information on the performance of digital technologies within the reviewed sources, including the lack of evidence of formal evaluation or assessments undertaken. This study highlights the need for further evaluative studies and oversight mechanisms moving forward.

Conclusions

In this study, scoping reviews of academic and nonacademic sources were used to obtain an overview of the evidence regarding implemented digital innovations to tackle the COVID-19 pandemic. This scoping review sought to gain an understanding of the characteristics of the literature reporting on digital technology use for COVID-19 and an understanding of the number, nature, and geographical distribution of digital innovations implemented during the first 10 months of the COVID-19 pandemic. This study built on the evidence base established by existing reviews by incorporating new sources and approaches to analysis. This study highlighted key trends related to the implementation settings, technologies used, and public health functions addressed by COVID-19–related digital innovations. This study also identified a wide-ranging set of barriers and risks that may affect the effective implementation of digital technologies for the COVID-19 pandemic. The existence of such barriers highlights the need for contextually appropriate technological interventions. Although this study did not critically evaluate the effectiveness of digital innovations, the findings from the broader literature point to the fact that technologies introduced as part of the COVID-19 pandemic response demonstrate varying levels of performance and that, in many cases, technologies have yet to be evaluated or comparatively assessed. These findings highlight the need for further evaluation and oversight mechanisms to balance opportunities and risks.

Authors' Contributions

JF, GCA, ERG, CF, KIM, HCG, and SG provided input to the development of the scoping review protocols, implementation of the scoping review, analysis of the results, and production of 2 reports on which this manuscript is based. Drafting of the manuscript was led by JF, with contributions from SG and a critical review by HCG. SG was the project leader for this study, with GCA and JF serving as project managers at different points in the project. HCG was the project owner and critically reviewed the manuscript. This study was funded by the European Centre for Disease Prevention and Control.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 29 KB - publichealth_v8i7e34605_app1.docx]

Multimedia Appendix 2
Search terms.
[DOCX File , 22 KB - publichealth_v8i7e34605_app2.docx]

Multimedia Appendix 3
Inclusion and exclusion criteria.
[DOC File , 18 KB - publichealth_v8i7e34605_app3.doc]

Multimedia Appendix 4
Extraction templates.
[DOCX File , 81 KB - publichealth_v8i7e34605_app4.docx]

Multimedia Appendix 5
Cross-analysis tables.
[DOCX File , 582 KB - publichealth_v8i7e34605_app5.docx]

References

1. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
2. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
3. Mbunge E, Akinnuwesi B, Fashoto SG, Metfula AS, Mashwama P. A critical review of emerging technologies for tackling COVID-19 pandemic. Hum Behav Emerg Technol 2021 Jan;3(1):25-39 [FREE Full text] [doi: 10.1002/hbe2.237] [Medline: 33363278]
4. Vargo D, Zhu L, Benwell B, Yan Z. Digital technology use during COVID-19 pandemic: a rapid review. Human Behav Emerg Technol 2021 Jan;3(1):13-24 [FREE Full text] [doi: 10.1002/hbe2.242]
5. Cohen IG, Gostin LO, Weitzner DJ. Digital smartphone tracking for COVID-19: public health and civil liberties in tension. JAMA 2020 Jun 16;323(23):2371-2372. [doi: 10.1001/jama.2020.8570] [Medline: 32459289]
6. Mbunge E. Integrating emerging technologies into COVID-19 contact tracing: opportunities, challenges and pitfalls. Diabetes Metab Syndr 2020;14(6):1631-1636 [FREE Full text] [doi: 10.1016/j.dsx.2020.08.029] [Medline: 32892060]
7. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
8. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016 Jul;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
9. Feedly. URL: https://feedly.com/ [accessed 2022-06-15]
10. Assessment of digital technologies for infectious disease surveillance, prevention and control. RAND Europe. 2020. URL: https://www.rand.org/randeurope/research/projects/using-digital-tech-for-disease-surveillance.html [accessed 2022-06-13]
11. Digital Single Market Glossary: Shaping Europe’s digital future. European Commission. URL: https://ec.europa.eu/digital-single-market/en/glossary [accessed 2021-02-22]
12. The 10 Essential Public Health Operations. World Health Organization Regional Office for Europe. 2021. URL: https://www.euro.who.int/en/health-topics/Health-systems/public-health-services/policy/the-10-essential-public-health-operations [accessed 2021-05-17]
13. 10 Essential Public Health Services. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/publichealthgateway/publichealthservices/essentialhealthservices.html [accessed 2021-05-17]
14. Danish Government launches digital solutions to support a safe and controlled COVID-19 reopening strategy. Health Tech Digital. 2020 Apr 30. URL: https://www.healthtechdigital.com/danish-government-launches-digital-solutions-to-support-a-safe-and-controlled-covid-19-reopening-strategy/ [accessed 2021-01-07]
15. Hoeksma J. Norway forced to backtrack on mass surveillance track and trace app. Digital Health. 2020. URL: https://www.digitalhealth.net/2020/06/norway-track-and-trace-app/ [accessed 2021-01-07]
16. Howard P, Neudert LM. AI can battle coronavirus, but privacy shouldn’t be a casualty. TechCrunch. 2020 May 27. URL: https://social.techcrunch.com/2020/05/26/ai-can-battle-coronavirus-but-privacy-shouldnt-be-a-casualty/ [accessed 2021-01-07]
17. Khalili J. Governments will use location data to map spread of coronavirus. TechRadar. 2020 Mar 18. URL: https://www.techradar.com/news/governments-will-use-location-data-to-map-spread-of-coronavirus [accessed 2021-01-06]
18. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. Chaos Solitons Fractals 2020 Oct;139:110059 [FREE Full text] [doi: 10.1016/j.chaos.2020.110059] [Medline: 32834612]
19. Lambrechts S. Australia's CovidSafe tracking app is now available – here's what you need to know. TechRadar. 2020 Apr 27. URL: https://www.techradar.com/news/australias-covidsafe-tracking-app-is-now-available-heres-what-you-need-to-know [accessed 2021-01-07]
20. Manancourt V, Delcker J, Scott M, Cerulus L. In fight against coronavirus, governments embrace surveillance. POLITICO. 2020 Mar 24. URL: https://www.politico.eu/article/coroanvirus-covid19-surveillance-data/ [accessed 2021-01-07]
21. Perez S. This Week in Apps: COVID-19 contact tracing apps, virtual dating on the rise, Quibi makes a debut. TechCrunch. 2020 Apr 11. URL: https://social.techcrunch.com/2020/04/11/this-week-in-apps-covid-19-contact-tracing-apps-virtual-dating-on-the-rise-quibi-makes-a-debut/ [accessed 2021-01-07]
22. Singh M. India’s contact-tracing app tops 100 million users in 41 days. TechCrunch. 2020 May 12. URL: https://techcrunch.com/2020/05/12/aarogya-setu-india-contact-tracing-app-10-crore/ [accessed 2021-02-23]
23. Ho HJ, Zhang ZX, Huang Z, Aung AH, Lim WY, Chow A. Use of a real-time locating system for contact tracing of health care workers during the COVID-19 pandemic at an Infectious Disease Center in Singapore: validation study. J Med Internet Res 2020 May 26;22(5):e19437 [FREE Full text] [doi: 10.2196/19437] [Medline: 32412416]
24. Pérez-Escoda A, Jiménez-Narros C, Perlado-Lamo-de-Espinosa M, Pedrero-Esteban LM. Social networks' engagement during the COVID-19 pandemic in Spain: health media vs. healthcare professionals. Int J Environ Res Public Health 2020 Jul 21;17(14):5261 [FREE Full text] [doi: 10.3390/ijerph17145261] [Medline: 32708231]

25. Kitchin R. Civil liberties or public health, or civil liberties and public health? Using surveillance technologies to tackle the spread of COVID-19. Space Polity 2020 Jun 03;24(3):362-381. [doi: 10.1080/13562576.2020.1770587]
26. Pollett S, Rivers C. Social media and the new world of scientific communication during the COVID-19 pandemic. Clin Infect Dis 2020 Nov 19;71(16):2184-2186 [FREE Full text] [doi: 10.1093/cid/ciaa553] [Medline: 32396623]
27. Whitelaw S, Mamas MA, Topol E, Van Spall HG. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health 2020 Aug;2(8):e435-e440 [FREE Full text] [doi: 10.1016/S2589-7500(20)30142-4] [Medline: 32835201]
28. Jin C, Chen W, Cao Y, Xu Z, Tan Z, Zhang X, et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. Nat Commun 2020 Oct 09;11(1):5088 [FREE Full text] [doi: 10.1038/s41467-020-18685-1] [Medline: 33037212]
29. Hernández-Orallo E, Calafate CT, Cano JC, Manzoni P. Evaluating the effectiveness of COVID-19 bluetooth-based smartphone contact tracing applications. Appl Sci 2020 Oct 13;10(20):7113. [doi: 10.3390/app10207113]
30. Hernandez-Orallo E, Manzoni P, Calafate CT, Cano JC. Evaluating how smartphone contact tracing technology can reduce the spread of infectious diseases: the case of COVID-19. IEEE Access 2020 May 27;8:99083-99097 [FREE Full text] [doi: 10.1109/ACCESS.2020.2998042] [Medline: 34192101]
31. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the mobile application rating scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
32. Gunashekar S, Gloinson ER. With Great Power Comes Great Responsibility - How to Harness Technologies in the Fight Against COVID-19. University of Cambridge. 2020 May 31. URL: https://www.csap.cam.ac.uk/Research-Policy-Engagement/spotlights-blogs/guest-post-salil-gunashekar-and-emily-ryen-gloinso/ [accessed 2021-05-14]

Abbreviations

AI: artificial intelligence
EU/EEA: European Union/European Economic Area
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by T Sanchez; submitted 01.11.21; peer-reviewed by L Maaß, C Guinemer; comments to author 09.12.21; revised version received 01.02.22; accepted 17.05.22; published 06.07.22.

Please cite as:
Francombe J, Ali GC, Gloinson ER, Feijao C, Morley KI, Gunashekar S, de Carvalho Gomes H
Assessing the Implementation of Digital Innovations in Response to the COVID-19 Pandemic to Address Key Public Health Functions: Scoping Review of Academic and Nonacademic Literature
JMIR Public Health Surveill 2022;8(7):e34605
URL: https://publichealth.jmir.org/2022/7/e34605
doi:10.2196/34605
PMID:35605152

Original Paper

Global Pattern and Trends in Penile Cancer Incidence: Population-Based Study

Leiwen Fu[1*], MD; Tian Tian[1*], PhD; Kai Yao[2*], PhD; Xiang-Feng Chen[3,4,5], PhD; Ganfeng Luo[1], MD; Yanxiao Gao[1], MD; Yi-Fan Lin[1], PhD; Bingyi Wang[1], MD; Yinghui Sun[1], MD; Weiran Zheng[1], MD; Peiyang Li[1], MD; Yuewei Zhan[1], MD; Christopher K Fairley[6,7], PhD; Andrew Grulich[8], PhD; Huachun Zou[1,8], PhD

[1]School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou, Guangdong, China

[2]Department of Urology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

[3]Center for Reproductive Medicine, Ren Ji Hospital, Shanghai Jiao Tong University, Shanghai, China

[4]Shanghai Key Laboratory for Assisted Reproduction and Reproductive Genetics, Shanghai, China

[5]Shanghai Human Sperm Bank, Shanghai, China

[6]Central Clinical School, Monash University, Melbourne, Australia

[7]Melbourne Sexual Health Centre, Alfred Health, Melbourne, Australia

[8]Kirby Institute, University of New South Wales, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Huachun Zou, PhD
School of Public Health (Shenzhen)
Sun Yat-sen University
Room 107, Unit 3, Building 3, Gezhi Garden
132 Waihuandong Road, Panyu District
Guangzhou, Guangdong, 511431
China
Phone: 86 20 8733 5651
Email: zouhuachun@mail.sysu.edu.cn

Abstract

Background: Penile cancer is a relatively rare genital malignancy whose incidence and mortality are rising in many countries.

Objective: This study aims to assess the recent incidence and mortality patterns and incidence trends of penile cancer.

Methods: The age-standardized incidence and mortality rates (ASIR and ASMR, respectively) of penile cancer in 2020 were estimated from the Global Cancer Registries (GLOBOCAN) database. Incidence trends of penile cancer from 1973 to 2012 were assessed in 44 populations from 43 countries using the Cancer Incidence in Five Continents plus (CI5*plus*) and the Nordic Cancer Registries (NORDCAN) databases. Average annual percentage change was calculated to quantify trends in ASIR using joinpoint regression.

Results: Globally, the estimated ASIR and ASMR of penile cancer were 0.80 (per 100,000) and 0.29 (per 100,000) in 2020, equating to 36,068 new cases and 13,211 deaths in 2020, respectively. There was no significant correlation between the ASIR (P=.05) or ASMR (P=.90) and Human Development Index. In addition, 15 countries saw increasing ASIR for penile cancer, 13 of which were from Europe (United Kingdom, Lithuania, Norway, Estonia, Finland, Sweden, Cyprus, Netherlands, Italy, Croatia, Slovakia, Russia, and the Czech), and 2 from Asia (China and Israel).

Conclusions: Although the developing countries still bear the higher incidence and mortality of penile cancer, the incidence is on the rise in most European countries. To mitigate the disease burden resulting from penile cancer, measures to lower the risk for penile cancers, including improving penile hygiene and male human papillomavirus vaccination, may be warranted.

(JMIR Public Health Surveill 2022;8(7):e34874) doi:10.2196/34874

KEYWORDS

global burden; penile cancer; incidence; average annual percentage change; epidemiology

Introduction

Penile cancer is rare and can occur anywhere on the penis, although most cases arise from the squamous epithelium of glans, coronal sulcus, and prepuce or foreskin. About 95% of penile cancer is classified as squamous cell carcinoma but penile cancer also includes sarcoma, melanoma, and basal cell carcinoma [1,2]. The estimated age-standardized incidence of penile cancer worldwide was 0.80 per 100,000 person-years in 2018, and the incidence is predicted to increase by more than 56% by 2040, according to the Global Cancer Registries (GLOBOCAN) Cancer Tomorrow prediction tool [3]. The change can be largely attributed to the increasing aging of the population, as penile cancer mostly affects older men with a peak in incidence in the sixth decade [4]. In certain Asian, African, and South American countries, the incidence of penile cancer constitutes up to 10% of malignancies in men [4,5]. The 5-year survival rate of penile cancer is about 65% but greater in countries with greater access to treatment [6]. The treatments for penile cancer can be disfiguring and affects the patient's quality of life and sexual function [4,7].

There are many causes of penile cancer. Factors that increase the risk include phimosis, poor personal hygiene, and persistent high-risk human papillomavirus (HPV) infection [8,9]. Phimosis can lead to poor sanitation under the preputial skin and accumulation of smegma, which has been proved to be carcinogenic in animals [4,10-12]. Phimosis may also aggravate balanitis and dermatitis, and is related to the development of aggressive penile malignancies [4,13,14]. Uncircumcised men with poor genital hygiene, even without phimosis, may also have retention of microorganisms and secretions [12,15,16]. A meta-analysis study reported that the pooled detection rate of HPV DNA among penile cancer cases was 50.8% [17]. HPV-16, HPV-6, and HPV-18 are the most common types involved [17-19]. Precancerous lesions associated with HPV infection increase the risk of invasive penile cancer, such as Bowen disease, erythroplasia of Queyrat, and bowenoid papulosis [4]. Furthermore, lack of circumcision, tobacco use, ultraviolet A phototherapy, lichen sclerosis, penile trauma, and low socioeconomic status are also found to be associated with penile cancer [8]. The incidence of penile cancer is negatively correlated with the Human Development Index (HDI) [20].

The incidence of penile cancer has been increasing in many areas in the past few decades [21-23]. There was a 21% increase, from 1.1 to 1.3 per 100,000, in penile cancer incidence in England between 1979 and 2009 [24]; in Norway, the incidence of penile cancer increased from 0.6 to 0.9 per 100,000 between 1956 and 2015 [21]; the incidence of penile cancer in Germany increased from 1.2 per 100,000 in 1961 to 1.8 per 100,000 in 2012 [22]. While previous studies have focused on the incidence trend of penile cancer in specific regions and populations, few reports are available on global patterns and long-term trends in the burden of penile cancer. Understanding the epidemiology of penile cancer can help shed light on factors underlying changing trends.

We aimed to examine the geographical variations in incidence and mortality patterns of penile cancer among 185 countries in 2020 and the long-term incidence trends in 43 countries with 44 populations during the period between1973 and 2012. Our objective is to inform future research and assist policymakers in adopting sound cancer control initiatives.

Methods

Data Source

The estimated data were extracted from the GLOBOCAN Database from the International Agency for Research on Cancer (IARC) [25] to assess the global burden of penile cancer in 185 countries and regions in 2020 [26]. The population-based penile cancer incidence data, with the requirement of at least 15 consecutive years of data, were from Cancer Incidence in 5 Continents (CI5) volumes [27], CI5*plus* [28], and the Nordic Cancer Registries (NORDCAN) database [29]. The quality of the data sources used in this paper has been evaluated in previous studies to assess the incidence trends of other cancers [30-32]. The CI5*plus* database contains updated annual incidence rates for 124 selected populations from 108 cancer registries published in CI5, representing 43 countries, for the period from 1973 to 2012 [32]. The NORDCAN database and program include detailed information and results on cancer incidence, mortality, and prevalence in each of the Nordic countries over 5 decades.

Four levels of HDI were used to further assess the cancer burden according to a binary proxy of development (low and medium HDI vs high and very high HDI) in GLOBOCAN 2020. The incidence data of Denmark, Finland, Iceland, Norway, and Sweden were extracted from the NORDCAN database for the years 1953-2016 [33]. The incidence data of Australia, Croatia, Czech Republic, New Zealand, and Russia were supplemented by their corresponding official national cancer data (Table 1) [34-38]. For volume XI, years 2008-2012 included plotting the overall age-standardized rate by country. Overall, incidence trends were evaluated for 44 populations from 43 countries. As many as 24 out of 43 countries were nationally representative and the representativeness of data in the remaining 19 countries has been verified in previous studies [39,40].

Statistical Analysis

The age-standardized rates were calculated using the World standard population [41]. Trends in incidence are shown with smoothed lines on fitting locally weighted regression (LOWESS) curves, and joinpoint regression (Joinpoint regression program 4.9.0.0, available through the Surveillance Research Program of the US National Cancer Institute) was used to assess temporal trends, which involves fitting a series of joined straight lines to age-standardized incidence rates (ASIRs) trends [42]. Changes in annual incidence rates of penile cancer were calculated as an annual percentage change (APC) in each segment. The joinpoint analysis provided the average annual percentage change (AAPC). To comprehensively estimate the magnitude and direction of trends, we calculated the AAPC and the corresponding 95% CI for the last available 15 years (1998-2012) and those available during the completed period in each country from the database. Correlation analysis was used to test the correlation between the ASIR or age-standardized mortality rate (ASMR) and HDI. All statistical analyses were done using R software 3.6.0 (R Core Team).

Results

Prediction of Incidence and Mortality Patterns in 2020

The global estimated ASIR of penile cancer was 0.8 (per 100,000) in 2020, with estimates indicating 36,068 newly diagnosed cases (Figure 1A and Multimedia Appendix 1). The ASIR of penile cancer varied among 5 continents, with higher ASIRs being observed in Southern Africa, South Asia, and South America. In 2020, the largest number of incident cases was estimated to have occurred in India (n=16,677), China (n=4628), and Brazil (n=1658). The highest ASIRs were found in Eswatini (7.0 per 100,000), Uganda (4.6 per 100,000), and Botswana (4.4 per 100,000), while the lowest were mostly concentrated in countries in Northern Africa, such as Nigeria and Libya (less than 0.01 per 100,000).

Estimates suggest that 13,211 men with penile cancer died in 2020 globally, corresponding to an ASMR of 0.29 cases per 100,000 (Figure 1B and Multimedia Appendix 2). Geographical patterns of ASMR were similar to those of ASIR, and the highest penile cancer ASMRs were noted in Eswatini (3.5 per 100,000) and Uganda (2.4 per 100,000). The largest number of deaths occurred in India (n=4760), China (n=1565), and Brazil (n=539) in 2020.

Both ASIR and ASMR in low- and middle-income countries were nearly twice as those in high-income countries (Multimedia Appendices 1 and 2). Low- and middle-income countries accounted for 40.20% (14,499/36,068) of the incidence and 47.73% (6305/13,211) of the deaths from penile cancer globally, respectively. There was no significant correlation between the ASIR or ASMR and HDI (ρ=1.43, *P*=.05; ρ=−0.01, *P*=.90; Figure 2).

Figure 1. Estimated age-standardized incidence rate (world) in 2020 for penile cancer (A) and estimated age-standardized mortality rate (world) in 2020 for penile cancer (B). (GLOBOCAN 2020 [25]). Data obtained from GLOBOCAN 2020. Map produced by the IARC and WHO. [26].

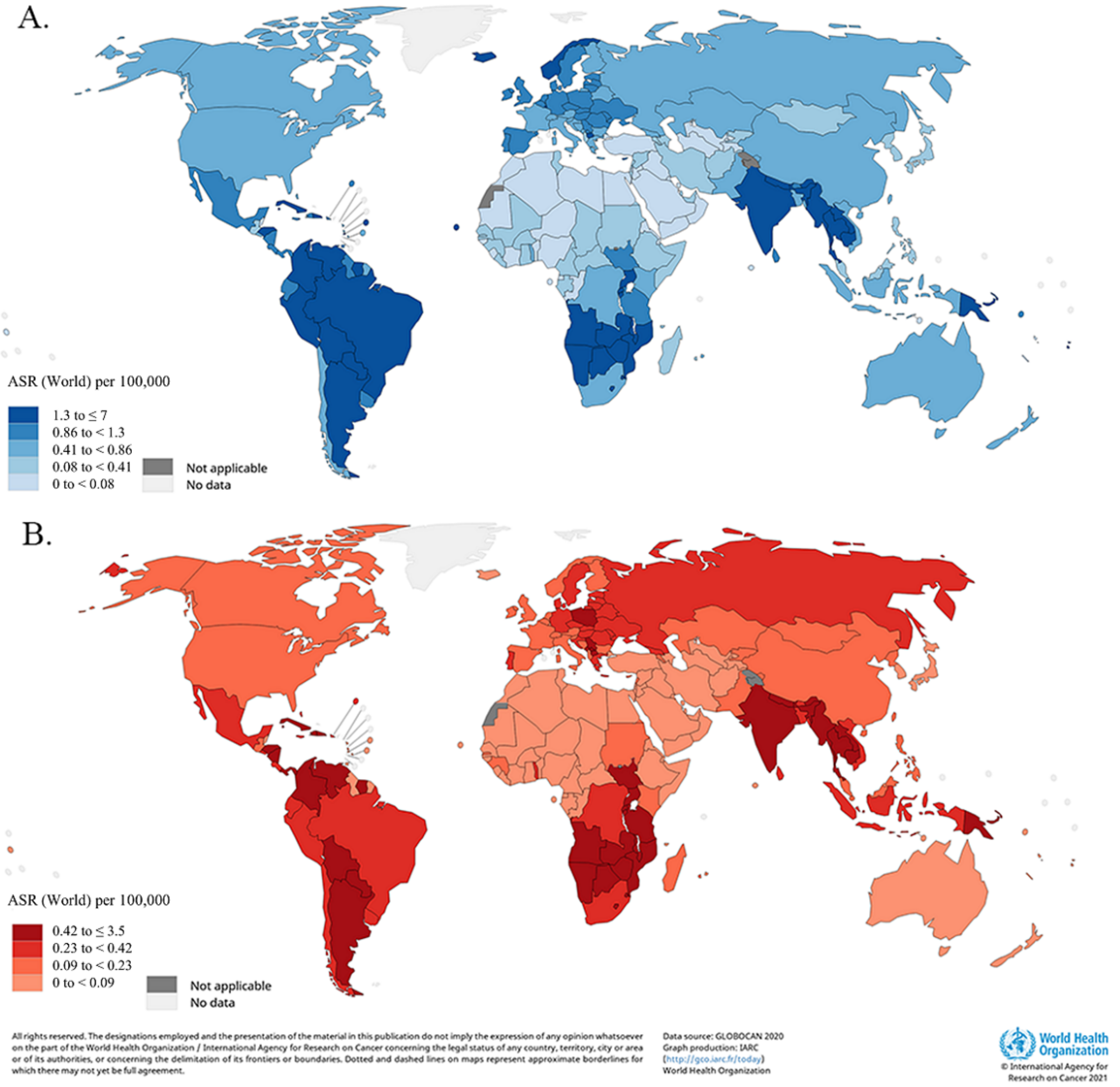
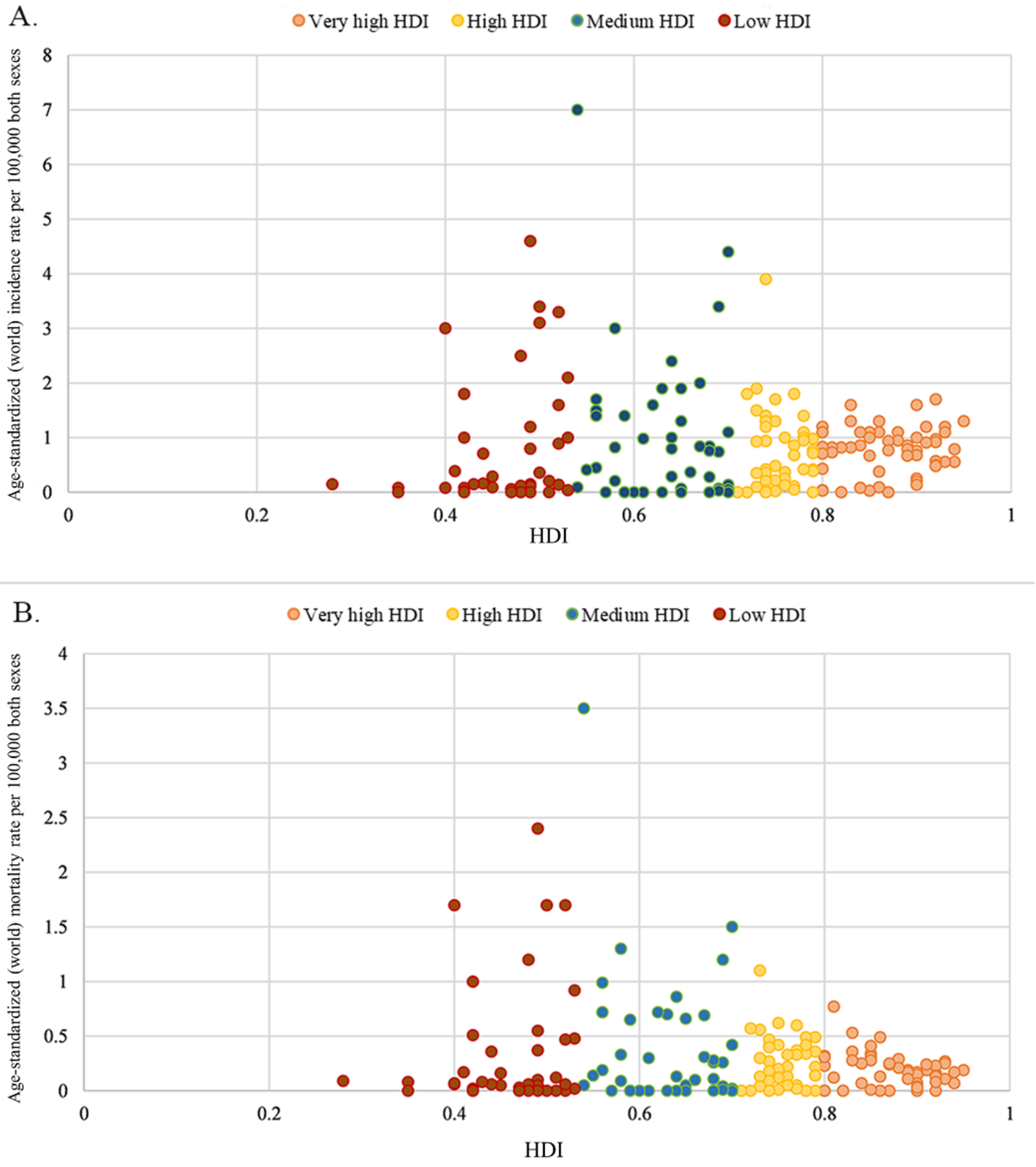


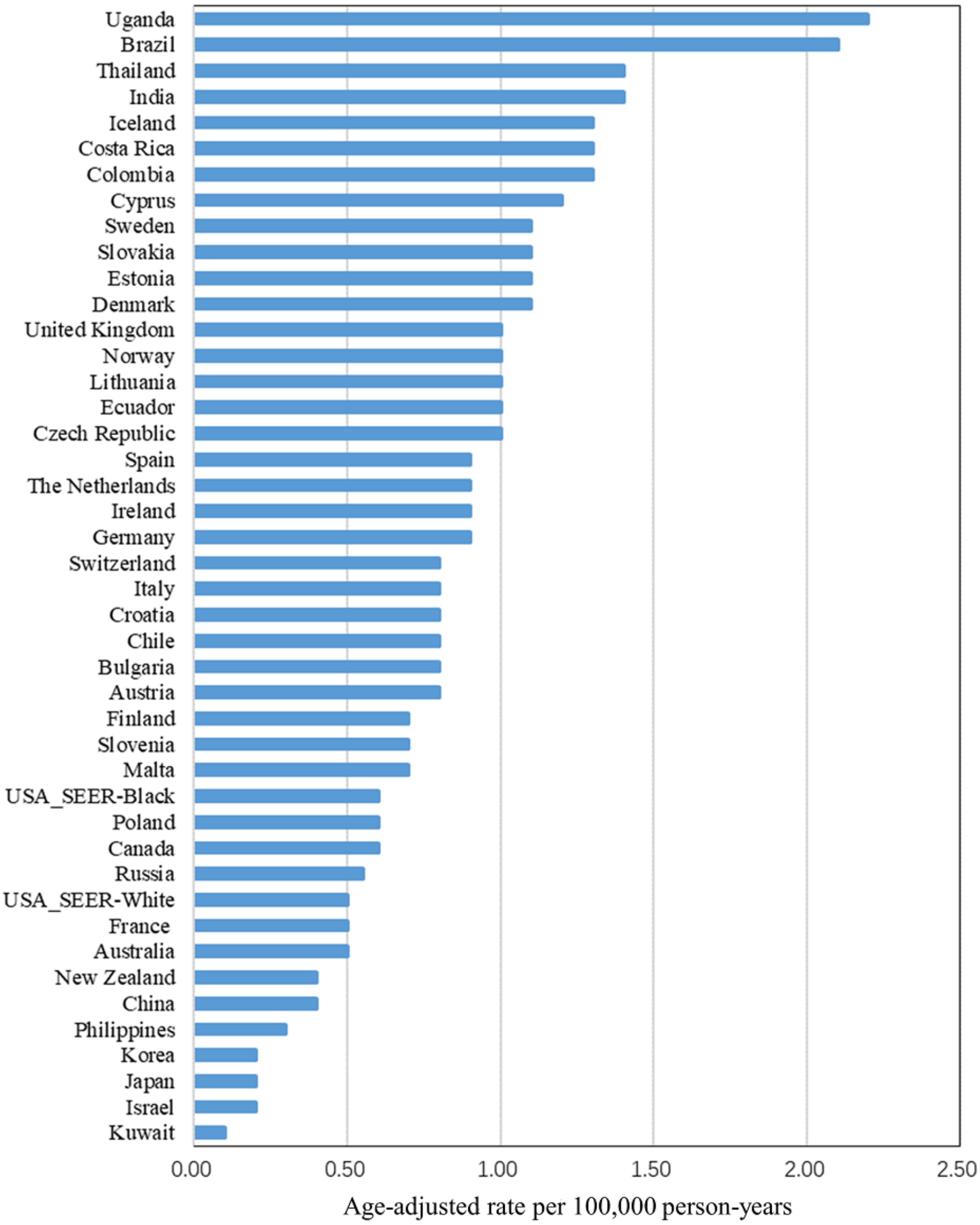
Figure 2. Distribution between (A) age-standardized incidence and (B) mortality rates of penile cancer and HDI (GLOBOCAN 2020). HDI: human development index.



Incidence Rates in 2008-2012

Among 44 populations from 43 countries included in the analysis, the highest ASIR of penile cancer between 2008 and 2012 was in Uganda (2.2 per 100,000), followed by Brazil (2.1 per 100,000), Thailand (1.4 per 100,000), and India (1.4 per 100,000) (Figure 3). The lowest ASIR was in Kuwait (0.1 per 100,000) and ASIRs were less than 0.5 per 100,000 in East Asia and West Asia.

Figure 3. Age-standardized (world standard population) rates of penile cancer incidence, 2008-2012. SEER: Surveillance, Epidemiology, and End Results Program.



Trends in Incidence

The trends in the ASIR of penile cancer between one year and another in 44 populations from 43 countries are displayed in Figure 4. Limited data in some populations meant we could calculate the AAPCs of penile cancer only in 40 populations from 39 countries (Table 1). The largest increase in ASIR was in Israel (AAPC 7.2, 95% CI 3.4 to 11.1; *P*=.001), followed by Cyprus (4.6, 95% CI 0.2 to 9.1; *P*=.04), Croatia (3.6, 95% CI 2.2 to 5.0; *P*<.001), and Lithuania (2.6, 95% CI 0.8 to 4.4; *P*=.007). ASIRs for penile cancer have significantly increased in 15 populations of which 7 were from Northern Europe (United Kingdom, *P*<.001; Lithuania, *P*=.007; Norway, *P*=.002; Estonia, *P*=.02; Finland, *P*=.001; Sweden, *P*=.006; and Cyprus, *P*=.04). In Uganda, the ASIR trend of penile cancer showed a rapid increase between 2007 and 2012. The corresponding APC was 53.3 (95% CI 12.4 to 109.0; *P*=.01).

ASIRs of penile cancer in 5 out of 40 populations, including 3 from Northern America and 2 from Asia, significantly decreased. These decreases were in the Philippines (–2.9, 95% CI –4.5 to –1.2; *P*=.002), India (–2.5, 95% CI –3.4 to –1.6; *P*<.001), the USA White (–1.9, 95% CI –3.1 to –0.6; *P*=.006), the USA Black (–0.8, 95% CI –1.4 to –0.3; *P*=.006), and Canada (–0.7, 95% CI –1.2 to –0.2; *P*=.004). In Thailand, the ASIR of penile cancer decreased between 1988 and 2012. The corresponding APC was –3.4 (95% CI –4.9 to –1.8; *P*<.001).

ASIRs of penile cancer in other parts of Europe, apart from France and Switzerland, increased over the 15-year period (1998-2012; Figure 5). These increases were significant in Russia (AAPC 1.1, 95% CI 0.3 to 1.9; *P*=.01), United Kingdom (1.8, 95% CI 0.8 to 2.8; *P*=.002), Finland (3.0, 95% CI 0.2 to 5.9; *P*=.006), Croatia (4.1, 95% CI 0.7 to 7.5; *P*=.02), Slovakia (4.7, 95% CI 3.2 to 6.2; *P*<.001), and Cyprus (4.6, 95% CI 0.2 to 9.1; *P*=.04). In Asia, significant increases were only observed in Israel (9.7, 95% CI 1.6 to 18.5; *P*=.02), whereas ASIR decreased in India (–3.0, 95% CI –5.2 to –0.8; *P*=.01) and the Philippines (–5.1, 95% CI –8.0 to –2.2; *P*=.003).

Figure 4. Age-standardized incidence rates of penile cancer. (A) Trends in penile cancer incidence increasing. (B) Trends in penile cancer incidence decreasing. ^{a}Regional data.

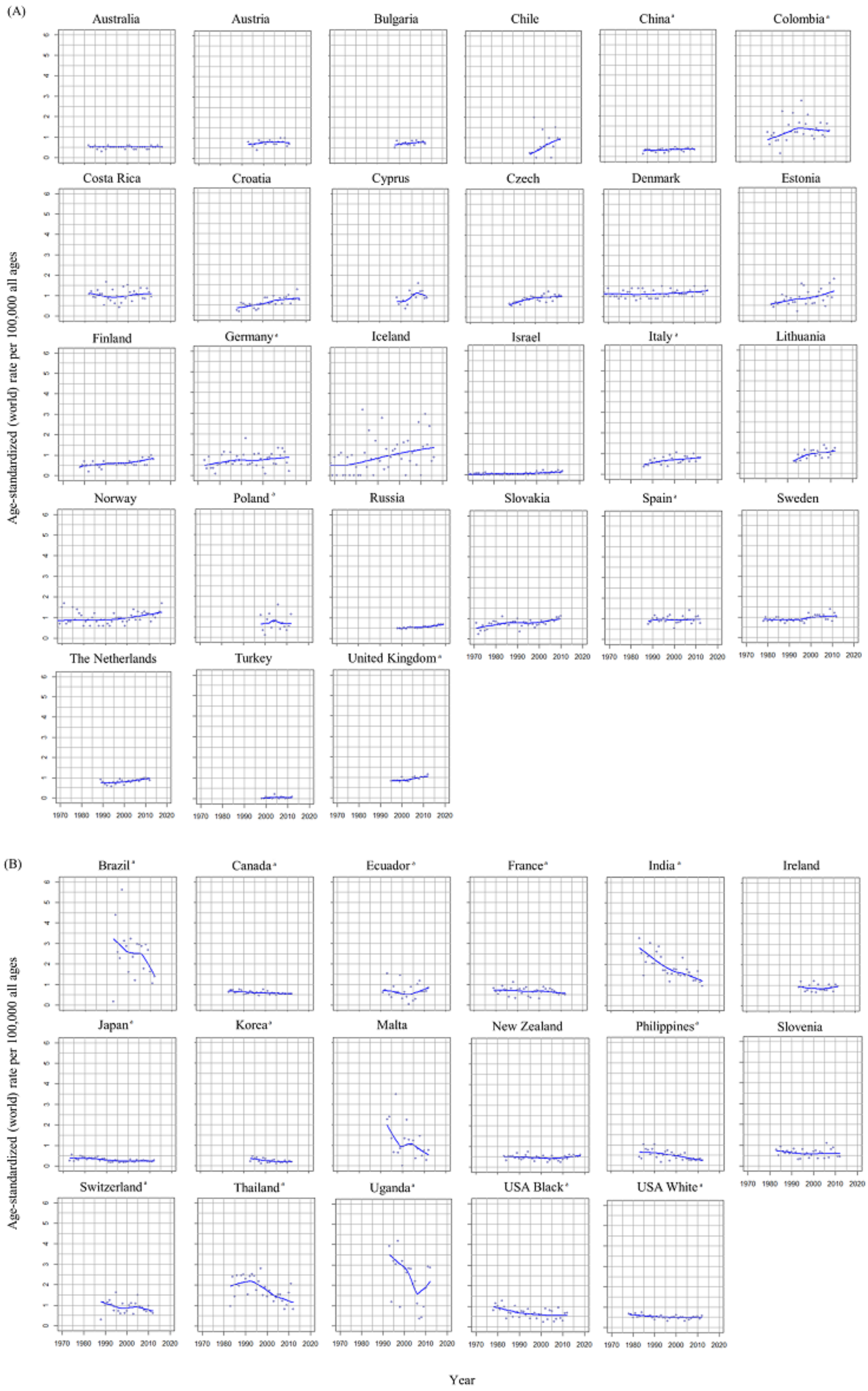


Table 1. International variation in carcinoma of penis incidence rates.

Countries	Registries	Database source	Period	APC[a]	AAPC (%)[b]	AAPC (95% CI)
Austria	National	[43]	1993-2012		0.9	–0.7 to 2.6
Australia	National	[34]	1982-2016		0.1	–0.5 to 0.6
Brazil	Goiania	CI5*plus*[c]	1993-2012		–0.0	–5.8 to 6.1
Bulgaria	National	CI5*plus*	1998-2012		1.3	–0.5 to 3.1
Canada	Alberta, British Columbia, Manitoba, Newfoundland, Nova Scotia, Ontario, Prince Edward Island, Saskatchewan	CI5*plus*	1983-2012		–0.7[d]	–1.2 to –0.2
China	Shanghai	CI5*plus*	1988-2012		1.6[d]	0.1 to 3.2
Colombia	Cali	CI5*plus*	1983-2012		1.8	–0.3 to 4.0
Costa Rica	National	CI5*plus*	1982-2011		0.2	–1.1 to 1.6
Croatia	National	[44]	1988-2017		3.6[d]	2.2 to 5.0
Cyprus	National	CI5*plus*	1998-2012		4.6[d]	0.2 to 9.1
Czech Republic	National	[45]	1977-2018		2.0[d]	1.6 to 2.4
Denmark	National	NORDCAN[e] database	1953-2016		0.1	–0.1 to 0.4
Ecuador	Quito	CI5*plus*	1991-2011		–1.3	–6.6 to 4.4
Estonia	National	CI5*plus*	1983-2012		2.2[d]	0.4 to 4.0
Finland	National	NORDCAN database	1953-1982	–1.9[d]		–3.1 to –0.7
		NORDCAN database	1982-2015	1.7[d]		0.7 to 2.7
		NORDCAN database	1953-2015		0.0	–0.4 to 0.5
France	Bas-Rhin, Calvados, Doubs, Isere	CI5*plus*	1979-2012		–0.5	–1.5 to 0.5
Germany	Saarland	CI5*plus*	1973-2012		0.9	–0.9 to 2.7
India	Chennai	CI5*plus*	1983-2012		–2.5[d]	–3.4 to –1.6
Ireland	National	CI5*plus*	1994-2012		–0.1	–1.8 to 1.6
Israel	National	CI5*plus*	1988-2012		7.2[d]	3.4 to 11.1
Italy	Biella, Naples, Parma, Romagna, Ragusa	CI5*plus*	1986-2012		2[d]	0.7 to 3.2
Japan	Miyagi Prefecture, Nagasaki, Osaka Prefecture	CI5*plus*	1973-1986	1.4		–1.9 to 4.8
		CI5*plus*	1986-1992	–10.6		–22.3 to 2.9
		CI5*plus*	1992-2012	1		–0.7 to 2.8
		CI5*plus*	1973-2012		–0.7	–3.2 to 1.8
Korea	Busan, Seoul, Gwangju, Incheon	CI5*plus*	1993-2012		–3.1	–6.1 to 0.1
Lithuania	National	CI5*plus*	1993-2012		2.6[d]	0.8 to 4.4
The Netherlands	National	CI5*plus*	1989-2012		1.3[d]	0.5 to 2.1
New Zealand	National	[37]	1983-2009	–1.3[d]		–2.6 to –0.1
		[37]	2009-2018	6.1		–0.3 to 12.9
		[37]	1983-2018		0.5	–1.2 to 2.3
Norway	National	NORDCAN database	1953-2016		0.6[d]	0.2 to 1.0
Philippines	Manila	CI5*plus*	1983-2012		–2.9[d]	–4.5 to –1.2
Poland	Kielce	CI5*plus*	1998-2012		1.7	–6.6 to 10.9

Countries	Registries	Database source	Period	APC[a]	AAPC (%)[b]	AAPC (95% CI)
Russia	National	[38]	1998-2019		1.6[d]	1.1 to 2.0
Slovakia	National	CI5*plus*	1971-2012		1.4[d]	0.6 to 2.1
Slovenia	National	CI5*plus*	1983-2012		–0.5	–1.9 to 0.9
Spain	Basque, Tarragona, Granada, Girona	CI5*plus*	1988-2012		0.2	–0.9 to 1.4
Sweden	National	NORDCAN database	1960-1989	–0.5		–1.1 to 0.1
		NORDCAN database	1989-2016	1.0[d]		0.3 to 1.6
		NORDCAN database	1960-2016		0.2	–0.1 to 0.4
Switzerland	Geneva, Neuchatel, Vaud	CI5*plus*	1988-2012		–0.2	–2.3 to 1.9
Thailand	Chiang Mai	CI5*plus*	1983-1988	13.8		–4.1 to 35.0
		CI5*plus*	1988-2012	–3.4[d]		–4.9 to –1.8
		CI5*plus*	1983-2012		–0.6	–3.6 to 2.5
Uganda	Kampala	CI5*plus*	1993-2004	1.2		–7.9 to 11.1
		CI5*plus*	2004-2007	–46.8		–86.7 to 112.9
		CI5*plus*	2007-2012	53.3[d]		12.4 to 109
		CI5*plus*	1993-2012		2	–17.8 to 26.5
United Kingdom	East England, East Midlands, London, Northeast, Northern Ireland, Northwest, Scotland, Southeast, Southwest, West Midlands, Yorkshire-Humber	CI5*plus*	1995-2012		1.6[d]	0.9 to 2.3
USA Black	SEER[f] (9 registries): Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah	CI5*plus*	1978-2012		–0.8[c]	–1.4 to –0.3
USA White	SEER (9 Registries): Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah	CI5*plus*	1978-2012		–1.9[c]	–3.1 to –0.6

[a] APC: annual percentage change.

[b] AAPC: average annual percentage change.

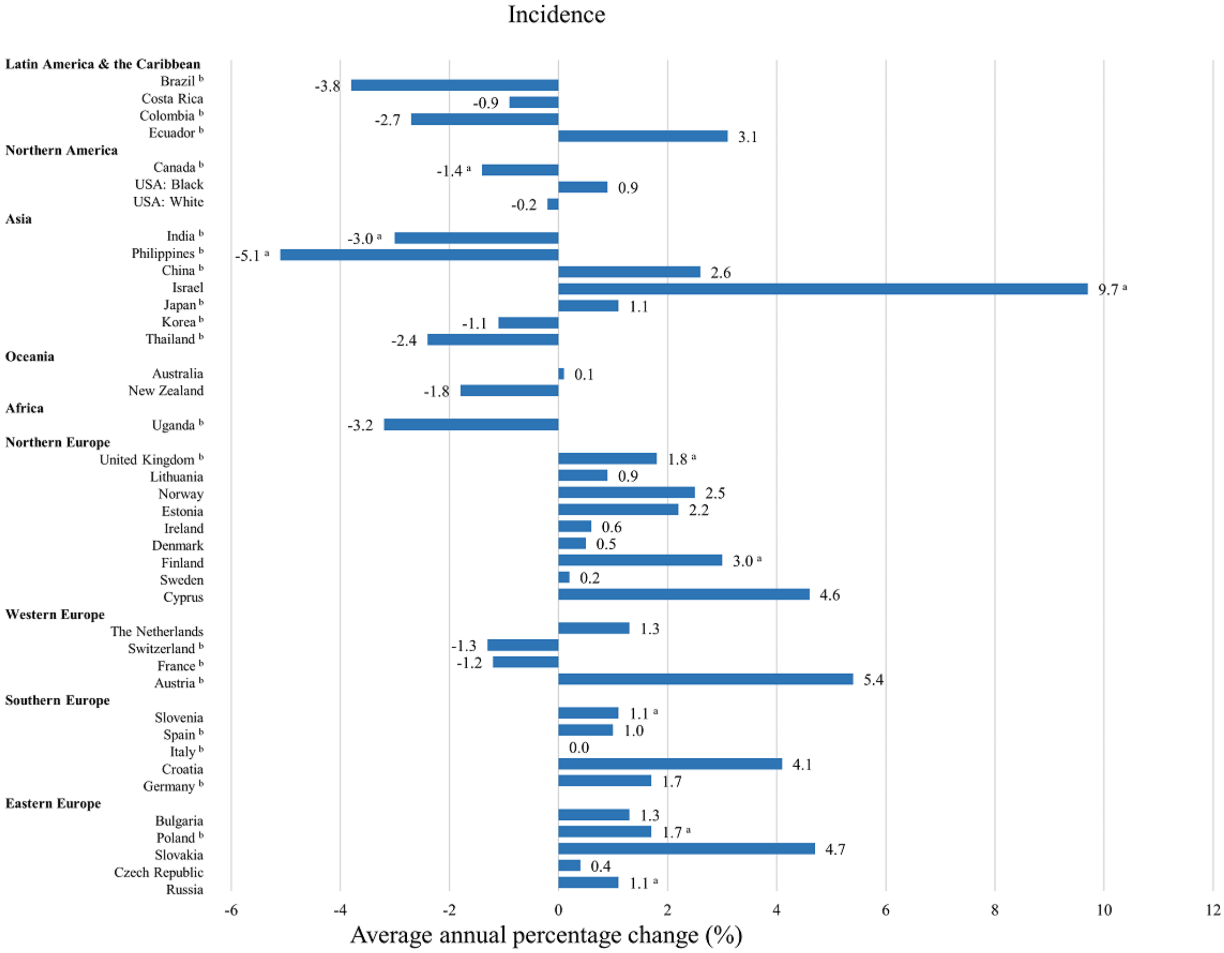
[c] CI5*plus*: Cancer Incidence in Five Continents plus.

[d] Statistically significant (*P*<.05).

[e] NORDCAN: Nordic Cancer Registries.

[f] SEER: Surveillance, Epidemiology, and End Results Program.

Figure 5. Average annual percentage change (AAPC) of penile cancer incidence in the recent 15 years (1998-2012). [a]statistically significant; [b]regional data (incidence).



Discussion

This study comprehensively describes the global pattern and incidence trend of penile cancer. We found that the higher incidence and mortality of penile cancer remain centered in developing settings, such as Southern Africa, South Asia, and South America. In examining temporal trends in incidence, we found that the ASIRs of penile cancer have increased in 15 of 40 populations, 13 of which were in Europe, and decreased in 5 populations.

Although penile cancer is a rare disease, its incidence varies greatly in different regions of the world. In this study, the highest ASIRs of penile cancer occurred in Southern Africa, especially in Eswatini (ASIR: 7.0 per 100,000) and Uganda (ASIR: 4.6 per 100,000) [46,47]. Human immunodeficiency virus (HIV) and HPV are the major public health problems in Southern Africa [48,49]. The immune system clearly plays an important role in the clearance and persistence of HPV infection and in the development of penile cancer [50]. Immunocompromised patients have a higher risk of malignant transformation of HPV lesions. Men who are HIV positive have a 2- to 3-fold increased risk for penile cancer compared with their negative counterparts [50,51]. A study of heterosexual men from Uganda showed that HPV prevalence in the penis was 90.7% among men who are HIV positive and 60.9% among men who are HIV negative [52]. The prevalence of HPV in penis could also be responsible for the high incidence of penile cancer in South American countries, such as Brazil, Colombia, and Argentina [53-56].

We found a significantly increasing trend in the ASIR of penile cancer among most European countries (Italy, the Netherlands, Croatia, Czech Republic, Slovakia, and Russia) during the study period, especially in Northern Europe (United Kingdom, Lithuania, Norway, Estonia, and Cyprus). Consistent with our results, an increasing trend in the ASIR of penile cancer was previously observed in Norway (1956-2015) [21], Netherlands (1989-2006) [57], and the UK (1979-2009) [24]. Similar to our findings, population-based studies reported a stable incidence of penile cancer in Australia during 1982-2005 and in France during 1989-2011 [23,58]. Two Finnish studies reported the decreasing trend of penile cancer ASIR in 1955-1977 and 1971-1995, respectively [59,60]. Moreover, the significantly increasing trend in the ASIR of penile cancer in Finland since 1998 has been described in this study. Our study found that the other 2 populations with a significant increase in the ASIR of penile cancer were China and Israel, which is consistent with the ASIR trend in China between 2005 and 2015, as described by Lu et al [61].

There are many reasons for the increasing trend in the ASIR of penile cancer observed in the aforesaid countries. Increased exposure of the population to HPV and decreasing rates of circumcision in children may play an important role. Childhood circumcision has a strong protective effect against penile cancer [4]. In the United Kingdom, the proportion of boys circumcised fell from 35% in the early 1930s to 6.5% by the mid-1980s; however, circumcision became much less popular after the mid-1940s [62]. The populations with an increasing trend of ASIR for penile cancer had lower rates of circumcision, except for Israel [63]. The incidence is negligible in Israel owing to the practice of religious neonatal circumcision. However, not all increasing trends can be explained by falling rates of childhood circumcision. In recent decades, there have been more immigrants from Muslim countries to European countries such as Russia, France, Norway, the Netherlands, and UK, and therefore, the number of men undergoing circumcision in some countries may rise. Smoking rates decreased substantially between 1970 and 2009 across Europe, which is unlikely to account for the increasing trends in ASIR of penile cancer [64]. The increase in HPV prevalence may explain why the incidence of some cancers that are attributed to high-risk HPV infections, such as anal cancer, cervical cancer, and oropharyngeal cancer, have risen [65-67]. Several studies have reported a strong association between HPV and a higher rate of partner change [51,68]. The significantly higher risk of HPV detection is associated with a younger age at first sexual intercourse and an increase in the number of lifelong female sexual partners. Both of these factors have changed in higher-income countries [69,70]. In the past 40 years, China's opening to the world has brought about economic recovery, but it has also led to changes in sexual behavior, which is reflected in the increase in the incidence of sexually transmitted diseases and changes in the pattern of HIV transmission [71,72].

We found that ASIR decreased in Brazil, Canada, the United States, and most Asian countries, including India, Japan, Korea, Philippines, and Thailand. Although male HPV vaccines are available in Brazil, the United States, and Canada, vaccination would not have had sufficient time to influence the rates of penile cancer in these countries. The major determinant of male circumcision in India is religion: Muslims practice male circumcision for cultural reasons, whereas the predominantly Hindu population does not. This hinders the national promotion of circumcision and is linked to the lower popularity of circumcision [73]. The improvement of personal hygiene might be responsible for the decrease in the incidence of penile cancer in some developing countries, which tend to have a large disease burden. Studies have shown that penile cancer cases in Brazil and India were mainly concentrated in areas that have the lowest HDI [74,75]. Consistent with the study of Goodman et al [76], the decreasing trends in the ASIR of penile cancer were observed in the United States, in both Whites and Blacks, which can be explained by the increasing rate of circumcision. A national probability sample of 1410 American men aged 18-59 years suggested a steady increase in the prevalence of circumcision from a low of 31% (1932) to 85% (1965) [77]. The prevalence of circumcision in the ethnic groups mentioned in this survey was negatively correlated with the incidence of penile cancer found in our analysis: Whites (81%) have a much higher circumcision rate than Blacks (65%).

The incidence trend of penile cancer observed in this study is similar to other long-lag HPV-related cancers, such as vulvar cancer and anal cancer [78,79]. The incidence trend of the other 2 cancers may not be directly comparable with penile cancer due to the different attributable risks. However, the increasing trend in some high-income countries is consistent, such as UK, Italy, and the Netherlands [78,79]. Currently, routine HPV vaccination of boys and men is implemented in several countries, such as Australia, Canada, the United States, and Austria [80]. Vaccination of boys and men may further reduce the incidence of penile cancer, anal cancer, and head and neck cancer; additionally, it may reduce the incidence of cervical cancer and its precursors by herd protection [81]. Expanding the benefits of HPV vaccination to boys and men in countries with a high burden of HPV infection should be evaluated as soon as possible.

The results in our study are enhanced by using 3 data sources (GLOBOCAN, CI5*plus*, and NORDCAN) that include the most recent data possible. Nonetheless, several limitations should be noted for this study. First, the estimates of incidence and mortality of penile cancer were obtained from GLOBOCAN, which is based on the best available data; however, in countries where the estimations are based on potentially biased, insufficient, and proxy data, the estimates should be interpreted with caution. Second, although the data for incidence trend analysis were extracted from a high-quality database (CI5*plus*), some regional data may not be representative of the entire country. Third, the analysis of data from several countries was based on small numbers, subject to substantial random variation, because of the rarity of the disease. Because of insufficient statistical power, we were unable to detect significant trends in smaller populations. Fourth, we were not able to describe trends in incidence by histological subtype or morphologic variant, nor perform age-period-cohort analysis in terms of risk factors associated with penile cancer due to data unavailability. This study could not demonstrate causality in the ASIR trend of penile cancer. Notwithstanding these weaknesses, these data are the largest currently available and allow comparisons across countries because of the uniform approach applied.

Further research is, however, needed to explain the observed regional differences. As a large proportion of penile cancer is attributable to HPV, the efficacy of HPV vaccines in high-risk groups should be assessed as soon as possible. Future research should also continue to explore the association of risk factors with prognosis in patients with penile cancer and to follow the evolution of incidence and survival of this cancer.

In conclusion, this study provides a comprehensive update on the global patterns and trends in the incidence of penile cancer. While the higher incidence and mortality of penile cancer remain in some developing countries, these have significantly increased in most European populations studied, but have also decreased in a few countries. Although there are many causes of penile cancer, HPV infection, poor penile hygiene, and lack of circumcision may play important roles. Improving penile hygiene and promoting the widespread use of male HPV vaccines should be part of prevention programs for penile cancer in the future.

Acknowledgments

The authors gratefully acknowledge all cancer registries and their staff who have contributed in sharing their data needed for this study. We thank Qianglin Fang, Chongguang Yang, Yawen Jiang, Siyang Liu, Huicui Meng, Jinqiu Yuan, and Yiqiang Zhan for their comments during the preparation of this manuscript. This study was supported by the Natural Science Foundation of China Excellent Young Scientists Fund (82022064), Natural Science Foundation of China International/Regional Research Collaboration Project (72061137001), Natural Science Foundation of China Young Scientist Fund (81703278), the National Science and Technology Major Project of China (2018ZX10721102), the Sanming Project of Medicine in Shenzhen (SZSM201811071), the High Level Project of Medicine in Longhua, Shenzhen (HLPM201907020105), the National Key Research and Development Program of China (2020YFC0840900), the Shenzhen Science and Technology Innovation Commission Basic Research Program (JCYJ20190807155409373), Special Support Plan for High-Level Talents of Guangdong Province (2019TQ05Y230), and the Fundamental Research Funds for the Central Universities (58000-31620005). All funding parties did not have any role in the design of the study or in the explanation of the data.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, HZ, upon reasonable request.

Conflicts of Interest

CKF owns shares in CSL Biotherapies. The other authors declare no conflicts of interest related to this work.

Multimedia Appendix 1
Estimated new cases number and age-standardized incidence rates for penile cancer.
[DOCX File , 32 KB - publichealth_v8i7e34874_app1.docx]

Multimedia Appendix 2
Estimated death cases number and age-standardized mortality rates for penile cancer.
[DOCX File , 30 KB - publichealth_v8i7e34874_app2.docx]

References

1. Ahmed ME, Khalil MI, Kamel MH, Karnes RJ, Spiess PE. Progress on Management of Penile Cancer in 2020. Curr. Treat. Options in Oncol 2020 Nov 23;22(1):4-4. [doi: 10.1007/s11864-020-00802-3]
2. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. Eur Urol 2016 Jul;70(1):93-105. [doi: 10.1016/j.eururo.2016.02.029] [Medline: 26935559]
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Sep 12:392-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
4. Thomas A, Necchi A, Muneer A, Tobias-Machado M, Tran ATH, Van Rompuy A, et al. Penile cancer. Nat Rev Dis Primers 2021 Feb 11;7(1):11-11. [doi: 10.1038/s41572-021-00246-5]
5. Bleeker MCG, Heideman DAM, Snijders PJF, Horenblas S, Dillner J, Meijer CJLM. Penile cancer: epidemiology, pathogenesis and prevention. World J Urol 2008 Jul 8;27(2):141-150. [doi: 10.1007/s00345-008-0302-z]
6. Verhoeven R, Janssen-Heijnen M, Saum K, Zanetti R, Caldarella A, Holleczek B, et al. Population-based survival of penile cancer patients in Europe and the United States of America: No improvement since 1990. European Journal of Cancer 2013 Apr;49(6):1414-1421. [doi: 10.1016/j.ejca.2012.10.029]
7. Dräger DL, Milerski S, Sievert KD, Hakenberg OW. Psychosoziale Auswirkungen bei Patienten mit Peniskarzinom. Urologe 2018 Feb 23;57(4):444-452. [doi: 10.1007/s00120-018-0603-9]
8. Douglawi A, Masterson TA. Updates on the epidemiology and risk factors for penile cancer. Transl. Androl. Urol 2017 Oct;6(5):785-790. [doi: 10.21037/tau.2017.05.19]
9. Douglawi A, Masterson TA. Penile cancer epidemiology and risk factors: a contemporary review. Curr Opin Urol 2019 Mar;29(2):145-149. [doi: 10.1097/MOU.0000000000000581] [Medline: 30562185]
10. Plaut A, Kohn-Speyer AC. The Carcinogenic Action of Smegma. Science 1947 Apr 11;105(2728):391-392. [doi: 10.1126/science.105.2728.391-b]
11. Pratt-Thomas HR, Heins HC, Latham E, Dennis EJ, McIvers FA. The carcinogenic effect of human smegma: An experimental study.I. Preliminary report. Cancer 1956 Jul;9(4):671-680. [doi: 10.1002/1097-0142(195607/08)9:4<671::aid-cncr2820090408>3.0.co;2-o]

12. Micali G, Nasca MR, Innocenzi D, Schwartz RA. Penile cancer. Journal of the American Academy of Dermatology 2006 Mar;54(3):369-391. [doi: 10.1016/j.jaad.2005.05.007]
13. Reuter S, Gupta SC, Chaturvedi MM, Aggarwal BB. Oxidative stress, inflammation, and cancer: How are they linked? Free Radical Biology and Medicine 2010 Dec;49(11):1603-1616. [doi: 10.1016/j.freeradbiomed.2010.09.006]
14. Elinav E, Nowarski R, Thaiss CA, Hu B, Jin C, Flavell RA. Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms. Nat Rev Cancer 2013 Oct 24;13(11):759-771. [doi: 10.1038/nrc3611]
15. Jin J. HPV Infection and Cancer. JAMA 2018 Mar 13;319(10):1058. [doi: 10.1001/jama.2018.0687]
16. Gregoire L, Cubilla AL, Reuter VE, Haas GP, Lancaster WD. Preferential Association of Human Papillomavirus With High-Grade Histologic Variants of Penile-Invasive Squamous Cell Carcinoma. JNCI Journal of the National Cancer Institute 1995 Nov 15;87(22):1705-1709. [doi: 10.1093/jnci/87.22.1705]
17. Olesen TB, Sand FL, Rasmussen CL, Albieri V, Toft BG, Norrild B, et al. Prevalence of human papillomavirus DNA and p16INK4a in penile cancer and penile intraepithelial neoplasia: a systematic review and meta-analysis. The Lancet Oncology 2019 Jan;20(1):145-158. [doi: 10.1016/s1470-2045(18)30682-x]
18. Backes DM, Kurman RJ, Pimenta JM, Smith JS. Systematic review of human papillomavirus prevalence in invasive penile cancer. Cancer Causes Control 2008 Dec 11;20(4):449-457. [doi: 10.1007/s10552-008-9276-9]
19. Hernandez B, Shvetsov Y, Goodman M, Wilkens L, Thompson P, Zhu X, et al. Reduced Clearance of Penile Human Papillomavirus Infection in Uncircumcised Men. J INFECT DIS 2010 May;201(9):1340-1343. [doi: 10.1086/651607]
20. Forman D, de MC, Lacey CJ, Soerjomataram I, Lortet-Tieulent J, Bruni L, et al. Global burden of human papillomavirus and related diseases. Vaccine 2012 Nov 20;30 Suppl 5:F12-F23. [doi: 10.1016/j.vaccine.2012.07.055] [Medline: 23199955]
21. Hansen BT, Orumaa M, Lie AK, Brennhovd B, Nygård M. Trends in incidence, mortality and survival of penile squamous cell carcinoma in Norway 1956-2015. Int. J. Cancer 2017 Dec 15;142(8):1586-1593. [doi: 10.1002/ijc.31194]
22. Schoffer O, Neumann A, Stabenow R, Schülein S, Böhm W, Gonsior A, et al. Penile cancer – Incidence, mortality, and survival in Saxony, Germany. Urologic Oncology: Seminars and Original Investigations 2019 Apr;37(4):295.e1-295.e8. [doi: 10.1016/j.urolonc.2018.12.003]
23. Daubisse-Marliac L, Colonna M, Trétarre B, Defossez G, Molinié F, Jéhannin-Ligier K, et al. Long-term trends in incidence and survival of penile cancer in France. Cancer Epidemiology 2017 Oct;50:125-131. [doi: 10.1016/j.canep.2017.08.014]
24. Arya M, Li R, Pegler K, Sangar V, Kelly JD, Minhas S, et al. Long-term trends in incidence, survival and mortality of primary penile cancer in England. Cancer Causes Control 2013 Oct 8;24(12):2169-2176. [doi: 10.1007/s10552-013-0293-y]
25. International Agency for Research on Cancer. URL: http://gco.iarc.fr/today [accessed 2022-03-01]
26. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 Feb 04:209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
27. Cancer Incidence in Five Continents. URL: http://ci5.iarc.fr/Default.aspx [accessed 2022-03-01]
28. CI5plus database. URL: http://ci5.iarc.fr/CI5plus/Default.aspx [accessed 2022-03-01]
29. Association of the Nordic Cancer Registries. URL: https://www.ancr.nu/cancer-data/nordcan-on-the-web/ [accessed 2022-03-01]
30. Bray F, Ferlay J, Laversanne M, Brewster D, Gombe Mbalawa C, Kohler B, et al. Cancer Incidence in Five Continents: Inclusion criteria, highlights from Volume X and the global status of cancer registration. Int. J. Cancer 2015 Aug 14;137(9):2060-2071. [doi: 10.1002/ijc.29670]
31. Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint Å, et al. NORDCAN – a Nordic tool for cancer information, planning, quality control and research. Acta Oncologica 2010 May 21;49(5):725-736. [doi: 10.3109/02841861003782017]
32. J. Ferlay, M. Colombet, F. Bray. Cancer incidence in five continents. IARC CancerBase No 9 Internet, International Agency for Research on Cancer, Lyon, France: CI5plus; 2018.
33. B. Danckert, J. Ferlay, G. Engholm. Cancer incidence, mortality, prevalence and survival in the Nordic countries. Association of the Nordic Cancer Registries, Danish Cancer Society: NORDCAN; 2019.
34. Australian Institute of Health and Welfare 2021. URL: http://www.aihw.gov.au/acim-books/ [accessed 2022-03-01]
35. Liu Y, Wang Z, Rader B, Li B, Wu C, Whittington JD, et al. Associations between changes in population mobility in response to the COVID-19 pandemic and socioeconomic factors at the city level in China and country level worldwide: a retrospective, observational study. The Lancet Digital Health 2021 Jun;3(6):e349-e359. [doi: 10.1016/s2589-7500(21)00059-5]
36. Xiao H, Dai X, Wagenaar BH, Liu F, Augusto O, Guo Y, et al. The impact of the COVID-19 pandemic on health services utilization in China: Time-series analyses for 2016–2020. The Lancet Regional Health - Western Pacific 2021 Apr;9:100122. [doi: 10.1016/j.lanwpc.2021.100122]
37. Ministry of Health, New Zealand Goverment. URL: http://www.health.govt.nz/nz-health-statistics/health-statistics-and-data-sets/cancer-data-and-stats [accessed 2022-03-01]
38. Russian Oncology Portal. URL: http://www.oncology.ru/service/statistics/malignant_tumors/ [accessed 2022-03-01]
39. Gurney JK, Florio AA, Znaor A, Ferlay J, Laversanne M, Sarfati D, et al. International Trends in the Incidence of Testicular Cancer: Lessons from 35 Years and 41 Countries. European Urology 2019 Nov;76(5):615-623. [doi: 10.1016/j.eururo.2019.07.002]

40. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut 2016 Jan 27;66(4):683-691. [doi: 10.1136/gutjnl-2015-310912]
41. Segi M. Cancer Mortality for Selected Sites in 24 Countries (1950–57). Tohoku University School of Public Health: Sendai, Japan; 1960.
42. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med 2000 Feb 15;19(3):335-351. [doi: 10.1002/(sici)1097-0258(20000215)19:3<335::aid-sim336>3.0.co;2-z] [Medline: 10649300]
43. European Cancer Information System. URL: http://eco.iarc.fr/EUREG/AnalysisT.aspx [accessed 2021-03-01]
44. Croatian Institute of Public Health. URL: https://www.hzjz.hr/sluzba-epidemiologija-prevencija-nezaraznih-bolesti/publikacije-odjel-za-maligne-bolesti/ [accessed 2021-03-01]
45. Epidemiology of Malignant Tumours in the Czech Republic. URL: http://www.svod.cz/?sec=aktuality&lang=en# [accessed 2021-03-01]
46. Pow-Sang MR, Ferreira U, Pow-Sang JM, Nardi AC, Destefano V. Epidemiology and Natural History of Penile Cancer. Urology 2010 Aug;76(2):S2-S6. [doi: 10.1016/j.urology.2010.03.003]
47. Lekoane KMB, Kuupiel D, Mashamba-Thompson TP, Ginindza TG. Evidence on the prevalence, incidence, mortality and trends of human papilloma virus-associated cancers in sub-Saharan Africa: systematic scoping review. BMC Cancer 2019 Jun 11;19(1):563-563. [doi: 10.1186/s12885-019-5781-3]
48. Chikandiwa A, Chimoyi L, Pisa PT, Chersich MF, Muller EE, Michelow P, et al. Prevalence of anogenital HPV infection, related disease and risk factors among HIV-infected men in inner-city Johannesburg, South Africa: baseline findings from a cohort study. BMC Public Health 2017 Jul 4;17(S3):425. [doi: 10.1186/s12889-017-4354-0]
49. Chikandiwa A, Pisa PT, Tamalet C, Muller EE, Michelow P, Chersich MF, et al. Prevalent, persistent anal HPV infection and squamous intraepithelial lesions: Findings from a cohort of men living with HIV in South Africa. PLoS ONE 2019 Dec 5;14(12):e0225571. [doi: 10.1371/journal.pone.0225571]
50. Gormley RH, Kovarik CL. Human papillomavirus–related genital disease in the immunocompromised host. Journal of the American Academy of Dermatology 2012 Jun;66(6):867.e1-867.e14. [doi: 10.1016/j.jaad.2010.12.050]
51. Christodoulidou M, Sahdev V, Houssein S, Muneer A. Epidemiology of penile cancer. Current Problems in Cancer 2015 May;39(3):126-136. [doi: 10.1016/j.currproblcancer.2015.03.010]
52. Tobian AAR, Grabowski MK, Kigozi G, Gravitt PE, Eaton KP, Serwadda D, et al. High-risk human papillomavirus prevalence is associated with HIV infection among heterosexual men in Rakai, Uganda. Sex Transm Infect 2012 May 24;89(2):122-127. [doi: 10.1136/sextrans-2012-050524]
53. Favorito LA, Nardi AC, Ronalsa M, Zequi SC, Sampaio FJB, Glina S. Epidemiologic study on penile cancer in Brazil. Int. braz j urol 2008 Oct;34(5):587-593. [doi: 10.1590/s1677-55382008000500007]
54. Soares A, de Carvalho IT, da Fonseca AG, Alencar AM, Leite CHB, Bastos DA, et al. Penile cancer: a Brazilian consensus statement for low- and middle-income countries. J Cancer Res Clin Oncol 2020 Oct 26;146(12):3281-3296. [doi: 10.1007/s00432-020-03417-1]
55. Munoz N, Castellsague X, Bosch FX, Tafur L, Sanjose SD, Aristizabal N, et al. Difficulty in Elucidating the Male Role in Cervical Cancer in Colombia, a High-Risk Area for the Disease. JNCI Journal of the National Cancer Institute 1996 Aug 07;88(15):1068-1075. [doi: 10.1093/jnci/88.15.1068]
56. Picconi MA, Eiján AM, Distéfano AL, Pueyo S, Alonio LV, Gorostidi S, et al. Human papillomavirus (HPV) DNA in penile carcinomas in Argentina: Analysis of primary tumors and lymph nodes. J. Med. Virol 2000 May;61(1):65-69. [doi: 10.1002/(sici)1096-9071(200005)61:1<65::aid-jmv10>3.0.co;2-z]
57. Graafland NM, Verhoeven RH, Coebergh JW, Horenblas S. Incidence trends and survival of penile squamous cell carcinoma in the Netherlands. Int. J. Cancer 2010 Mar 25;128(2):426-432. [doi: 10.1002/ijc.25355]
58. Grulich AE, Jin F, Conway EL, Stein AN, Hocking J. Cancers attributable to human papillomavirus infection. Sex. Health 2010;7(3):244. [doi: 10.1071/sh10020]
59. Pukkala E, Weiderpass E. Socio-economic differences in incidence rates of cancers of the male genital organs in Finland, 1971-95. Int. J. Cancer 2002 Nov 22;102(6):643-648. [doi: 10.1002/ijc.10749]
60. Maiche AG. Epidemiological aspects of cancer of the penis in Finland. European Journal of Cancer Prevention 1992;1(2):153-158. [doi: 10.1097/00008469-199202000-00008]
61. Lu Y, Li P, Luo G, Liu D, Zou H. Cancer attributable to human papillomavirus infection in China: Burden and trends. Cancer 2020 Jun 02;126(16):3719-3732. [doi: 10.1002/cncr.32986]
62. Sewell J, Ranasinghe W, De Silva D, Ayres B, Ranasinghe T, Hounsome L, et al. Trends in penile cancer: a comparative study between Australia, England and Wales, and the US. SpringerPlus 2015 Aug 14;4(1):420. [doi: 10.1186/s40064-015-1191-4]
63. Morris BJ, Wamai RG, Henebeng EB, Tobian AAR, Klausner JD, Banerjee J, et al. Erratum to: Estimation of country-specific and global prevalence of male circumcision. Popul Health Metrics 2016 Apr 4;14(1):4. [doi: 10.1186/s12963-016-0080-6]
64. Marcon A, Pesce G, Calciano L, Bellisario V, Dharmage SC, Garcia-Aymerich J, et al. Trends in smoking initiation in Europe over 40 years: A retrospective cohort study. PLoS ONE 2018 Aug 22;13(8):e0201881. [doi: 10.1371/journal.pone.0201881]

65. Alemany L, Saunier M, Alvarado-Cabrero I, Quirós B, Salmeron J, Shin H, et al. Human papillomavirus DNA prevalence and type distribution in anal carcinomas worldwide. Int. J. Cancer 2014 May 30;136(1):98-107. [doi: 10.1002/ijc.28963]
66. Parkin DM, Bray F. Chapter 2: The burden of HPV-related cancers. Vaccine 2006 Aug;24:S11-S25. [doi: 10.1016/j.vaccine.2006.05.111]
67. Braakhuis BJM, Leemans CR, Visser O. Incidence and survival trends of head and neck squamous cell carcinoma in the Netherlands between 1989 and 2011. Oral Oncol 2014 Jul;50(7):670-675 [FREE Full text] [doi: 10.1016/j.oraloncology.2014.03.008] [Medline: 24735546]
68. Maden C, Sherman KJ, Beckmann AM, Hislop TG, Teh C, Ashley RL, et al. History of Circumcision, Medical Conditions, and Sexual Activity and Risk of Penile Cancer. JNCI Journal of the National Cancer Institute 1993 Jan 06;85(1):19-24. [doi: 10.1093/jnci/85.1.19]
69. Bajos N, Bozon M, Beltzer N, Laborde C, Andro A, Ferrand M, et al. Changes in sexual behaviours: from secular trends to public health policies. AIDS 2010 May 15;24(8):1185-1191. [doi: 10.1097/QAD.0b013e328336ad52] [Medline: 20299962]
70. Mercer CH, Tanton C, Prah P, Erens B, Sonnenberg P, Clifton S, et al. Changes in sexual attitudes and lifestyles in Britain through the life course and over time: findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). Lancet 2013 Nov 30;382(9907):1781-1794 [FREE Full text] [doi: 10.1016/S0140-6736(13)62035-8] [Medline: 24286784]
71. Wu Z, Chen J, Scott SR, McGoogan JM. History of the HIV Epidemic in China. Curr HIV/AIDS Rep 2019 Nov 26;16(6):458-466. [doi: 10.1007/s11904-019-00471-4]
72. Gil VE, Wang MS, Anderson AF, Lin GM, Wu ZO. Prostitutes, prostitution and STD/HIV transmission in Mainland China. Social Science & Medicine 1996 Jan;42(1):141-152. [doi: 10.1016/0277-9536(95)00064-x]
73. Sinha A, Chandhiok N, Sahay S, Deb S, Bharat S, Gupta A, et al. Male circumcision for HIV prevention in India: emerging viewpoints and practices of health care providers. AIDS Care 2015 May 22;27(9):1196-1198. [doi: 10.1080/09540121.2015.1039957]
74. Coelho RWP, Pinho JD, Moreno JS, Garbis DVEO, do Nascimento AMT, Larges JS, et al. Penile cancer in Maranhão, Northeast Brazil: the highest incidence globally? BMC Urol 2018 May 29;18(1):29. [doi: 10.1186/s12894-018-0365-0]
75. Nandakumar A, Gupta PC, Gangadharan P, Visweswara RN, Parkin DM. Geographic pathology revisited: Development of an atlas of cancer in India. Int. J. Cancer 2005 Sep 20;116(5):740-754. [doi: 10.1002/ijc.21109]
76. Goodman MT, Hernandez BY, Shvetsov YB. Demographic and pathologic differences in the incidence of invasive penile cancer in the United States, 1995-2003. Cancer Epidemiol Biomarkers Prev 2007 Sep;16(9):1833-1839. [doi: 10.1158/1055-9965.EPI-07-0221] [Medline: 17855702]
77. Laumann EO, Masi CM, Zuckerman EW. Circumcision in the United States. Prevalence, prophylactic effects, and sexual practice. JAMA 1997 Apr 02;277(13):1052-1057. [Medline: 9091693]
78. Bray F, Laversanne M, Weiderpass E, Arbyn M. Geographic and temporal variations in the incidence of vulvar and vaginal cancers. Int. J. Cancer 2020 Jul 06;147(10):2764-2771. [doi: 10.1002/ijc.33055]
79. Islami F, Ferlay J, Lortet-Tieulent J, Bray F, Jemal A. International trends in anal cancer incidence rates. Int. J. Epidemiol 2016 Oct 27:dyw276. [doi: 10.1093/ije/dyw276]
80. Stanley M. HPV vaccination in boys and men. Hum Vaccin Immunother 2014;10(7):2109-2111 [FREE Full text] [doi: 10.4161/hv.29137] [Medline: 25424825]
81. Schmeler KM, Sturgis EM. Expanding the benefits of HPV vaccination to boys and men. Lancet 2016 Apr 30;387(10030):1798-1799. [doi: 10.1016/S0140-6736(16)30314-2] [Medline: 27203488]

Abbreviations

AAPC: average annual percentage change
APC: annual percentage change
ASIR: age-standardized incidence rate
ASMR: age-standardized mortality rate
CI5: Cancer Incidence in Five Continents
GLOBOCAN: Global Cancer Registries
HDI: Human Development Index
HIV: human immunodeficiency virus
HPV: human papillomavirus
IARC: International Agency for Research on Cancer
LOWESS: locally weighted regression
NORDCAN: Nordic Cancer Registries
SEER: Surveillance, Epidemiology, and End Results Program

Edited by G Eysenbach, H Bradley; submitted 11.11.21; peer-reviewed by S Pesälä, H Akram; comments to author 24.02.22; revised version received 01.03.22; accepted 26.05.22; published 06.07.22.

Please cite as:
Fu L, Tian T, Yao K, Chen XF, Luo G, Gao Y, Lin YF, Wang B, Sun Y, Zheng W, Li P, Zhan Y, Fairley CK, Grulich A, Zou H
Global Pattern and Trends in Penile Cancer Incidence: Population-Based Study
JMIR Public Health Surveill 2022;8(7):e34874
URL: https://publichealth.jmir.org/2022/7/e34874
doi:10.2196/34874
PMID:35793140

Original Paper

Linguistic Pattern–Infused Dual-Channel Bidirectional Long Short-term Memory With Attention for Dengue Case Summary Generation From the Program for Monitoring Emerging Diseases–Mail Database: Algorithm Development Study

Yung-Chun Chang[1,2], PhD; Yu-Wen Chiu[1,3], BA; Ting-Wu Chuang[3], PhD

[1]Graduate Institute of Data Science, Taipei Medical University, Taipei, Taiwan
[2]Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei, Taiwan
[3]Department of Molecular Parasitology and Tropical Diseases, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Ting-Wu Chuang, PhD
Department of Molecular Parasitology and Tropical Diseases
School of Medicine, College of Medicine
Taipei Medical University
No 250
Wu-Hsing Street
Taipei, 110
Taiwan
Phone: 886 27361661 ext 3123
Email: chtingwu@tmu.edu.tw

Abstract

Background: Globalization and environmental changes have intensified the emergence or re-emergence of infectious diseases worldwide, such as outbreaks of dengue fever in Southeast Asia. Collaboration on region-wide infectious disease surveillance systems is therefore critical but difficult to achieve because of the different transparency levels of health information systems in different countries. Although the Program for Monitoring Emerging Diseases (ProMED)–mail is the most comprehensive international expert–curated platform providing rich disease outbreak information on humans, animals, and plants, the unstructured text content of the reports makes analysis for further application difficult.

Objective: To make monitoring the epidemic situation in Southeast Asia more efficient, this study aims to develop an automatic summary of the alert articles from ProMED-mail, a huge textual data source. In this paper, we proposed a text summarization method that uses natural language processing technology to automatically extract important sentences from alert articles in ProMED-mail emails to generate summaries. Using our method, we can quickly capture crucial information to help make important decisions regarding epidemic surveillance.

Methods: Our data, which span a period from 1994 to 2019, come from the ProMED-mail website. We analyzed the collected data to establish a unique Taiwan dengue corpus that was validated with professionals' annotations to achieve almost perfect agreement (Cohen κ=90%). To generate a ProMED-mail summary, we developed a dual-channel bidirectional long short-term memory with attention mechanism with infused latent syntactic features to identify key sentences from the alerting article.

Results: Our method is superior to many well-known machine learning and neural network approaches in identifying important sentences, achieving a macroaverage F1 score of 93%. Moreover, it can successfully extract the relevant correct information on dengue fever from a ProMED-mail alerting article, which can help researchers or general users to quickly understand the essence of the alerting article at first glance. In addition to verifying the model, we also recruited 3 professional experts and 2 students from related fields to participate in a satisfaction survey on the generated summaries, and the results show that 84% (63/75) of the summaries received high satisfaction ratings.

Conclusions: The proposed approach successfully fuses latent syntactic features into a deep neural network to analyze the syntactic, semantic, and contextual information in the text. It then exploits the derived information to identify crucial sentences in the ProMED-mail alerting article. The experiment results show that the proposed method is not only effective but also outperforms the compared methods. Our approach also demonstrates the potential for case summary generation from ProMED-mail alerting articles. In terms of practical application, when a new alerting article arrives, our method can quickly identify the relevant case information, which is the most critical part, to use as a reference or for further analysis.

(JMIR Public Health Surveill 2022;8(7):e34583) doi:10.2196/34583

KEYWORDS

ProMED-mail; natural language processing; dengue; dual channel; bidirectional long short-term memory

Introduction

Background

Globalization and climate change have exacerbated the frequency and virulence of infectious diseases worldwide [1-3]. Climate and environmental changes play an undeniable role in changing disease ecology and transmission dynamics [4-6], with transboundary transmission also being frequently linked to international transportation [7]. Monitoring the region-wide or global infectious disease transmission patterns relies on intercountry collaborations to share disease surveillance information. The Program for Monitoring Emerging Diseases (ProMED)–mail [8] was launched by the International Society for Infectious Diseases in 1994 to collect global disease outbreak information on humans, animals, and plants [9,10]. Currently, ProMED-mail is the largest unofficial infectious diseases platform based on volunteer reporting, and it receives disease outbreak reports or research findings from different users (including individual scientists and governmental agencies) around the world. The EpiCore program, involving a worldwide network of public health professionals, was added in 2014 to scrutinize and verify the reported information [11].

Each report’s quality is enhanced through an expert-review process that includes reducing data redundancy and errors, which are common in social media reports. Reports from social media platforms such as Twitter and Facebook have been used to detect disease outbreaks in previous works, with Google Trends being a good example of web data providing early warning messages regarding influenza [12]. However, social media reports have at least three main limitations. The first involves unclear definitions. Many infectious diseases might share very similar clinical symptoms, making it difficult to differentiate them from simple keyword searches by users. Second, with the passage of time, the attention that people pay to disease outbreaks wanes. The third limitation relates to social media accessibility in different countries or groups of people: when researchers conduct long-term pattern analysis or multinational analysis of disease outbreaks, social media will introduce bias. These issues have been addressed by a few studies [13,14]. Social media and ProMED-mail might play different roles regarding dengue detection and analysis: whereas social media can be used to detect the emergence of dengue in the early stage, ProMED-mail can provide richer, more correct, and reliable epidemiological information, as well as continuous monitoring of that information [15-17].

Well-known contributions of ProMED-mail are the early reports of suspected cases of severe acute respiratory syndrome in China in 2002 and the Middle East respiratory syndrome coronavirus in Saudi Arabia in 2012 [11,18]. More recently, ProMED-mail data have been used to analyze a cholera outbreak in Africa, a vector-borne disease outbreak amid violent conflict in Syria, and global avian influenza outbreaks [19-21]. Thanks to >25 years of effort, huge amounts of disease outbreak information have been accumulated in the ProMED-mail database; however, the unstructured text format of the ProMED-mail report hampers the efficiency of scientific analysis. Most previous studies using the ProMED-mail database usually relied on labor-intensive review processes, making it difficult to analyze multiple diseases and broader study areas. However, natural language processing (NLP) can help because it is a powerful technique for extracting information from unstructured clinical or health records [22-24].

With the outbreak of COVID-19, sources of epidemic surveillance have received more and more attention, prompting the publication of several research papers. Nonetheless, few studies have taken advantage of NLP technology for the development or analysis of data from ProMED websites. Carrion and Madoff [11] have noted that every season, ProMED would publish a word cloud of epidemics in various regions to show epidemics that have been particularly severe in each region in the current season. Taking 2016 as an example, alerting reports from all over the world were processed by NLP technology to produce a visualized word cloud in which dengue is the keyword for entire Southeast Asia. In addition, Kim et al [25] developed a deep learning approach to automatically recognize the relevant information that is necessary to deal with potential disease outbreaks; this is consistent with our view. Their study used 2 approaches—convolutional neural network (CNN) and bidirectional long short-term memory (BiLSTM) [26,27]—to classify the sources of the texts about infectious diseases in the alerting articles, and they achieved an overall accuracy rate of 92.9%.

In recent years, factors such as climate, weather, and culture have stimulated global epidemic monitoring, from which it has become clear that dengue fever in Southeast Asia still poses a serious threat [28,29]. Dengue incidence has increased significantly around the world in the last 2 decades [30], with an estimated 390 million infections per year, of which 96 million exhibit clinical symptoms. Each year, approximately 3.9 billion people are at risk of infection with dengue viruses, with 2 million severe cases and 2100 deaths [31]. Although dengue infection is prevalent in 129 countries, 70% of all cases are located in Asia. The reported number of dengue cases has increased >8-fold over the last 2 decades, and most deaths have occurred in younger age groups. As there is currently no effective vaccine or treatment available for dengue infection, dengue surveillance information is imperative for disease control and prevention.

For these reasons, our research focuses on Southeast Asia. As the alerting articles from ProMED-mail are lengthy (an average of 1872 words and 82 sentences in Asian-related dengue fever alerting articles), it is important to develop a summary generation system to assist relevant researchers to become more proficient in monitoring the pandemic. In general, case information is related to the outbreak location, time, and patient [32]. The combination of these 3 types of information constitutes the coincidence of occurrence of case information. However, although NLP research for information extraction has flourished [33-36], it is not easy to determine whether all the important case information is contained in a single ProMED-mail alerting article. For example, as digits usually represent the number of cases, the appearance of important case information is often accompanied by the appearance of numbers. However, in the sentence “CDC deputy director-general Chuang Jen-Hsiang said on Wednesday [August 7, 2019] that the patient also has underlying diseases, which is why he was only diagnosed with dengue after 2 screenings and multiple hospital visits,” the digit “2” represents frequency, rather than the number of cases. It thus cannot become one of the sentences in the summary, although it contains dengue-related keywords such as disease, dengue, and diagnose. Moreover, location and time are also relevant information for important cases. We thus assume intuitively that if the location and time are mentioned in the same sentence, it may more likely describe relevant information about the case. However, although the sentence “Since 18 Nov [2006], Kaohsiung County City health authorities reinforced implementation of the mosquito-elimination campaign” mentions location and time, the content obviously does not contain information related to any infection outbreak.

Objectives

The specific aim of this research was therefore to extract abstract sentences that contain important epidemiological information on dengue incidence in Southeast Asia. Our proposed method can enhance the decision-making efficiency of epidemic monitoring units by quickly and automatically generating summaries of alerting articles. This is particularly important during the current COVID-19 pandemic. Specifically, our method first decomposes an alerting article from ProMED-mail into sentences. Next, the sentences are classified into 2 categories in accordance with their syntactic characteristics. Finally, the proposed deep neural network–based method integrates linguistic patterns and latent syntactic features to identify important sentences as the basic unit of summary generation. The experiment results based on real-world data sets demonstrate that the proposed method successfully exploits the syntactic, semantic, and contextual information relevant to epidemiological information on dengue. Consequently, our method not only outperforms many well-known information extraction methods but also achieves a satisfaction rating of 84% for the abstractive sentence summarization of dengue alerting article data.

Methods

Data Corpus

The aim of this study was to develop a method to extract sentences that convey dengue case information in ProMED-mail alerting articles and then automatically generate summaries. However, to the best of our knowledge, there is no official data set for crucial sentences extraction with regard to dengue. For this reason, we compiled our own data corpus for method development and performance evaluation. To do this, we first collected all the articles from 1994 to 2019 on the ProMED-mail website as a preliminary corpus. Next, we used the country name to further extract Southeast Asia articles and then verified that the title contained “Dengue/DHF (dengue hemorrhagic fever) updates” to indicate a series of dengue fever alerting articles. To facilitate the efficiency of the corpus construction, we sampled 15% of the instances from the data set for annotation for a total of 129 articles that contained 965 sentences. Next, the data set was annotated by 2 experts who are medical professionals with high English proficiency. Before the annotation stage, we conducted a training to ensure that the annotators had a common understanding of what defines crucial sentences in a dengue summary. We provided each annotator with 20 instances of both positive and negative cases during the training stage. A third expert acted as an arbiter for verifying the annotations. During the actual labeling process, the annotators labeled 246 instances as crucial sentences (Cohen κ=0.895, which indicates that the interobserver agreement of our annotated data corpus is reliable [37-39]). The final annotation results were used for the performance evaluation of the proposed model. It is worth noting that 80.1% (197/246) of the crucial sentences were composed of multiple clauses with complex syntactic structures (Table 1). Furthermore, 34.6% (85/246) of the sentences contained digits that did not convey case information (eg, “Dengue virus circulating type 1” and “20-34 mosquitoes or mosquito larvae found in every 100 households”). In light of these sentence statistics, it is clearly a challenging task to extract sentences that convey case information. The corpus has been released to promote further research and is available at the Taipei Medical University Dengue Case Corpus [40].

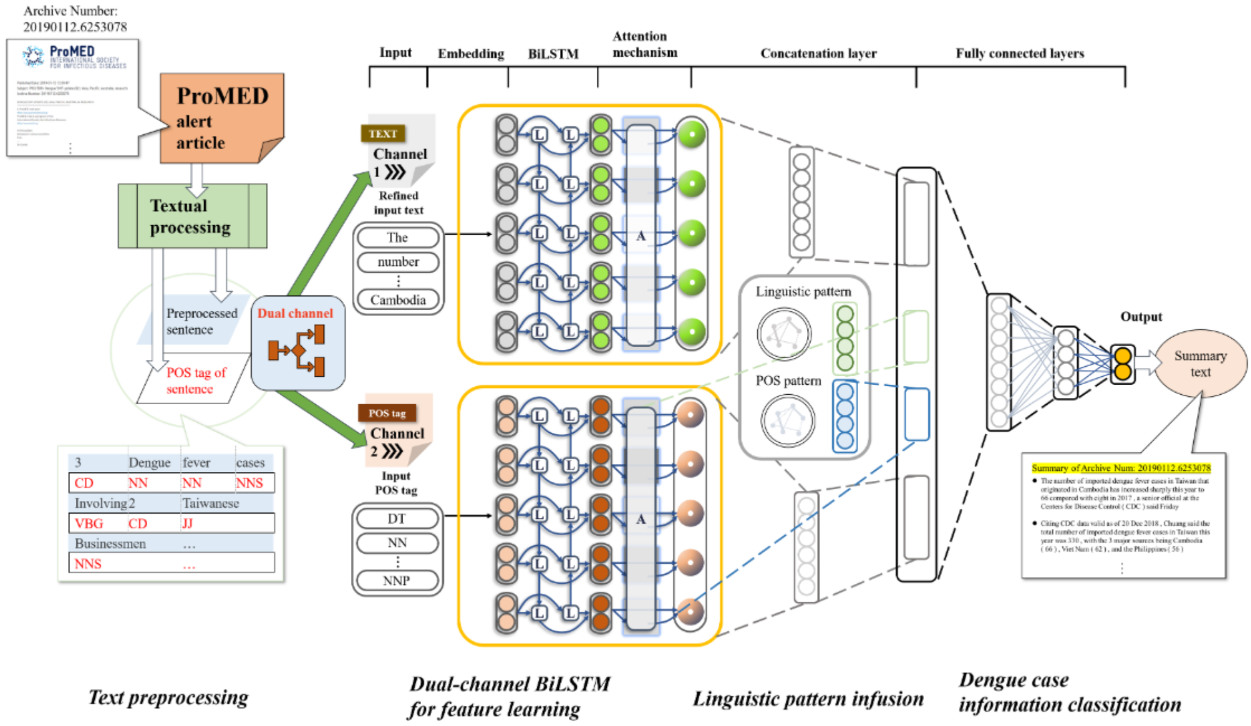
Table 1. The statistics of our corpus (N=965)[a].

	Number of sentences, n (%)	
	Single-clausal sentences	Multiclausal sentences
Dengue case sentences (n=246)	49 (19.9)	197 (80.1)
Non-dengue case sentences (n=719)	407 (56.6)	312 (43.4)

[a] Number of paragraphs: 129, number of sentences: 965, number of single-clausal sentences: 456, and number of multiclausal sentences: 509.

Figure 1 presents the overview of the proposed framework, which automatically detects summary sentences that contain case-related information in a set of documents of the dengue alerting series written in English. In this research, we used a linguistic pattern–infused BiLSTM with an attention mechanism neural network for dengue case information extraction. That is, to extract sentences that convey dengue case information from alerting articles from ProMED-mail, we treated dengue case information extraction as a binary classification problem that can be formulated as follows. Let $W = \{w_1,…, w_k\}$ be a set of words and $S = \{s_1,…, s_m\}$ be a set of sentences from a set of alerting articles. Each sentence s comprises a set of words such that sW. The goal of this task was to decide whether a sentence s_j expresses dengue case information. Our framework consisted of 4 main procedures: text preprocessing, dual-channel BiLSTM (DuBiLSTM), linguistic pattern infusion, and dengue case information classification. Further details of each procedure are provided in the following sections.

Figure 1. Overview of the proposed framework. A: attention layer; BiLSTM: bidirectional long short-term memory; CD: cardinal number; DT: determiner; JJ: adjective; L: forward long short-term memory layer and backward long short-term memory layer; NN: noun, singular or mass; NNS: noun, plural; POS: parts of speech; ProMED: Program for Monitoring Emerging Diseases; VBG: verb, gerund, or present participle.



3	Dengue	fever	cases
CD	NN	NN	NNS
Involving	2	Taiwanese	
VBG	CD	JJ	
Businessmen		…	
NNS		…	

Summary of Archive Num: 20190112.6253078

- The number of imported dengue fever cases in Taiwan that originated in Cambodia has increased sharply this year to 66 compared with eight in 2017 , a senior official at the Centers for Disease Control (CDC) said Friday
- Citing CDC data valid as of 20 Dec 2018 , Chuang said the total number of imported dengue fever cases in Taiwan this year was 330 , with the 3 major sources being Cambodia (66) , Viet Nam (62) , and the Philippines (56)

⋮

Text Preprocessing

Preprocessing is a critical task that needs to be performed before feeding data into a neural network. When a report document d_n is entered, we first decompose it into a set of paragraphs $P = \{p_1,..., p_i\}$ and obtain the sentence collections $S = \{s_1,..., s_j\}$ through sentence segmentation for each paragraph. Next, we break a sentence into tokens and tag the parts of speech (POS) $T = \{t_1,..., t_k\}$ using the Natural Language Toolkit package [41]. As dengue case information can be narrated in a sequence of clauses, we recognize 2 types of sentences, namely single-clausal sentences and multiclausal sentences. Moreover, as frequently used words are generally not helpful for identifying dengue case information, we removed the stop words as well as punctuation marks (commas and semicolons) in the sentences.

DuBiLSTM With Attention Mechanism

In this research, we developed a DuBiLSTM with an attention mechanism neural network to learn latent semantic features behind both alert article texts and shallow paring information. The embedding layer is first used to transform the input tokens and POS tags of a sentence into 300D vectors. We used Global Vectors for Word Representation pretrained word embeddings (ie, glove.6B) to transform sentences of the alerting articles into 300D vectors. The POS embeddings are learned from the embedding layer of our model with continuous bag-of-words mode. Specifically, a sentence is represented by $s = \{w_1,..., w_k\}\{pos_1,..., pos_k\}$, its corresponding word vector is $v_w = \{v_{w\_1},..., v_{w\_n}\}$, and its POS vector is $v_{pos} = \{v_{pos\_1},..., v_{pos\_n}\}$, which are the inputs of the model.

Compared with the original recurrent neural network, the reason for the improvement of long short-term memory (LSTM) is its special design. LSTM defines and maintains an internal memory cell state throughout the life cycle to establish temporal connections. This internal memory cell state is the most important element of LSTM's structure. The LSTM model consists of a series of identical timing modules. In addition to the original input, LSTM has 3 designs—forget gate, input gate, and output gate—that determine whether the input is important enough to be remembered and whether it can be output.

The details are described herein. Suppose there are 3 element-wise functions that help to calculate the next moment by the previous moment and this moment where σ(.) is a sigmoid function, tanh(.) is a hyperbolic tangent function, and is the product. We also have $x_t \in {}^d$ and $h_t \in {}^h$ denoting the input vector and the hidden state vector at moment t, respectively, whereas $U \in {}^{h\times h}$ and $W \in {}^{h\times d}$ indicate the weight metrices of gates or cells for input vector x_t and hidden state vector h_t, respectively, and $b \in {}^{h}$ indicates the weight metrices of gates or cells for the bias vector, where the superscripts *d* and *h* refer to the number of input features and number of hidden units, respectively. The forget gate at the moment *t* $f_t \in {}^{h}$ determines the information to be forgotten by outputting a number in (0, 1), in line with the following equation:

$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f)$ **(1)**

With regard to the second mechanism, the input gate of LSTM then decides what new information input should be kept by calculating $i_t \in {}^{h}$ and $\tilde{c}_t \in {}^{h}$ and combining the 2 parameters in the light of the following equations:

$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i)$ **(2)**

$\tilde{c}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c)$ **(3)**

$c_t = f_t \; c_{t-1} + i_t \; \tilde{c}_t$ **(4)**

The third special mechanism is the output gate. This represents which parts of the cell state should be outputted based on the following equations, where $h_t \in {}^{h}$ represents the hidden state vector, also known as the output vector:

$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o)$ **(5)**

$h_t = o_t \; \tanh(c_t)$ **(6)**

However, the information in the LSTM network is a 1-way transmission, and LSTM can only use past information, not future information. BiLSTM can consider both past and future data information by connecting 2 LSTM networks with opposite timings in the same output. The forward andbackward LSTMs can obtain, respectively, the past and future data information of the input sequence. The hidden state H_t of BiLSTM at time *t* includes forward and backward :

In addition, as the attention mechanism can allocate more attention to important information and less to other information, the receiving sensitivity and processing speed of information in the focused attention area are greatly improved. The attention function can softly map the combination of query *Q* and a set of key-value pairs $\{K,V\}$ to some notable outputting results, where $Q = \{Q_1,..., Q_N\}$ and $\{K,V\} = \{(K_1,V_1),..., (K_M,V_M)\}$. Furthermore, the multihead attention mechanism would convert *Q*, *K*, and *V* into *H* subspaces in the first step, with various and learnable linear projections, as the following transforming equation shows:

where $\{Q^h, K^h, V^h\}$ are the input query, key, and value of the *n*th head, respectively; represents the parameter matrices at the same time; and *d* and d_k indicate, respectively, the dimension of the model and its subspace. In the second step, *H* attention functions are implemented in parallel to generate the output states $O = \{O^1,..., O^H\}$, where any O^h in *O* is defined in the following equation:

where Att^h is the attention distribution that comes from the *h*th attention head. Finally, the output states *O* are concatenated with each other and then connected with distinctively generated features for the next stage.

Linguistic Pattern Infusion

The human perception of a dengue case information alert involves identifying a relevant lexicon or semantic content to rapidly narrow down the scope of possible candidates. For instance, when an expression contains strongly correlated words such as "dengue" and "total" at the same time, it is natural to conclude that this is probably an expression about a dengue case. These lexical indicators can help explain how humans can skim through an article to quickly capture the dengue case information expressions. Therefore, we used log-likelihood ratio (LLR); it is an effective feature selection method that can generate representative patterns from sequences of dengue case information expressions [42-44]:

Given a training data set composed of binary labels for representing sentences that describe dengue case information (*D*) or not (¬*D*), we pair words from sentences to generate a set of co-occurring word pairs $WP = \{wp_1,..., wp_f\}$ and POS pairs $PP = \{pp_1,..., pp_g\}$. The LLR uses the following mechanism to calculate the likelihood that the occurrence of a word pair and a POS pair in the dengue case information is not random. To illustrate, we take the LLR calculation for a word pair, where $N(D)$ and $N(\neg D)$ are the numbers of positive and negative sentences, respectively. $N(wp$^D), which is denoted as *k*, is the number of alert articles containing *wp* and *D* simultaneously, whereas $N(wp\wedge\neg D)$, which is denoted as *l*, is the number of negative sentences that include *wp*. To further simplify the formula, we also define $m = N(D) - k$ as the number of sentences containing *D* without the word pair *wp* and $n = N(\neg D) - l$, which means the number of sentences with neither *D* nor *wp*. A maximum likelihood estimation is conducted to obtain probabilities $p(wp)$, $p(w|D)$, and $p(w|{\wedge}D)$ by calculating the log-likelihood of the hypothesis that the presence of *wp* in set *D* is not random. A word pair with a large LLR value is therefore closely associated with the expression of dengue case information. We rank all the word pairs by every LLR value in the training data, and the top 50 word pairs that describe dengue case information and those that do not are selected as linguistic patterns for positive and negative sentences, respectively. The same procedure is adopted to calculate the LLR value for the POS pairs for the compilation of POS patterns.

Next, we integrate the generated linguistic patterns and POS patterns into a DuBiLSTM with an attention mechanism by concatenating both positive and negative vectors (which are composed of 60 and 25 dimensions, respectively) with the LLR value of the matched patterns (ie, 170D pattern vectors). As the LLR values of linguistic patterns and POS patterns indicate the weight associated with positive and negative sentences, merging the linguistic pattern and POS pattern features into a DuBiLSTM with an attention mechanism is discriminative.

Finally, the direct splicing strategy is used to fuse pattern features with latent semantic features from a DuBiLSTM with an attention mechanism neural network. The calculation formula for this is as follows:

$$F = PF \oplus LSF \quad \mathbf{(13)}$$

where PF=pattern features and LSF=latent semantic features.

Dengue Case Information Classifier

The final step in our framework involves constructing a classifier to predict the labels through the 3 fully connected layers and the activation layer and then to output the distribution probability of the labels. In the fully connected layer, the model maps the fused feature vector to the instance label space. In the output layer, the softmax function is used for normalization, and the output of the fully connected layer is converted into the approximate probability value y for each category. The calculation formula is as follows:

where M is the parameter matrix of the connection layer, F is the characterization of the fusion-distributed characteristics, b is the bias, and softmax is a normalization function. Although the 3 fully connected layers increase the computational cost, the classifier efficiently learns weights through the neuron layer [45-47]. The neurons in each layer will be connected to the neurons in the next layer. Considering the convergence rate, the rectified linear unit (ReLU) function is used as an activation function for nonlinear operation. This can easily cause overfitting in model learning; therefore, to avoid this, we use the dropout mechanism to correct for overfitting [48]. The 2 probabilities are predicted for negative and positive, and the larger probability will be taken out to become the final prediction result through the softmax function.

In addition, the Adam optimizer [49] was chosen to optimize the loss function of the network. The model parameters are fine-tuned by the Adam optimizer, which has been shown to be an effective and efficient backpropagation algorithm. We use the cross-entropy function as the loss function because it can reduce the risk of a gradient disappearance during the process of stochastic gradient descent; this is why it often performs better than the classification error rate or the mean square error [50]. The loss rate of the model can be calculated using the following equation:

where N is the number of training samples, y is the label of the sample, and is the output of the model.

Comparative Analysis Models

To conduct a comprehensive evaluation of the proposed method, we also developed baselines of the machine learning model and deep neural networks to estimate the significance of our approach and the performance variation in different classification systems. Our first baseline uses a tokenized representation evaluated on the radial basis function kernel–based support vector machine (SVM) [51]. This system learns the statistical relevance of each token in a clinical record within different syntactic and semantic contexts. Next, 2 ensemble learning approaches were also implemented for comparison. The first is random forest (RF), an ensemble learning method for classification that constructs a multitude of decision trees (DTs) adopting term frequency–inverse document frequency text representation. The other model is the classifier extreme gradient boosting (XGB), which is a gradient boosting DT that integrates multiple learners for classification problems [52]. We included XGB in this study because it has been validated on real-life large-scale imbalanced data sets and solves many data science–related problems in a fast and accurate way [53]. Finally, the 3 deep neural networks were compared for performance evaluation. The first deep learning model is a class of feedforward artificial neural networks called multilayer perceptron (MLP) [54]. We constructed an MLP through 3 fully connected dense layers afterward the input layer (the model was constructed by adding 3 hidden layers between the input layer and the output layer, where hidden layers are responsible for feature extraction to help output layer to do classification work). In addition, we also adopted a whole-text multikernel CNN model using static word embeddings [55,56] of instances (CNN for text) as another baseline. The last deep learning model that we included in the comparative analysis is the LSTM recurrent neural network, which is capable of learning order dependence in a text sequence and is widely used in NLP research.

To examine the incremental performance that benefits from the proposed method, the BiLSTM and DuBiLSTM are also listed as comparators. To serve as a basis for comparison, we also included naïve Bayes [51,57-59] and DT [51] as baselines.

Results

Evaluation Metrics

In our experiments, the performance evaluation metrics included precision, recall, and F1 score. In general, there is a trade-off between precision and recall. As the 2 metrics evaluate system performance from different perspectives, a single metric that balances (ie, averages) the trade-off is essential. The F1 score is the harmonic mean of precision and recall, and as it is generally close to the minimum of the 2 values, it can be considered an attempt to find the best possible compromise (balance) between precision and recall [51]. The F1 score is also deemed a conservative metric that prevents the possible overestimation of system performance because the harmonic mean is always less than, or equal to, the arithmetic mean and geometric mean. For this reason, the F1 score is extensively used to judge the superiority of information systems [51]. We used the macroaverage to compute the average performance, and to obtain reliable verification results, we adopted a 10-fold cross-validation approach [60]. Our model was implemented with Keras [61] under the following configurations: the dropout probability was set at 0.35 after each layer, loss function was categorized as cross-entropy, ReLU activation was applied to the dense layer, and training was set at 40 epochs.

Model Comparisons

The model performance is shown in Table 2. First, the naïve Bayes classifier, which is a conditional probability–based approach using bag-of-words feature space with term frequency–inverse document frequency term weighting, only achieved a mediocre performance. As this classifier only considers surface word weightings, it has difficulty representing interword relations, and its overall F1 score is only 70.34%. By contrast, the DT further learns keyword weighting for representation through an entropy-based feature selection method. Hence, the DT is able to obtain significant improvement. The XGB obtained better performances because it is able to integrate multiple machine learning algorithms using the gradient boosting mechanism to optimize the loss function. Likewise, the RF integrates multiple DTs through ensemble learning to optimize prediction results. Therefore, the overall performance of the XGB and RF are similar, with both achieving F1 scores of approximately 85%. It is worth mentioning that the prediction performance of an SVM achieves an F1 score of approximately 90%. This is because the SVM can solve nonlinear obstacles and build models based on learned word and phrase correlations in context, thereby enhancing classification performance. When comparing the deep learning approaches, the performance of an MLP is similar to the ensemble learning–based approaches. The redundancy and inefficiency might be caused by the large number of parameters in the fully connected neural network structure. The learned weighting is therefore unreliable and leads to the lowest performance among all neural models considered in this study. In contrast to the MLP, CNN for text and LSTM have better performances that achieve F1 scores of approximately 90%. This indicates that both deep neural network models efficiently represent textual information and learn the context of the ProMED alerting articles to identify dengue case information. It is worth noting that our method can extract latent linguistic features from ProMED alerting reports because learned word and POS embeddings are adopted to represent syntactic and context relations. Moreover, we used discriminative patterns to encode the characteristics of collocation relationships to capture descriptors of dengue within the alert reports. Consequently, our method achieves the best overall precision, recall, and F1 score among the compared methods.

Table 2. The performance results of the compared methods.

System	Negative, precision; recall; F1 score (%)	Positive, precision; recall; F1 score (%)	Macroaverage, precision; recall; F1 score (%)	*P* value
NB[a]	81.69; 99.30; 89.64	94.51; 34.96; 51.04	88.10; 67.13; 70.34[b]	<.001
DT[c]	95.79; 85.40; 90.29	67.59; 89.02; 76.84	81.69; 87.21; 83.57[b]	<.001
RF[d]	96.53; 88.87; 92.54	73.60; 90.65; 81.24	85.06; 89.76; 86.89[b]	<.001
SVM[e]	94.12; 95.69; 94.90	86.75; 82.52; 84.58	90.43; 89.10; 89.74[b]	<.001
XGB[f]	92.55; 91.52; 92.03	75.98; 78.46; 77.20	84.26; 84.99; 84.61[b]	<.001
MLP[g]	94.64; 90.82; 92.69	76.00; 84.96; 80.23	85.32; 87.89; 86.46[b]	<.001
CNN[h] for text	94.47; 94.99; 94.73	85.12; 83.74; 84.43	89.80; 89.37; 89.58[b]	<.001
LSTM[i]	94.72; 94.85; 94.79	84.90; 84.55; 84.73	89.81; 89.70; 89.76[b]	<.001
BiLSTM[j]	95.74; 93.88; 94.80	83.08; 87.80; 85.38	89.41; 90.84; 90.09[k]	.94
DuBiLSTM[l]	95.89; 94.16; 95.02	83.78; 88.21; 85.94	89.84; 91.18; 90.48[k]	.95
Our method	97.72; 95.27; 96.48	87.12; 93.50; 90.20	92.42; 94.38; 93.34	—[m]

[a] NB: naïve Bayes.

[b] *P*<.001 (a chi-square test was applied to determine whether our method significantly improves performance in comparison with other methods).

[c] DT: decision tree.

[d] RF: random forest.

[e] SVM: support vector machine.

[f] XGB: extreme gradient boosting.

[g] MLP: multilayer perceptron.

[h] CNN: convolutional neural network.

[i] LSTM: long short-term memory.

[j] BiLSTM: bidirectional long short-term memory.

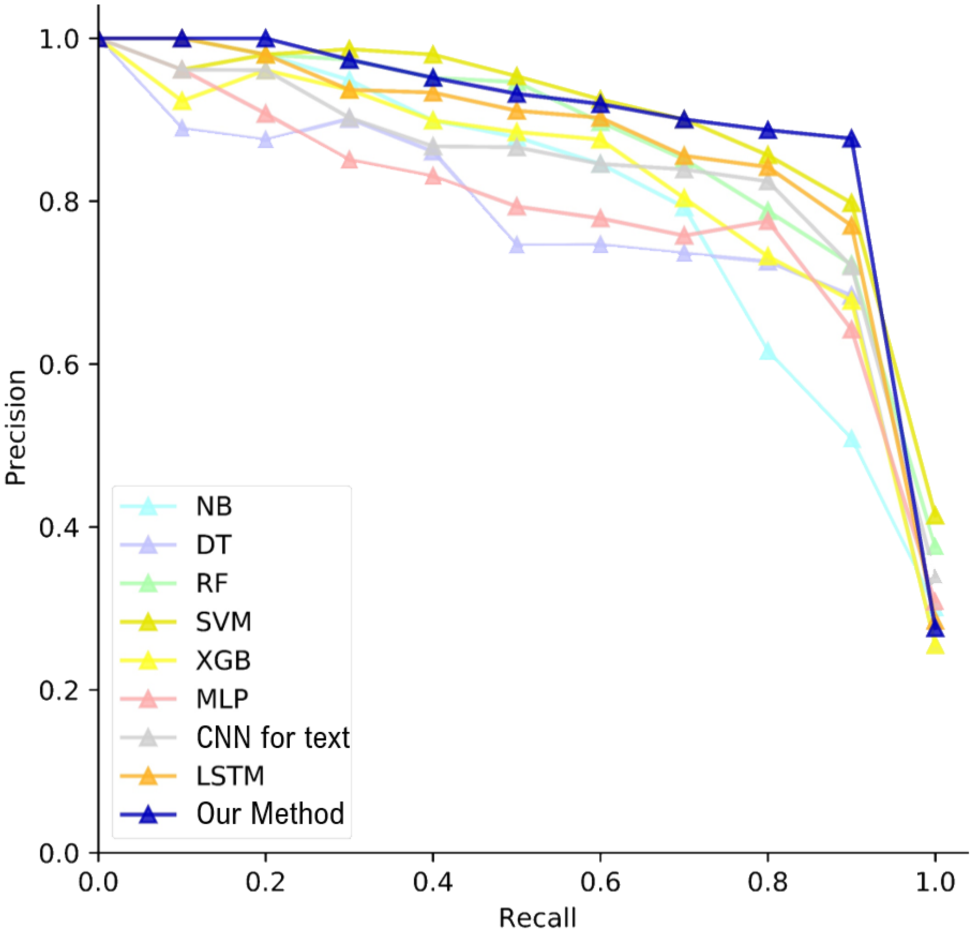
[k] *P*>.05 (a chi-square test was applied to determine whether our method significantly improves performance in comparison with other methods).

[l] DuBiLSTM: dual-channel bidirectional long short-term memory.

[m] Not available.

Finally, we evaluated the performances of the compared methods using 11-point precision recall curves [62]. To plot these curves, the evaluated sentences were sorted according to their prediction scores. Figure 2 shows that the precision scores of our method at the 11 recall levels are superior to those of the compared methods. In other words, our method is able to most accurately extract sentences that convey dengue case information.

Figure 2. The precision recall curves of the compared methods. CNN: convolutional neural network; DT: decision tree; LSTM: long short-term memory; MLP: multilayer perceptron; NB: naïve Bayes; RF: random forest; SVM: support vector machine; XGB: extreme gradient boosting.



To summarize, the DuBiLSTM is able to learn the latent syntactic and semantic information of a text, and the attention mechanism can further highlight the important elements in context. The linguistic patterns are successfully integrated into the neural network to capture discriminative collocation of latent features. Consequently, our method significantly outperforms the compared methods and achieves a remarkable dengue case information extraction performance.

Discussion

Principal Findings

This study describes a new method for identifying dengue case information by using BiLSTM with an attention mechanism enriched with linguistic patterns. As the results show, BiLSTM can consider contextual information more efficiently. Through its bidirectional mechanism, the output for any current moment is not only related to a previous state but may also be related to a future state. The DuBiLSTM can yield an even more slightly improved overall performance because of the benefits accruing from the enhancement of the precision and recall of both positive and negative categories. This indicates that the dual-channel framework is able to generate more shallow linguistic features for BiLSTM. It is noteworthy that our method achieves the best performance. As the generated linguistic pattern can examine the content of sentences to identify dengue case information, it does not conflict with the DuBiLSTM, which analyzes syntactic and semantic information in the sentences.

As a consequence, combining BiLSTM and DuBiLSTM improves the system performance and achieves a remarkable performance on the Taipei Medical University Dengue Case Corpus. The high proportion of dengue case information expressions can be identified by the generated linguistic patterns. For instance, the positive sentence “Taiwan recorded another 7 cases of dengue fever [Wednesday, September 5, 2018], bringing the total number so far this year to 81 and prompting stronger calls by the relevant authorities for greater public cooperation to prevent the spread of the mosquito-borne disease in the peak season.” is correctly detected as dengue case information sentence through the successful match of the generated pattern *[total]-[dengue]* and *[reported]-[dengue]*. It shows that identifying sentences with matched patterns can enhance the performance to discriminate the dengue case information extraction.

As shown in Figures 3 and 4, we visualized the collocation of words and POS for further observation, where nodes and edges represent the linguistic pattern, with the depth of the edge denoting the weight value (ie, LLR) of the collocation. We can observe the appearance of linguistic patterns in Figure 3, such as *[dengue]-[reported]* and *[locally]-[acquired]*, indicating that the sentence is more likely to be crucial. In addition, we also noticed that the word *case* has radial edges, which suggests that many discriminative linguistic patterns are composed of the word *case*. This is because case-related information often mentions the term *case*. For instance, "Dengue [reported] 100 cases locally acquired; Municipality most affected: Kaohsiung City," which is a very typical example containing accurate case information. In addition, from Figure 4, we observe from the POS pattern network that the more important POS collocations are *[JJS]-[CD]*, where *JJS* stands for *adjective, superlative* and *CD* stands for *cardinal number*, and *[JJS]-[NN]*, where *NN* stands for *noun, singular or mass* (the detailed meanings of POS tags are provided in the URL [63]. This is because the case numbers mostly occur in digit form, and in this situation, the POS tag belongs to *CD*, which is then combined with adjectives and nouns to complete the description of the case information.

Table 3 lists the errors in the sentences of single-clausal and multiclausal types. As shown in the table, the total error rate of the proposed method is 6% (58/965). The individual error rates of single-clausal and multiclausal sentence types are 3.1% (14/456) and 8.6% (44/509), respectively. This indicates that dengue case information in multiclausal sentences is difficult to detect. This is because the syntactic structures of multiclausal sentences are so complex that they confuse the pattern-matching process. As a result, the matched linguistic and POS patterns are prone to errors that affect the correctness of pattern representation and the performance of the corresponding detection. We also observed from the results that a vast proportion of false positives, that is, negative instances incorrectly identified as positive, occurred because the sentences expressed global dengue case information instead of expressing information from the observed country of the alert. For example, our model incorrectly classified "This alarming, particularly considering 96 million cases of symptomatic dengue year worldwide." as a positive sentence. However, the sentence conveys the aggregation of an annual and global pandemic situation, rather than representing a piece of detailed case information, although it contains the highly associated words *case* and *dengue*. For the false negatives, we noticed that a sentence may be split into several fragments because of the writing style. This can result in an incomplete context and thus unclear semantics behind the text. For instance, the sentence "Locality affected: Tainan 32 past week," the digit "32" actually represents the number of cases. However, because the word *cases* is omitted from the text, our model missed the positive sentence, which increased the false-negative rate.

Figure 3. The network visualization for generated linguistic patterns. CDC: Taiwan Centers for Disease Control.

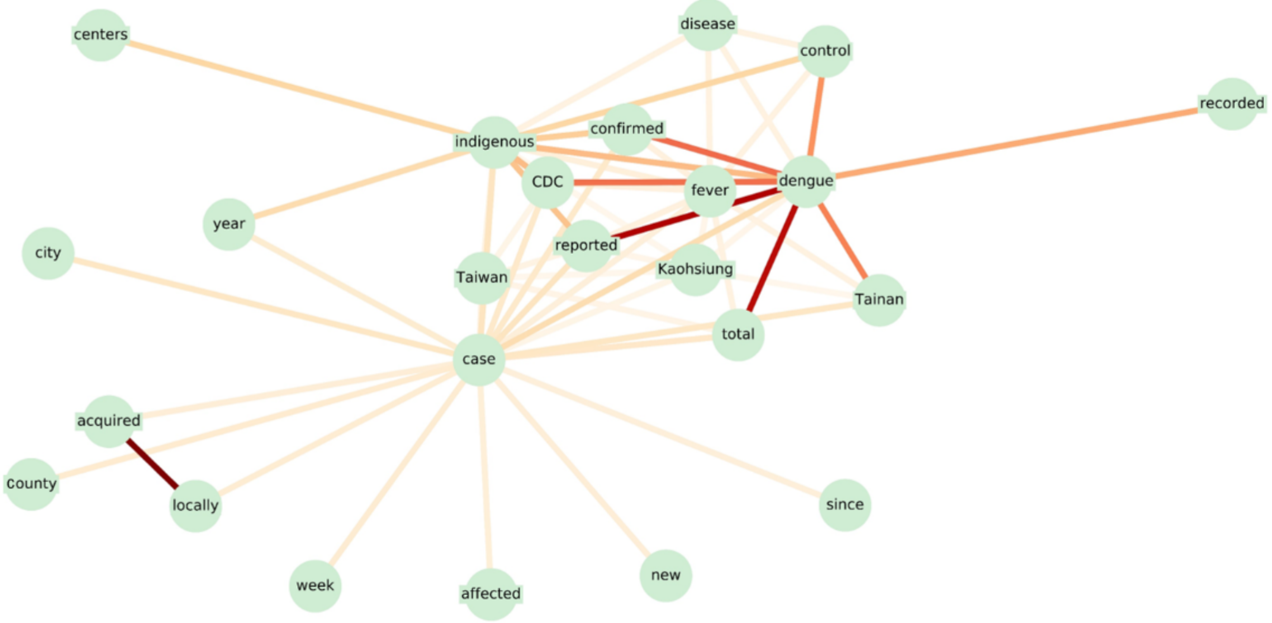


Figure 4. The network visualization for generated parts-of-speech (POS) patterns. CC: coordinating conjunction; CD: cardinal number; DT: determiner; EX: existential there; IN: preposition or subordinating conjunction; JJ: adjective; JJS: adjective, superlative; NN: noun, singular or mass; NNP: proper noun, singular; NNPS: proper noun, plural; NNS: noun, plural; PRP$: possessive pronoun; RB: adverb; RBS: adverb, superlative; TO: to; VBD: verb, past tense; VBG: verb, gerund, or present participle; VBN: verb, past participle; VBP: verb, nonthird person singular present; VBZ: verb, third person singular present; WDT: wh-determiner.

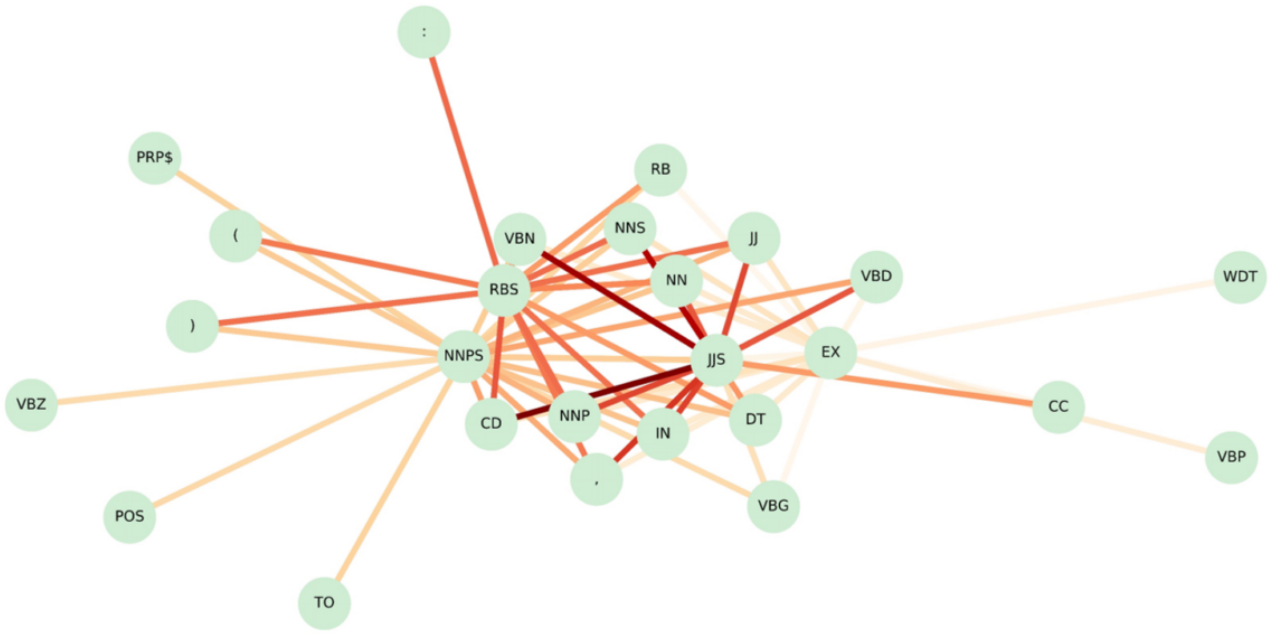


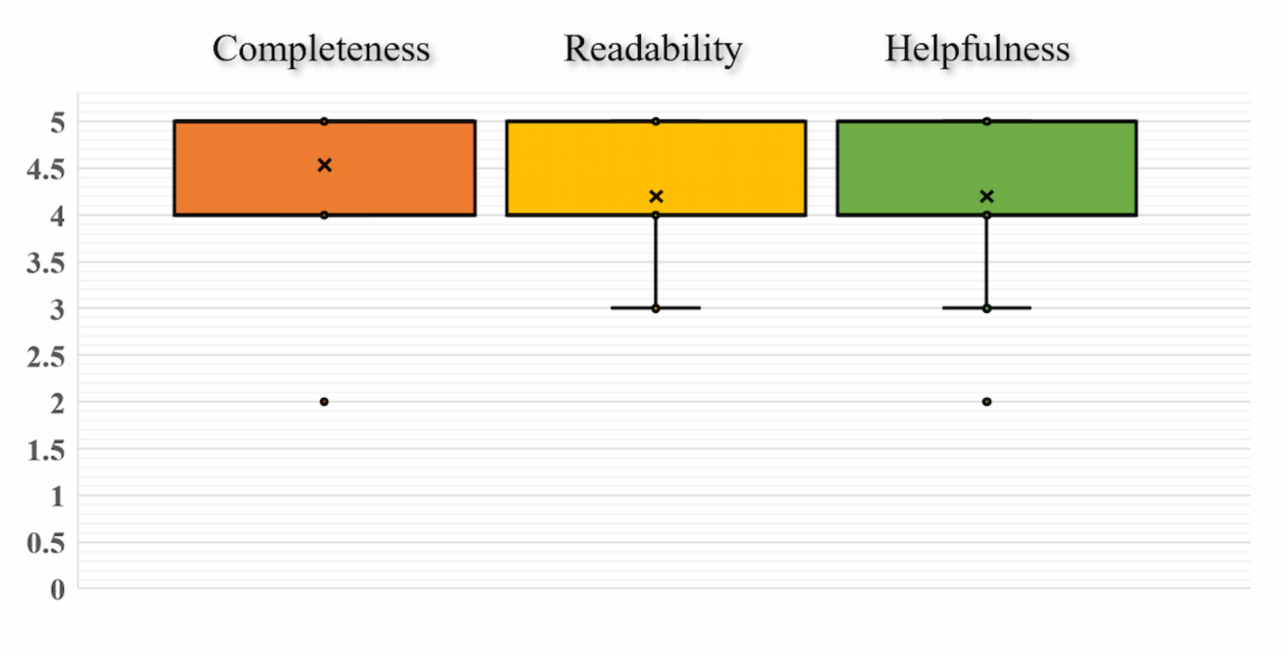
Table 3. Error distribution of dengue case information detection.

Clause type	False positive, n (%)	False negative, n (%)	Error rate, n (%)
Single-clausal (n=456)	7 (1.5)	7 (1.5)	14 (3.1)
Multiclausal (n=509)	29 (5.7)	15 (2.9)	44 (8.6)
Corpus (n=965)	36 (3.7)	22 (2.3)	58 (6)

The goal of this research was to automatically generate a summary of the alerting articles from ProMED-mail to help researchers reduce reading effort and more quickly comprehend the main topic. Given an alerting report, our model can extract crucial sentences that express dengue case information, and these extracted sentences can be combined to form a summary. To estimate the practicality of the proposed model for epidemic monitoring, we conducted a satisfaction analysis experiment to assess the acceptability of the summaries by end users from the medical science field. The survey participants were 2 female students from the School of Public Health, Taipei Medical University; 2 male faculty members from the department of parasitology and tropical diseases, Taipei Medical University; and 1 male internal medicine physician from a Taipei Medical University–affiliated hospital; their ages ranged from 23 to 55 years. They evaluated on a 5-point Likert scale the quality of the summaries generated by our method. The summaries were randomly sampled from 25% (5/20) of the Southeast Asia dengue alerting reports from January 2019 to December 2020. We estimated the average number of words and sentences from both the original alerting reports and the generated summaries and then derived the compression rates at the word level (2.8) and sentence level (3.7). We included three questionnaire items: (1) completeness—*completeness of the generated summary content*, (2) readability—*fluent and easy to read*, and (3) helpfulness—*helps to improve analysis efficiency or reduce text reading time*.

To analyze the bias of the scoring distribution, we used a box plot to illustrate the distribution of the scores in the satisfaction survey (Figure 5). The extended range of the boxed image (including box whiskers) represents the highest to lowest distribution of the 5 scores, and the symbol *x* in the box indicates the mean value of the scores. As we can see, the item *Completeness* has the highest score, which indicates that all the epidemiologists were satisfied with the quality and accuracy of the summary text. The average scores of the remaining 2 items are also >4.2, which demonstrates that our automatic summaries are of high quality. These findings are evidence of the usefulness of this research. Nevertheless, there are still a few outliers with scores of 2 in the satisfaction questionnaire, which means that the epidemiologists were not entirely satisfied with these case summaries. On the basis of further analysis of the results, it was found that this was mainly due to incorrect results of our prediction of multiclausal sentences, which is the main type of error in the proposed model. However, multiclausal sentences typically entail rich information. Therefore, as the quality of the summary is sensitively affected by the incorrect identification of multiclausal sentences, one of our directions for future research is to improve the accuracy of multiclausal sentence prediction.

Figure 5. Box plot of expert assessment on a 5-point Likert scale of the quality of generated summaries.



To summarize, the experiment results from the satisfaction analysis demonstrate that our summary system is helpful for experts and scholars to quickly read and effectively analyze a large number of briefings.

Limitations and Future Directions

This study includes some limitations. The approaches developed in this study mainly focus on extracting sentence information for summarizing or, more specifically, extracting qualitative information from unstructured content. However, this approach is currently unable to acquire precise quantitative information. For instance, the number of incidence cases (newly infected) and cumulative cases cannot be reliably identified using current algorithms. Our future work will therefore focus on this issue by integrating date, location, and identification of the number of cases to retrieve important quantities of disease information. This information can be applied to more advanced spatial and temporal analyses in the future.

The second limitation is that the reporting effort and frequency in ProMED-mail are not consistent because of its volunteer-oriented design. This problem could be overcome by integrating other outbreak-reporting platforms such as the HealthMap project, which provides a visualized platform for various disease alerts [53,64]. Collecting epidemiological surveys from the scientific literature is another approach that can be used to enrich the data set.

Conclusions

The combination of high rates of international travel and rapid environmental changes makes region-wide collaboration in monitoring emerging and re-emerging infectious diseases necessary. The current COVID-19 pandemic has also had a huge impact on the surveillance and control of other infectious diseases [65]. In addition, because ProMED-mail records and follows up undiagnosed diseases in different countries [66], this abundant disease surveillance information is unstructured and is thus not able to be efficiently used by public health workers or scientists. Our proposed deep neural network provides a good way to extract outbreak information from unstructured text, which can then be further analyzed.

In summary, our study built a prototype of an NLP algorithm to retrieve sentence summarizations from the ProMED-mail database. This approach can help medical scientists and public health workers to save more time on content summarization and analysis. Further work will continue to optimize the algorithm to extract more important quantification information.

Acknowledgments

This research was supported by the Ministry of Science and Technology of Taiwan (MOST 108-2638-H-002-002-MY2 and MOST 109-2410-H-038-012-MY2).

Authors' Contributions

YCC and TWC helped to design the study and conceived the research question. YCC, TWC, and YWC conducted the experiments and statistical analyses and reviewed and interpreted the findings. YCC, TWC, and YWC wrote the manuscript, reviewed it, and noted the points of revision.

Conflicts of Interest

None declared.

References

1. Johansson MA, Cummings DA, Glass GE. Multiyear climate variability and dengue--El Niño southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis. PLoS Med 2009 Nov;6(11):e1000168 [FREE Full text] [doi: 10.1371/journal.pmed.1000168] [Medline: 19918363]
2. Kovats RS, Bouma MJ, Hajat S, Worrall E, Haines A. El Niño and health. Lancet 2003 Nov 01;362(9394):1481-1489. [doi: 10.1016/S0140-6736(03)14695-8] [Medline: 14602445]
3. Petersen LR, Jamieson DJ, Powers AM, Honein MA. Zika Virus. N Engl J Med 2016 Apr 21;374(16):1552-1563. [doi: 10.1056/NEJMra1602113] [Medline: 27028561]
4. Patz JA, Olson SH. Malaria risk and temperature: influences from global climate change and local land use practices. Proc Natl Acad Sci U S A 2006 Apr 11;103(15):5635-5636 [FREE Full text] [doi: 10.1073/pnas.0601493103] [Medline: 16595623]
5. Ebi KL, Nealon J. Dengue in a changing climate. Environ Res 2016 Nov;151:115-123 [FREE Full text] [doi: 10.1016/j.envres.2016.07.026] [Medline: 27475051]
6. Chuang T, Soble A, Ntshalintshali N, Mkhonta N, Seyama E, Mthethwa S, et al. Assessment of climate-driven variations in malaria incidence in Swaziland: toward malaria elimination. Malar J 2017 Jun 01;16(1):232 [FREE Full text] [doi: 10.1186/s12936-017-1874-0] [Medline: 28571572]
7. Wesolowski A, Qureshi T, Boni MF, Sundsøy PR, Johansson MA, Rasheed SB, et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. Proc Natl Acad Sci U S A 2015 Sep 22;112(38):11887-11892 [FREE Full text] [doi: 10.1073/pnas.1504964112] [Medline: 26351662]
8. ProMED homepage. ProMED. URL: https://promedmail.org/ [accessed 2022-07-04]
9. Morse SS, Rosenberg BH, Woodall J. ProMED global monitoring of emerging diseases: design for a demonstration program. Health Policy 1996 Dec;38(3):135-153 [FREE Full text] [doi: 10.1016/0168-8510(96)00863-9] [Medline: 10162418]
10. Chase V. ProMED: a global early warning system for disease. Environ Health Perspect 1996 Jul;104(7):699 [FREE Full text] [doi: 10.1289/ehp.104-1469400] [Medline: 8841752]
11. Carrion M, Madoff LC. ProMED-mail: 22 years of digital surveillance of emerging infectious diseases. Int Health 2017 May 01;9(3):177-183 [FREE Full text] [doi: 10.1093/inthealth/ihx014] [Medline: 28582558]
12. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
13. Cai O, Sousa-Pinto B. United States influenza search patterns since the emergence of COVID-19: infodemiology study. JMIR Public Health Surveill 2022 Mar 03;8(3):e32364 [FREE Full text] [doi: 10.2196/32364] [Medline: 34878996]
14. Husnayain A, Chuang T, Fuad A, Su EC. High variability in model performance of Google relative search volumes in spatially clustered COVID-19 areas of the USA. Int J Infect Dis 2021 Aug;109:269-278 [FREE Full text] [doi: 10.1016/j.ijid.2021.07.031] [Medline: 34273513]
15. Castillo-Salgado C. Trends and directions of global public health surveillance. Epidemiol Rev 2010;32(1):93-109. [doi: 10.1093/epirev/mxq008] [Medline: 20534776]
16. Madoff LC. ProMED-mail: an early warning system for emerging diseases. Clin Infect Dis 2004 Jul 15;39(2):227-232. [doi: 10.1086/422003] [Medline: 15307032]
17. Lorthe TS, Pollack MP, Lassmann B, Brownstein JS, Cohn E, Divi N, et al. Evaluation of the EpiCore outbreak verification system. Bull World Health Organ 2018 May 01;96(5):327-334 [FREE Full text] [doi: 10.2471/BLT.17.207225] [Medline: 29875517]
18. Milne-Price S, Miazgowicz KL, Munster VJ. The emergence of the Middle East respiratory syndrome coronavirus. Pathog Dis 2014 Jul;71(2):121-136 [FREE Full text] [doi: 10.1111/2049-632X.12166] [Medline: 24585737]
19. Lessler J, Moore SM, Luquero FJ, McKay HS, Grais R, Henkens M, et al. Mapping the burden of cholera in sub-Saharan Africa and implications for control: an analysis of data across geographical scales. Lancet 2018 May 12;391(10133):1908-1915 [FREE Full text] [doi: 10.1016/S0140-6736(17)33050-7] [Medline: 29502905]
20. Tarnas MC, Desai AN, Lassmann B, Abbara A. Increase in vector-borne disease reporting affecting humans and animals in Syria and neighboring countries after the onset of conflict: a ProMED analysis 2003-2018. Int J Infect Dis 2021 Jan;102:103-109 [FREE Full text] [doi: 10.1016/j.ijid.2020.09.1453] [Medline: 33002614]
21. Chatziprodromidou IP, Arvanitidou M, Guitian J, Apostolou T, Vantarakis G, Vantarakis A. Global avian influenza outbreaks 2010-2016: a systematic review of their distribution, avian species and virus subtype. Syst Rev 2018 Jan 25;7(1):17 [FREE Full text] [doi: 10.1186/s13643-018-0691-z] [Medline: 29368637]
22. Izquierdo JL, Ancochea J, Savana COVID-19 Research Group, Soriano JB. Clinical characteristics and prognostic factors for intensive care unit admission of patients with COVID-19: retrospective study using machine learning and natural language processing. J Med Internet Res 2020 Oct 28;22(10):e21801 [FREE Full text] [doi: 10.2196/21801] [Medline: 33090964]

23. Jiao Y, Sharma A, Ben Abdallah A, Maddox TM, Kannampallil T. Probabilistic forecasting of surgical case duration using machine learning: model development and validation. J Am Med Inform Assoc 2020 Dec 09;27(12):1885-1893 [FREE Full text] [doi: 10.1093/jamia/ocaa140] [Medline: 33031543]
24. Manning C, Schutze H. Foundations of Statistical Natural Language Processing. Cambridge, Massachusetts, United States: MIT press; 1999.
25. Kim M, Chae K, Lee S, Jang H, Kim S. Automated classification of online sources for infectious disease occurrences using machine-learning-based natural language processing approaches. Int J Environ Res Public Health 2020 Dec 17;17(24):9467 [FREE Full text] [doi: 10.3390/ijerph17249467] [Medline: 33348764]
26. Fernandes P, Allamanis M, Brockschmidt M. Structured neural summarization. In: Proceedings of the Structured neural summarization. 2019 Presented at: The International Conference on Learning Representations (ICLR); May 6 - 9, 2019; New Orleans URL: https://arxiv.org/abs/1811.01824
27. Liu G, Guo J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing 2019 Apr;337:325-338. [doi: 10.1016/j.neucom.2019.01.078]
28. Murray NE, Quam MB, Wilder-Smith A. Epidemiology of dengue: past, present and future prospects. Clin Epidemiol 2013;5:299-309 [FREE Full text] [doi: 10.2147/CLEP.S34440] [Medline: 23990732]
29. Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ, et al. Dengue: a continuing global threat. Nat Rev Microbiol 2010 Dec;8(12 Suppl):S7-16 [FREE Full text] [doi: 10.1038/nrmicro2460] [Medline: 21079655]
30. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature 2013 Apr 25;496(7446):504-507 [FREE Full text] [doi: 10.1038/nature12060] [Medline: 23563266]
31. Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, Hoen AG, et al. Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS Negl Trop Dis 2012;6(8):e1760 [FREE Full text] [doi: 10.1371/journal.pntd.0001760] [Medline: 22880140]
32. Medical record. Wikipedia. URL: https://en.wikipedia.org/wiki/Medical_record [accessed 2022-07-04]
33. Zhai C, Lafferty J. A study of smoothing methods for language models applied to ad hoc information retrieval. SIGIR Forum 2017 Aug 02;51(2):268-276. [doi: 10.1145/3130348.3130377]
34. Krallinger M, Rabal O, Lourenço A, Oyarzabal J, Valencia A. Information retrieval and text mining technologies for chemistry. Chem Rev 2017 Jun 28;117(12):7673-7761. [doi: 10.1021/acs.chemrev.6b00851] [Medline: 28475312]
35. Marcos-Pablos S, García-Peñalvo FJ. Information retrieval methodology for aiding scientific database search. Soft Comput 2018 Oct 12;24(8):5551-5560. [doi: 10.1007/s00500-018-3568-0]
36. Ricardo BY, Berthier RN. Modern Information Retrieval. London, United Kingdom: Pearson Education; 1999.
37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310]
38. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012;22(3):276-282. [doi: 10.11613/bm.2012.031]
39. Guba E, Lincoln Y. Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approaches. San Francisco, California, United States: Jossey-Bass; Sep 03, 2016.
40. TMUDCC: Taipei Medical University Dengue Case Corpus. Taipei Medical University. URL: http://nlp.tmu.edu.tw/ProMED/TMUDCC/index.html [accessed 2022-07-04]
41. Natural Language Toolkit. NLTK. URL: https://www.nltk.org/ [accessed 2022-07-04]
42. Schütze H, Manning C, Raghavan P. Introduction to Information Retrieval. Cambridge: Cambridge University Press; 2008.
43. Cheng Y, Chen Y, Yeh W, Chang Y. Valence and arousal-infused bi-directional LSTM for sentiment analysis of government social media management. Applied Sci 2021 Jan 19;11(2):880. [doi: 10.3390/app11020880]
44. Chen C, Warikoo N, Chang YC, Chen JH, Hsu WL. Medical knowledge infused convolutional neural networks for cohort selection in clinical trials. J Am Med Inform Assoc 2019 Nov 01;26(11):1227-1236 [FREE Full text] [doi: 10.1093/jamia/ocz128] [Medline: 31390470]
45. Wang H, Shi H, Lin K, Qin C, Zhao L, Huang Y, et al. A high-precision arrhythmia classification method based on dual fully connected neural network. Biomedical Signal Process Control 2020 Apr;58:101874. [doi: 10.1016/j.bspc.2020.101874]
46. Hussain S, Mokhtar M, Howe JM. Sensor failure detection, identification, and accommodation using fully connected cascade neural network. IEEE Trans Ind Electron 2015 Mar;62(3):1683-1692. [doi: 10.1109/tie.2014.2361600]
47. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986 Oct;323(6088):533-536. [doi: 10.1038/323533a0]
48. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(1):1929-1958. [doi: 10.5555/2627435.2670313]
49. Kingma D, Ba JL. Adam: a method for stochastic optimization. In: Proceedings of the ICLR 2015. 2015 Presented at: ICLR 2015; May 7-9, 2015; San Diego, CA, USA.
50. Shore J, Johnson R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. IEEE Trans Inform Theory 1980 Jan;26(1):26-37. [doi: 10.1109/tit.1980.1056144]
51. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. Burlington, Massachusetts, United States: Morgan Kaufmann; 2012.

52. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13 - 17, 2016; San Francisco, California, USA URL: https://dl.acm.org/doi/10.1145/2939672.2939785 [doi: 10.1145/2939672.2939785]
53. Fan J, Wang X, Wu L, Zhou H, Zhang F, Yu X, et al. Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: a case study in China. Energy Conversion Manag 2018 May;164:102-111. [doi: 10.1016/j.enconman.2018.02.087]
54. Rossi F, Conan-Guez B. Functional multi-layer perceptron: a non-linear tool for functional data analysis. Neural Netw 2005 Jan;18(1):45-60. [doi: 10.1016/j.neunet.2004.07.001] [Medline: 15649661]
55. Liao S, Wang J, Yu R, Sato K, Cheng Z. CNN for situations understanding based on sentiment analysis of twitter data. Procedia Comput Sci 2017;111:376-381. [doi: 10.1016/j.procs.2017.06.037]
56. Kim Y. Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); Oct 25–29, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1181]
57. Rish I. An empirical study of the naive Bayes classifier. In: Proceedings of the IJCAI 2001 workshop on empirical methods in artificial intelligence. 2001 Presented at: IJCAI 2001 workshop on empirical methods in artificial intelligence; Aug 4 - 10, 2001; Seattle, Washington, USA URL: https://dominoweb.draco.res.ibm.com/db24eb109a77428785256aff005d3df2.html
58. Hand DJ, Yu K. Idiot's Bayes—not so stupid after all? Int Stat Rev 2001 Dec;69(3):385-398. [doi: 10.1111/j.1751-5823.2001.tb00465.x]
59. Webb GI, Boughton JR, Wang Z. Not so Naive Bayes: aggregating one-dependence estimators. Mach Learn 2005 Jan;58(1):5-24. [doi: 10.1007/s10994-005-4258-6]
60. Mitchell T. Machine Learning. New York, United States: McGraw-Hill Education; 1997.
61. Keras homepage. Keras. URL: https://keras.io/ [accessed 2022-07-04]
62. Hull D. Using statistical testing in the evaluation of retrieval experiments. In: Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval. 1993 Presented at: SIGIR93: 16th International ACM/SIGIR '93 Conference on Research and Development in Information Retrieval; Jun 27 - Jul 1, 1993; Pittsburgh Pennsylvania USA. [doi: 10.1145/160688.160758]
63. Alphabetical list of part-of-speech tags used in the Penn Treebank Project. Penn Treebank. URL: https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html [accessed 2022-07-04]
64. Nelson R. HealthMap: the future of infectious diseases surveillance? Lancet Infectious Diseases 2008 Oct 21;8(10):596. [doi: 10.1016/S1473-3099(08)70219-6]
65. Valentin S, Mercier A, Lancelot R, Roche M, Arsevska E. Monitoring online media reports for early detection of unknown diseases: insight from a retrospective study of COVID-19 emergence. Transbound Emerg Dis 2021 May;68(3):981-986 [FREE Full text] [doi: 10.1111/tbed.13738] [Medline: 32683774]
66. Rolland C, Lazarus C, Giese C, Monate B, Travert A, Salomon J. Early detection of public health emergencies of international concern through undiagnosed disease reports in ProMED-Mail. Emerg Infect Dis 2020 Feb;26(2):336-339 [FREE Full text] [doi: 10.3201/eid2602.191043] [Medline: 31961311]

Abbreviations

BiLSTM: bidirectional long short-term memory
CNN: convolutional neural network
DHF: dengue hemorrhagic fever
DT: decision tree
DuBiLSTM: dual-channel bidirectional long short-term memory
LLR: log-likelihood ratio
LSTM: long short-term memory
MLP: multilayer perceptron
NLP: natural language processing
POS: parts of speech
ProMED: Program for Monitoring Emerging Diseases
ReLU: rectified linear unit
RF: random forest
SVM: support vector machine
XGB: extreme gradient boosting

Edited by H Bradley; submitted 30.10.21; peer-reviewed by K Chih Hao, S Doan; comments to author 23.02.22; revised version received 15.04.22; accepted 27.05.22; published 13.07.22.

Please cite as:
Chang YC, Chiu YW, Chuang TW
Linguistic Pattern–Infused Dual-Channel Bidirectional Long Short-term Memory With Attention for Dengue Case Summary Generation From the Program for Monitoring Emerging Diseases–Mail Database: Algorithm Development Study
JMIR Public Health Surveill 2022;8(7):e34583
URL: https://publichealth.jmir.org/2022/7/e34583
doi:10.2196/34583
PMID:35830225

Original Paper

Cost-Effectiveness of Lung Cancer Screening Using Low-Dose Computed Tomography Based on Start Age and Interval in China: Modeling Study

Zixuan Zhao[1*], PhD; Lingbin Du[2*], MPH, PhD; Yuanyuan Li[3], MD, PhD; Le Wang[2], PhD; Youqing Wang[2], MD; Yi Yang[1], PhD; Hengjin Dong[1], PhD

[1]Department of Science and Education of the Fourth Affiliated Hospital, Center for Health Policy Studies, School of Public Health, Zhejiang University School of Medicine, Hangzhou, China

[2]Department of Cancer Prevention, Cancer Hospital of the University of Chinese Academy of Sciences, Zhejiang Cancer Hospital, Hangzhou, China

[3]Department for Science and Education, Hangzhou Ninth People's Hospital, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Hengjin Dong, PhD
Department of Science and Education of the Fourth Affiliated Hospital, Center for Health Policy Studies
School of Public Health
Zhejiang University School of Medicine
No. 866 Yuhangtang Road, Xihu District
Hangzhou, 310058
China
Phone: 86 13221076129
Email: donghj@zju.edu.cn

Related Article:

This is a corrected version. See correction statement: https://publichealth.jmir.org/2022/10/e43025

Abstract

Background: Lung cancer is the most commonly diagnosed cancer and the leading cause of cancer-related death in China. The effectiveness of screening for lung cancer has been reported to reduce lung cancer–specific and overall mortality, although the cost-effectiveness, optimal start age, and screening interval remain unclear.

Objective: This study aimed to assess the cost-effectiveness of lung cancer screening among heavy smokers in China by incorporating start age and screening interval.

Methods: A Markov state-transition model was used to assess the cost-effectiveness of a lung cancer screening program in China. The evaluated screening strategies were based on a screening start age of 50-74 years and a screening interval of once or annually. Transition probabilities were obtained from the literature and validated, while cost parameters were derived from databases of local medical insurance bureaus. A societal perspective was adopted. The outputs of the model included costs, quality-adjusted life years (QALYs), and lung cancer–specific mortality, with future costs and outcomes discounted by 5%. A currency exchange rate of 1 CNY=0.1557 USD is applicable. The incremental cost-effectiveness ratio (ICER) was calculated for different screening strategies relative to nonscreening.

Results: The proposed model suggested that screening led to a gain of 0.001-0.042 QALYs per person as compared with the findings in the nonscreening cohort. Meanwhile, one-time and annual screenings were associated with reductions in lung cancer–related mortality of 0.004%-1.171% and 6.189%-15.819%, respectively. The ICER ranged from 119,974.08 to 614,167.75 CNY per QALY gained relative to nonscreening. Using the World Health Organization threshold of 212,676 CNY per QALY gained, annual screening from a start age of 55 years and one-time screening from the age of 65 years can be considered as cost-effective in China. Deterministic and probabilistic sensitivity analyses were conducted.

Conclusions: This economic evaluation revealed that a population-based lung cancer screening program in China for heavy smokers using low-dose computed tomography was cost-effective for annual screening of smokers aged 55-74 years and one-time screening of those aged 65-74 years. Moreover, annual lung cancer screening should be promoted in China to realize the benefits of a guideline-recommended screening program.

(JMIR Public Health Surveill 2022;8(7):e36425) doi:10.2196/36425

KEYWORDS

cost-effectiveness analysis; low-dose computed tomography; screening; lung cancer; China

Introduction

Lung cancer is a leading cause of death in China and globally. The incidence of lung cancer has recently increased dramatically, both in urban and rural areas, and it is currently the most common form of cancer in China. According to the National Central Cancer Registry of China, in 2015, the incidence of lung cancer was 57.26 cases/100,000 persons and the associated mortality rate was 45.87 deaths/100,000 persons, accounting for 20% and 27% of the values for all cancers, respectively [1]. At present, about 70%-75% of lung cancer patients are diagnosed in the middle or advanced stage of the disease [2]. Although there has been remarkable progress in treatment, the 5-year survival rate of patients with advanced lung cancer (stage IV) remains poor, at only 4.2% [3]. A previous study reported that surgical resection in the early stage of lung cancer (stage I) could significantly improve the 10-year survival rate to 92% [4]. Moreover, the disease burden of lung cancer in China is expected to substantially increase labor costs and medical expenditure in the near future. Therefore, promoting prevention, early diagnosis, and timely treatment can improve the prognosis and reduce the disease burden of lung cancer in China.

The effectiveness of low-dose computed tomography (LDCT) for the screening of lung cancer has been confirmed by the National Lung Screening Trial conducted at 33 medical centers in the United States [5]; the UK Lung Cancer RCT Pilot Screening Trial [6], a randomized controlled trial of LDCT screening for lung cancer versus usual care; and the Detection of Lung Cancer Through Low-dose CT Screening Trial conducted by the Dutch Cancer Society [7]. Several other studies have been conducted to explore the cost-effectiveness of lung cancer screening, although most were conducted in the United States and Europe. These studies reported notable differences in disease burden and treatment costs as compared with the findings in China. For example, the incremental cost-effectiveness ratio (ICER) has been reported to be €19,302 (US $22,542) per life year gained and €30,291 (US $35,377) per quality-adjusted life year (QALY) gained in Germany [8], while the ICER has been reported to be US $52,000 per life year gained and US $81,000 per QALY gained in the United States [9]. However, there has been only 1 similar study conducted in China, but this was limited to early versus nonearly lung cancer, which could have underestimated screening effectiveness [10]. In addition, risk factors and the epidemiology of lung cancer differ among countries. Therefore, the aim of this study was to evaluate the cost-effectiveness of LDCT for the screening of lung cancer in China from a societal perspective.

Methods

Study Design

This study was conducted in 2 steps. In the first step, a Markov state-transition model with a lifetime horizon was used to mimic the natural progression of lung cancer and assess the potential impact of LDCT screening compared with a lack of screening in a Chinese cohort aged 50 to 74 years. In the second step, the Markov state-transition model combined with real-world data was used to estimate the ICER of each specific screening strategy as compared with nonscreening. A discount rate of 5% was applied to the costs of both strategies. Important assumptions in this study are summarized in Textbox 1.

Textbox 1. Summary of key assumptions.

Description of assumptions

- A simulated cohort of heavy smokers at a start age of 50-74 years was assumed to be followed up until the age of 79 years (mean life expectancy in China) or death.
- A heavy smoker in this study was defined as a current smoker who smokes at least 20 pack-years.
- Individuals in the screened cohort were assumed to undergo screening by low-dose computed tomography once or annually, and those with positive screening results were assumed to have undergone diagnostic biopsies.
- While in the maintenance cancerous stages, the maintenance cost by stage was assumed to be 10% of the treatment cost.
- All costs were expressed in CNY (2021; 1 CNY=0.1557 USD).
- Future costs and effectiveness were discounted by 5%.
- Adherence to screening and follow-up was assumed to be 100%.

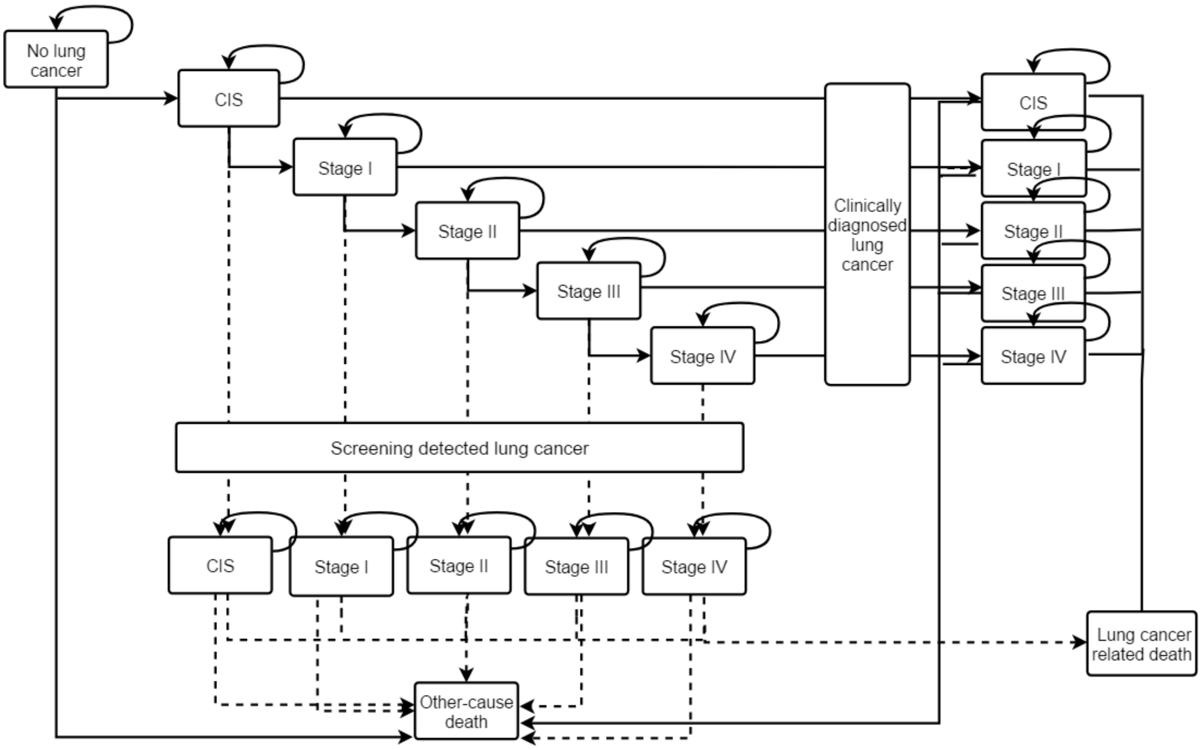
Study Population

The model simulated a cohort of 100,000 heavy smokers in China aged 50 to 74 years until the age of 79 years or death. A heavy smoker was defined as a current smoker who smokes at least 20 pack-years according to the China National Lung Cancer Screening Guidelines with LDCT (2018 version) [11] and the Cost-effectiveness Evaluation of the 2021 US Preventive Services Task Force Recommendation for Lung Cancer Screening [12].

Markov Model and Transition Probabilities

Lung cancer is assumed to progress sequentially from less advanced to more advanced preclinical stages, as depicted in Figure 1. The following 5 stages are distinguished based on the American Joint Committee on Cancer (AJCC) Cancer Staging Manual, 8th edition: carcinoma in situ (CIS), stage I, stage II, stage III, and stage IV [13]. In this study, stages IA, IB, IIIA, and IIIB were not considered because data were not available for clinical practice in population-based cancer registries in China [14]. The probability of deterioration from a healthy state to all-cause death was retrieved from the 2010 Population Census of the People’s Republic of China [15]. The probability of lung cancer–specific death was retrieved from the published literature [16]. Parameters of disease progression from a healthy state to lung cancer were based on the incidence of lung cancer among smokers in China [17]. The incidence of smokers was modeled as a multiplicative function of smoking rate, age, and sex-specific parameters. The incidence of lung cancer in the general population by sex and age (I_G) served as the baseline incidence. Specifically, the incidence of lung cancer for smokers (I_S) was modeled as $I_S = OR * I_G / (1 + (OR – 1)) * R_S$, and the incidence of lung cancer for nonsmokers (I_N) was modeled as $I = I_S \times R + I_N \times (1 – R)$, where OR is the odds ratio for the incidence of lung cancer in smokers, which was extracted from a previous publication [16], R is the proportion of smokers by sex and age reported in the Global Adult Tobacco Survey [18], and I is the incidence of lung cancer in the general population of China. Finally, the incidence of lung cancer (I_{20}) among smokers in China was modeled as $I_{20} = I_N \times RR$, where the relative risk (RR) of lung cancer (>20 pack-years) attributable to smoking was derived from the published literature [19].

Figure 1. Schematic diagram of natural history for lung cancer screening. CIS: carcinoma in situ.



Individuals in the nonscreened cohort were diagnosed based on symptoms. The probability of progression to a more advanced stage of lung cancer or a clinical diagnosis, as described by Ten Haaf et al [20] and a hospital-based multi-center retrospective clinical epidemiological survey in China [21], is detailed in Table 1 [15-19,21-26]. Overall, 19.0% of lung cancer cases were clinically detected in stage I, 16.5% in stage II, 34.7% in stage III, and 29.9% in stage IV.

It was assumed that patients in the screened cohort underwent screening by LDCT at least once or annually and those with positive results underwent additional testing, including biopsy. The positive result rate and proportion of lung cancer by stage were derived from the Wenling lung cancer screening program, which was initiated in 2018 to conduct annual LDCT screening of local high-risk populations over a 3-year period. Of 10,175 asymptomatic individuals who were screened in 2018, 65 (0.64%) were diagnosed with lung cancer (Table 1). Annual screening was conducted in accordance with the protocol of the Cancer Screening Program in Urban China to determine the morphology and size of nodules [22]. The specificity and sensitivity of LDCT for screening of lung cancer were derived from the results of the Multicenter Italian Lung Detection trial [23]. The probability of progression to a more advanced stage or a maintenance state is detailed by stage in Table 1, as described in previous studies [16,22,24]. The proposed model was validated by comparing key outcomes to external empirical data that were not used for model development (Multimedia Appendix 1).

Table 1. Input parameters of the Markov model for lung cancer screening.

Variable	Base case value			Distribution	Source
	Male	Female	Overall		
Lung cancer incidence in the general population (per 100,000 persons) by age in years					
50-54	81.0559	89.6626	N/A[a]	Beta	[17]
55-59	162.0833	112.4574	N/A	Beta	[17]
60-64	256.0943	154.6871	N/A	Beta	[17]
65-69	373.6808	190.2521	N/A	Beta	[17]
70-74	498.0681	242.6310	N/A	Beta	[17]
Smoking rate in the general population					
50-64	0.60	0.04	N/A	Beta	[18]
65-74	0.45	0.07	N/A	Beta	[18]
RR[b] (>20 pack-years)	N/A	N/A	3.87	Beta	[19]
Proportion of lung cancer by stage (nonscreened cohort)					
CIS[c]	N/A	N/A	0.000	Beta	[21]
I	N/A	N/A	0.190	Beta	[21]
II	N/A	N/A	0.165	Beta	[21]
III	N/A	N/A	0.346	Beta	[21]
IV	N/A	N/A	0.299	Beta	[21]
Proportion of lung cancer by stage (LDCT[d] screened cohort)					Wenling lung cancer screening program
CIS	N/A	N/A	0.0370	Beta	N/A
I	N/A	N/A	0.6852	Beta	N/A
II	N/A	N/A	0.0370	Beta	N/A
III	N/A	N/A	0.1852	Beta	N/A
IV	N/A	N/A	0.0556	Beta	N/A
Sensitivity of LDCT (%)	N/A	N/A	79	Beta	[23]
Specificity of LDCT (%)	N/A	N/A	81	Beta	[23]
Mortality of all-cause death (%) by age group					
50-54	N/A	N/A	3.59	Beta	[15]
55-59	N/A	N/A	4.73	Beta	[15]
60-64	N/A	N/A	8.19	Beta	[15]
65-69	N/A	N/A	12.99	Beta	[15]
70-74	N/A	N/A	21.08	Beta	[15]
Lung cancer mortality rate in the general population (per 100,000 persons) by age group					
50-54	N/A	N/A	28.81	Beta	[16]
55-59	N/A	N/A	52.86	Beta	[16]
60-64	N/A	N/A	101.93	Beta	[16]
65-69	N/A	N/A	153.34	Beta	[16]
70-74	N/A	N/A	248.57	Beta	[16]
Transition probabilities (1 year)					
Lung cancer stage CIS to I	N/A	N/A	0.0980	Beta	[24]

Variable	Base case value			Distribution	Source
	Male	Female	Overall		
Lung cancer stage I to II	N/A	N/A	0.3682	Beta	[22]
Lung cancer stage I to III	N/A	N/A	0.0328	Beta	[22]
Lung cancer stage I to IV	N/A	N/A	0.0745	Beta	[22]
Lung cancer stage II to III	N/A	N/A	0.2260	Beta	[22]
Lung cancer stage II to IV	N/A	N/A	0.1510	Beta	[22]
Lung cancer stage III to IV	N/A	N/A	0.1455	Beta	[22]
Lung cancer stage CIS to death	N/A	N/A	0.00	Beta	[16]
Lung cancer stage I to death	N/A	N/A	0.04	Beta	[16]
Lung cancer stage II to death	N/A	N/A	0.07	Beta	[16]
Lung cancer stage III to death	N/A	N/A	0.13	Beta	[16]
Lung cancer stage IV to death	N/A	N/A	0.18	Beta	[16]
Utility by stage					
CIS	N/A	N/A	0.87	Beta	[25]
I	N/A	N/A	0.84	Beta	[26]
II	N/A	N/A	0.84	Beta	[26]
III	N/A	N/A	0.87	Beta	[26]
IV	N/A	N/A	0.75	Beta	[26]
Cost (CNY[e])					Survey data
Direct screening cost	N/A	N/A	245.86	Gamma	N/A
Indirect screening cost	N/A	N/A	23.07	Gamma	N/A
Prediagnosis cost	N/A	N/A	628.36	Gamma	N/A
Biopsy diagnosis cost	N/A	N/A	1232.44	Gamma	N/A
Treatment cost by stage					
CIS	N/A	N/A	47,341.85	Gamma	N/A
I	N/A	N/A	53,344.51	Gamma	N/A
II	N/A	N/A	83,365.95	Gamma	N/A
III	N/A	N/A	90,643.18	Gamma	N/A
IV	N/A	N/A	116,471.34	Gamma	N/A

[a] N/A: not applicable.

[b] RR: relative risk.

[c] CIS: carcinoma in situ.

[d] LDCT: low-dose computed tomography.

[e] A currency exchange rate of 1 CNY=0.1557 USD is applicable.

Cost Data

The total cost of the screening program included direct expenses (ie, public advertising, management of screening invitations, salaries of staff members, and depreciation of screening equipment) and indirect expenses (ie, transportation and wages for missed work). In addition, the cost of diagnostic biopsies for participants with positive results after initial LDCT was considered. Screening-related costs were retrieved from data provided by the Wenling lung cancer screening program. Costs of treatment of lung cancer by stage were derived from a database of local medical insurance bureaus, which included 4947 patients and 107,248 relevant records. The cost of maintenance by stage accounted for 10% of the total treatment cost. All costs in this study are expressed in Chinese yuan (CNY) at a discount of 5% of rates in 2018. A currency exchange rate of 1 CNY=0.1557 USD is applicable.

Quality of Life

The putative benefit of cancer screening for early diagnosis was assumed to be a difference in life expectancy and QALY after treatment. As the severity and responsiveness to treatment vary according to stage, the specified utility score for each stage was used for calculation [25,27]. The utility score was 0.84 for lung cancer stage I/II, 0.87 for CIS and stage III, and 0.75 for stage Ⅳ (Table 1).

Evaluation Strategies

As the scheduled screening program included several key characteristics, different combinations of screening intervals and start ages, as well as a nonscreening cohort, were evaluated (Table 2). In order to achieve more realistic economic evaluation outcomes, one-time screening was applied in this study because no periodic screening program has been implemented nationwide in China and most of the study participants were screened for lung cancer only once. Therefore, the rationale of one-time screening was based on limited financial support for lung cancer screening programs in China. Moreover, strategies with annual screening from different start ages were simulated to determine whether efforts are needed to promote periodic screening programs in China in order to realize relative benefits based on current guidelines.

Table 2. Evaluation strategies.

Scenario	Screening tool	Screening interval	Start age (years)
LDCT[a]#1	LDCT	Annual	50, 55, 60, 65, and 70
LDCT#2	LDCT	One time	50, 55, 60, 65, and 70
Nonscreening	N/A[b]	N/A	50, 55, 60, 65, and 70

[a]LDCT: low-dose computed tomography.

[b]N/A: not applicable.

Outcomes and Cost-Effectiveness

The main outcomes of the cost-effectiveness analysis for each strategy were QALYs and total costs. The ICER was calculated by dividing the incremental costs by the incremental QALYs gained for each screening strategy as compared to nonscreening. In China, there is no regulated or published cost-effectiveness threshold. Hence, the threshold recommended by the World Health Organization (WHO) is commonly used. Given that 3 times the gross domestic product per capita was used as a reference point, a tentative threshold value of 212,676 CNY was adopted in this study.

Sensitivity Analysis

The Markov state-transition model was developed using TreeAge Pro 2021 software (TreeAge Software, Inc). The parameters of direct screening cost, maintenance cost, discount rate, consumer price index (CPI) rate, incidence rate of heavy smokers, and specificity and sensitivity of LDCT uncertainty were investigated by 1-way deterministic sensitivity analyses. The costs of direct screening, as well as maintenance costs, CPI rate, and incidence rate of heavy smokers, were set to vary by 30% as compared to base values. The discount rate was set to range from 0% to 8%, and the sensitivity and specificity of LDCT were set to range from 0.63 to 0.95 and 0.65 to 0.97, respectively. Input parameters were randomly drawn from beta or gamma distributions (Table 1).

Results

The results of the model suggested that the QALYs of the screening cohort increased by 0.001 to 0.042 as compared to that of the nonscreening cohort. The reduction in lung cancer–associated mortality ranged from 0.004% to 1.171% for one-time screening and from 6.189% to 15.819% for annual screening (Table 3). The average costs per person in the nonscreening cohort, one-time screening cohort, and annual screening cohort were 24,896.93, 25,521.61, and 34,105.70 CNY, respectively, at a start age of 50 years, which seemed to be the most noncost-effective among the 5 age groups. Conversely, the most cost-effective start age was 70 years, with ICERs in the one-time screening and annual screening cohorts of 180,280.19 and 119,974.08 CNY per QALY gained. As compared to the nonscreening cohort, the ICER of the screening cohort, regardless of the screening interval, ranged from 119,974.08 to 614,167.75 CNY per QALY gained. Using the WHO threshold of 212,676 CNY per QALY gained, annual screening at a start age of 55-74 years was determined to be the most cost-effective in China. For one-time screening, the cost-effective start age was 65-74 years.

The sensitivity of the model for the above-mentioned parameters is shown in Figure 2. Generally, the model results were robust with no variation exceeding 212,676 CNY per QALY gained at a start age of 65-74 years. The highest sensitivity was observed for the rate of newly developed lung cancer in heavy smokers. The accuracy parameters of LDCT (ie, sensitivity and specificity) and the direct cost of the screening program had relatively high influences on the ICERs, while variations in discount rates had relatively little influence. After 10,000 repetitions, Monte Carlo simulation revealed that the average ICER ranged from 143,253.62 to 776,678.97 CNY, which was greater than the base ICER (Figure 3).

Table 3. Base case results with different screening settings (per 100,000 persons).

Start age and strategy[a]	Cost (CNY,[b] millions)	QALYs[c] (10,000 years)	Lung cancer mortality reduction vs nonscreening (%)	ICER[d] Scr vs Non_scr	ICER Scr_annu vs Scr_once
50 years					
Non_scr	2489.69	135.92	N/A[e]	N/A	N/A
Scr_once	2552.16	135.93	0.0041	614,167.75	N/A
Scr_annu	3410.57	136.30	6.1886	245,746.19	235,467.06
55 years					
Non_scr	2380.25	121.21	N/A	N/A	N/A
Scr_once	2448.97	121.23	0.0145	365,289.96	N/A
Scr_annu	3176.64	121.62	6.7044	192,119.62	183,886.78
60 years					
Non_scr	2154.69	104.08	N/A	N/A	N/A
Scr_once	2230.61	104.11	0.0467	263,083.31	N/A
Scr_annu	2808.69	104.50	7.7816	154,401.89	146,456.38
65 years					
Non_scr	1773.45	84.40	N/A	N/A	N/A
Scr_once	1860.84	84.45	0.1997	192,574.66	N/A
Scr_annu	2260.66	84.77	10.0628	131,284.57	122,745.38
70 years					
Non_scr	1184.22	61.56	N/A	N/A	N/A
Scr_once	1279.79	61.61	1.1705	180,280.19	N/A
Scr_annu	1476.25	61.80	15.8193	119,974.08	103,182.45

[a] Non_scr: nonscreening; Scr_once: one-time screening; Scr_annu: annual screening.

[b] A currency exchange rate of 1 CNY=0.1557 USD is applicable.

[c] QALY: quality-adjusted life year.

[d] ICER: incremental cost-effectiveness ratio.

[e] N/A: not applicable.

Figure 2. Tornado diagrams. The tornado diagrams illustrate the change in the incremental cost-effectiveness ratio (ICER). The blue column shows the impact of decreasing the input parameters on the results. Similarly, the red column shows the impact of increasing the input parameters on the results. CPI: consumer price index; EV: expected value; LDCT: low-dose computed tomography.

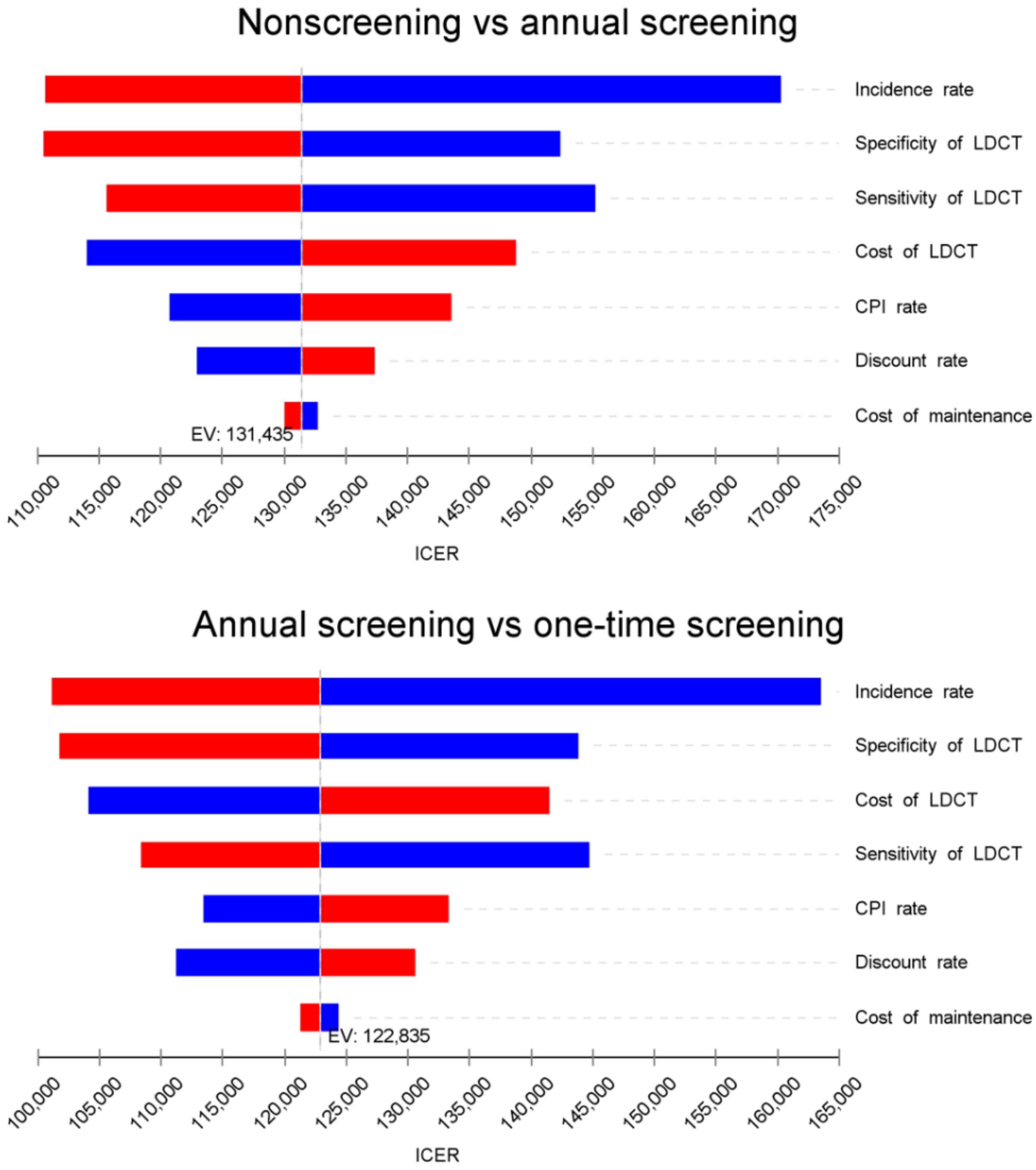
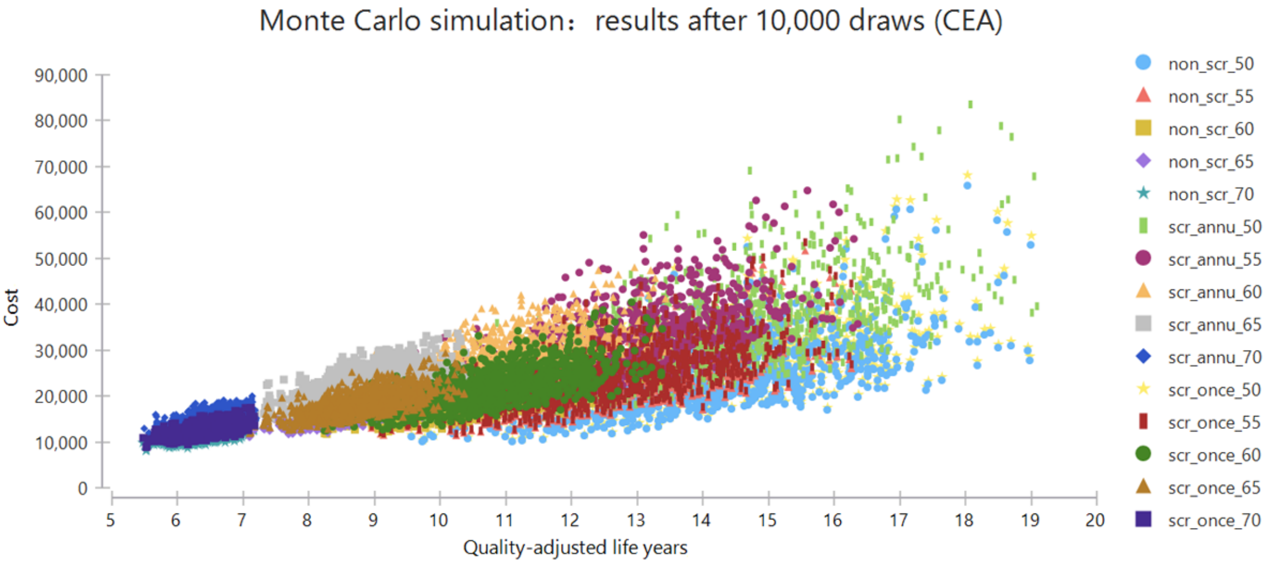


Figure 3. Probabilistic sensitivity analyses. The screening strategies are labeled as follows: screening or not screening interval_start age. CEA: cost-effectiveness analysis; non_scr: nonscreening; scr_annu: annual screening; scr_once: one-time screening.



Discussion

Principal Findings

This is the first cost-effectiveness analysis of a lung cancer screening program with different start ages and screening intervals using real-world data in China. In summary, using a lifetime societal perspective for one-time or annual LDCT for screening of heavy smokers, the annual screening strategy with a start age of 55-74 years showed strong dominance as compared with the nonscreening strategy. These results were sensitive to the rate of newly developed lung cancer and the specificity of LDCT. As compared with the nonscreening strategy, the one-time screening strategy was cost-effective for patients aged 65-74 years, using a cost-effectiveness threshold of 212,676 CNY per QALY gained. This finding is consistent with that in the UK Lung Screen trial, which demonstrated a long-term benefit from a single screen and provided potentially important data for inclusion in future modeling studies to optimize the screening interval [26]. All simulated results of probabilistic sensitivity analysis were robust when the main input parameters were varied.

Although the analytical approach was somewhat similar to that in a previous study by Yuan et al [28], the strategies were enriched by adding screening intervals and thus arrived at different conclusions. First, Yuan et al predicted that the ICERs of all screening strategies with a start age of 40-74 years were 3-fold lower than the gross domestic product per capita. However, this result is consistent with only part of the strategies in this study. Second, Yuan et al predicted a minimum ICER at a start age of 65 years, whereas the results of this study demonstrated a decreasing trend in ICER per QALY gained from a start age of 50-74 years, regardless of the screening interval. These differences may have resulted from a combination of several factors. For example, Yuan et al used a discount rate of 3%, while a rate of 5% was adopted in this study in accordance with the China Guidelines for Pharmacoeconomic Evaluations [29], and staging was simplified in the Markov model by ignoring the CIS stage. For comparison between annual screening and nonscreening, the ICER of 119,974 to 245,746 CNY in this study is comparable to previous estimates of US $24,934, US $49,200-96,700, and US $33,825 per QALY gained reported by studies conducted in New Zealand, the United States, and Canada, respectively [30-32].

Regarding the implications of policies related to lung cancer screening, the China National Lung Cancer Screening Guidelines with LDCT (2018 version) [11] were partially confirmed by the recommendations for lung cancer screening and early diagnosis and treatment guidelines in China [33] from a health economic perspective. Though the updated recommendations for lung cancer screening and early diagnosis and treatment guidelines raised the minimum cumulative smoking exposure from 20 to 30 pack-years relative to the 2018 version, the results were robust according to deterministic 1-way sensitivity analysis. In addition to the low utilization of lung cancer screening programs in China, there is a need to improve the accessibility and affordability of population-based screening programs to better capture the full extent of benefits associated with lung cancer screening. Annual lung cancer screening for heavy smokers at a start age of 55-74 years is considered cost-effective in China. Although screening from the age of 70 years had the lowest ICER per QALY gained as compared to nonscreening, it is unreasonable to simply use a start age of 70 years. An older start age is associated with fewer QALYs obtained.

Limitations

There were several limitations to this study that should be addressed. First, like most mathematical models, the model used in this study to estimate the incidence of lung cancer in heavy smokers was a simplification of the biological complexity of lung carcinogenesis and neglected the influence of various endogenous and exogenous risk factors, such as family history and residential/occupational exposure to radon, which may have led to underestimation of the incidence of lung cancer in the targeted population. Further, as heavy smokers are more likely to die from other diseases (eg, cardiovascular diseases and other cancers), its application to estimate the general probability of all-cause death in this population might have slightly underestimated the mortality rate in this work. Nevertheless, the use of this nomothetic approach has aided the development of prevention and control strategies against lung cancer in the United States [34]. Second, the cumulative burden of radiation from annual screening with LDCT was not considered. Albert et al reported that annual LDCT would result in additional radiation exposure of 1.5 mSv per year [35]. Still, recent studies have reported that the potential benefit of lung cancer screening to prevent death was greater than the potential harm of increased radiation exposure [36,37]. Third, smoking cessation events and other health-related behavioral changes due to screening participation were not incorporated in the model due to the lack of relevant data. Further research may benefit from the incorporation of patient-level data extracted from on-going randomized controlled trials with microsimulation models for cost-effectiveness analysis of lung cancer screening.

Conclusion

This economic evaluation revealed that a population-based lung cancer screening program in China for heavy smokers using LDCT could result in more QALYs, although with greater expense than nonscreening. Using the WHO threshold for cost-effectiveness analysis, the annual screening strategy from 55 to 74 years and one-time screening strategy from 65 to 74 years can be considered cost-effective. Moreover, annual screening was the most promising; thus, annual screening should be promoted in China to realize actual benefits.

Authors' Contributions

Conceptualization: HD and ZZ; methodology: ZZ and LD; software: YL and LW; investigation: YW and YY; data curation: HD; writing-original draft preparation: ZZ; writing-review and editing: ZZ and LD; supervision: HD; funding acquisition: LD. All authors have read and agree to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Validation of the natural history model of lung cancer.
[DOCX File , 366 KB - publichealth_v8i7e36425_app1.docx]

References

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66(2):115-132 [FREE Full text] [doi: 10.3322/caac.21338] [Medline: 26808342]
2. China Lung Cancer Prevention and Treatment Alliance. Chinese expert consensus on early diagnosis of primary bronchial lung cancer. Chinese Journal of Tuberculosis and Respiratory Diseases 2014;37(3):172-176. [doi: 10.3760/cma.j.issn.1001-0939.2014.03.005]
3. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019 Sep;69(5):363-385 [FREE Full text] [doi: 10.3322/caac.21565] [Medline: 31184787]
4. International Early Lung Cancer Action Program Investigators, Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, et al. Survival of patients with stage I lung cancer detected on CT screening. N Engl J Med 2006 Oct 26;355(17):1763-1771. [doi: 10.1056/NEJMoa060476] [Medline: 17065637]
5. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011 Aug 04;365(5):395-409 [FREE Full text] [doi: 10.1056/NEJMoa1102873] [Medline: 21714641]
6. Field JK, Duffy SW, Baldwin DR, Whynes DK, Devaraj A, Brain KE, et al. UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. Thorax 2016 Feb;71(2):161-170 [FREE Full text] [doi: 10.1136/thoraxjnl-2015-207140] [Medline: 26645413]
7. Horeweg N, Scholten ET, de Jong PA, van der Aalst CM, Weenink C, Lammers JJ, et al. Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers. The Lancet Oncology 2014 Nov;15(12):1342-1350. [doi: 10.1016/S1470-2045(14)70387-0] [Medline: 25282284]
8. Hofer F, Kauczor H, Stargardt T. Cost-utility analysis of a potential lung cancer screening program for a high-risk population in Germany: A modelling approach. Lung Cancer 2018 Oct;124:189-198 [FREE Full text] [doi: 10.1016/j.lungcan.2018.07.036] [Medline: 30268459]
9. Black WC, Gareen IF, Soneji SS, Sicks JD, Keeler EB, Aberle DR, National Lung Screening Trial Research Team. Cost-effectiveness of CT screening in the National Lung Screening Trial. N Engl J Med 2014 Nov 06;371(19):1793-1802 [FREE Full text] [doi: 10.1056/NEJMoa1312547] [Medline: 25372087]

10. Sun C, Zhang X, Guo S, Liu Y, Zhou L, Shi J, et al. Determining cost-effectiveness of lung cancer screening in urban Chinese populations using a state-transition Markov model. BMJ Open 2021 Jul 01;11(7):e046742 [FREE Full text] [doi: 10.1136/bmjopen-2020-046742] [Medline: 34210726]
11. Zhou Q, Fan Y, Wang Y, Qiao Y, Wang G, Huang Y, et al. [China National Lung Cancer Screening Guideline with low-dose computed tomography (2018 version)]. Zhongguo Fei Ai Za Zhi 2018 Feb 20;21(2):67-75 [FREE Full text] [doi: 10.3779/j.issn.1009-3419.2018.02.01] [Medline: 29526173]
12. Toumazis I, de Nijs K, Cao P, Bastani M, Munshi V, Ten Haaf K, et al. Cost-effectiveness evaluation of the 2021 US Preventive Services Task Force recommendation for lung cancer screening. JAMA Oncol 2021 Dec 01;7(12):1833-1842. [doi: 10.1001/jamaoncol.2021.4942] [Medline: 34673885]
13. Rami-Porta R, Asamura H, Travis WD, Rusch VW. Lung cancer - major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA Cancer J Clin 2017 Mar;67(2):138-155 [FREE Full text] [doi: 10.3322/caac.21390] [Medline: 28140453]
14. Wei W, Zeng H, Zheng R, Zhang S, An L, Chen R, et al. Cancer registration in China and its role in cancer prevention and control. The Lancet Oncology 2020 Jul;21(7):e342-e349. [doi: 10.1016/S1470-2045(20)30073-5] [Medline: 32615118]
15. Population Census Office of the State Council of the People's Republic of China, Department of Population and Employment Statistics of the National Bureau of Statistics of the People's Republic of China. Tabulation on the 2010 Population Census of the People's Republic of China. Beijing, China: China Statistics Press; 2010:112-113.
16. Zhang M, Wu C, Gong Y, Peng P, Gu K, Shi L, et al. Survival analysis of patients with lung cancer in Shanghai. China Oncology 2017;27(5):326-333. [doi: 10.19401/j.cnki.1007-3639.2017.05.002]
17. He J. China Cancer Registry Annual Report 2018. Beijing, China: People's Medical Publishing House; 2019:214-215.
18. CDC China. Global Adult Tobacco Survey (GATS) China 2015 Report. Shanghai, China: CDC China; 2016:93-96.
19. Osaki Y, Okamoto M, Kaetsu A, Kishimoto T, Suyama A. Retrospective cohort study of smoking and lung cancer incidence in rural prefecture, Japan. Environ Health Prev Med 2007 Jul;12(4):178-182 [FREE Full text] [doi: 10.1007/BF02897988] [Medline: 21432062]
20. Ten Haaf K, van Rosmalen J, de Koning HJ. Lung cancer detectability by test, histology, stage, and gender: estimates from the NLST and the PLCO trials. Cancer Epidemiol Biomarkers Prev 2015 Jan;24(1):154-161 [FREE Full text] [doi: 10.1158/1055-9965.EPI-14-0745] [Medline: 25312998]
21. Shi J, Wang L, Wu N, Li J, Hui Z, Liu S, LuCCRES Group. Clinical characteristics and medical service utilization of lung cancer in China, 2005-2014: Overall design and results from a multicenter retrospective epidemiologic survey. Lung Cancer 2019 Feb;128:91-100. [doi: 10.1016/j.lungcan.2018.11.031] [Medline: 30642458]
22. Chen W, Li N, Cao MM, Ren JS, Shi JF, Chen HD, et al. Preliminary analysis of cancer screening program in urban China from 2013 to 2017. China Cancer 2020;29:1-6 [FREE Full text]
23. Sozzi G, Boeri M, Rossi M, Verri C, Suatoni P, Bravi F, et al. Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: A correlative MILD trial study. JCO 2014 Mar 10;32(8):768-773. [doi: 10.1200/jco.2013.50.4357]
24. Banerjee AK. Preinvasive lesions of the bronchus. J Thorac Oncol 2009 Apr;4(4):545-551 [FREE Full text] [doi: 10.1097/JTO.0b013e31819667bd] [Medline: 19279508]
25. Chen S. Study on economic burden and quality of life of lung cancer patients. Wanfang. 2016. URL: https://d.wanfangdata.com.cn/thesis/ChJUaGVzaXNOZXdTMjAyMjA1MjYSCUQwMTAyNTY5MBoIbXp0dnh3NHI%253D [accessed 2022-06-25]
26. Field JK, Vulkan D, Davies MP, Baldwin DR, Brain KE, Devaraj A, et al. Lung cancer mortality reduction by LDCT screening: UKLS randomised trial results and international meta-analysis. Lancet Reg Health Eur 2021 Nov;10:100179 [FREE Full text] [doi: 10.1016/j.lanepe.2021.100179] [Medline: 34806061]
27. Sturza J. A review and meta-analysis of utility values for lung cancer. Med Decis Making 2010;30(6):685-693. [doi: 10.1177/0272989X10369004] [Medline: 20448248]
28. Yuan J, Sun Y, Wang K, Wang Z, Li D, Fan M, et al. Cost effectiveness of lung cancer screening with low-dose CT in heavy smokers in China. Cancer Prev Res (Phila) 2022 Jan;15(1):37-44. [doi: 10.1158/1940-6207.CAPR-21-0155] [Medline: 34580085]
29. Liu G, Hu S, Wu J, Wu J, Yang L, Li H. China guidelines for pharmacoeconomic evaluations. Beijing, China: China Market Press; 2020:16-17.
30. McLeod M, Sandiford P, Kvizhinadze G, Bartholomew K, Crengle S. Impact of low-dose CT screening for lung cancer on ethnic health inequities in New Zealand: a cost-effectiveness analysis. BMJ Open 2020 Sep 24;10(9):e037145 [FREE Full text] [doi: 10.1136/bmjopen-2020-037145] [Medline: 32973060]
31. Criss SD, Cao P, Bastani M, Ten Haaf K, Chen Y, Sheehan DF, et al. Cost-effectiveness analysis of lung cancer screening in the United States: A comparative modeling study. Ann Intern Med 2019 Dec 03;171(11):796-804. [doi: 10.7326/M19-0322] [Medline: 31683314]
32. Ten Haaf K, Tammemägi MC, Bondy SJ, van der Aalst CM, Gu S, McGregor SE, et al. Performance and cost-effectiveness of computed tomography lung cancer screening scenarios in a population-based setting: A microsimulation modeling

analysis in Ontario, Canada. PLoS Med 2017 Feb;14(2):e1002225 [FREE Full text] [doi: 10.1371/journal.pmed.1002225] [Medline: 28170394]

33. He J, Li N, Chen W, Wu N, Shen H, Jiang Y, et al. China lung cancer screening and early diagnosis and early treatment guidelines (2021, Beijing). Chinese Journal of Oncology 2021;43(3):243-268 [FREE Full text]
34. Meza R, Jeon J, Toumazis I, Ten Haaf K, Cao P, Bastani M, et al. Evaluation of the benefits and harms of lung cancer screening with low-dose computed tomography: Modeling study for the US Preventive Services Task Force. JAMA 2021 Mar 09;325(10):988-997 [FREE Full text] [doi: 10.1001/jama.2021.1077] [Medline: 33687469]
35. Albert JM. Radiation risk from CT: Implications for cancer screening. American Journal of Roentgenology 2013 Jul;201(1):W81-W87. [doi: 10.2214/ajr.12.9226]
36. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and harms of CT screening for lung cancer: a systematic review. JAMA 2012 Jun 13;307(22):2418-2429 [FREE Full text] [doi: 10.1001/jama.2012.5521] [Medline: 22610500]
37. McCunney RJ, Li J. Radiation risks in lung cancer screening programs: a comparison with nuclear industry workers and atomic bomb survivors. Chest 2014 Mar 01;145(3):618-624. [doi: 10.1378/chest.13-1420] [Medline: 24590022]

Abbreviations

CIS: carcinoma in situ
CPI: consumer price index
ICER: incremental cost-effectiveness ratio
LDCT: low-dose computed tomography
OR: odds ratio
QALY: quality-adjusted life year
RR: relative risk
WHO: World Health Organization

Edited by H Bradley; submitted 14.01.22; peer-reviewed by L Guo, H Yutong; comments to author 10.05.22; revised version received 21.05.22; accepted 27.05.22; published 06.07.22.

Please cite as:
Zhao Z, Du L, Li Y, Wang L, Wang Y, Yang Y, Dong H
Cost-Effectiveness of Lung Cancer Screening Using Low-Dose Computed Tomography Based on Start Age and Interval in China: Modeling Study
JMIR Public Health Surveill 2022;8(7):e36425
URL: https://publichealth.jmir.org/2022/7/e36425
doi:10.2196/36425
PMID:35793127

Original Paper

Using Social Media to Predict Food Deserts in the United States: Infodemiology Study of Tweets

Nekabari Sigalo[1], BSc, MPH; Beth St Jean[1], BA, MSc, PhD; Vanessa Frias-Martinez[1,2], BSc, MSc, PhD

[1]College of Information Studies, University of Maryland, College Park, MD, United States

[2]University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, MD, United States

Corresponding Author:
Nekabari Sigalo, BSc, MPH
College of Information Studies
University of Maryland
4130 Campus Dr
College Park, MD, 20742
United States
Phone: 1 301 405 2033
Email: nsigalo@umd.edu

Abstract

Background: The issue of food insecurity is becoming increasingly important to public health practitioners because of the adverse health outcomes and underlying racial disparities associated with insufficient access to healthy foods. Prior research has used data sources such as surveys, geographic information systems, and food store assessments to identify regions classified as food deserts but perhaps the individuals in these regions unknowingly provide their own accounts of food consumption and food insecurity through social media. Social media data have proved useful in answering questions related to public health; therefore, these data are a rich source for identifying food deserts in the United States.

Objective: The aim of this study was to develop, from geotagged Twitter data, a predictive model for the identification of food deserts in the United States using the linguistic constructs found in food-related tweets.

Methods: Twitter's streaming application programming interface was used to collect a random 1% sample of public geolocated tweets across 25 major cities from March 2020 to December 2020. A total of 60,174 geolocated food-related tweets were collected across the 25 cities. Each geolocated tweet was mapped to its respective census tract using point-to-polygon mapping, which allowed us to develop census tract–level features derived from the linguistic constructs found in food-related tweets, such as tweet sentiment and average nutritional value of foods mentioned in the tweets. These features were then used to examine the associations between food desert status and the food ingestion language and sentiment of tweets in a census tract and to determine whether food-related tweets can be used to infer census tract–level food desert status.

Results: We found associations between a census tract being classified as a food desert and an increase in the number of tweets in a census tract that mentioned unhealthy foods (*P*=.03), including foods high in cholesterol *(P*=.02*)* or low in key nutrients such as potassium (*P*=.01). We also found an association between a census tract being classified as a food desert and an increase in the proportion of tweets that mentioned healthy foods (*P*=.03) and fast-food restaurants (*P*=.01) with positive sentiment. In addition, we found that including food ingestion language derived from tweets in classification models that predict food desert status improves model performance compared with baseline models that only include socioeconomic characteristics.

Conclusions: Social media data have been increasingly used to answer questions related to health and well-being. Using Twitter data, we found that food-related tweets can be used to develop models for predicting census tract food desert status with high accuracy and improve over baseline models. Food ingestion language found in tweets, such as census tract–level measures of food sentiment and healthiness, are associated with census tract–level food desert status.

(JMIR Public Health Surveill 2022;8(7):e34285) doi:10.2196/34285

KEYWORDS

social media; Twitter; food deserts; food insecurity

Introduction

Background

Healthy food is vital to everyday life. However, healthy food is not equally accessible to everyone [1]. Food insecurity refers to an individual's lack of sufficient and consistent access to healthy foods that are both affordable and good in quality because of the lack of financial and other resources [2]. In 2018, the United States Department of Agriculture (USDA) estimated that 14.3 million households (11.1%) in the United States were food insecure [2].

Geographic location is one of the most important contributing factors to food insecurity and access to healthy foods [3]. *Food deserts* can be broadly defined as geographic regions where residents do not have sufficient access to fresh fruits, vegetables, and other essential ingredients for healthy eating [4]. Access to healthy foods can be limited because of low availability of grocery stores, low access to sustainable transportation, abundance of perceivably cheaper but unhealthy fast-food options, or a combination of such reasons [5,6]. Food deserts are prevalent in rural as well as urban regions, implying that regions with an abundance of food options can still be considered food deserts based on the definition of *healthy* food [7].

Identifying Food Deserts

The disparities in healthy food access among underserved communities have fueled the interest of public health practitioners, researchers, and community activists in not only identifying regions that are currently food deserts but also regions that are at risk for becoming food deserts in the future. The Economic Research Service at the USDA uses various indicators for the official identification of food deserts in the United States at the census tract-level. A review of the literature determined that other frequently used measures to assess food access are as follows: (1) geographic information systems (GIS) technology, where researchers use geocoding to map resources and create density maps that illustrate differences in food security and access in various locations [8]; (2) food store assessments, which may include both objective and subjective assessments of the food environment [9-13]; and (3) consumer surveys, which allow researchers to gather data from randomly selected households—data regarding household food expenditures and consumption over a specified period [4].

Although each of these food desert identification methods have been widely used and have provided rich insights into food insecurity in the United States, each method comes with unique challenges. For example, GIS technology comes with the risk of misidentification of food stores in the GIS and mapping fails to provide information about food consumption behavior [10]. Food store assessments may be associated with high costs and small, nonrandom sample sizes, as well as significant time spent conducting assessments [13]. Consumer surveys have been found to reflect self-reporting inaccuracies [14]. Each of the challenges to the state-of-the-art approaches present room for another novel approach that uses an alternative, more modern data source. This study examines the use of food ingestion language found on social media, specifically tweets, for predicting food desert status among census tracts in the United States.

Social Media for Public Health Research

Researchers have increasingly looked to social media data as a means of measuring population health and well-being in a less intrusive and more scalable manner [15]. Social media data have proved useful in predicting health outcomes in many studies; therefore, these data may prove to be a very rich source for yet another health-related issue: food insecurity. Using social media data to predict the emergence of food deserts provides a people-centered approach for identifying food deserts by allowing for the examination of the dietary consumption and habits of individuals who reside in food deserts versus those who do not reside in food deserts [16].

Prior studies have successfully extracted information from social media to address various types of health-related outcomes, relying on the naturalistic observations deduced from social media data to answer questions related to health and well-being [17]. For example, in a study that sought to predict depression among Twitter users, researchers leveraged behavioral cues found in tweets to develop a classifier for depression [17]. In a study that considered Twitter data for various public health applications, researchers conducted syndromic surveillance of serious illnesses, measured behavioral risk factors, and mapped illnesses to various geographic regions [18]. Another study used Twitter to monitor and predict influenza prevalence in the United States by conducting a network analysis of Twitter users and demonstrating the association of social ties and colocation of people who were symptomatic with one's risk of contracting influenza [19]. A study that sought to develop a publicly available neighborhood-level data set with indicators related to health behaviors and well-being also examined the associations between these Twitter-derived indicators and key neighborhood demographics [20]. Another study examined Instagram posts to understand dietary choices and nutritional challenges in food deserts [4]. The study by Gore et al [21] examined the relationship between the obesity rate in urban areas and the expressions of happiness, diet, and physical activity in tweets.

As seen in this study, several other studies similarly leveraged natural language processing methods such as sentiment analysis, emotion analysis, and topic modeling to use social media to answer public health research questions. For example, some studies [22-26] collected tweets over the course of the COVID-19 pandemic to examine public sentiments and opinions regarding COVID-19 vaccines. Researchers [27,28] conducted topic modeling and emotion analyses to identify the themes and emotions related to the COVID-19 vaccines to aid public health officials in the battle against COVID-19.

Study Overview

In this study, we leveraged the linguistic constructs in food-related tweets to develop a classification model for food deserts in the United States. We considered both tweet sentiment and overall nutritional values of foods found in tweets to identify associations between living in a food desert and food consumption.

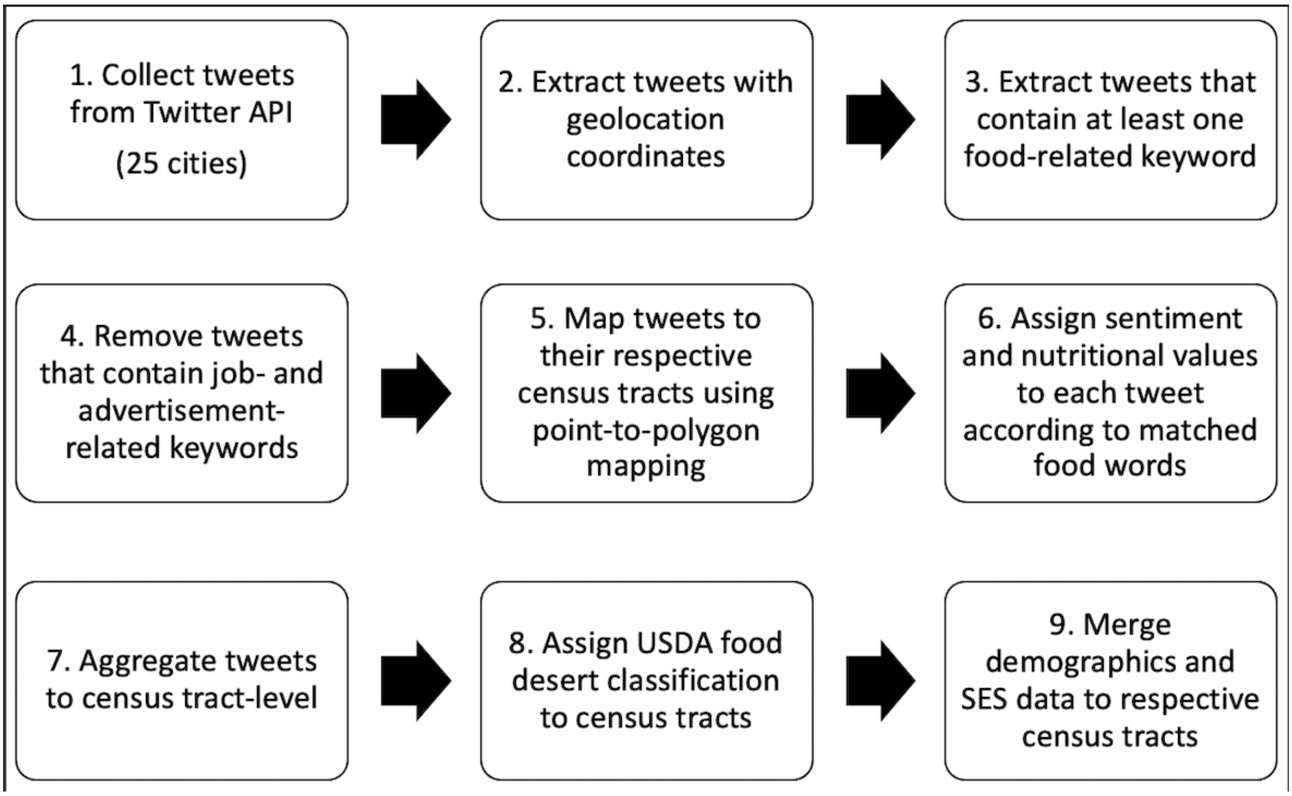
To our knowledge, this is the first study to develop a model for inferring food desert status among census tracts in the United States using Twitter data. The main objective of this study was to examine the linguistic constructs found in food-related tweets to evaluate the differences in food nutritional value and food consumption behavior of individuals in food deserts versus those in non–food deserts. Our key hypotheses are as follows: (1) living in a food desert is associated with positive mentions of unhealthy foods, such as tweets that mention foods that are high in caloric content or low in vital nutrients such as fiber and calcium, and (2) food ingestion language among Twitter users in a census tract can be used to infer census tract–level food desert status.

Methods

Overview

An overview of the entire data collection and preparation process is illustrated in Figure 1 and described in the following subsections.

Figure 1. Twitter data collection process. API: application programming interface; SES: socioeconomic status; USDA: United States Department of Agriculture.



Data Collection

Twitter Data

From March 2020 to December 2020, the Twitter streaming application programming interface (API), which provides access to a random sample of 1% of publicly available tweets, was used to collect tweets (including retweets and quoted tweets) from 25 of the most populated cities in the United States (Textbox 1) [29]. The 25 cities included in this analysis are among the top 50 most populated cities in the United States. However, we decided not to go with the most populated cities (such as New York City, Los Angeles, and Houston) because we wanted to understand whether the framework we developed would be beneficial for smaller cities that are not typically the focus of these types of infodemiology studies. Public health resources directed at improving population health are historically limited and can vary from one public health jurisdiction to the next [30]. Heavily populated cities such as New York City, Los Angeles, and Houston likely have an abundance of resources (both financial and personnel) that can be used to conduct food desert identification using more traditional (expensive) methods. Although the framework outlined in this study would also be useful for heavily populated cities, we believe that less populated cities with fewer resources would benefit the most from this type of study.

When a location-based search is specified, the Twitter API extracts tweets tied to a certain location based on two criteria that are not mutually exclusive: (1) the user has their location enabled for all tweets, in which case these tweets will have specific GPS coordinates, or (2) the user has location information in their profile, such as the city and state they live in, in which case all tweets associated with this user will be tied to this location but without specific geocoordinates. In both cases, these location-tagged tweets are eligible for selection by the Twitter API when a location-based search is specified [31].

As this analysis sought to assign individual tweets to their respective census tracts, all tweets in our sample were required to have specific geolocation information (latitude and longitude GPS coordinates). A parsing module was created to filter out tweets without specific geolocation information. Next, to extract tweets related to food ingestion, tweets were further filtered by a list of 1787 food-related words from the USDA FoodData Central Database (examples are presented in Textbox 2) [32]. Names of popular fast-food restaurants extracted from Wikipedia [33] were also included in this list, as was done in the study by Vydiswaran et al [16].

Tweets related to job postings and advertisements were filtered out by excluding tweets with hashtags and keywords such as "#jobs," "#hiring," and "#ad." For the purposes of this research, we assumed that the tweets in our sample, which, at minimum, contained at least one of 1787 food-related keywords, were related to food consumption, as was done in the study by Nguyen et al [20]. To assess the impact of this assumption, a random sample of 1000 tweets was selected for manual classification as food related or not food related. Among the 1000 tweets in the random sample, 770 (77%) were classified as food-related, whereas 230 (23%), although they contained food keywords, were classified as not food-related. Tweets that matched to food words but were not related to food consumption included tweets related to, for example, Apple products (eg, "I went to the Apple Store to purchase an iPhone") and common city nicknames (eg, New York City, aka "The Big Apple").

Textbox 1. Targeted cities for Twitter data collection.

Targeted cities for data collection

- Albuquerque, New Mexico
- Dallas, Texas
- Atlanta, Georgia
- Baltimore, Maryland
- Colorado Springs, Colorado
- Fresno, California
- Kansas City, Missouri
- Las Vegas, Nevada
- Long Beach, California
- Louisville, Kentucky
- Mesa, Arizona
- Miami, Florida
- Milwaukee, Wisconsin
- Minneapolis, Minnesota
- New Orleans, Louisiana
- Oakland, California
- Oklahoma City, Oklahoma
- Omaha, Nebraska
- Portland, Oregon
- Raleigh, North Carolina
- Sacramento, California
- Tucson, Arizona
- Tulsa, Oklahoma
- Virginia Beach, Virginia
- Wichita, Kansas

Textbox 2. Examples of food-related keywords.

Food-related keywords

- Healthy
- Acai
- Apple
- Apricot
- Avocado
- Banana
- Blackberries
- Blueberries
- Cantaloupe
- Cherries
- Clementine
- Unhealthy
- Cheesecake
- Cupcake
- Donut
- Pepsi
- Sprite
- Sunkist
- Red velvet cake
- Chicken McNuggets
- Double cheeseburger
- Zinger burger
- Fast-food restaurants
- Jack in the Box
- Chick-fil-A
- Burger King
- Dairy Queen
- Del Taco
- Taco Bell
- Bojangles
- Checkers
- Popeyes
- Whataburger

Twitter-Derived Features

We referred to similar work conducted by Nguyen et al [20] to classify each food item as healthy or unhealthy. The classification of foods as healthy or unhealthy was subjective and conducted by 2 different annotators (Daniela Nganjo and Pauline Comising). Fruits and vegetables were classified as healthy food items. Unhealthy food items included fried foods, fast-food items, and other food items commonly considered to be unhealthy. The following nutritional values, per 100 g, were obtained for each food item in the list using the USDA FoodData Central Database: calories, calcium, carbohydrates, cholesterol, energy, fat, fiber, iron, potassium, protein, fatty acids, sodium, sugar, vitamin A, and vitamin C.

To measure the healthiness of foods mentioned in tweets, the overall nutritional values of the foods mentioned in each tweet were calculated. To calculate the nutritional values of foods mentioned in each tweet, regular expression matching was used to compare the words in each tweet to the items described in the aforementioned food list (Textbox 2). The keyword-matching algorithm first searched the tweet text for matches to food keywords containing multiple words, then searched the tweet text for matches to food keywords with fewer words. Using this method, the tweet text was searched for keywords with 3 words, for example, before searching for keywords with 2 words, and the tweet text was searched for keywords with 2 words, before searching for keywords with 1 word. For example, both “Burger King” and “burger” were included in the food list. Using this keyword-matching algorithm, a tweet was searched for the keyword “Burger King” first to avoid an incorrect match to the keyword “burger” alone. Once this match was made, the keyword “Burger King” was removed from the tweet text and the remaining tweet text was searched for single-word keywords such as “burger.”

Next, the respective nutritional values for each matched food word were then calculated for the corresponding tweet. For tweets having >1 match to food names in the food list, the assigned nutritional value was equal to the average of the nutritional values for all matched food items in the tweet.

Sentiment Analysis

To capture the attitudes toward foods mentioned in tweets, we conducted a sentiment analysis of all tweets using the bing lexicon from the tidytext package in R [34]. The bing lexicon provides a label of *negative* or *positive* for thousands of words in the English language. To label the overall sentiment of a tweet, positive expression words were assigned a value of 1, negative expression words were assigned a value of –1, and neutral expression words were assigned a value of 0. An overall *sentiment score* was assigned to each tweet by summing the values assigned to all expression words present in a tweet. Tweets with a positive sentiment score were labeled as having overall positive sentiment, tweets with a negative sentiment score were labeled as having an overall negative sentiment, and tweets with a score of 0 were labeled as having overall neutral sentiment. The resulting tweet sentiment assignments were then used to flag the following types of tweets: tweets that mentioned healthy foods with positive sentiment; tweets that mentioned healthy foods with negative sentiment; tweets that mentioned unhealthy foods with positive sentiment; tweets that mentioned unhealthy foods with negative sentiment; tweets that mentioned fast-food restaurants with positive sentiment; and tweets that mentioned fast-food restaurants with negative sentiment. These tweet-level indicators were later aggregated to the census tract-level to produce neighborhood-specific features related to the proportion of tweets that expressed positive or negative sentiment toward healthy foods, unhealthy foods, and fast-food restaurants.

Mapping Tweets to Census Tracts

As this analysis examined food desert status at the census tract-level, for all census tracts in the 25 cities listed in Textbox 1, each tweet was then mapped to its respective census tract using point-to-polygon mapping of the latitude and longitude coordinates of the geolocated tweet to the bounding box of the respective census tract [35]. Once each tweet was mapped to a census tract, the tweets were aggregated to the census tract-level and the average nutritional content per food item mentioned in tweets within each census tract was calculated. Additional census tract–level food-related Twitter-derived features included the following: (1) percentage of all tweets in a census tract that mention the following with either positive or negative sentiment: healthy foods, unhealthy foods, and fast-food restaurants, and (2) average number of healthy food, unhealthy food, and fast-food mentions per tweet. Tweets with neutral sentiment were not excluded from the analysis sample, but we did not consider neutral sentiment as an independent feature. A complete list of food-related census tract–level features derived from Twitter can be found in Textbox 3.

Textbox 3. Twitter-derived food features.

Food features

- Percentage of tweets that mention healthy foods with positive sentiment
- Percentage of tweets that mention healthy foods with negative sentiment
- Percentage of tweets that mention unhealthy foods with positive sentiment
- Percentage of tweets that mention unhealthy foods with negative sentiment
- Percentage of tweets that mention fast-food restaurants with positive sentiment
- Percentage of tweets that mention fast-food restaurants with negative sentiment
- Average number of healthy food mentions
- Average number of unhealthy food mentions
- Average number of fast-food mentions
- Average number of calories per food item (per 100 g)
- Average calcium per food item (per 100 g)
- Average carbohydrates per food item (per 100 g)
- Average cholesterol per food item (per 100 g)
- Average energy per food item (per 100 g)
- Average fiber per food item (per 100 g)
- Average iron per food item (per 100 g)
- Average potassium per food item (per 100 g)
- Average fat per food item (per 100 g)
- Average protein per food item (per 100 g)
- Average saturated fatty acids per food item (per 100 g)
- Average sodium per food item (per 100 g)
- Average sugar per food item (per 100 g)
- Average trans fatty acids per food item (per 100 g)
- Average unsaturated fatty acids per food item (per 100 g)
- Average vitamin A per food item (per 100 g)
- Average vitamin C per food item (per 100 g)
- Average number of calories per healthy food item (per 100 g)
- Average number of calories per unhealthy food item (per 100 g)

Food Desert Status

Once all data were collected and aggregated to the census tract-level, each census tract was classified as a food desert or not a food desert, according to the USDA Food Access Research Atlas classification of low-income and low-access tracts measured at 1 mile for urban areas and 10 miles for rural areas. The USDA classifies low-income tracts using the following criteria: (1) at least 20% of the residents live below the federal poverty level; (2) median family income is, at most, 80% of the median family income for the state in which the census tract lies; or (3) the census tract is in a metropolitan area and the median family income is, at most, 80% of the median family income for the metropolitan area in which the census tract lies [36]. Low-access census tracts are classified by a significant share (≥500 individuals or at least 33%) of individuals in the census tract being far from a supermarket or grocery store [36].

In total, 7.52% (299/3978) of census tracts with geolocated food-related tweets were classified as low-income, low-access food deserts, measured at 1 mile for urban areas and 10 miles for rural areas.

Demographics and Socioeconomic Status Features

Demographic and socioeconomic status (SES) characteristics at the census tract-level were pulled from the 2019 American Community Survey and merged onto the census tract–level tweets data set. The demographic variables used in this analysis are presented in Textbox 4.

Textbox 4. Census tract–level demographic and socioeconomic status features extracted from the 2019 American Community Survey.

Demographic variables

- Percentage White and non-Hispanic
- Percentage Black or African American
- Percentage other race
- Percentage Asian
- Percentage American Indian or Alaska Native
- Percentage owner-occupied housing units
- Percentage of population living below the federal poverty line
- Number of housing units
- Number of households
- Median family income (US $, 2019)
- Median age (years)
- Population

Data Analysis

Analyses were performed using R software (version 3.5.1; The R Foundation for Statistical Computing) and Python (version 3.8).

Evaluating the Association Between Living in a Food Desert and Food Ingestion Language on Twitter

To test the hypothesis that living in a food desert is associated with the food ingestion language of Twitter users, adjusted linear regression was conducted using food desert status as a treatment and the SES features listed in Textbox 4 as control features to analyze which Twitter-derived features presented in Textbox 3 were statistically significantly different between food deserts and non–food deserts. Each Twitter-derived feature (Textbox 3) was designated as the outcome variable in individual linear regression models, as specified in the following equation:

$$y_{Twitter} = \beta_0 + \beta_{FD}x_{FD} + \beta_{SES1}x_{SES1} + \ldots + \beta_{SES12}x_{SES12} + Error$$

where $y_{Twitter}$ = Individual Twitter – derived food feature; β_0 = y – intercept (constant); β_{FD}=food desert classification; and β_{SES1}-β_{SES2}=each of the 12 demographic and socioeconomic features listed in Textbox 4.

Twitter-derived features that were found to have individual, significant associations with food desert status were later used as features in the classification model for predicting food desert status to test the hypothesis that key food ingestion language found in tweets can be used to infer census tract–level food desert status.

Predicting Food Desert Status

To test the hypothesis that food ingestion language found in tweets can be used to infer census tract–level food desert status, classification models were developed using the Twitter-derived food-related nutritional features listed in Textbox 3. We developed 5 different classification models with different sets of features that would allow us to determine which models, if any, show improvements over a baseline model (Table 1). The first model, which was considered the baseline model, included demographics and SES features previously found to be strong predictors of food desert status in prior studies [37]; the second model included the demographics and SES features from the baseline model, plus the Twitter-derived food-related nutritional features presented in Textbox 3; the third model included the demographics and SES features from the baseline model, plus the tweet sentiment features; the fourth model included all the features (from models 2 and 3 combined); and the fifth model included the demographics and SES features from the baseline model, plus all Twitter-derived food-related features found to have a statistically significant association with census tract–level food desert status.

All features were standardized using minimum-maximum normalization, a method that standardizes data by rescaling the range of individual features to (0, 1), as described in the study by Cao et al [38]. The data were divided into a 70:30 training data and testing data split. Each of the models were built using 5-fold cross-validation to keep computation time to a minimum. Using the *caret* package in R, each model described in Table 1 was run using several different classification methods: adaptive boosting, gradient boosting, logistic regression, and ensemble methods [39]. The ensemble model combined adaptive boosting, gradient boosting, and logistic regression as base methods. Ensemble modeling is a process that aggregates the predictions of many different modeling algorithms and uses the results of the base models as inputs into a logistic regression model. The ensemble performs as a single model, reducing the generalization error of the prediction compared with the base models alone. The results of each classification method, regardless of performance, are presented in this paper.

Table 1. Classification models for predicting food desert status.

Model	Description	Features
1	Demographics and SES[a] only (baseline)	Demographics and SES features (Textbox 4)
2	Demographics and SES+nutritional values	Demographics and SES features (Textbox 4) and Twitter-derived food-related nutritional features (Textbox 3)
3	Demographics and SES+Twitter mentions sentiment	Demographics and SES features (Textbox 4) and sentiment analysis of Twitter mentions features (Textbox 3)
4	Demographics and SES+nutritional values+Twitter mentions sentiment	Demographics and SES features (Textbox 4), Twitter-derived food-related nutritional features (Textbox 3), and sentiment analysis of Twitter mentions features (Textbox 3)
5	Demographics and SES+statistically significant features	Demographics and SES features (Textbox 4) and Twitter-derived food-related features found to have a statistically significant association with census tract–level food desert status

[a] SES: socioeconomic status.

Ethics Approval

The University of Maryland College Park institutional review board has determined that this project does not meet the definition of human participant research under the purview of the institutional review board according to federal regulations.

Results

Overview

A total of 60,174 geolocated food-related tweets were collected during the data collection period. Across the 25 cities in our sample, 3978 census tracts had at least one geolocated food-related tweet, with a median of 4 (IQR 8) geolocated food-related tweets per census tract. Long Beach, California, had the largest representation of tweets (17,303/60,174, 28.75%), as well as the largest representation of users (5189/17,978, 28.86%; Table 2). Fresno, California, had the smallest representation of tweets (421/60,174, 0.7%), and Wichita, Kansas, had the smallest representation of users (132/17,978, 0.73%). The maximum number of tweets by a single individual was 1277 (from a user in Long Beach, California). On average, there were 6686 (SD 3629) tweets collected from 3264 (SD 1385) users each month. The remaining tweet and user statistics can be found in Table 2.

Table 3 displays descriptive statistics of the census tract–level Twitter-derived food features. On average, there was a higher percentage of tweets that mentioned healthy foods with positive sentiment (34%) versus negative sentiment (20%), a higher percentage of tweets that mentioned unhealthy foods with positive sentiment (34%) versus negative sentiment (17%), and a higher percentage of tweets that mentioned fast-food restaurants with positive sentiment (21%) versus negative sentiment (12%).

Table 4 displays descriptive statistics of census tract–level demographics and SES features among census tracts represented in this analysis. Across the represented census tracts, 62.7% (10,682,930/17,038,167) of all residents were White and non-Hispanic, 15.6% (2,657,954/17,038,167) were Black or African American, and 8.9% (1,516,397/17,038,167) identified as other race. The median family income across census tracts was approximately US $82,000, and the median age was approximately 37 years.

Table 2. Number of tweets (N=60,174) and users (N=17,978) by city.

City	Number of tweets, n (%)	Number of users, n (%)
Albuquerque, New Mexico	839 (1.39)	224 (1.26)
Atlanta, Georgia	4936 (8.2)	1739 (9.67)
Baltimore, Maryland	2521 (4.19)	684 (3.8)
Colorado Springs, Colorado	847 (1.41)	268 (1.49)
Dallas, Texas	2472 (4.11)	782 (4.35)
Fresno, California	421 (0.7)	153 (0.85)
Kansas City, Missouri	1651 (2.74)	532 (2.96)
Las Vegas, Nevada	2336 (3.88)	872 (4.85)
Long Beach, California	17,303 (28.75)	5189 (28.86)
Louisville, Kentucky	1246 (2.07)	406 (2.26)
Mesa, Arizona	1888 (3.14)	616 (3.43)
Miami, Florida	2576 (4.28)	1080 (6.01)
Milwaukee, Wisconsin	1578 (2.62)	388 (2.16)
Minneapolis, Minnesota	1282 (2.13)	471 (2.62)
New Orleans, Louisiana	2144 (3.56)	641 (3.57)
Oakland, California	2601 (4.32)	614 (3.42)
Oklahoma City, Oklahoma	1143 (1.9)	371 (2.06)
Omaha, Nebraska	742 (1.23)	198 (1.1)
Portland, Oregon	5528 (9.19)	928 (5.16)
Raleigh, North Carolina	1588 (2.64)	454 (2.53)
Sacramento, California	1721 (2.86)	565 (3.14)
Tucson, Arizona	794 (1.32)	250 (1.39)
Tulsa, Oklahoma	622 (1.03)	209 (1.16)
Virginia Beach, Virginia	960 (1.6)	212 (1.18)
Wichita, Kansas	435 (0.72)	132 (0.73)

Table 3. Descriptive statistics of Twitter-derived food features from geolocated food-related tweets.

Twitter-derived food features	Values, mean (SD)
Percentage of tweets that mention healthy foods, positive sentiment	33.8 (0.4)
Percentage of tweets that mention healthy foods, negative sentiment	19.8 (0.3)
Percentage of tweets that mention unhealthy foods, positive sentiment	33.5 (0.4)
Percentage of tweets that mention unhealthy foods, negative sentiment	17.1 (0.3)
Percentage of tweets that mention fast-food restaurants, positive sentiment	21.2 (0.3)
Percentage of tweets that mention fast-food restaurants, negative sentiment	11.7 (0.3)
Average number of healthy food mentions	0.2 (0.3)
Average number of unhealthy food mentions	0.4 (0.4)
Average number of fast-food mentions	0.1 (0.3)
Average number of calories per food item (per 100 g)	155.1 (96.3)
Average calcium per food item (per 100 g)	74 (91.3)
Average carbohydrates per food item (per 100 g)	23.2 (10.9)
Average cholesterol per food item (per 100 g)	57.3 (284.4)
Average energy per food item (per 100 g)	285.1 (115.7)
Average fat per food item (per 100 g)	10.4 (6.9)
Average fiber per food item (per 100 g)	1.7 (1.4)
Average iron per food item (per 100 g)	1.7 (8.5)
Average potassium per food item (per 100 g)	194.5 (93)
Average protein per food item (per 100 g)	7 (4.1)
Average saturated fatty acids per food item (per 100 g)	3.6 (2.5)
Average sodium per food item (per 100 g)	524.7 (962.7)
Average sugar per food item (per 100 g)	11.8 (8.3)
Average trans fatty acids per food item (per 100 g)	0.1 (0.2)
Average unsaturated fatty acids per food item (per 100 g)	2.6 (4)
Average vitamin A per food item (per 100 g)	548.8 (734.5)
Average vitamin C per food item (per 100 g)	7.1 (15.8)
Average number of calories per healthy food item (per 100 g)	67.4 (61.5)
Average number of calories per unhealthy food item (per 100 g)	189.8 (125.9)

Table 4. Descriptive statistics of census tract–level demographics and socioeconomic status features extracted from the 2019 American Community Survey.

Characteristic	Values, mean (SD)
Percentage White and non-Hispanic	62.7 (23.4)
Percentage Black or African American	15.6 (21.0)
Percentage other race	8.9 (12.3)
Percentage Asian	7.4 (9.2)
Percentage American Indian or Alaska Native	1.0 (1.9)
Percentage owner-occupied housing units	49.3 (24.8)
Percentage of population living below the federal poverty line	16.2 (12.1)
Number of housing units	1788.4 (863.5)
Number of households	1628.0 (799.1)
Median family income (US $, 2019)	82,371.4 (42,680.1)
Median age (years)	37.0 (6.8)
Population	4283.1 (2243.6)

Hypothesis 1: Living in a Food Desert Is Associated With the Food Ingestion Language and Sentiments of Tweets Observed Among Twitter Users

The adjusted linear regression models confirmed this hypothesis, revealing significant associations between food desert status and 5 of the Twitter-derived food characteristics (Table 5). The results show that a census tract being classified as a food desert was associated with an increase in the average cholesterol concentration (per 100 g; *P*=.02*)* per food item mentioned in tweets, a decrease in the average potassium concentration (per 100 g) per food item mentioned in tweets (*P*=.01), and an increase in the average number of unhealthy foods mentioned per tweet (*P*=.03). A census tract being classified as a food desert was also associated with an increase in the proportion of tweets that mentioned healthy foods as well as the proportion of tweets that mentioned fast-food restaurants with positive sentiment (*P*=.03 and *P*=.01, respectively).

Although we did not expect to see an association between living in a food desert and an increase in mentions of healthy foods with positive sentiment, we hypothesize that such an association might reflect aspirational tweets of individuals who long for healthy food that is not present in their neighborhood (for example, the positive sentiment does not reflect food consumption but rather a wish to increase accessibility).

Table 5. Adjusted linear regression model results examining the associations between living in a food desert and food ingestion language of Twitter users.

Twitter-derived food features	β coefficient	*P* value	SE	R-squared
Percentage of tweets that mention healthy foods, positive sentiment	.077	.03	0.036	0.003
Percentage of tweets that mention healthy foods, negative sentiment	.023	.44	0.031	3.45×10^{-5}
Percentage of tweets that mention unhealthy foods, positive sentiment	–0.051	.06	0.027	0.001
Percentage of tweets that mention unhealthy foods, negative sentiment	.022	.32	0.022	3.98×10^{-4}
Percentage of tweets that mention fast-food restaurants, positive sentiment	.096	.01	0.039	0.005
Percentage of tweets that mention fast-food restaurants, negative sentiment	.010	.74	0.032	8.88×10^{-5}
Average number of healthy food mentions	–0.002	.54	0.003	9.57×10^{-5}
Average number of unhealthy food mentions	.014	.03	0.006	0.001
Average number of fast-food mentions	–0.003	.76	0.010	2.45×10^{-5}
Average number of calories per food item (per 100 g)	.005	.58	0.009	7.93×10^{-5}
Average calcium per food item (per 100 g)	–0.001	.60	0.002	7.36×10^{-5}
Average carbohydrates per food item (per 100 g)	–0.009	.19	0.007	4.46×10^{-4}
Average cholesterol per food item (per 100 g)	.005	.02	0.002	0.001
Average energy per food item (per 100 g)	.004	.60	0.007	7.37×10^{-5}
Average fat per food item (per 100 g)	–0.005	.69	0.012	4.27×10^{-5}
Average fiber per food item (per 100 g)	–0.014	.10	0.008	7.26×10^{-4}
Average iron per food item (per 100 g)	$–6.44\times10^{-4}$	.56	0.001	9.04×10^{-5}
Average potassium per food item (per 100 g)	–0.008	.01	0.003	0.002
Average protein per food item (per 100 g)	–0.002	.88	0.010	6.11×10^{-6}
Average saturated fatty acids per food item (per 100 g)	.007	.31	0.007	2.70×10^{-4}
Average sodium per food item (per 100 g)	–0.005	.06	0.002	9.13×10^{-4}
Average sugar per food item (per 100 g)	–0.005	.35	0.005	2.29×10^{-4}
Average trans fatty acids per food item (per 100 g)	–0.002	.79	0.007	1.78×10^{-5}
Average unsaturated fatty acids per food item (per 100 g)	.002	.72	0.006	3.39×10^{-5}
Average vitamin A per food item (per 100 g)	.004	.58	0.007	8.19×10^{-5}
Average vitamin C per food item (per 100 g)	$–5.53\times10^{-4}$	.71	0.002	3.52×10^{-5}
Average number of calories per healthy food item (per 100 g)	9.58×10^{-4}	.95	0.017	1.92×10^{-6}
Average number of calories per unhealthy food item (per 100 g)	.007	.64	0.015	8.42×10^{-5}

Hypothesis 2: Food Ingestion Language Among Twitter Users in a Census Tract Can Be Used to Infer Census Tract–Level Food Desert Status

To test the hypothesis that food ingestion language found in tweets can be used to infer census tract–level food desert status, we used various machine learning methods to compare the performance of 5 classification models (Table 6). In this paper, we evaluated model performance by comparing each model's area under the receiver operating characteristic curve (AUC) metric, which measures how well each model can distinguish a non–food desert census tract from a food desert census tract. We used this metric, instead of accuracy, for evaluating model performance because this metric is better suited to measure model performance on class-imbalanced data [40], as is the case with the imbalanced food desert classification outcome in our sample data (of the 3978 census tracts, 299, 7.52%, were food desert census tracts). Model 3, which included sentiment features related to food mentions, showed an improvement over the baseline model AUC, using the gradient boosting classification method, by >7%. This was also the best performing model (AUC 0.823). Model 4, which included all Twitter-derived food-related features, showed an improvement over the baseline model AUC, using the logistic regression classification method, of nearly

19%. These results confirm hypothesis 2, suggesting that the best performing models involve the inclusion of Twitter-derived food ingestion language.

Table 6. Model performance.

Method and model[a]	AUC[b]
Adaptive boosting	
1 (baseline)	0.759
2	0.749
3	0.738
4	0.650
5	0.723
Gradient boosting	
1 (baseline)	0.766
2	0.797
3	0.823
4	0.777
5	0.699
Logistic regression	
1 (baseline)	0.682
2	0.720
3	0.777
4	0.809
5	0.663
Ensemble method	
1 (baseline)	0.769
2	0.771
3	0.760
4	0.641
5	0.740

[a] Model descriptions (refer to Table 1)—1: demographics and socioeconomic status only (baseline); 2: demographics and socioeconomic status+nutritional values; 3: demographics and socioeconomic status+Twitter mentions sentiment; 4: demographics and socioeconomic status+nutritional values+Twitter mentions sentiment; 5: demographics and socioeconomic status+statistically significant features.

[b] AUC: area under the receiver operating characteristic curve.

Discussion

Principal Findings

In this study, we sought to address two key hypotheses: (1) living in a food desert is associated with positive mentions of unhealthy foods, such as tweets that mention foods that are high in caloric content or low in vital nutrients such as fiber and calcium, and (2) food ingestion language among Twitter users in a census tract can be used to infer census tract–level food desert status. The study found significant associations between living in a food desert and tweeting about unhealthy foods, including foods high in cholesterol content or low in key nutrients such as potassium. We also found that supplementing classification models with features derived from food ingestion language found in tweets, such as positive sentiment toward mentions of healthy foods and fast-food restaurants, improves baseline models that only include demographic and SES features by up to 19%, with AUC scores >0.8.

Study Findings in Context

Assessing and understanding the food environment in neighborhoods is key to addressing the issue of food insecurity in the United States. The USDA conducts the official identification of food deserts in the United States but this assessment is infrequent and the latest assessment from 2015 is outdated. Other methods such as GIS technology, surveys, and food store assessments, although effective, can be costly and time consuming. Although conducting assessments of food stores provides important insights into the food environment, this study suggests that perhaps residents of census tracts unknowingly provide important information regarding the food environment on Twitter through the food ingestion language found in tweets. Using social media data for food insecurity research allows researchers to examine food consumption in various regions, allowing a comparison of how food ingestion differs between areas where residents have sufficient access to healthy foods and areas where residents do not have sufficient access to healthy foods.

The findings of this study contribute to the literature on food insecurity in the United States by examining the potential effects of living in a food desert on food consumption using Twitter-derived food ingestion features as a proxy to examine food consumption. In this study, we found that food desert status is associated with not only the sentiment toward the types of foods mentioned in tweets but also the nutritional content of foods mentioned in tweets. More specifically, a census tract being classified as a food desert was associated with an increase in the average cholesterol concentration and a decrease in the average potassium concentration (per 100 g) per food item mentioned in tweets, as well as an increase in the proportion of tweets that mention unhealthy foods. A census tract classified as a food desert was also associated with an increase in the proportion of tweets that mentioned healthy foods and fast-food restaurants with positive sentiment. These findings support prior studies that also found associations between neighborhood characteristics, such as food desert status or fast-food density, and the *healthiness* of tweets in a census tract [20]. These findings also echo the findings in the study by Gore et al [21], which revealed that the prevalence of tweets containing terms related to fruit and vegetables was correlated with lower obesity rates in cities.

This study makes further contributions by examining the predictive ability of food ingestion language derived from tweets on census tract food desert status. This builds upon a similar study that used Instagram posts to understand dietary choices and nutritional challenges in food deserts [4]. In this study, we investigated to what extent ingestion language extracted from Instagram posts was able to infer a census tract's food desert status. This study yielded a model with high accuracy (>80%).

Other similar studies that sought to examine food consumption using tweets across various geographic regions suggest that many of the food-related tweets in an area may be an artifact of visitors to the area, not residents. For example, a study conducted by Mitchell et al [41] showed that travel destinations such as Hawaii have an abundance of tweets with food-related terms. Similarly, the World Happiness Report [42] showed that a larger number of food-related words in tweets were used by users who regularly travel large distances, such as tourists. Although these studies suggest that the tweets we collected may have been from residents or from people who were simply visiting an area, in our study, we decided to consider all tweets under the premise that tweets from nonresidents can also reflect their food consumption experiences when they are in that neighborhood, which still provides some information regarding the local food environment. It is also important to note that because the data collection period for this study occurred during the height of the COVID-19 pandemic (particularly during travel restrictions and quarantine mandates), this might have allowed us to better capture local movement and tweets from actual residents in these areas because people were being encouraged to stay closer to home and not travel to other areas [43].

Developing an algorithm that predicts food deserts by extracting information from tweets allows researchers to monitor food insecurity more frequently than current methods allow. The use of tweets for research related to food insecurity provides researchers with more frequently updated information, thereby addressing the "lag between capturing information about newly opened and recently closed food retail businesses" [4]. This framework also has implications for policy making and advocacy. On the basis of the results presented in this paper, we recommend the use of similar algorithms by public health officials to encourage the allocation of food resources to census tracts that have been identified as food deserts using the algorithm, especially if these neighborhoods are not currently identified as food deserts according to the USDA's classifications. Public health officials may also leverage this framework to advocate for policy interventions that either prevent food deserts from emerging or increase access to healthy foods in neighborhoods identified as food deserts using the algorithm, minimize the impacts of limited food access, support data-driven decision-making, and encourage grocery store chains to expand into neighborhoods based on need rather than potential profit.

Limitations

Although prior research has proved social media to be a rich data source, it does have some limitations. The ability to pull millions of tweets from a single data source is an attractive characteristic of Twitter data, but a study conducted by Pew Research Center showed that Twitter users are more likely to be younger than the general population (29% of Twitter users are aged 18 to 29 years compared with 21% of the general population in the United States), more highly educated (42% of Twitter users are college graduates compared with 31% of the general population in the United States), have higher incomes (41% of Twitter users earn at least US $75,000 per year compared with 32% of the general population in the United States), and are more likely to consider themselves Democrats (36% of Twitter users consider themselves Democrats compared with 30% of the general population in the United States) [44]. These demographics raise some concerns in terms of bias in study results and suggest limited ability to generalize results to the larger population.

Adding to the lack of representation among Twitter users is the disparity in Twitter activity among Twitter users. The median number of tweets for Twitter users is only 2 tweets per month. Just 10% of Twitter users account for 80% of the tweets across users in the United States [44]. In studies that use Twitter data, this disparity suggests that a large sample of tweets may only reflect, in reality, a much smaller sample of individuals.

Tweets were collected using the Twitter streaming API, which is limited to a random sample of 1% of all tweets sent by Twitter users at any given time. Of this limited sample of tweets, studies have shown that only approximately 1% to 2% of the tweets from the Twitter streaming API include geolocation information [20]. Because of the nature of this study, our analysis required geolocated tweets, significantly reducing the number of tweets allowed in our sample. As a result, we excluded many census tracts in the 25 cities from our sample because of a lack of geolocated tweets that were also food related. In addition, census tracts that did contain geolocated food-related tweets may have had only a small number of tweets and these tweets may not be representative of the tweets of all Twitter users who reside in a particular census tract. As our analysis is limited to geolocated tweets, there is also the potential for tweets without location information to differ significantly from tweets with geolocation information, which may suggest biased results because of unknown underlying factors.

Despite these limitations, the results of this study confirm both our hypotheses, demonstrating that food ingestion language found in tweets provides a signal that differentiates food deserts from non–food deserts.

Conclusions

The issue of food insecurity is an important public health issue because of the adverse health outcomes and underlying racial and economic disparities that are associated with insufficient access to healthy foods [4]. Social media data have been increasingly used to answer questions related to health and well-being. Prior research has used various data sources for identifying regions classified as food deserts [4], but this study suggests that perhaps the individuals in these regions unknowingly provide their own accounts of food consumption and food insecurity on social media. In this study, we demonstrated that food desert status is associated with food ingestion language found on Twitter and that food ingestion language can be used to predict and assess the food environment in American neighborhoods.

Acknowledgments

The authors thank Nhat Pham, Daniela Nganjo, and Pauline Comising for their assistance with quality control activities associated with the Twitter data.

Conflicts of Interest

None declared.

References

1. Azétsop J, Joy TR. Access to nutritious food, socioeconomic individualism and public health ethics in the USA: a common good approach. Philos Ethics Humanit Med 2013 Oct 29;8(1):16 [FREE Full text] [doi: 10.1186/1747-5341-8-16] [Medline: 24165577]
2. Definitions of food security. U.S. Department of Agriculture. URL: https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/definitions-of-food-security/ [accessed 2021-05-05]
3. Rose D. Access to healthy food: a key focus for research on domestic food insecurity. J Nutr 2010 Jun;140(6):1167-1169 [FREE Full text] [doi: 10.3945/jn.109.113183] [Medline: 20410087]
4. De Choudhury M, Sharma S, Kiciman E. Characterizing dietary choices, nutrition, and language in food deserts via social media. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 2016 Presented at: CSCW '16: Computer Supported Cooperative Work and Social Computing; Feb 27- Mar 2, 2016; San Francisco California USA. [doi: 10.1145/2818048.2819956]
5. Crowe J, Lacy C, Columbus Y. Barriers to food security and community stress in an urban food desert. Urban Sci 2018 May 31;2(2):46. [doi: 10.3390/urbansci2020046]
6. Sansom G, Hannibal B. Disparate access to nutritional food; place, race and equity in the United States. BMC Nutr 2021 Jun 29;7(1):29 [FREE Full text] [doi: 10.1186/s40795-021-00434-2] [Medline: 34183071]
7. Murrell A, Jones R. Erratum: Murrell, a., et al. Measuring food insecurity using the food abundance index: implications for economic, health and social well-being. 2020, 17, 2434. Int J Environ Res Public Health 2020 Sep 07;17(18):6495 [FREE Full text] [doi: 10.3390/ijerph17186495] [Medline: 32906601]
8. Block JP, Scribner RA, DeSalvo KB. Fast food, race/ethnicity, and income: a geographic analysis. Am J Prev Med 2004 Oct;27(3):211-217. [doi: 10.1016/j.amepre.2004.06.007] [Medline: 15450633]
9. Freedman DA. Local food environments: they're all stocked differently. Am J Community Psychol 2009 Dec;44(3-4):382-393. [doi: 10.1007/s10464-009-9272-6] [Medline: 19838789]
10. Freedman DA, Bell BA. Access to healthful foods among an urban food insecure population: perceptions versus reality. J Urban Health 2009 Nov 7;86(6):825-838 [FREE Full text] [doi: 10.1007/s11524-009-9408-x] [Medline: 19898939]
11. Lewis LB, Sloane DC, Nascimento LM, Diamant AL, Guinyard JJ, Yancey AK, et al. African Americans’ access to healthy food options in South Los Angeles Restaurants. Am J Public Health 2005 Apr;95(4):668-673. [doi: 10.2105/ajph.2004.050260]
12. Morland K, Filomena S. Disparities in the availability of fruits and vegetables between racially segregated urban neighbourhoods. Public Health Nutr 2007 Dec 01;10(12):1481-1489. [doi: 10.1017/s1368980007000079]
13. Glanz K, Sallis JF, Saelens BE, Frank LD. Nutrition Environment Measures Survey in stores (NEMS-S): development and evaluation. Am J Prev Med 2007 Apr;32(4):282-289. [doi: 10.1016/j.amepre.2006.12.019] [Medline: 17383559]
14. Inagami S, Cohen DA, Finch BK, Asch SM. You are where you shop: grocery store locations, weight, and neighborhoods. Am J Prev Med 2006 Jul;31(1):10-17. [doi: 10.1016/j.amepre.2006.03.019] [Medline: 16777537]

15. Food Access Research Atlas. U.S. Department of Agriculture. URL: https://www.ers.usda.gov/data-products/food-access-research-atlas/about-the-atlas/#definitions [accessed 2021-05-05]
16. Vydiswaran VG, Romero DM, Zhao X, Yu D, Gomez-Lopez I, Lu JX, et al. Uncovering the relationship between food-related discussion on Twitter and neighborhood characteristics. J Am Med Inform Assoc 2020 Feb 01;27(2):254-264 [FREE Full text] [doi: 10.1093/jamia/ocz181] [Medline: 31633756]
17. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. In: Proceedings of the International AAAI Conference on Web and Social Media. 2021 Presented at: International AAAI Conference on Web and Social Media; Jun 7-10, 2021; Virtual.
18. Paul M, Dredze M. You are what you tweet: analyzing twitter for public health. In: Proceedings of the International AAAI Conference on Web and Social Media. 2021 Presented at: International AAAI Conference on Web and Social Media; Jun 7-10, 2021; Virtual.
19. Sadilek A, Kautz H, Silenzio V. Modeling spread of disease from social interactions. In: Proceedings of the International AAAI Conference on Web and Social Media. 2021 Presented at: International AAAI Conference on Web and Social Media; 2021; Virtual.
20. Nguyen QC, Li D, Meng H, Kath S, Nsoesie E, Li F, et al. Building a national neighborhood dataset from geotagged Twitter data for indicators of happiness, diet, and physical activity. JMIR Public Health Surveill 2016 Oct 17;2(2):e158 [FREE Full text] [doi: 10.2196/publichealth.5869] [Medline: 27751984]
21. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
22. Sattar NS, Arifuzzaman S. COVID-19 vaccination awareness and aftermath: public sentiment analysis on Twitter data and vaccinated population prediction in the USA. Applied Sci 2021 Jun 30;11(13):6128. [doi: 10.3390/app11136128]
23. Liu S, Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: a sentiment analysis. Vaccine 2021 Sep 15;39(39):5499-5505 [FREE Full text] [doi: 10.1016/j.vaccine.2021.08.058] [Medline: 34452774]
24. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial intelligence-enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study. J Med Internet Res 2021 Apr 05;23(4):e26627 [FREE Full text] [doi: 10.2196/26627] [Medline: 33724919]
25. Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. Int J Infect Dis 2021 Jul;108:256-262 [FREE Full text] [doi: 10.1016/j.ijid.2021.05.059] [Medline: 34052407]
26. Ansari MT, Khan NA. Worldwide COVID-19 vaccines sentiment analysis through twitter content. Electron J Gen Med 2021;18(6). [doi: 10.29333/ejgm/11316]
27. Lyu JC, Han EL, Luli GK. COVID-19 vaccine-related discussion on Twitter: topic modeling and sentiment analysis. J Med Internet Res 2021 Jun 29;23(6):e24435 [FREE Full text] [doi: 10.2196/24435] [Medline: 34115608]
28. Jiang LC, Chu TH, Sun M. Characterization of vaccine tweets during the early stage of the COVID-19 outbreak in the United States: topic modeling analysis. JMIR Infodemiology 2021 Sep 14;1(1):e25636 [FREE Full text] [doi: 10.2196/25636] [Medline: 34604707]
29. City and town population totals: 2010-2019. United States Census Bureau. URL: https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-cities-and-towns.html [accessed 2022-01-29]
30. Committee on Public Health Strategies to Improve Health; Institute of Medicine. For the Public’s Health: Investing in a Healthier Future. Washington, DC: National Academies Press; 2012.
31. Filtering Tweets by location. Twitter. URL: https://developer.twitter.com/en/docs/tutorials/filtering-tweets-by-location [accessed 2022-02-17]
32. FoodData Central. U.S. Department of Agriculture. URL: https://fdc.nal.usda.gov/ [accessed 2022-04-19]
33. List of fast food restaurant chains. Wikipedia. URL: https://en.wikipedia.org/w/index.php?title=List_of_fast_food_restaurant_chains&oldid=1082336902 [accessed 2022-04-13]
34. Silge J, Robinson D. tidytext: text mining and analysis using tidy data principles in R. J Open Source Softw 2016 Jul;1(3):37. [doi: 10.21105/joss.00037]
35. Zhang Z, Yin D, Virrantaus K, Ye X, Wang S. Modeling human activity dynamics: an object-class oriented space–time composite model based on social media and urban infrastructure data. Comput Urban Sci 2021 May 06;1(1). [doi: 10.1007/s43762-021-00006-x]
36. Documentation. U.S. Department of Agriculture. URL: https://www.ers.usda.gov/data-products/food-access-research-atlas/documentation/#lowaccess [accessed 2022-04-25]
37. Characteristics and influential factors of food deserts. U.S. Department of Agriculture. URL: https://www.ers.usda.gov/publications/pub-details/?pubid=45017 [accessed 2021-05-05]
38. Cao XH, Stojkovic I, Obradovic Z. A robust data scaling algorithm to improve classification accuracies in biomedical data. BMC Bioinformatics 2016 Sep 09;17(1):359 [FREE Full text] [doi: 10.1186/s12859-016-1236-x] [Medline: 27612635]
39. The caret Package. GitHub. URL: https://topepo.github.io/caret/ [accessed 2022-04-13]
40. Halimu C, Kasem A, Shah Newaz SH. Empirical comparison of area under ROC curve (AUC) and Mathew Correlation Coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification. In:

Proceedings of the 3rd International Conference on Machine Learning and Soft Computing. 2019 Presented at: ICMLSC 2019: 2019 the 3rd International Conference on Machine Learning and Soft Computing; Jan 25 - 28, 2019; Da Lat Viet Nam. [doi: 10.1145/3310986.3311023]

41. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place. PLoS One 2013 May 29;8(5):e64417 [FREE Full text] [doi: 10.1371/journal.pone.0064417] [Medline: 23734200]
42. World Happiness Report. Columbia University. URL: https://www.earth.columbia.edu/sitefiles/file/Sachs%20Writing/2012/World%20Happiness%20Report.pdf [accessed 2022-04-13]
43. Xu P, Dredze M, Broniatowski DA. The twitter social mobility index: measuring social distancing practices with geolocated tweets. J Med Internet Res 2020 Dec 03;22(12):e21499 [FREE Full text] [doi: 10.2196/21499] [Medline: 33048823]
44. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2021-04-15]

Abbreviations

API: application programming interface
AUC: area under the receiver operating characteristic curve
GIS: geographic information systems
SES: socioeconomic status
USDA: United States Department of Agriculture

Edited by H Bradley; submitted 14.10.21; peer-reviewed by N Cesare, M Alvarez de Mon, R Gore; comments to author 23.02.22; revised version received 16.05.22; accepted 27.05.22; published 05.07.22.

Please cite as:
Sigalo N, St Jean B, Frias-Martinez V
Using Social Media to Predict Food Deserts in the United States: Infodemiology Study of Tweets
JMIR Public Health Surveill 2022;8(7):e34285
URL: https://publichealth.jmir.org/2022/7/e34285
doi:10.2196/34285
PMID:35788108

Original Paper

Investigating the Impact of the New York State Flavor Ban on e-Cigarette–Related Discussions on Twitter: Observational Study

Yankun Gao[1], PhD; Zidian Xie[1], PhD; Dongmei Li[1], PhD

University of Rochester Medical Center, Rochester, NY, United States

Corresponding Author:
Dongmei Li, PhD
University of Rochester Medical Center
265 Crittenden Boulevard CU 420708
Rochester, NY, 14642-0708
United States
Phone: 1 5852767285
Email: Dongmei_Li@urmc.rochester.edu

Abstract

Background: On May 18, 2020, the New York State Department of Health implemented a statewide flavor ban to prohibit the sales of all flavored vapor products, except for tobacco or any other authorized flavor.

Objective: This study aims to investigate the discussion changes in e-cigarette–related tweets over time with the implementation of the New York State flavor ban.

Methods: Through the Twitter streaming application programming interface, 59,883 e-cigarette–related tweets were collected within the New York State from February 6, 2020, to May 17, 2020 (period 1, before the implementation of the flavor ban), May 18, 2020-June 30, 2020 (period 2, between the implementation of the flavor ban and the online sales ban), July 1, 2020-September 15, 2020 (period 3, the short term after the online sales ban), and September 16, 2020-November 30, 2020 (period 4, the long term after the online sales ban). Sentiment analysis and topic modeling were conducted to investigate the changes in public attitudes and discussions in e-cigarette–related tweets. The popularity of different e-cigarette flavor categories was compared before and after the implementation of the New York State flavor ban.

Results: Our results showed that the proportion of e-cigarette–related tweets with negative sentiment significantly decreased (4305/13,246, 32.5% vs 3855/14,455, 26.67%, $P<.001$), and tweets with positive sentiment significantly increased (5246/13,246, 39.6% vs 7038/14,455, 48.69%, $P<.001$) in period 4 compared to period 3. “Teens and nicotine products” was the most frequently discussed e-cigarette–related topic in the negative tweets. In contrast, “nicotine products and quitting” was more prevalent in positive tweets. The proportion of tweets mentioning mint and menthol flavors significantly increased right after the flavor ban and decreased to lower levels over time. The proportions of fruit and sweet flavors were most frequently mentioned in period 1, decreased in period 2, and dominated again in period 4.

Conclusions: The proportion of e-cigarette–related tweets with different attitudes and frequently discussed flavor categories changed over time after the implementation of the New York State ban of flavored vaping products. This change indicated a potential impact of the flavor ban on public discussions of flavored e-cigarettes.

(JMIR Public Health Surveill 2022;8(7):e34114) doi:10.2196/34114

KEYWORDS

New York State flavor ban; e-cigarettes; twitter; topic modeling; sentiment analysis

Introduction

Electronic cigarettes (e-cigarettes) are battery-powered devices that allow users to inhale substances by heating vaporized liquid, which usually contains nicotine, flavoring agents, propylene glycol/vegetable glycerin, and other additives [1]. In recent years, e-cigarette use has increased rapidly among youth and young adults in the United States. From 2014 to 2020, e-cigarettes were the most commonly used tobacco product among high school and middle school students [2]. The JUUL e-cigarette system’s compact design, high nicotine content, and myriad of flavor options allowed it to dominate the US e-cigarette market by 2018, and it has become increasingly popular among teens since mid-2015 [2,3]. Multiple studies have suggested that e-cigarette use is associated with adverse health consequences, such as respiratory disorders, mental health problems, cognitive issues, and cancers [4-9]. In addition, e-cigarette use is unlikely to help adult smokers quit smoking [10]. By contrast, young people who use e-cigarettes are more likely to initiate smoking than those who do not [11], and the nicotine in e-cigarettes can increase the risk of addiction to other substances [12].

Flavored tobacco products can hide the harsh taste of tobacco, and the various flavors available have become the most attractive feature of e-cigarettes for youth and young adults [13,14]. An estimated 72.2% of high school and 59.2% of middle school students reported using flavored e-cigarettes, with the most popular flavors being fruit, menthol or mint, and various sweets [2]. Compared to adults, youth were more likely to use multiple flavor categories, and the most reported combination was fruit and candy [15]. While many flavors contained in e-liquids are used as food additives and scents, there are concerns about their safety in the lung [16,17]. Studies have suggested that the flavoring chemicals in e-cigarettes can harm lung tissue by impairing the cilia function in the airway epithelium and imposing inflammatory and oxidative responses in lung cells [18-20]. The inhalation of cinnamaldehyde, a flavoring agent commonly used in e-cigarettes, may increase the risk of respiratory infections in e-cigarette users [21]. The presence of vanillin is related to higher toxicity values, and the concentration of vanillin and cinnamaldehyde has been correlated with toxicity [22]. In addition, the variety of e-cigarette flavors has rapidly grown. While 7764 unique e-cigarette flavors were reported on the internet in January 2014, there was a net increase of 242 new flavors monthly in the 17 months that followed [23].

Due to their potential adverse health consequences, various policies have been announced or implemented to protect young people from flavored e-cigarettes. On November 15, 2018, the US Food and Drug Administration (FDA) implemented age restrictions on the sale of flavored e-cigarettes in physical locations and heightened age verification procedures for online sales [24]. On June 25, 2019, San Francisco banned the sale and distribution of e-cigarettes in the city to keep them away from young people [25]. On September 17, 2019, New York State announced its first-in-the-nation policy on flavored e-cigarettes [26]. That same year, multiple states and counties passed similar bans on the sale of flavored vaping products [3,27-30]. On January 2, 2020, the FDA announced their flavor enforcement policy to restrict flavored vaping products other than tobacco and menthol flavors, which were less popular among teenagers, and implemented the policy on February 6, 2020 [31]. On May 18, 2020, the New York State Department of Health implemented a flavor ban to prohibit the sales of all flavored vapor products, except for tobacco flavor and any flavored product that received a premarket approval order from the FDA, and banned the online sale of any vapor products on July 1, 2020 [32,33].

In this study, we aim to investigate sentiment changes in e-cigarette–related tweets over time with the implementation of the New York State flavor ban. We used Twitter data for this research since global tweets are updated continuously, allowing us to track public opinion in real time, which traditional surveys cannot generally provide. Twitter is a popular social media platform in the United States, with many users being teenagers and young adults [34]. In addition, previous studies successfully used Twitter data to study public perceptions of e-cigarette and e-cigarette–related policies [35,36]. In this study, we compared the sentiment and topic changes in e-cigarette–related tweets over time. In addition, we aimed to examine the changes in e-cigarette flavors mentioned on Twitter with the implementation of the New York State flavor ban. The findings of this study can provide insights into the potential impact of the flavor ban, which can be helpful for other policies on the regulation of flavored e-cigarettes.

Methods

Data Collection

E-cigarette–related tweets were collected through the Twitter streaming application programming interface (API) using e-cigarette–related keywords, including “e-cig,” “e-cigs,” “ecig,” “ecigs,” “electroniccigarette,” “ecigarette,” “ecigarettes,” “vape,” “vapers,” “vaping,” “vapes,” “e-liquid,” “ejuice,” “eliquid,” “e-juice,” “vapercon,” “vapeon,” “vapefam,” “vapenation,” and “juul” [37-39]. Twitter data were collected during 4 time periods: February 6, 2020-May 17, 2020 (before the implementation of the New York State flavor ban), May 18, 2020-June 30, 2020 (between the implementation of the flavor ban and the online sales ban), July 1, 2020-September 15, 2020 (the short term after the online sales ban), and September 16, 2020-November 30, 2020 (the long term after the online sales ban). September 15, 2020, was the midpoint of the data collected after implementing the flavor ban and therefore used as the cutoff between periods 3 and 4. As a result, a total of 2,929,784 unique e-cigarette–related tweets were collected.

To remove e-cigarette promotion tweets, Twitter posts and IDs were filtered out using promotion-related keywords, including “dealer,” “deal,” “customer,” “promotion,” “promo,” “promos,” “discount,” “sale,” “free shipping,” “sell,” “$,” “%,” “dollar,” “offer,” “percent off,” “store,” “save,” “price,” and “wholesale” [40]. After the promotion filtering, the data set contained 2,298,791 unique e-cigarette–related tweets. To investigate e-cigarette–related tweets within the state of New York State, geolocation keywords that contained city names related to the state, such as “New York,” “NY,” “Syracuse,” “Buffalo,” and so forth were used to filter users' geolocations. As a result, 59,883 e-cigarette–related tweets within the state of New York were obtained, with period 1 having 24,976 unique tweets, period 2 having 7206 unique tweets, period 3 having 13,246 unique tweets, and period 4 having 14,455 unique tweets.

Sentiment Analysis

The Valence Aware Dictionary and Sentiment Reasoner (VADER) was used as the sentiment analyzer to compute Twitter users’ attitudes [41]. A sentiment score was calculated for each tweet, ranging from −1.00 to +1.00. The attitudes of tweets were defined as positive if sentiment scores were in the range of +0.05 to +1.00, neutral if sentiment scores were in the range of −0.05 to +0.05, and negative if sentiment scores were in the range of −1.00 to −0.05. A score of −0.05 was included in the negative sentiment group and that of +0.05 was included in the positive sentiment group. To compare the sentiments between different periods, the number of tweets with different sentiments was normalized by the total number of tweets within each period. The 2-proportion *Z* test was used to compare the proportions of positive, neutral, and negative tweets between the different periods [42].

Topic Modeling

Topic modeling, specifically the latent Dirichlet allocation (LDA) model, was conducted to analyze the Twitter text content and determine the most frequently discussed topics. LDA is a generative model for text modeling, in which each word is assigned to a topic, and topics are generated with keywords and their corresponding weights [43]. LDA modeling was applied to tweets with different attitudes in the 4 periods. First, uppercase characters were converted to lowercase, and all punctuation, stop words, and white spaces were removed. Then, the Python library Genism was used to identify some frequent bigrams and trigrams [44]. Words were then lemmatized using spaCy to make all tenses present, and only nouns, verbs, adjectives, and adverbs were kept [44]. After all the data cleaning processing procedures, coherence scores were calculated, and the maximum coherence score was used to determine the number of topics [45]. Finally, the keywords and the percentage distribution of each topic were visualized with the pyLDAvis package [46].

Flavor Frequency

Further filtering was conducted using 1198 e-liquid flavor names from 129 e-liquid brands to identify tweets mentioning flavors [40]. As a result, 3714 tweets were collected in period 1, 1764 in period 2, 2027 in period 3, and 3544 in period 4. The e-cigarette flavors were grouped into 8 categories, including fruit, sweets, beverage, mint, menthol, tobacco, mixed, and other [40]. The proportions of different flavor categories were calculated by dividing the number of tweets mentioning each category by the total number of e-cigarette–related tweets in the same period. The proportions of each flavor category were compared between different periods by using the 2-proportion *Z* test with a significance level of 5%.

Ethics Approval

Only publicly available tweets were used for this study. There was no identifying information on Twitter users in this study. In addition, this study was reviewed and approved by the Office for Human Subject Protection Research Subjects Review at the University of Rochester (study ID STUDY00006570).

Results

Public Attitudes in e-Cigarette–Related Tweets in New York

To investigate whether the New York flavor ban could potentially affect public sentiments in e-cigarette–related tweets, the proportions of positive, neutral, and negative e-cigarette–related tweets before and after the flavor ban were compared at different periods using 2-sided two proportion *Z* tests (Multimedia Appendix 1). When comparing period 2 (May 18, 2020-June 30, 2020) to period 1 (February 6, 2020-May 17, 2020), we found no significant difference in the proportions for all attitudes (2738/7206, 38% vs 9342/24,976, 37.4% for positive, *P*=.35; 2054/7206, 28.5% vs 7381/24,976, 29.55% for neutral, *P*=.07; and 2414/7206, 33.5% vs 8253/24,976, 33.04% for negative, *P*=.43). From period 2 (May 18, 2020-June 30, 2020) to period 3 (July 1, 2020-September 15, 2020), the proportion of positive discussions significantly increased (2738/7206, 38% vs 5246/13,246, 39.6%, *P*=.025), while the proportions of neutral and negative tweets did not show significant changes (2054/7206, 28.5% vs 3694/13,246, 27.89%, *P*=.36 and 2414/7206, 33.5% vs 4305/13,246, 32.5%, *P*=.15, respectively). However, the proportion of both neutral (3694/13,246, 27.89% vs 3562/14,455, 24.64%, *P*<.001) and negative (4305/13,246, 32.5% vs 3855/14,455, 26.67%, *P*<.001) tweets significantly decreased, and the proportion of positive tweets (5246/13,246, 39.6% vs 7038/14,455, 48.69%, *P*<.001) significantly increased in period 4 (September 16, 2020-November 30, 2020) compared to period 3 (July 1, 2020-September 15, 2020).

Topics Discussed in e-Cigarette–Related Tweets in Different Periods

To determine if there was any change in e-cigarette–related topics discussed before and after the implementation of the New York State flavor ban, we applied the LDA topic modeling to e-cigarette–related Twitter posts from New York State. The popular topics discussed in the positive and negative tweets at different time periods are summarized in Table 1. We noticed that the popular topics in tweets with a positive sentiment were similar in all 4 periods. The majority of tweets focused on topics related to “nicotine products (include vaping and smoking) and quitting.” Moreover, 2 other topics, “nicotine products and health” and “nicotine products and people behavior,” were also frequently discussed in the positive tweets. However, while the “nicotine products and quitting” and “nicotine products and people behavior” topics also appeared in the tweets with a negative sentiment, “teens and nicotine products” was the most frequently discussed topic in the negative tweets in periods 1, 2, and 4. In period 2, “mint flavor vape in New York” became one of the most popular topics in the negative tweets.

Table 1. Top topics related to e-cigarettes before and after the New York State flavor ban.

Periods	Positive sentiment		Negative sentiment	
	Topics (% of tokens)	Keywords	Topics (% of tokens)	Keywords
February 6, 2020, to May 17, 2020[a] (24,976 posts)	Nicotine products and quitting (40.8)	vap, vape, nicotine, smoking, quit, cigarette, smoker, smoke, help, people, product, amp, tobacco, vaper, vaping	Teens and nicotine products (55.4)	vap, cigarette, nicotine, vape, amp, smoke, product, smoking, people, not, lung, quit, tobacco, teen, vaper
	Nicotine products and people behavior (36.1)	vape, juul, vap, get, go, smoke, good, not, be, make, pod, love, want, hit, do	Nicotine products and quitting (35.2)	juul, vape, get, go, vap, smoke, stop, pod, lose, not, be, fuck, shift, make, hit
May 18,2020, to June 30, 2020[b] (7206 posts)	Nicotine products and quitting (47.4)	vape, vap, juul, smoke, get, not, good, nicotine, go, be, people, quit, do, cigarette, smoking	nicotine products and quitting (54.9)	vap, vape, nicotine, smoke, amp, cigarette, tobacco, smoking, not, people, say, quit, year, product, vaper
	Nicotine products and health (26.9)	vap, smoker, vaper, smoke, nicotine, smoking, vape, product, quit, amp, help, tobacco, worldvapeday, health, study	mint flavor vape in New York (30.1)	vape, juul, fuck, pod, cana, anything, get, mint, new_York, month, people, get_rid, hit, police_force, like_two
July 1,2020, to September 15, 2020[c] (13,246 posts)	Nicotine products and people behavior (42.6)	vape, juul, get, vap, smoke, good, be, hit, go, not, make, amp, know, day, pen	Teens and nicotine products (48.3)	vap, stop, nicotine, white_people, pray, whole_time, downfall, mom_asking, vape, cigarette, smoke, smoking, vaper, tobacco, amp, teen
	Nicotine products and quitting (41.8)	vap, nicotine, vape, smoking, smoke, cigarette, quit, tobacco, smoker, product, map, vaping, vaper, not, help	Nicotine products and addiction (45.4)	vape, juul, get, vap, fuck, addict, go, smoke, drink, not, be, s, shit, pod, mad
September 16, 2020, to November 30, 2020[d] (14,455 posts)	Nicotine products and quitting (64%)	vape, vap, nicotine, smoke, juul, cigarette, good, get, smoking, quit, amp, smoker, people, be, go	Teens and nicotine products (54.2)	vap, vape, nicotine, smoke, not, smoking, teen, cigarette, quit, people, tobacco, do, amp, smoker, vaper
	Nicotine products and people behavior (26%)	vape, smoke, xbox, blow, series, not, say, can, believe, get, buy, bottom, feel_proud, twitter, created_post	Nicotine products and people behavior (36.3)	vape, juul, vap, go, get, fuck, pen, ita, shit, hit, be, lose, bad, take, dona

[a] Before implementation of the New York State flavor ban.

[b] Between the implementation of the New York State flavor ban and the online sales ban.

[c] Short term after the online sales ban.

[d] Long term after the online sales ban.

e-Cigarette Flavors Mentioned on Twitter

To investigate the potential effects of the New York State flavor ban on e-cigarette flavors discussed on Twitter, we compared the proportions of the e-cigarette flavor categories over time (Multimedia Appendix 2). Compared to period 1, in period 2, the proportions of tweets mentioning mint (296/3714, 8% vs 321/1764, 18%, *P*<.001) and menthol (312/3714, 8.4% vs 340/1764, 19%, *P*<.001) flavors significantly increased, while the proportions of tweets mentioning fruit (1559/3714, 41% vs 674/1764, 38%, *P*<.05), sweet (829/3714, 22.3% vs 253/1764, 14%, *P*<.001), beverage (456/3714, 12.3% vs 103/1764, 6%, *P*<.001), and tobacco (164/3714, 4.4% vs 39/1764, 2%, *P*<.001) flavors significantly decreased. After the ban of the online sale of all vapor products, the proportion of tweets mentioning mint (321/1764, 18% vs 246/2027, 12%, *P*<.001) and menthol (340/1764, 19% vs 278/2027, 14%, *P*<.001) significantly decreased in period 3 compared to period 2 but was still significantly higher than that in period 1 (246/2027, 12% vs 296/3714, 8% and 278/2027, 14% vs 312/3714, 8.4%, respectively, *P*<.001 for both flavors). At the same time, the proportions of tweets mentioning sweets (253/1764, 14% vs 340/2,027, 17%, *P*<.05), beverage (103/1,764, 6% vs 216/2,027, 11%, *P*<.001), and tobacco (39/1764, 2% vs 80/2027, 4%, *P*<.001) flavors significantly increased in period 3 compared to period 2. In period 4, a few months after the implementation of the New York State flavor ban, the proportion of tweets mentioning mint (246/2027, 12% vs 118/3544, 3%, *P*<.001), menthol (278/2027, 14% vs 160/3544, 5%, *P*<.001), beverage (216/2027, 11% vs 160/3544, 5%, *P*<.001), and tobacco (80/2027, 4% vs 68/3544, 2%, *P*<.001) significantly decreased, while the proportion of tweets mentioning fruit (819/2027, 40% vs 2004/3544, 57%, *P*<.001) and sweet flavors (340/2027, 17% vs 976/3544, 28%, *P*<.001) significantly increased compared to period 3.

Discussion

Principal Findings

To reduce the use of flavored e-cigarettes by young people and limit the appeal of flavored e-cigarette products, New York State implemented a flavor ban on May 18, 2020, and an online sales ban on July 1, 2020 [32,47]. In this study, we showed that after the implementation of the flavor ban, the proportion of positive e-cigarette–related tweets in New York State significantly increased over time, while the proportion of tweets with neutral or negative sentiments significantly decreased. The main topics in positive tweets were nicotine products, quitting, and health in all 4 periods, while the negative tweets focused on teens and nicotine products. Among all the tweets mentioning flavors, the proportion of tweets mentioning mint or menthol flavors significantly increased right after the implementation of the New York flavor ban and then gradually decreased.

Comparison With Prior Work

In this study, we showed that the most frequently mentioned flavors before implementing the New York State flavor ban were fruit (1559/3714, 42%) and sweets (829/3714, 22.3%), which was consistent with previous studies that fruit and sweet flavors are the most popular e-cigarette flavor categories [40]. Before the New York State flavor ban, the FDA implemented a flavor enforcement policy to restrict closed system devices containing flavored liquids other than tobacco and menthol flavors, which were less preferred by teenagers [31]. Different types of e-cigarettes (eg, disposable e-cigarettes [48]) might have partially contributed to the high proportions of fruit and sweet flavors discussed in period 1.

Right after the implementation of the New York State flavor ban, the percentage of tweets mentioning menthol increased to 19% (340/1764) in period 2, compared to 8.4% (312/3714) in period 1 (Multimedia Appendix 2). The ban limited the sales of menthol-flavored vapor products, which the FDA enforcement policy allowed. Therefore, many discussions around the newly banned menthol flavor appeared on Twitter in period 2, such as “I had a dream i went to the gas station and my menthol juul pods were back” and “They should bring back the Menthol juul pod so people can stay off of Newports cuz Newports kill the hood that's facts...” The proportion of menthol flavor–related discussion declined to 5% (160/3544) by period 4, which indicated the potentially reduced availability of menthol-flavored vapor products in New York State.

Topics related to nicotine products and quitting were the most prevalent in tweets with positive sentiments, such as “I’m so much happier ever since i decided to be a healthier person and quit vaping and smoking weed.” However, in negative tweets, the topic of teens and nicotine products was actively discussed, such as “All e-cigarette advertising is anti-smoking advertising. And smoking actually kills people. A LOT of people. Nicotine vaping does not. Anti-vaping advertising is well-funded and well-orchestrated. Their ads increase teen vaping and discourage smokers from quitting.” In period 4, the proportion of negative e-cigarette–related tweets significantly decreased (3855/14455, 26.67%) compared to period 3 (4305/13,246, 32.5%). These dramatic changes in the proportions of positive and negative tweets did not occur until period 4, a few months after the implementation of the flavor ban, which suggest an association between the policy and sentiment changes in e-cigarette–related tweets. “Teens and nicotine products” was a major topic discussed in both periods 3 and 4. The declined proportion indicated that public concerns about teen vaping might have reduced over time after the New York State flavor ban.

The youth initiation of e-cigarette use was associated with flavored e-cigarette products, and fruit and sweet flavors are the most popular flavor categories among youth in the United States [15,40,49,50]. Mint flavor is one of the most often used JUUL e-cigarettes flavors in middle school and high school students [51]. However, after the sale of all unauthorized flavors was banned in New York State, there was a significantly increased proportion of mint flavor–related tweets (321/1764,18% in period 2 vs 296/3714, 8% in period 1), while the proportions of tweets mentioning fruit, sweet, and beverage flavors decreased. Consistent with this result, mint-related discussions became a major topic in the negative tweets in period 2. In this period, there was one popular tweet, “How did New York get rid of mint juul pods in like two months but can’t do anything about the exceedingly racist police force?” This tweet mentioned mint flavor and other political topics and has since been retweeted virally, which might explain why the mint flavor became prominent among all the banned flavors in period 2.

Tobacco flavor is less preferred by US youth [51]. The New York State flavor ban prohibited the sales of all flavors other than tobacco. Multiple studies show that e-cigarette consumers are willing to shift to different flavors when certain popular flavors are restricted [50,52]. However, we did not observe an increase in tobacco flavor–related discussions after the New York State flavor ban. In contrast, discussions about fruit and sweet flavors still dominated the tweets mentioning flavors. One reason might be that the amount of discussions on social media does not necessarily reflect the amount of e-cigarette use in real life. Another possible explanation might be that people turn to other sources to get the flavored e-cigarettes they like, such as buying from other states or the black market.

Limitations

This study has several limitations. First, we collected tweets from New York State using the geographical location of users for our analysis. However, most Twitter users are not willing to share their locations in their tweets [53], which might have introduced some biases to our study. Second, we aimed to analyze the potential effects of the New York flavor ban on public attitudes and user behaviors. However, the FDA enforcement policy was implemented a few months before the ban. Therefore, the changes we observed in this study might have been a combination of the 2 policies other than the flavor ban alone. Moreover, e-cigarette bans were not new to the United States at the time of the New York State ban, and attitudes collected via Twitter may have been a reaction to this long-term trend rather than an acute event. Third, we investigated the changes in attitudes of e-cigarette–related tweets before and after the New York State flavor ban, which only reflects general discussion rather than self-reports of e-cigarette usage. Therefore, whether people quit vaping or switched to other available nicotine products after the flavor ban remains unknown and will be further pursued in future studies. Fourth, our study only showed the association between flavor ban and the attitude changes in e-cigarette–related tweets, which cannot determine the causal effects of the New York flavor ban on public attitudes. Finally, there were only 5 months of data collected after the implementation of the flavor ban. As a result, public discussions of e-cigarettes in the longer term remain unknown.

Conclusions

Using Twitter data, this study showed the changes in public attitudes in e-cigarette–related discussions and changes in the proportion of e-cigarette flavor categories mentioned before and after the New York State flavor ban. Our results indicated that the public might have less concerns about teen vaping after the implementation of the flavor ban. In addition, our results showed that while the mentions of some banned flavors (eg, mint and menthol) temporally increased right after the flavor ban, the most popular flavors (eg, fruit and sweets) dominated discussions again over time. These results indicated that stricter nationwide policies are required to prohibit flavored nicotine products. The findings of this study provide some preliminary evidence of public responses to the New York State flavor ban as well as valuable insights for policy makers to further regulate flavored nicotine products.

Acknowledgments

The research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health (NIH) and the Food and Drug Administration (FDA) Center for Tobacco Products (award number U54CA228110). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or FDA.

Data Availability

The data and scripts used for analysis and creating figures are available upon request from the corresponding author.

Authors' Contributions

ZX and DL conceived and designed the study, assisted with interpretation of analyses, and edited the manuscript. YG analyzed the data and wrote the manuscript. All authors have approved the final article.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sentiment changes in e-cigarette–related tweets with the implementation of the New York State flavor ban. * indicates $P<.05$, ** indicates $P<.01$. February 6, 2020-May 17, 2020, before implementation of the flavor ban; May 18, 2020-June 30, 2020, between the implementation of the flavor ban and the online sales ban; July 1, 2020-September 15, 2020, the short term after the online sales ban; September 16, 2020-November 30, 2020, the long term after the online sales ban.

[PNG File , 74 KB - publichealth_v8i7e34114_app1.png]

Multimedia Appendix 2

Proportions of tweets mentioning e-cigarette flavor categories over time before and after the New York State flavor ban. February 6, 2020-May 17, 2020, before implementation of the flavor ban; May 18, 2020-June 30, 2020, between the implementation of the flavor ban and the online sales ban; July 1, 2020-September 15, 2020, the short term after the online sales ban; September 16, 2020-November 30, 2020, the long term after the online sale ban.

[PNG File , 76 KB - publichealth_v8i7e34114_app2.png]

References

1. Laestadius LI, Wahl MM, Cho YI. #Vapelife: an exploratory study of electronic cigarette use and promotion on instagram. Subst Use Misuse 2016 Oct 14;51(12):1669-1673. [doi: 10.1080/10826084.2016.1188958] [Medline: 27484191]
2. Cullen KA, Gentzke AS, Sawdey MD, Chang JT, Anic GM, Wang TW, et al. e-cigarette use among youth in the United States, 2019. JAMA 2019 Dec 03;322(21):2095-2103 [FREE Full text] [doi: 10.1001/jama.2019.18387] [Medline: 31688912]
3. Michigan's ban of flavored nicotine vaping products to protect children's health effective immediately with rules filing. Michigan.gov. 2019. URL: https://www.michigan.gov/mdhhs/0,5885,7-339--507760--,00.html [accessed 2022-06-10]
4. McConnell R, Barrington-Trimis JL, Wang K, Urman R, Hong H, Unger J, et al. Electronic cigarette use and respiratory symptoms in adolescents. Am J Respir Crit Care Med 2017 Apr 15;195(8):1043-1049. [doi: 10.1164/rccm.201604-0804oc]
5. Rubenstein DA, Hom S, Ghebrehiwet B, Yin W. Tobacco and e-cigarette products initiate Kupffer cell inflammatory responses. Mol Immunol 2015 Oct;67(2):652-660. [doi: 10.1016/j.molimm.2015.05.020] [Medline: 26072673]

6. MacDonald A, Middlekauff HR. Electronic cigarettes and cardiovascular health: what do we know so far? VHRM 2019 Jun;15:159-174. [doi: 10.2147/vhrm.s175970]
7. Obisesan OH, Mirbolouk M, Osei AD, Orimoloye OA, Uddin SMI, Dzaye O, et al. Association between e-cigarette use and depression in the behavioral risk factor surveillance system, 2016-2017. JAMA Netw Open 2019 Dec 02;2(12):e1916800 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.16800] [Medline: 31800073]
8. Grana R, Benowitz N, Glantz SA. E-cigarettes. Circulation 2014 May 13;129(19):1972-1986. [doi: 10.1161/circulationaha.114.007667]
9. Chen L, Lu X, Yuan J, Luo J, Luo J, Xie Z, et al. A social media study on the associations of flavored electronic cigarettes with health symptoms: observational study. J Med Internet Res 2020 Jun 22;22(6):e17496 [FREE Full text] [doi: 10.2196/17496] [Medline: 32568093]
10. Weaver SR, Huang J, Pechacek TF, Heath JW, Ashley DL, Eriksen MP. Are electronic nicotine delivery systems helping cigarette smokers quit? Evidence from a prospective cohort study of U.S. adult smokers, 2015-2016. PLoS One 2018 Jul 9;13(7):e0198047 [FREE Full text] [doi: 10.1371/journal.pone.0198047] [Medline: 29985948]
11. Hair EC, Barton AA, Perks SN, Kreslake J, Xiao H, Pitzer L, et al. Association between e-cigarette use and future combustible cigarette use: Evidence from a prospective cohort of youth and young adults, 2017-2019. Addict Behav 2021 Jan;112:106593. [doi: 10.1016/j.addbeh.2020.106593] [Medline: 32927247]
12. E-Cigarette Use Among Youth and Young Adults. A Report of the Surgeon General.: U.S. Department of Health and Human Services; 2016. URL: https://www.ncbi.nlm.nih.gov/books/NBK538680/ [accessed 2022-06-16]
13. King BA, Tynan MA, Dube SR, Arrazola R. Flavored-little-cigar and flavored-cigarette use among US middle and high school students. J Adolesc Health 2014 Jan;54(1):40-46 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.07.033] [Medline: 24161587]
14. Gentzke AS, Wang TW, Jamal A, Park-Lee E, Ren C, Cullen KA, et al. Tobacco product use among middle and high school students - United States, 2020. MMWR Morb Mortal Wkly Rep 2020 Dec 18;69(50):1881-1888 [FREE Full text] [doi: 10.15585/mmwr.mm6950a1] [Medline: 33332300]
15. Schneller LM, Bansal-Travers M, Goniewicz ML, McIntosh S, Ossip D, O'Connor RJ. Use of flavored e-cigarettes and the type of e-cigarette devices used among adults and youth in the US-results from wave 3 of the population assessment of tobacco and health study (2015-2016). Int J Environ Res Public Health 2019 Aug 20;16(16):2991 [FREE Full text] [doi: 10.3390/ijerph16162991] [Medline: 31434229]
16. Klager S, Vallarino J, MacNaughton P, Christiani DC, Lu Q, Allen JG. Flavoring chemicals and aldehydes in e-cigarette emissions. Environ Sci Technol 2017 Sep 19;51(18):10806-10813. [doi: 10.1021/acs.est.7b02205] [Medline: 28817267]
17. Gotts JE, Jordt S, McConnell R, Tarran R. What are the respiratory effects of e-cigarettes? BMJ 2019 Sep 30;366:l5275 [FREE Full text] [doi: 10.1136/bmj.l5275] [Medline: 31570493]
18. Gerloff J, Sundar IK, Freter R, Sekera ER, Friedman AE, Robinson R, et al. Inflammatory response and barrier dysfunction by different e-cigarette flavoring chemicals identified by gas chromatography-mass spectrometry in e-liquids and e-vapors on human lung epithelial cells and fibroblasts. Appl In Vitro Toxicol 2017 Mar 01;3(1):28-40 [FREE Full text] [doi: 10.1089/aivt.2016.0030] [Medline: 28337465]
19. Muthumalage T, Prinz M, Ansah KO, Gerloff J, Sundar IK, Rahman I. Inflammatory and oxidative responses induced by exposure to commonly used e-cigarette flavoring chemicals and flavored e-liquids without nicotine. Front Physiol 2017 Jan 11;8:1130 [FREE Full text] [doi: 10.3389/fphys.2017.01130] [Medline: 29375399]
20. Park H, O'Sullivan M, Vallarino J, Shumyatcher M, Himes BE, Park J, et al. Transcriptomic response of primary human airway epithelial cells to flavoring chemicals in electronic cigarettes. Sci Rep 2019 Feb 01;9(1):1400 [FREE Full text] [doi: 10.1038/s41598-018-37913-9] [Medline: 30710127]
21. Clapp PW, Lavrich KS, van Heusden CA, Lazarowski ER, Carson JL, Jaspers I. Cinnamaldehyde in flavored e-cigarette liquids temporarily suppresses bronchial epithelial cell ciliary motility by dysregulation of mitochondrial function. Am J Physiol Lung Cell Mol Physiol 2019 Mar 01;316(3):L470-L486 [FREE Full text] [doi: 10.1152/ajplung.00304.2018] [Medline: 30604630]
22. Sassano MF, Davis ES, Keating JE, Zorn BT, Kochar TK, Wolfgang MC, et al. Evaluation of e-liquid toxicity using an open-source high-throughput screening assay. PLoS Biol 2018 Mar 27;16(3):e2003904 [FREE Full text] [doi: 10.1371/journal.pbio.2003904] [Medline: 29584716]
23. Zhu S, Sun JY, Bonnevie E, Cummins SE, Gamst A, Yin L, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. Tob Control 2014 Jul 16;23 Suppl 3(suppl 3):iii3-iii9 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051670] [Medline: 24935895]
24. Scott Gottlieb. Statement from FDA Commissioner Scott Gottlieb, M. D., on proposed new steps to protect youth by preventing access to flavored tobacco products and banning menthol in cigarettes. 2018. URL: https://tinyurl.com/bddrvbxw [accessed 2022-06-16]
25. Restricting the sale, manufacture, and distribution of tobacco products, including electronic cigarettes. SFGov Legistar. URL: https://sfgov.legistar.com/View.ashx?M=F&ID=7111897&GUID=7C3912E3-BB56-420D-896E-6FEA1391287D [accessed 2022-06-10]

26. Governor Cuomo announces New York State implements first-in-the-nation ban on flavored e-cigarettes. Governer.NY.Gov. URL: https://www.governor.ny.gov/news/governor-cuomo-announces-new-york-state-implements-first-nation-ban-flavored-e-cigarettes [accessed 2022-06-10]
27. Emergency Notice of Changes in Regulations That May Affect Your Business. Health.ri.gov. 2019. URL: https://health.ri.gov/publications/letters/20191004_ERECigFlavorBan_LetterToRetailers_EnglishAndSpanish.pdf [accessed 2022-06-16]
28. OHA, OLCC file rules banning flavored vaping sales, including online. Oregon.gov. 2019. URL: https://www.oregon.gov/oha/ERD/Pages/OLCCFileRulesBanningFlavoredVapingSalesIncludingOnline.aspx [accessed 2022-06-16]
29. Governor Bullock directs ban on flavored e-cigarettes to address public health emergency. Montana.gov. 2019. URL: https://dphhs.mt.gov/aboutus/news/2019/flavorede-cigban [accessed 2021-10-06]
30. Washington State bans vapor products containing vitamin e acetate. Washington State Department of Health. 2019. URL: https://www.doh.wa.gov/Newsroom/Articles/ID/1038/Washington-state-bans-vapor-products-containing-vitamin-E-acetate [accessed 2021-10-06]
31. FDA finalizes enforcement policy on unauthorized flavored cartridge-based e-cigarettes that appeal to children, including fruit and mint. US Food and Drug Administration. 2020. URL: https://www.fda.gov/news-events/press-announcements/fda-finalizes-enforcement-policy-unauthorized-flavored-cartridge-based-e-cigarettes-appeal-children [accessed 2021-10-06]
32. New York State Department of Health announces statewide ban of flavored nicotine vapor products takes effect today. Health.NY.gov. 2020. URL: https://health.ny.gov/press/releases/2020/2020-05-18_fl_nicotine_vapor_products_ban.htm [accessed 2021-10-06]
33. Diane Caruana. New York's flavour ban goes into effect on July 1st. Vaping Post. 2020. URL: https://www.vapingpost.com/2020/05/25/new-yorks-flavour-ban-goes-into-effect-on-july-1st/ [accessed 2021-10-06]
34. Omnicoreagency. 2021. URL: https://www.omnicoreagency.com/periscope-statistics/ [accessed 2021-10-06]
35. Lu X, Sun L, Xie Z, Li D. Perception of the Food and Drug Administration electronic cigarette flavor enforcement policy on Twitter: observational study. JMIR Public Health Surveill 2022 Mar 29;8(3):e25697 [FREE Full text] [doi: 10.2196/25697] [Medline: 35348461]
36. Sun L, Lu X, Xie Z, Li D. Public reactions to the New York State policy on flavored electronic cigarettes on Twitter: observational study. JMIR Public Health Surveill 2022 Feb 03;8(2):e25216 [FREE Full text] [doi: 10.2196/25216] [Medline: 35113035]
37. Gao Y, Xie Z, Sun L, Xu C, Li D. Electronic cigarette–related contents on Instagram: observational study and exploratory analysis. JMIR Public Health Surveill 2020 Nov 05;6(4):e21963 [FREE Full text] [doi: 10.2196/21963] [Medline: 33151157]
38. Gao Y, Xie Z, Li D. Electronic cigarette users' perspective on the COVID-19 pandemic: observational study using Twitter data. JMIR Public Health Surveill 2021 Jan 05;7(1):e24859 [FREE Full text] [doi: 10.2196/24859] [Medline: 33347422]
39. The Streaming APIs. 2021. URL: https://developer.twitter.com/en/docs [accessed 2021-10-06]
40. Lu X, Chen L, Yuan J, Luo J, Luo J, Xie Z, et al. User perceptions of different electronic cigarette flavors on social media: observational study. J Med Internet Res 2020 Jun 24;22(6):e17280 [FREE Full text] [doi: 10.2196/17280] [Medline: 32579123]
41. Hutto C, Gilbert E. VADER: parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media. 2014 Presented at: International AAAI Conference on Web and Social Media; 2014; Ann Arbor, Michigan p. 216-225 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14550 [doi: 10.4135/9781452244723.n466]
42. Schumacker R. Chapter 12. In: Z Test for Differences in Proportions. Learning Statistics Using R. Newbury Park, California: Sage Publishing; 2015:1-24.
43. Blei D, Ng A, Jordan M. Latent Dirichlet allocation. J Mach Learn Res 2003 Mar 01:993-1022.
44. Ghanoum T. Topic modelling in Python with spaCy and Gensim. Towards Data Science. 2021 Dec 20. URL: https://towardsdatascience.com/topic-modelling-in-python-with-spacy-and-gensim-dc8f7748bdbf [accessed 2022-06-16]
45. Korenčić D, Ristov S, Šnajder J. Document-based topic coherence measures for news media text. Expert Syst Appl 2018 Dec;114:357-373. [doi: 10.1016/j.eswa.2018.07.063]
46. Tran K. pyLDAvis: topic modelling exploration tool that every NLP data scientist should know. Neptune. 2021 Nov 15. URL: https://neptune.ai/blog/pyldavis-topic-modelling-exploration-tool-that-every-nlp-data-scientist-should-know [accessed 2022-06-16]
47. Gentzke AS, Wang TW, Jamal A, Park-Lee E, Ren C, Cullen KA, et al. Tobacco product use among middle and high school students - United States, 2020. MMWR Morb Mortal Wkly Rep 2020 Dec 18;69(50):1881-1888 [FREE Full text] [doi: 10.15585/mmwr.mm6950a1] [Medline: 33332300]
48. Sun L, Tao C, Xie Z, Li D. Promotion of disposable electronic cigarette flavors and topics on Twitter. Int J Environ Res Public Health 2020 Dec 10;17(24):9221 [FREE Full text] [doi: 10.3390/ijerph17249221] [Medline: 33321714]
49. Leventhal A, Goldenson NI, Cho J, Kirkpatrick MG, McConnell RS, Stone MD, et al. Flavored e-cigarette use and progression of vaping in adolescents. Pediatrics 2019 Nov;144(5):e20190789 [FREE Full text] [doi: 10.1542/peds.2019-0789] [Medline: 31659004]
50. Liber A, Cahn Z, Larsen A, Drope J. Flavored e-cigarette sales in the United States under self-regulation from January 2015 through October 2019. Am J Public Health 2020 Jun;110(6):785-787. [doi: 10.2105/ajph.2020.305667]

51. Leventhal AM, Miech R, Barrington-Trimis J, Johnston LD, O'Malley PM, Patrick ME. Flavors of e-cigarettes used by youths in the United States. JAMA 2019 Dec 03;322(21):2132-2134 [FREE Full text] [doi: 10.1001/jama.2019.17968] [Medline: 31688891]
52. Diaz MC, Donovan EM, Schillo BA, Vallone D. Menthol e-cigarette sales rise following 2020 FDA guidance. Tob Control 2021 Nov 23;30(6):700-703. [doi: 10.1136/tobaccocontrol-2020-056053] [Medline: 32967985]
53. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015 Sep 2;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]

Abbreviations

API: application programming interface
FDA: Food and Drug Administration
NIH: National Institutes of Health
LDA: latent Dirichlet allocation
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by H Bradley; submitted 06.10.21; peer-reviewed by N Cesare, J Trevino; comments to author 21.02.22; revised version received 08.04.22; accepted 10.05.22; published 08.07.22.

Please cite as:
Gao Y, Xie Z, Li D
Investigating the Impact of the New York State Flavor Ban on e-Cigarette–Related Discussions on Twitter: Observational Study
JMIR Public Health Surveill 2022;8(7):e34114
URL: https://publichealth.jmir.org/2022/7/e34114
doi:10.2196/34114
PMID:35802417

Original Paper

HIV and Sexually Transmitted Infection Testing Among Substance-Using Sexual and Gender Minority Adolescents and Young Adults: Baseline Survey of a Randomized Controlled Trial

Jayelin N Parker[1], MPH; Seul Ki Choi[2], MPH, PhD; Jose A Bauermeister[2], MPH, PhD; Erin E Bonar[1,3], PhD; Adam W Carrico[4], PhD; Rob Stephenson[1,5], MA, MSc, PhD

[1]Center for Sexuality and Health Disparities, University of Michigan, Ann Arbor, MI, United States
[2]Program for Sexuality, Technology, and Action Research, University of Pennsylvania, Philadelphia, PA, United States
[3]Addiction Center, Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States
[4]School of Public Health, University of Miami, Miami, FL, United States
[5]Department of Systems, Populations and Leadership, School of Nursing, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Jayelin N Parker, MPH
Center for Sexuality and Health Disparities
University of Michigan
400 N Ingalls
Room 3340
Ann Arbor, MI, 48109
United States
Phone: 1 7347644646
Email: jayelinp@umich.edu

Abstract

Background: Gay, bisexual, and other men who have sex with men and transgender individuals are more heavily affected by HIV and other sexually transmitted infections (STIs) than their cisgender, heterosexual peers. In addition, sexual and gender minorities who use substances are often at a further increased risk of HIV and other STIs. Increasing testing for HIV and other STIs allows this hardly reached population to receive early intervention, prevention, and education.

Objective: We explored HIV and STI testing patterns among 414 sexual and gender minority adolescents and young adults aged 15 to 29 years who self-reported substance use and lived in southeastern Michigan.

Methods: We analyzed data from the baseline survey of a 4-arm randomized controlled trial that aimed to examine the efficacy of a brief substance use intervention for creating gains in engagement in HIV prevention. We fit multinomial logistic regression models to 2 categorical HIV and STI testing variables (lifetime and previous 12 months) based on self-reports of testing (never, STIs only, HIV only, or both). In addition, we compared HIV and STI testing behaviors across demographic characteristics, structural factors, psychosocial barriers, substance use, and sexual behaviors.

Results: Our findings showed that 35.5% (147/414) of adolescents and young adults reported not being tested for either HIV or STIs in the previous year, and less than half (168/414, 40.6%) of the sample achieved the Centers for Disease Control and Prevention recommendation of HIV and STI testing once per year. We observed HIV and STI testing disparities across sociodemographic (eg, sexual identity, education, and income) and health (eg, substance use) correlates. Specifically, cisgender gay men who have sex with men were more likely to report being tested for HIV compared with bisexual men and transgender individuals, who were more likely to be tested for STIs.

Conclusions: This study illustrates the results of an HIV prevention intervention in southeastern Michigan showing the need for HIV prevention interventions that leverage structural factors, psychosocial barriers, and substance use as key drivers to achieve HIV and STI testing rates to meet the Centers for Disease Control and Prevention guidelines.

Trial Registration: ClinicalTrials.gov NCT02945436; http://clinicaltrials.gov/ct2/show/NCT02945436

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.9414

(JMIR Public Health Surveill 2022;8(7):e30944) doi:10.2196/30944

KEYWORDS

testing; substance use; sexual minority; social determinants

Introduction

Background

In the United States, sexual and gender minority populations continue to experience disproportionate burdens of HIV and other sexually transmitted infections (STIs) compared with heterosexual and cisgender populations. Although <0.5% of the adult American population lives with HIV [1], it is estimated that the prevalence of HIV is significantly higher among sexual and gender minorities—14.1% of transgender women and 3.2% of transgender men live with HIV [1], and 16.6% of gay, bisexual, and other men who have sex with men (GBMSM) live with HIV [2]. In 2018, 69% of new HIV diagnoses were among GBMSM [3] and, despite efforts to reduce HIV, the incidence rates among GBMSM aged <30 years continue to increase [4]. Although transgender individuals comprise only 2% of new HIV diagnoses, in 2018, the largest percentage (27%) of diagnoses of HIV infection among transgender persons was for transgender female adults and adolescents aged 25 to 29 years, followed by transgender female adults and adolescents aged 20 to 24 years (25%) [5]. A 2019 systematic review and meta-analysis found that an estimated 14% of transgender women live with HIV, with strong disparities by race and ethnicity—an estimated 44% of Black or African American transgender women, 26% of Hispanic and Latina transgender women, and 7% of White transgender women live with HIV [6].

GBMSM and transgender populations have also been shown to be at an increased risk of STIs other than HIV [7-10]. Although women in the United States have a higher disease burden of chlamydia than men, from 2014 to 2018, there was an increase of 37.8% in STI infection rates among men, a rise primarily credited to increasing incident STI rates among GBMSM in particular [11]. Across 18 STI clinics within 9 US STI surveillance regions, the median site-specific chlamydia prevalence for GBMSM was 5.8% in urethral cases, 16.1% in rectal cases, and 2.7% in pharyngeal cases in 2018 [11]. A recent systematic review found prevalence ranges as follows among transgender women: syphilis (1.4%-50.4%), gonorrhea (2.1%-19.1%), and chlamydia (2.7%-24.7%) [12].

Within sexual and gender minority populations, the rates of HIV and STIs are significantly higher among substance using individuals [13,14]. In a recent cross-sectional sample of 2216 youth (aged 12-26 years) living with HIV, 32.9% of participants reported weekly or more frequent tobacco use, 27.5% reported marijuana use, 21.3% reported alcohol use, and 22.5% reported other illicit drug use [15]. Concurrently addressing substance use and HIV and STI prevention and treatment has the potential to create significant gains in reducing rates of HIV and STIs among sexual and gender minority communities [13,16]. Although there is an abundance of literature demonstrating the high rates of HIV among young GBMSM [17-21], young transgender individuals [16,22,23], and people who use or misuse substances [24-26], little research attention has focused on the intersection of these marginalized groups and has sought to understand the patterns of HIV and STI testing among young, substance using, sexual and gender minority individuals.

Testing for HIV and other STIs is a critical entry point into the HIV and STI care continuums, allowing people to be aware of their status, obtain treatment, and take appropriate action to prevent further transmission [27]; however, testing rates are often low among young, substance using, sexual and gender minority communities [28-33]. A recent study with transgender individuals showed that, of those at risk of contracting HIV, 23% had never been tested for HIV [34]. In addition, there are wide variations in testing within sexual and gender minority populations by race and ethnicity—for example, Latinx GBMSM are tested less frequently than other GBMSM [35]. In a recent study of testing behaviors among GBMSM, 8% of HIV tests delivered were to first-time testers, and 70.7% of first-time testers were among racial and ethnic minorities [36]. The low rates of HIV and STI testing among sexual and gender minorities have been attributed to anticipated stigma, fear, and perception of individual risk, which often act as barriers to willingness and ability to be tested for HIV and STIs [37-41]. In addition, structural barriers such as anticipated or previous poor interaction with or trust in medical providers and stigma often prevent young, sexual and gender minority individuals from seeking testing for HIV and other STIs [39,42,43].

This study focused on understanding the factors associated with HIV and STI testing among sexual and gender minority adolescents and young adults (AYA) in Detroit, Michigan. The Detroit Metro Area contains the most heavily concentrated group of people living with HIV in the state of Michigan, where the highest HIV and STI infection rates are among those living in the city of Detroit [44]. In 2018, in the Detroit Metro Area, 75.5% of people living with HIV were people of color, 55% were GBMSM, and 18.3% were between the ages of 13 and 29 years [45]. Young African American or Black GBMSM (aged 15-19 years) were the only group of people to experience an increase in new HIV infections over the last 2 decades in Michigan, where only 0.1% of the population are young GBMSM [45]. The transgender community accounted for 1.8% of people living with HIV in Detroit, where 82.6% of transgender individuals living with HIV are Black or African American [45]. GBMSM account for 22% of new HIV infections [45] such that, in 2018, young GBMSM were 270 times more likely to be diagnosed with HIV than young non-GBMSM in the Detroit Metro Area [45].

Objectives

This study used baseline data from a randomized controlled trial with substance using, sexual and gender minority AYA recruited in southeastern Michigan to understand patterns of testing for HIV and STIs. The primary aim of this analysis was to understand HIV and STI testing behaviors among substance using sexual and gender minority AYA, a group for whom it is critical to develop interventions that can increase testing uptake. Understanding the social determinants (eg, gender identity, sexual identity, and education) associated with HIV and STI testing among this vulnerable population has the potential to inform the development of interventions tailored to the unique needs of a population experiencing substance use and multiple structural and interpersonal barriers to engaging in routine testing.

Methods

Study Procedures and Participants

We analyzed data from the baseline survey of a 4-arm randomized controlled trial of substance using, sexual and gender minority AYA (aged 15-29 years; N=414) that examined the efficacy of a brief substance use intervention for creating gains in engagement in HIV prevention. The eligibility criteria were as follows: being aged 15 to 29 years at the time of screening; living in southeastern Michigan (based on eligible zip codes); identifying as a man, male, or transgender person; having had condomless oral or anal sex at least once in the 6 months before screening; and having had at least one binge drinking or substance misuse experience (≥5 standard alcoholic beverages in a single setting) in the last 3 months before screening. The study recruitment period was from April 2017 to August 2019. Participants were walked through consent by a staff member, ensuring that they could opt out of the study at any time and were not forced to participate.

Ethics Approval

The study was approved by the Institutional Review Board of the University of Michigan (HUM00105125). The trial is also registered at ClinicalTrials.gov (NCT02945436). More information regarding the randomized controlled trial is available in a detailed protocol paper [46].

Measures

HIV and STI Testing

The participants responded whether they had ever been tested for HIV or STIs in their lifetime. On the basis of the participants' responses, we created a categorical variable of lifetime HIV and STI testing with 4 groups: never tested for either HIV or STIs, tested for STIs only, tested for HIV only, or tested for both HIV and STIs. If they had been tested in the past, the participants responded whether they had been tested in the 12 months before the survey. Similar to lifetime testing, we created a categorical variable of previous–12-month HIV and STI testing: never tested for HIV and STIs, tested for STIs only, tested for HIV only, or tested for both HIV and STIs.

Individual Factors

We included demographic characteristics (ie, age, race, ethnicity [Hispanic or Latinx], gender identity, sexual identity, and disability) in the analysis. Age was dichotomized as 15 to 21 years or 22 to 29 years. The participants were asked about their gender identity using 4 response options (cisgender, transgender men, transgender women, or nonbinary). Owing to the small number of responses in each category, transgender men, transgender women, and nonbinary individuals were combined into *other*, creating a dichotomized variable for gender identity (cisgender or other). Sexual identity was categorized as gay, bisexual, and other.

Structural Factors

The participants self-reported incarceration history (never incarcerated, incarcerated in their lifetime but not incarcerated in the last 12 months, and incarcerated in the last 12 months), employment (full-time or other), health insurance enrollment (currently insured or not insured), and housing stability (residing in stable housing or not residing in stable housing). The participants responded with the highest level of education that they had completed (some high school, graduated high school or obtained General Educational Development, and some college and higher). The participants reported their household income in the previous year from all sources before tax, which we divided into <US $15,000, ≥US $15,000 but <US $40,000, and ≥US $40,000.

HIV-Related Characteristics

The participants indicated how likely they felt they were to contract HIV with responses on a 4-point Likert scale from *very likely* to *very unlikely*. The participants also responded how likely they felt they were to contract HIV compared with their peers in the next 10 years using the same response options. The participants reported their awareness of pre-exposure prophylaxis (PrEP; yes or no) and their use of PrEP (never used, past use, or current use). We merged these 2 variables and created a PrEP continuum variable with 3 response options (unaware or aware but never used, past use, and current use).

Mental Health

The 7-item General Anxiety Disorder scale assessed frequency of anxiety symptoms in the previous 2 weeks; symptomatology comprised (1) minimal, (2) mild, (3) moderate, and (4) severe [47]. Depressive symptoms in the previous week were assessed using a shortened version of the Center for Epidemiological Studies-Depression Scale [48]. The frequency of symptomatology comprised (1) never, (2) hardly ever, (3) some of the time, and (4) most of the time. A cutoff score of ≥16 indicates clinically relevant depressive symptomatology.

Substance Use

We used the Alcohol Use Disorders Identification Test [49], a 10-item screener to measure hazardous drinking in the previous 3 months, which is defined by a total score of ≥8. Items based on the Alcohol, Smoking, and Substance Involvement Screening Test [50] were used to characterize previous–3-month substance consumption among the participants. This resulted in the creation of dichotomous variables (yes or no) representing previous–3-month smoking, use of cannabis, stimulants (cocaine, cocaine powder, crack cocaine, or methamphetamine), sedatives (prescribed Xanax, Klonopin, or Valium), club drugs (ecstasy, ketamine, or gamma-hydroxybutyrate opioids), opioids (prescribed opioids or heroin), hallucinogens (mushrooms, lysergic acid diethylamide, or phencyclidine), and amyl-nitrites (poppers or rush). Given the low frequency of drugs used other than cannabis (and tobacco), we created two composite drug use variables with the remaining 6 individual drugs assessed: (1) other drug use (yes or no) and (2) the number of other drugs used in the previous 3 months (continuous, range 0-6).

Sexual Risk Behaviors

The participants reported the total incidence of receptive or insertive condomless anal intercourse in the previous 3 months. In addition, the participants reported the total incidence of condomless vaginal intercourse in the previous 3 months.

Statistical Analysis

Descriptive analyses were conducted to characterize the sample (eg, means, SDs, and proportions). Bivariate multinomial logistic regression models were fitted to the 2 categorical outcomes measuring lifetime and previous–12-month HIV and STI testing. Among the 4 different HIV and STI testing outcomes (never, STIs only, HIV only, or both), the largest group was selected as the reference group. Each model compared HIV and STI testing categories according to demographic characteristics, structural factors, psychosocial barriers, and substance use and sexual behaviors. Dummy variables were created for gender and sexual identity, and ordinal variables (education, income, incarceration, HIV likely, PrEP awareness and use, and anxiety) were considered continuous variables in the modeling. Models were fit using SAS statistical software (version 9.4; SAS Institute Inc).

Results

Sample Description

Multimedia Appendix 1 summarizes the descriptive information of the participants in detail. Briefly, the mean age of the participants was 22.5 (SD 3.22) years. More than half identified as White (285/414, 68.8%), cisgender male (331/414, 80%), and gay (270/414, 65.2%). Most had secure housing (261/414, 63%) and health insurance (347/414, 83.8%) and had completed at least some college (295/414, 71.3%).

Regarding *lifetime* testing, 62.6% (259/414) had been tested for both HIV and STIs, 7% (29/414) had been tested only for STIs, 10.4% (43/414) had been tested only for HIV, and 20% (83/414) had never been tested for either. In the previous *12 months*, 40.6% (168/414) had been tested for both HIV and STIs, 15.5% (64/414) had been tested only for HIV, 8.5% (35/414) had been tested only for STIs, and 35.5% (147/414) had not been tested for either HIV or STIs. One-fourth of the participants (110/414, 26.6%) reported a lifetime diagnosis of an STI (49/414, 11.8%) in the previous *12 months*. Most participants (355/414, 85.7%) were either unaware of PrEP or had not used it in the past.

Approximately half of the participants reported moderate to severe anxiety symptoms (175/414, 42.3%) and depressive symptoms (262/414, 63.3%). Most participants (284/414, 68.6%) reported experience with cannabis, and nearly half of the participants (178/414, 43%) reported experience with other drugs in the previous 3 months. More than half of the participants (257/414, 62.1%) reported condomless anal intercourse, and 14% (58/414) had had condomless vaginal intercourse in the previous 3 months.

Associations With HIV and STI Testing

Overview

Multimedia Appendices 2 and 3 show sample characteristics across HIV and STI testing in their lifetime and in the previous 12 months, respectively. Multimedia Appendices 4 and 5 show the results of the multinomial regression models demonstrating the relationships between independent variables and HIV and STI testing (lifetime and previous 12 months, respectively).

Demographic Characteristics

Older participants were more likely to have been tested for both HIV and STIs in their lifetime (odds ratio [OR] 1.27, 95% CI 1.16-1.38) and in the previous 12 months (OR 1.12, 95% CI 1.04-1.21). In addition, older participants were less likely to have only been tested for STIs in their lifetime (OR 0.82, 95% CI 0.72-0.94), whereas older participants were more likely to have been tested for HIV in the previous 12 months (OR 1.12, 95% CI 1.02-1.23). Participants who identified as cisgender were 77% less likely to have been tested only for STIs (OR 0.23, 95% CI 0.10-0.50) than to have been tested for both HIV and STIs in their lifetime, whereas cisgender participants were 5 times more likely to have only been tested for HIV (OR 5.16, 95% CI 1.53-17.45) in the previous 12 months. Participants who identified as gay were less likely to have been tested only for STIs in their lifetime (OR 0.29, 95% CI 0.13-0.64) and in the previous 12 months (OR 0.45, 95% CI 0.21-0.94) than those who had been tested for both HIV and STIs. However, participants who identified as bisexual were more likely to have been tested only for STIs in their lifetime (OR 4.24, 95% CI 1.87-9.59) and to have not been tested for either HIV or STIs in the previous 12 months (OR 1.97, 95% CI 1.10-3.52). Participants with higher levels of education were more likely to have been tested for HIV and STIs in their lifetime (OR 2.19, 95% CI 1.48-3.25) and in the previous 12 months (OR 1.76, 95% CI 1.19-2.56) than to have never been tested for either HIV or STIs. Those with higher incomes were less likely to have only been tested for STIs in their lifetime (OR 0.52, 95% CI 0.28-0.96) than to have been tested for both HIV and STIs.

HIV-Related Characteristics

Participants who reported higher awareness and use of PrEP were less likely to have been tested only for HIV in their lifetime (OR 0.25, 95% CI 0.07-0.83) and in the previous 12 months (OR 0.36, 95% CI 0.19-0.70).

Psychosocial Barriers

Participants with greater anxiety symptoms were more likely to have only been tested for STIs in their lifetime (OR 1.42, 95% CI 1.01-1.99), although depressive symptomatology was not associated with HIV and STI testing.

Substance Use Behaviors

Tobacco use and hazardous drinking were not associated with HIV and STI testing in their lifetime and in the previous 12 months; however, sedative and opioid use were associated with HIV and STI testing in their lifetime, and cannabis, opioid, and amyl-nitrite use was associated with previous-year HIV and STI testing. Participants who used sedatives (OR 2.05, 95% CI 1.04-4.03) and opioids (OR 4.38, 95% CI 1.58-12.16) were more likely to not have been tested for either HIV or STIs in their lifetime compared with those who had been tested for both. AYA who used opioids were more likely to have not been tested for HIV or STIs in the previous 12 months (OR 3.64, 95% CI 1.15-11.56), whereas AYA who used amyl-nitrites were less likely to have been tested for either HIV or STIs in the previous 12 months (OR 0.52, 95% CI 0.27-0.97). Participants who used cannabis were less likely to have only been tested for HIV (OR 0.54, 95% CI 0.29-0.98) in the previous 12 months.

Sexual Behaviors

Participants who reported condomless anal intercourse were more likely to have been tested for both HIV and STIs in their lifetime (OR 2.89, 95% CI 1.74-4.79) and in the previous 12 months (OR 2.76, 95% CI 1.73-4.40) compared with those who had not been tested for either HIV or STIs. Specifically, receptive condomless anal intercourse was associated with HIV and STI testing in their lifetime (OR 0.50, 95% CI 0.30-0.82) and in the previous 12 months (OR 0.46, 95% CI 0.29-0.72). Similarly, participants who reported condomless anal intercourse (OR 0.26, 95% CI 0.12-0.58), especially receptive condomless anal intercourse (OR 0.27, 95% CI 0.12-0.64), in the previous 3 months were less likely to have been tested only for STIs in their lifetime. However, participants who reported condomless vaginal intercourse (OR 3.19, 95% CI 1.34-7.62), especially receptive condomless vaginal intercourse (OR 3.30, 95% CI 1.20-9.07), were more likely to have been tested for only an STI compared with testing for both HIV and STIs in their lifetime.

Discussion

Principal Findings

In this sample of substance using sexual and gender minority AYA, approximately two-thirds of the participants reported being tested for both HIV and STIs in their lifetime (259/414, 62.6%); however, 20% (83/414) of the sample had not been tested for either in the previous year, a concerning rate given the Centers for Disease Control and Prevention (CDC) recommendation that all sexually active, young, substance using, sexual and gender minority individuals should be tested for STIs annually and for HIV every 3 to 6 months [51]. Our study findings show that 40.6% (168/414) of the sample followed the recommendation of annual testing for both HIV and STIs, which differs from previous studies—national HIV testing rates for young men who have sex with men varied from 15% to 30% [52,53], 15% of transgender youth aged 15 to 24 years were tested for HIV and 23% were tested for STIs in the previous year [54], and 87% of Black sexual minority men aged 18 to 30 years were tested for HIV in the previous year [55]. Our sample showed greater adherence to testing behaviors than previous studies, although 59.5% (246/414) of the sample did not follow the CDC's recommendations for routine HIV and STI testing. Previous studies have illustrated that young GBMSM report a lack of awareness of the need for HIV testing, fear of results and rejection, physical and economic access issues, testing-related stigma, and unfriendly testing environments [56]. Participants with higher levels of PrEP awareness and use were more likely to engage in HIV and STI testing, likely a function of the need for routine testing for PrEP users. Consistent with previous studies [57,58], participants with lower educational attainment and income were less likely to engage in HIV and STI testing, pointing to a lack of social and economic capital as a significant barrier to engaging in testing. However, each of these barriers to testing is malleable, and interventions that work directly to eliminate barriers to HIV and STI testing and promote knowledge of its benefits should be considered a research and programmatic priority for substance using sexual and gender minority AYA [59-61].

Transgender individuals and bisexual men reported having been tested for STIs but not HIV, whereas cisgender gay men were more likely to have been tested for HIV than for STIs. These patterns of HIV and STI testing may be attributed to differing perceived susceptibility to HIV or STIs across sexual and gender minority communities but also to the targeting that is often used in HIV and STI programs. HIV testing has long focused on the need for GBMSM to be tested regularly, and only recently in the epidemic have promotional materials begun to include the transgender community [62,63]. However, STI programs have focused on the general population, in particular adolescents, and it is possible that bisexual and transgender populations are receiving more messaging about the need for STI testing or resonate more clearly with the material used to promote STI testing, if any. Similarly, those who reported condomless vaginal intercourse were more likely to have been tested only for STIs, and those reporting condomless anal intercourse were less likely to have been tested for HIV and STIs compared with those who had been tested for both HIV and STIs. Just as programs may be targeting testing differentially by demographic characteristics, they may be targeting their programs based on risk, targeting HIV and STI testing more toward GBMSM, who are more likely to engage in condomless anal intercourse. This result indicates the need for HIV and STI programs to think critically about how they promote testing and to ensure the inclusion of all sexual and gender minority AYA and their unique prevention needs in their promotional efforts.

Both self-reported mental health and substance use were associated with HIV and STI testing. Participants with higher levels of anxiety were more likely to have only been tested for STIs, perhaps suggesting that anxiety is a barrier to engaging in HIV testing, although further research is warranted to understand the pathways between anxiety and engagement in testing. Opioid- and sedative using participants were more likely to not have been tested for either HIV or STIs compared with those who had been tested for both HIV and STIs in their lifetime. In general, substance using AYA may have specific barriers to engaging in testing—for example, fear of having to report their substance use, which prevents them from seeking testing services [4], whereas other samples of substance using AYA report barriers such as not feeling at risk or not being offered a test [64]. However, it is important to note that patterns could vary based on misuse versus medical use, which was not distinguished in our sample and is a direction for future research. However, in this sample, inhalant using participants were more likely to have been tested for both HIV and STIs in the previous year. Previous studies have demonstrated associations between amyl-nitrate use and risky sexual behavior, and it is plausible that these participants engage in testing in recognition of their risk-taking behaviors [65], especially when amyl-nitrites are almost exclusively used to facilitate anal intercourse [66]. The results indicate that there is no single relationship between substance use and engagement in HIV and STI testing and that it is important to look critically at the relationships between individual types of substance use (ie, opioids vs cannabis vs amyl-nitrites) when attempting to design programs that promote testing for substance using AYA.

Limitations

There are several limitations to this analysis. First, the cross-sectional nature of the data precludes any inference of causality, but it does provide a foundation for further examinations to draw from. Second, the small number of participants reporting noncisgender identities precluded a thorough examination of associations between gender identity and HIV and STI testing, a common limitation of the literature, which often aggregates gender identities [67-69]. Third, participants may be susceptible to demand characteristics or underreporting of sensitive behaviors, although this limitation is tempered by the use of confidentiality assurances and computerized self-report, which enhances the validity of self-report for substance use (which is reliable and valid) [70] and sexual risk behaviors, and there is a wealth of previous studies that have shown the validity of self-reported measures of substance use [71,72]. Finally, this trial recruited substance using GBMSM; however, there were 23 participants who reported discrepant responses in substance use and drinking between the screener and baseline survey. Of these 23 participants, 15 (65%) reported drinking but not binge drinking, and 8 (35%) reported drug use at some point in their lives but not in the previous 3 months.

Conclusions

The goal of this study was to understand the HIV and STI testing behaviors of substance using sexual and gender minority AYA living in southeastern Michigan. Our results show that, contrary to CDC testing guidelines, 35.5% (147/414) of AYA had not been tested for either HIV or STIs in the previous year, and less than half (168/414, 40.6%) meet the CDC testing recommendations. To meet the CDC's guidelines on testing, HIV and STI testing interventions need to recognize the specific barriers to engaging in testing experienced by substance using sexual and gender minority AYA, who experience multiple layers of potential stigma grounded in their age, sexual and gender identities, and substance use behavior. Central to this is the recognition that not all substances are linked to testing behaviors in the same way, and research and programmatic efforts need to consider the differential testing needs, attitudes, and barriers of different types of substance using AYA.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Demographic and behavioral characteristics of a sample of substance-using sexual and gender minority adolescents and young adults in the Detroit Metro Area (aged 15-29 years; N=414).

[PDF File (Adobe PDF File), 169 KB - publichealth_v8i7e30944_app1.pdf]

Multimedia Appendix 2

Distribution of demographic characteristics, structural factors, psychosocial barriers, and substance use and sexual behaviors by lifetime HIV testing among substance-using sexual and gender minority adolescents and young adults (N=414).

[PDF File (Adobe PDF File), 179 KB - publichealth_v8i7e30944_app2.pdf]

Multimedia Appendix 3

Distribution of demographic characteristics, structural factors, psychosocial barriers, and substance use and sexual behaviors by previous-year HIV and sexually transmitted infection testing among substance-using sexual and gender minority adolescents and young adults (N=414).

[PDF File (Adobe PDF File), 179 KB - publichealth_v8i7e30944_app3.pdf]

Multimedia Appendix 4

Odds of lifetime HIV and sexually transmitted infection testing by demographic characteristics, structural factors, psychosocial barriers, and substance use and sexual behaviors among substance-using sexual and gender minority adolescents and young adults (N=414).

[PDF File (Adobe PDF File), 199 KB - publichealth_v8i7e30944_app4.pdf]

Multimedia Appendix 5

Odds of previous-year HIV and sexually transmitted infection testing by demographic characteristics, structural factors, psychosocial barriers, and substance use and sexual behaviors among substance-using sexual and gender minority adolescents and young adults (N=414).

[PDF File (Adobe PDF File), 199 KB - publichealth_v8i7e30944_app5.pdf]

References

1. HIV and transgender communities. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/hiv/pdf/policies/data/cdc-hiv-policy-issue-brief-transgender.pdf [accessed 2020-07-30]
2. HIV News Media Resources. Centers for Disease Control and Prevention. 2017 Jul 27. URL: https://www.cdc.gov/nchhstp/newsroom/hiv-resources.html [accessed 2021-04-02]
3. Basic Statistics. Division of HIV Prevention, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/hiv/basics/statistics.html [accessed 2020-07-30]
4. Lauby J, Zhu L, Milnamow M, Batson H, Bond L, Curran-Groome W, et al. Get real: evaluation of a community-level HIV prevention intervention for young MSM who engage in episodic substance use. AIDS Educ Prev 2017 Jun;29(3):191-204. [doi: 10.1521/aeap.2017.29.3.191] [Medline: 28650230]
5. U.S. Statistics. HIV.gov. 2021 Mar 17. URL: https://www.hiv.gov/hiv-basics/overview/data-and-trends/statistics [accessed 2021-04-02]
6. Becasen JS, Denard CL, Mullins MM, Higa DH, Sipe TA. Estimating the prevalence of HIV and sexual behaviors among the US transgender population: a systematic review and meta-analysis, 2006-2017. Am J Public Health 2019 Jan;109(1):e1-e8 [FREE Full text] [doi: 10.2105/AJPH.2018.304727] [Medline: 30496000]
7. Johnson AS, Hall HI, Hu X, Lansky A, Holtgrave DR, Mermin J. Trends in diagnoses of HIV infection in the United States, 2002-2011. JAMA 2014 Jul;312(4):432-434 [FREE Full text] [doi: 10.1001/jama.2014.8534] [Medline: 25038362]
8. Frieden TR, Foti KE, Mermin J. Applying public health principles to the HIV epidemic--how are we doing? N Engl J Med 2015 Dec 03;373(23):2281-2287. [doi: 10.1056/NEJMms1513641] [Medline: 26624243]
9. Brewer TH, Schillinger J, Lewis FM, Blank S, Pathela P, Jordahl L, et al. Infectious syphilis among adolescent and young adult men: implications for human immunodeficiency virus transmission and public health interventions. Sex Transm Dis 2011 May;38(5):367-371. [doi: 10.1097/OLQ.0b013e3181ffa7b0] [Medline: 21150816]
10. Su JR, Beltrami JF, Zaidi AA, Weinstock HS. Primary and secondary syphilis among black and Hispanic men who have sex with men: case report data from 27 States. Ann Intern Med 2011 Aug 02;155(3):145-151. [doi: 10.7326/0003-4819-155-3-201108020-00004] [Medline: 21810707]
11. Sexually transmitted disease surveillance 2018. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/std/stats18/STDSurveillance2018-full-report.pdf [accessed 2021-07-30]
12. Van Gerwen OT, Jani A, Long DM, Austin EL, Musgrove K, Muzny CA. Prevalence of sexually transmitted infections and human immunodeficiency virus in transgender persons: a systematic review. Transgend Health 2020 Jun 8;5(2):90-103 [FREE Full text] [doi: 10.1089/trgh.2019.0053] [Medline: 32656353]
13. Volkow ND, Montaner J. Enhanced HIV testing, treatment, and support for HIV-infected substance users. JAMA 2010 Apr 14;303(14):1423-1424 [FREE Full text] [doi: 10.1001/jama.2010.421] [Medline: 20388900]
14. Pantin M, Leonard NR, Hagan H. Sexual HIV/HSV-2 risk among drug users in New York City: an HIV testing and counseling intervention. Subst Use Misuse 2013 Apr;48(6):438-445 [FREE Full text] [doi: 10.3109/10826084.2013.778279] [Medline: 23528143]
15. Gamarel KE, Brown L, Kahler CW, Fernandez MI, Bruce D, Nichols S, Adolescent Medicine Trials Network for HIV/AIDS Intervention. Prevalence and correlates of substance use among youth living with HIV in clinical settings. Drug Alcohol Depend 2016 Dec 01;169:11-18 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.10.002] [Medline: 27750182]
16. Jadwin-Cakmak L, Reisner SL, Hughto JM, Salomon L, Martinez M, Popoff E, et al. HIV prevention and HIV care among transgender and gender diverse youth: design and implementation of a multisite mixed-methods study protocol in the U.S. BMC Public Health 2019 Dec 15;19(1):1531 [FREE Full text] [doi: 10.1186/s12889-019-7605-4] [Medline: 31730450]
17. Kelley CF, Kraft CS, de Man TJ, Duphare C, Lee HW, Yang J, et al. The rectal mucosa and condomless receptive anal intercourse in HIV-negative MSM: implications for HIV transmission and prevention. Mucosal Immunol 2017 Jul;10(4):996-1007 [FREE Full text] [doi: 10.1038/mi.2016.97] [Medline: 27848950]
18. Figueroa JP, Cooper CJ, Edwards JK, Byfield L, Eastman S, Hobbs MM, et al. Understanding the high prevalence of HIV and other sexually transmitted infections among socio-economically vulnerable men who have sex with men in Jamaica. PLoS One 2015 Feb 6;10(2):e0117686 [FREE Full text] [doi: 10.1371/journal.pone.0117686] [Medline: 25659122]
19. Garofalo R, Hotton AL, Kuhns LM, Gratzer B, Mustanski B. Incidence of HIV infection and sexually transmitted infections and related risk factors among very young men who have sex with men. J Acquir Immune Defic Syndr 2016 May 01;72(1):79-86 [FREE Full text] [doi: 10.1097/QAI.0000000000000933] [Medline: 26745827]
20. Hoenigl M, Chaillon A, Morris SR, Little SJ. HIV infection rates and risk behavior among young men undergoing community-based testing in San Diego. Sci Rep 2016 May 16;6:25927 [FREE Full text] [doi: 10.1038/srep25927] [Medline: 27181715]
21. Stephenson R, White D, Darbes L, Hoff C, Sullivan P. HIV testing behaviors and perceptions of risk of HIV infection among MSM with main partners. AIDS Behav 2015 Mar;19(3):553-560 [FREE Full text] [doi: 10.1007/s10461-014-0862-4] [Medline: 25081599]

22. Fisher CB, Fried AL, Desmond M, Macapagal K, Mustanski B. Perceived barriers to HIV prevention services for transgender youth. LGBT Health 2018;5(6):350-358 [FREE Full text] [doi: 10.1089/lgbt.2017.0098] [Medline: 30070960]
23. Reisner SL, Jadwin-Cakmak L, White Hughto JM, Martinez M, Salomon L, Harper GW. Characterizing the HIV prevention and care continua in a sample of transgender youth in the U.S. AIDS Behav 2017 Dec;21(12):3312-3327 [FREE Full text] [doi: 10.1007/s10461-017-1938-8] [Medline: 29138982]
24. Shoptaw S, Montgomery B, Williams CT, El-Bassel N, Aramrattana A, Metsch L, et al. Not just the needle: the state of HIV-prevention science among substance users and future directions. J Acquir Immune Defic Syndr 2013 Jul;63 Suppl 2(0 2):S174-S178 [FREE Full text] [doi: 10.1097/QAI.0b013e3182987028] [Medline: 23764632]
25. Mitchell CG, Oltean A. Integrating HIV prevention into substance user treatment: current practices and challenges. Subst Use Misuse 2007;42(14):2173-2182. [doi: 10.1080/10826080701663190] [Medline: 18097998]
26. Elkbuli A, Polcz V, Dowd B, McKenney M, Prado G. HIV prevention intervention for substance users: a review of the literature. Subst Abuse Treat Prev Policy 2019 Jan 03;14(1):1 [FREE Full text] [doi: 10.1186/s13011-018-0189-7] [Medline: 30606266]
27. DiNenno EA, Prejean J, Irwin K, Delaney KP, Bowles K, Martin T, et al. Recommendations for HIV screening of gay, bisexual, and other men who have sex with men - United States, 2017. MMWR Morb Mortal Wkly Rep 2017 Aug 11;66(31):830-832 [FREE Full text] [doi: 10.15585/mmwr.mm6631a3] [Medline: 28796758]
28. Sanchez TH, Zlotorzynska M, Sineath RC, Kahle E, Tregear S, Sullivan PS. National trends in sexual behavior, substance use and HIV testing among United States men who have sex with men recruited online, 2013 through 2017. AIDS Behav 2018 Aug;22(8):2413-2425. [doi: 10.1007/s10461-018-2168-4] [Medline: 29948340]
29. Eklund AC, Dillon FR, Ebersole RC. Individual and structural factors predicting HIV testing among Latinx MSM: substance use as a moderator. AIDS Care 2020 Feb;32(2):193-201 [FREE Full text] [doi: 10.1080/09540121.2019.1619664] [Medline: 31104481]
30. Starks TJ, Dellucci TV, Gupta S, Robles G, Stephenson R, Sullivan PS, et al. A pilot randomized trial of intervention components addressing drug use in Couples HIV Testing and Counseling (CHTC) with male couples. AIDS Behav 2019 Sep;23(9):2407-2420 [FREE Full text] [doi: 10.1007/s10461-019-02455-2] [Medline: 30887190]
31. Morgan E, Skaathun B, Lancki N, Jimenez AD, Ramirez-Valles J, Bhatia R, et al. Trends in HIV risk, testing, and treatment among MSM in Chicago 2004-2014: implications for HIV elimination planning. J Urban Health 2017 Oct;94(5):699-709 [FREE Full text] [doi: 10.1007/s11524-017-0175-9] [Medline: 28631059]
32. Nerlander LM, Hoots BE, Bradley H, Broz D, Thorson A, Paz-Bailey G, NHBS Group. HIV infection among MSM who inject methamphetamine in 8 US cities. Drug Alcohol Depend 2018 Sep 1;190:216-223. [doi: 10.1016/j.drugalcdep.2018.06.017] [Medline: 30055426]
33. Pitasi MA, Oraka E, Clark H, Town M, DiNenno EA. HIV testing among transgender women and men - 27 states and Guam, 2014-2015. MMWR Morb Mortal Wkly Rep 2017 Aug 25;66(33):883-887 [FREE Full text] [doi: 10.15585/mmwr.mm6633a3] [Medline: 28837547]
34. Sevelius JM, Poteat T, Luhur WE, Reisner SL, Meyer IH. HIV testing and PrEP use in a national probability sample of sexually active transgender people in the United States. J Acquir Immune Defic Syndr 2020 Aug 15;84(5):437-442 [FREE Full text] [doi: 10.1097/QAI.0000000000002403] [Medline: 32692101]
35. Dillon FR, Eklund A, Ebersole R, Ertl MM, Martin JL, Verile MG, et al. Heterosexual self-presentation and other individual- and community-based correlates of HIV testing among Latino men who have sex with men. Psychol Men Masc 2019 Apr;20(2):238-251 [FREE Full text] [doi: 10.1037/men0000166] [Medline: 31592191]
36. Clark HA, Oraka E, DiNenno EA, Wesolowski LG, Chavez PR, Pitasi MA, et al. Men who have sex with men (MSM) who have not previously tested for HIV: results from the MSM testing initiative, United States (2012-2015). AIDS Behav 2019 Mar;23(2):359-365 [FREE Full text] [doi: 10.1007/s10461-018-2266-3] [Medline: 30173345]
37. Golub SA, Gamarel KE. The impact of anticipated HIV stigma on delays in HIV testing behaviors: findings from a community-based sample of men who have sex with men and transgender women in New York City. AIDS Patient Care STDS 2013 Nov;27(11):621-627 [FREE Full text] [doi: 10.1089/apc.2013.0245] [Medline: 24138486]
38. Jürgensen M, Tuba M, Fylkesnes K, Blystad A. The burden of knowing: balancing benefits and barriers in HIV testing decisions. a qualitative study from Zambia. BMC Health Serv Res 2012 Jan 05;12:2 [FREE Full text] [doi: 10.1186/1472-6963-12-2] [Medline: 22222028]
39. Prestage G, Brown G, Keen P. Barriers to HIV testing among Australian gay men. Sex Health 2012 Dec;9(5):453-458. [doi: 10.1071/SH12033] [Medline: 23380195]
40. Schwarcz S, Richards TA, Frank H, Wenzel C, Hsu LC, Chin CS, et al. Identifying barriers to HIV testing: personal and contextual factors associated with late HIV testing. AIDS Care 2011 Jul;23(7):892-900. [doi: 10.1080/09540121.2010.534436] [Medline: 21424942]
41. Goldenberg T, Jadwin-Cakmak L, Popoff E, Reisner SL, Campbell BA, Harper GW. Stigma, gender affirmation, and primary healthcare use among black transgender youth. J Adolesc Health 2019 Oct;65(4):483-490 [FREE Full text] [doi: 10.1016/j.jadohealth.2019.04.029] [Medline: 31303554]

42. Logie CH, Lacombe-Duncan A, Brien N, Jones N, Lee-Foon N, Levermore K, et al. Barriers and facilitators to HIV testing among young men who have sex with men and transgender women in Kingston, Jamaica: a qualitative study. J Int AIDS Soc 2017 Apr 04;20(1):21385 [FREE Full text] [doi: 10.7448/IAS.20.1.21385] [Medline: 28406274]
43. Malebranche DJ, Peterson JL, Fullilove RE, Stackhouse RW. Race and sexual identity: perceptions about medical culture and healthcare among Black men who have sex with men. J Natl Med Assoc 2004 Jan;96(1):97-107. [Medline: 14746359]
44. Gupta GR, Parkhurst JO, Ogden JA, Aggleton P, Mahal A. Structural approaches to HIV prevention. Lancet 2008 Aug 30;372(9640):764-775. [doi: 10.1016/S0140-6736(08)60887-9] [Medline: 18687460]
45. Detroit metro area HIV surveillance report new diagnoses and prevalence tables. Michigan Department of Health and Human Services. 2019 Jul 1. URL: https://www.michigan.gov/documents/mdhhs/Detroit_Metro_Area_HIV_Surveillance_Report_July_2019_660528_7.pdf [accessed 2021-07-30]
46. Stephenson R, Bonar EE, Carrico A, Hunter A, Connochie D, Himmelstein R, et al. Intervention to increase HIV testing among substance-using young men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Apr 30;7(4):e114 [FREE Full text] [doi: 10.2196/resprot.9414] [Medline: 29712625]
47. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
48. Kohout FJ, Berkman LF, Evans DA, Cornoni-Huntley J. Two shorter forms of the CES-D (Center for Epidemiological Studies Depression) depression symptoms index. J Aging Health 1993 May;5(2):179-193. [doi: 10.1177/089826439300500202] [Medline: 10125443]
49. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jul;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
50. WHO ASSIST Working Group. The alcohol, smoking and substance involvement screening test (ASSIST): development, reliability and feasibility. Addiction 2002 Sep;97(9):1183-1194. [doi: 10.1046/j.1360-0443.2002.00185.x] [Medline: 12199834]
51. Recommendations for HIV prevention with adults and adolescents with HIV in the United States, 2014. Centers for Disease Control and Prevention, Health Resources and Services Administration, National Institutes of Health, American Academy of HIV Medicine, Association of Nurses in AIDS Care, International Association of Providers of AIDS Care, the National Minority AIDS Council, and Urban Coalition for HIV/AIDS Prevention Services. 2014 Dec 11. URL: http://stacks.cdc.gov/view/cdc/26062 [accessed 2021-07-30]
52. Phillips 2nd G, Ybarra ML, Prescott TL, Parsons JT, Mustanski B. Low rates of human immunodeficiency virus testing among adolescent gay, bisexual, and queer men. J Adolesc Health 2015 Oct;57(4):407-412 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.06.014] [Medline: 26318671]
53. Kann L, McManus T, Harris WA, Shanklin SL, Flint KH, Hawkins J, et al. Youth risk behavior surveillance - United States, 2015. MMWR Surveill Summ 2016 Jun 10;65(6):1-174. [doi: 10.15585/mmwr.ss6506a1] [Medline: 27280474]
54. Sharma A, Kahle E, Todd K, Peitzmeier S, Stephenson R. Variations in testing for HIV and other sexually transmitted infections across gender identity among transgender youth. Transgend Health 2019 Feb 20;4(1):46-57 [FREE Full text] [doi: 10.1089/trgh.2018.0047] [Medline: 30805557]
55. Rosengren AL, Davy-Mendez T, Hightow-Weidman LB. Online sex partner seeking and HIV testing frequency among young Black sexual minority men. J HIV AIDS Soc Serv 2020;19(1):42-54 [FREE Full text] [doi: 10.1080/15381501.2019.1692749] [Medline: 32372886]
56. Pharr JR, Lough NL, Ezeanolue EE. Barriers to HIV testing among young men who have sex with men (MSM): experiences from Clark County, Nevada. Glob J Health Sci 2015 Nov 03;8(7):9-17 [FREE Full text] [doi: 10.5539/gjhs.v8n7p9] [Medline: 26925893]
57. Nelson KM, Pantalone DW, Gamarel KE, Carey MP, Simoni JM. Correlates of never testing for HIV among sexually active Internet-recruited gay, bisexual, and other men who have sex with men in the United States. AIDS Patient Care STDS 2018 Jan;32(1):9-15 [FREE Full text] [doi: 10.1089/apc.2017.0244] [Medline: 29232170]
58. Bauermeister JA, Eaton L, Andrzejewski J, Loveluck J, VanHemert W, Pingel ES. Where you live matters: structural correlates of HIV risk behavior among young men who have sex with men in metro Detroit. AIDS Behav 2015 Dec;19(12):2358-2369 [FREE Full text] [doi: 10.1007/s10461-015-1180-1] [Medline: 26334445]
59. Stephenson R, Metheny N, Sharma A, Sullivan S, Riley E. Providing home-based HIV testing and counseling for transgender youth (Project Moxie): protocol for a pilot randomized controlled trial. JMIR Res Protoc 2017 Nov 28;6(11):e237 [FREE Full text] [doi: 10.2196/resprot.8562] [Medline: 29183868]
60. Bauermeister JA, Golinkoff JM, Horvath KJ, Hightow-Weidman LB, Sullivan PS, Stephenson R. A multilevel tailored Web app-based intervention for linking young men who have sex with men to quality care (get connected): protocol for a randomized controlled trial. JMIR Res Protoc 2018 Aug 02;7(8):e10444 [FREE Full text] [doi: 10.2196/10444] [Medline: 30072358]
61. Kuhns LM, Garofalo R, Hidalgo M, Hirshfield S, Pearson C, Bruce J, et al. A randomized controlled efficacy trial of an mHealth HIV prevention intervention for sexual minority young men: MyPEEPS mobile study protocol. BMC Public Health 2020 Jan 15;20(1):65 [FREE Full text] [doi: 10.1186/s12889-020-8180-4] [Medline: 31941475]

62. Fisher CB, Fried AL, Macapagal K, Mustanski B. Patient-provider communication barriers and facilitators to HIV and STI preventive services for adolescent MSM. AIDS Behav 2018 Oct;22(10):3417-3428 [FREE Full text] [doi: 10.1007/s10461-018-2081-x] [Medline: 29546468]
63. Martin-Smith HA, Okpo EA, Bull ER. Exploring psychosocial predictors of STI testing in university students. BMC Public Health 2018 May 29;18(1):664 [FREE Full text] [doi: 10.1186/s12889-018-5587-2] [Medline: 29843658]
64. Bonar EE, Cunningham RM, Chermack ST, Blow FC, Barry KL, Booth BM, et al. Prescription drug misuse and sexual risk behaviors among adolescents and emerging adults. J Stud Alcohol Drugs 2014 Mar;75(2):259-268 [FREE Full text] [doi: 10.15288/jsad.2014.75.259] [Medline: 24650820]
65. Hambrick HR, Park SH, Schneider JA, Mayer KH, Carrico AW, Sherman SE, et al. Poppers and PrEP: use of pre-exposure prophylaxis among men who have sex with men who use inhaled nitrites. AIDS Behav 2018 Nov;22(11):3658-3672 [FREE Full text] [doi: 10.1007/s10461-018-2139-9] [Medline: 29744767]
66. Schwartz C, Fast D, Knight R. Poppers, queer sex and a Canadian crackdown: examining the experiences of alkyl nitrite use among young sexual minority men. Int J Drug Policy 2020 Mar;77:102670. [doi: 10.1016/j.drugpo.2020.102670] [Medline: 32143965]
67. Chandwani S, Abramowitz S, Koenig LJ, Barnes W, D'Angelo L. A multimodal behavioral intervention to impact adherence and risk behavior among perinatally and behaviorally HIV-infected youth: description, delivery, and receptivity of adolescent impact. AIDS Educ Prev 2011 Jun;23(3):222-235. [doi: 10.1521/aeap.2011.23.3.222] [Medline: 21696241]
68. Hoth AB, Shafer C, Dillon DB, Mayer R, Walton G, Ohl ME. Iowa TelePrEP: a public-health-partnered telehealth model for human immunodeficiency virus preexposure prophylaxis delivery in a rural state. Sex Transm Dis 2019 Aug;46(8):507-512. [doi: 10.1097/OLQ.0000000000001017] [Medline: 31295217]
69. Johnson MC, Chung R, Leung SY, Edelstein Z, Yuan Y, Flavin SM. Combating stigma through HIV self-testing: New York state's HIV home test giveaway program for sexual minorities. J Public Health Manag Pract 2022;28(2):174-183 [FREE Full text] [doi: 10.1097/PHH.0000000000001138] [Medline: 32015253]
70. Simons JS, Wills TA, Emery NN, Marks RM. Quantifying alcohol consumption: self-report, transdermal assessment, and prediction of dependence symptoms. Addict Behav 2015 Dec;50:205-212 [FREE Full text] [doi: 10.1016/j.addbeh.2015.06.042] [Medline: 26160523]
71. Williams RJ, Nowatzki N. Validity of adolescent self-report of substance use. Subst Use Misuse 2005;40(3):299-311. [doi: 10.1081/ja-200049327] [Medline: 15776978]
72. Nichols SL, Lowe A, Zhang X, Garvie PA, Thornton S, Goldberger BA, et al. Concordance between self-reported substance use and toxicology among HIV-infected and uninfected at risk youth. Drug Alcohol Depend 2014 Jan 01;134:376-382 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.11.010] [Medline: 24309297]

Abbreviations

AYA: adolescents and young adults
CDC: Centers for Disease Control and Prevention
GBMSM: gay, bisexual, and other men who have sex with men
OR: odds ratio
PrEP: pre-exposure prophylaxis
STI: sexually transmitted infection

Edited by H Bradley; submitted 03.06.21; peer-reviewed by S Nagavally, J Park, F Cagayan; comments to author 21.02.22; revised version received 01.03.22; accepted 10.05.22; published 01.07.22.

Please cite as:
Parker JN, Choi SK, Bauermeister JA, Bonar EE, Carrico AW, Stephenson R
HIV and Sexually Transmitted Infection Testing Among Substance-Using Sexual and Gender Minority Adolescents and Young Adults: Baseline Survey of a Randomized Controlled Trial
JMIR Public Health Surveill 2022;8(7):e30944
URL: https://publichealth.jmir.org/2022/7/e30944
doi:10.2196/30944
PMID:35776441

Original Paper

Strategies to Identify and Reach Young Women Who Sell Sex With HIV Prevention and Care Services: Lessons Learnt From the Implementation of DREAMS Services in Two Cities in Zimbabwe

Sungai T Chabata[1,2], MSc; Rumbidzo Makandwa[1], MSc; Bernadette Hensen[3], MSc, PhD; Phillis Mushati[1], MSc; Tarisai Chiyaka[1], MSc; Sithembile Musemburi[1], MSc; Joanna Busza[4], MSc; Sian Floyd[5], MSc; Isolde Birdthistle[6], MSc, PhD; James R Hargreaves[4], MSc, PhD; Frances M Cowan[1,7], MSc, MD

[1]Centre for Sexual Health and HIV/AIDS Research (CeSHHAR), Harare, Zimbabwe
[2]Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands
[3]Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, United Kingdom
[4]Centre for Evaluation, London School of Hygiene and Tropical Medicine, London, United Kingdom
[5]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom
[6]Department of Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom
[7]Department of International Public Health, Liverpool School of Tropical Medicine, Liverpool, United Kingdom

Corresponding Author:
Sungai T Chabata, MSc
Centre for Sexual Health and HIV/AIDS Research (CeSHHAR)
4 Bath Road
Belgravia
Harare
Zimbabwe
Phone: 263 773577686
Email: sungaichabata@gmail.com

Abstract

Background: Young women who sell sex (YWSS), are underserved by available HIV prevention and care services. The Determined, Resilient, Empowered, AIDS-free, Mentored and Safe (DREAMS) Partnership aimed to reduce the risk of HIV acquisition among vulnerable populations of adolescent girls and young women, including YWSS, in 10 sub-Saharan African countries. We describe 2 methods, respondent-driven sampling (RDS) and peer outreach, used to refer YWSS for DREAMS services in Zimbabwe, and compare the characteristics and engagement of YWSS referred to these services by each method. We hypothesized that RDS would identify YWSS at higher risk of HIV and those who were less engaged with HIV prevention and care services than peer outreach.

Objective: We aimed to compare respondent-driven sampling and peer outreach in recruiting and referring high-risk populations for HIV prevention and care services.

Methods: We used RDS, a sampling method designed to reach a representative sample of the network of key populations, and peer outreach, a programmatic approach to identify, reach, and refer YWSS for DREAMS between April and July 2017, and January 2017 and July 2018, respectively, in 2 cities in Zimbabwe. For RDS, we conducted detailed mapping to understand sex work typology and geography, and then purposively selected 10 “seed” participants in each city to initiate RDS. For peer outreach, we initiated recruitment through 18 trained and age-matched peer educators using youth-tailored community mobilization. We described the characteristics and service engagement of YWSS who accessed DREAMS services by each referral approach and assessed the association of these characteristics with referral approach using the chi-square test. Analysis was performed with and without restricting the period when RDS took place. We estimated the relative incremental costs of recruiting YWSS using each strategy for referral to DREAMS services.

Results: Overall, 5386 and 1204 YWSS were referred for DREAMS services through peer outreach and RDS, respectively. YWSS referred through RDS were more likely to access DREAMS services compared to YWSS referred through peer outreach

(501/1204, 41.6% vs 930/5386, 17.3%; *P*<.001). Regardless of referral approach, YWSS who accessed DREAMS had similar education levels, and a similar proportion tested HIV negative and reported not using a condom at the last sex act. A higher proportion of YWSS accessing DREAMS through RDS were aged 18-19 years (167/501, 33.3% vs 243/930, 26.1%; *P*=.004) and more likely to be aware of their HIV status (395/501, 78.8% vs 396/930, 42.6%; *P*<.001) compared to those accessing DREAMS services through peer outreach. The incremental cost per young woman who sells sex recruited was US $7.46 for peer outreach and US $52.81 for RDS.

Conclusions: Peer outreach and RDS approaches can reach and refer high-risk but different groups of YWSS for HIV services, and using both approaches will likely improve reach.

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-018-5085-6

(JMIR Public Health Surveill 2022;8(7):e32286) doi:10.2196/32286

KEYWORDS

respondent-driven sampling; peer outreach; female sex worker; young women who sell sex; HIV prevention; Zimbabwe; sub-Saharan Africa

Introduction

Achieving the ambitious goal to end the AIDS epidemic by 2030 requires that programs identify and reach populations at high risk of HIV. Adolescent girls and young women with high-risk sexual lifestyles, such as young women who sell sex, remain a priority [1-3]. In many countries, including Zimbabwe, selling sex is criminalized, making young women who sell sex particularly hidden and hard to reach with HIV prevention and care services.

Strategies to identify, reach, and engage young women who sell sex are critical. The approaches most commonly used are venue-based approaches, including community-based outreach [4], and network-based approaches, including peer outreach and enhanced peer outreach [5]. Community-based outreach involves the use of peer educators and outreach workers to access and engage with hidden populations in communities where they congregate. Peer outreach is based on the reasoning that by engaging with peers who match the desired population, one may reach unidentified, harder-to-reach, high-risk populations [6]. Enhanced peer outreach is an incentivized version of peer outreach similar in design to respondent-driven sampling (RDS), [7] where outreach workers give coupons to peer mobilizers to recruit other peers within their network. Peer mobilizers are selected from the hidden population based on factors, such as their network size, communication skills, risk behaviors, age, location, and knowledge of peers who engage in high-risk behaviors or have never accessed HIV services [6]. The difference between enhanced peer outreach and RDS is that enhanced peer outreach is a programmatic approach where the number of coupons given to each mobilizer is not restricted, while RDS is a sampling method designed to recruit a representative sample of the network of the population in question, and thus, the number of coupons given to each recruiter is limited to reduce the likelihood of overrepresentation of participants with more extensive networks [7].

Between 2016 and 2019, the Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe (DREAMS) Partnership aimed to reduce the risk of HIV acquisition among the most vulnerable adolescent girls and young women, including young women who sell sex, in 10 sub-Saharan African countries by providing a package of biomedical, social, and economic interventions [8,9]. To ensure that young women who sell sex were reached during implementation of the DREAMS Partnership in Zimbabwe, we used peer outreach, a programmatic strategy to identify and refer young women who sell sex for DREAMS services. Separately, we used RDS, a research sampling strategy, to recruit young women who sell sex to a research cohort to determine the impact of DREAMS on HIV incidence, with cohort participants offered onward referral to DREAMS services [3,10]. DREAMS was implemented and the research cohorts were recruited in 2 Zimbabwean cities. Here, we compared the characteristics and engagement of young women who sell sex, who were referred to DREAMS services for young women who sell sex [11], using each method. We assessed which strategy reached more young women who sell sex, hypothesizing that RDS would identify young women who sell sex at higher risk of HIV and who were less engaged with HIV prevention and care services than peer outreach.

Methods

The DREAMS Partnership

The DREAMS partnership offered a comprehensive package of evidence-based social and clinical interventions to reduce the rate of new HIV infections and other outcomes among adolescent girls and young women. In Zimbabwe, DREAMS was implemented in 6 districts in Bulawayo, Chipinge, Gweru, Makoni, Mazowe, and Mutare [12]. Six implementing partners delivered the DREAMS package, which included social protection services, gender-based violence prevention and care services, and HIV prevention services, including an offer of pre-exposure prophylaxis to young women who sell sex aged 18 to 24 years. The overall coordination of DREAMS implementing partners differed by region. To increase economic well-being, young women who sell sex were offered economic empowerment programs, including vocational skills training, financial literacy training, savings and lending support, and services to support access to education, including cash transfers and educational subsidies.

The Sisters With a Voice Program

DREAMS services for young women who sell sex were accessible through the Sisters with a Voice program (Sisters), in addition to other implementing partners. As described elsewhere [11,13], Sisters is a national sex worker program that was established in Zimbabwe in 2009 on behalf of the Ministry of Health and Child Care, and the National AIDS Council. Briefly, Sisters provides free condoms and contraception, HIV testing and counseling, syndromic management of sexually transmitted infections, health education, and legal advice supported by a network of peer educators.

Study Location and Participants

This study was conducted in 2 districts/cities (Bulawayo and Mutare) that were selected as part of DREAMS impact evaluation from the 6 districts where DREAMS was being implemented [9]. A cohort of young women who sell sex was recruited using RDS and followed-up for 24 months to determine the impact of DREAMS on HIV incidence among young women who sell sex in these 2 cities compared to the incidence among young women who sell sex in 4 towns where DREAMS was not implemented and young women who sell sex only had access to Sisters services [9,10]. Young women who sell sex were defined as adolescent girls and young women aged 18 to 24 who reported exchanging sex with someone because they gave them money, a gift, or material support (important was that sex would not have happened without that exchange). Young women who sell sex included those who self-identified as female sex workers or not. Informal discussions about sex work environment at each location were conducted to ensure that only young women who sell sex were recruited in the study.

Recruiting Young Women Who Sell Sex for DREAMS Services

Peer Outreach

We identified and enrolled young women who sell sex to DREAMS services through a network of 18 (9 in each city) trained and age-matched peer educators supported and directly supervised by the Sisters program outreach team. Peer educators received a monthly incentive of US $15. As described in detail elsewhere [1], peer educators are trained to build rapport with other young women who sell sex and mobilize them for HIV prevention and care services, using a range of community mobilization activities and materials that were specifically tailored for reaching young women who sell sex aged <25 years. Community mobilization materials included 21 activity packs, organized into 6 modules relating to social asset building, HIV prevention, condom promotion and provision, gender norms, basic financial literacy, and sexual violence prevention [1]. These materials were designed in consultation with young women who sell sex and experts, and were piloted and evaluated among a sample of young women who sell sex using in-depth interviews and focus group discussions [1].

Peer educators were recruited in 2016 and worked in the DREAMS program until 2018. Peer educators included 10 who self-identified as female sex workers and 8 who did not. All were 18 to 20 years old and had some secondary school education or higher. Their duties involved identifying hotspots and adolescent girls and young women through word of mouth. Hotspots included secondary schools, colleges, streets, parties, bars, and homes. Peer educators would meet with young women, talk to them about the nature of DREAMS services and where they could access the services, and, if they were interested, give them a referral slip to be produced the first time a young woman who sells sex accessed DREAMS services. Peer outreach was conducted between January 2017 and July 2018.

Respondent-Driven Sampling

For the evaluation cohort, we recruited young women who sell sex, using RDS. First, we conducted detailed social mapping as described elsewhere [14]. A team of field workers was trained to engage with young women who sell sex at different locations whether they self-identified as female sex workers or not [14]. They visited hotspots where young women who sell sex were likely to solicit clients as informed by discussions with local peer educators, selecting the 3 busiest days of the week, Thursday, Friday, and Saturday, over 2 consecutive weekends [14]. During site visits, the research team used direct observation, group discussions, and informal interviews.

Mapping helped to understand sex work typology and geography, and was used to purposively select 10 “seed” participants to initiate RDS recruitment at each site. Seeds represented each identified sex work typology, geography, and target age group. Of the 20 seeds selected across the 2 cities, 80% (16/20) were 20 to 24 years old, 95% (19/20) had some secondary school education or higher, 70% (14/20) were single or never married, and 75% (15/20) self-identified as female sex workers. Young women who sell sex were eligible to participate in RDS surveys if they were aged 18 to 24 years. Each “seed” provided written informed consent to participate, and was interviewed and given 2 recruitment coupons to pass on to young women who sell sex in her social network [9]. Young women who sell sex, who received a coupon attended the survey site, provided written informed consent to participate, and, on completion of survey activities, were given 2 coupons to pass on to 2 further young women who sell sex they knew, who sold sex in that location and who had not previously been recruited to the survey. The recruitment process happened over 6 waves until the desired sample size of about 600 young women who sell sex was achieved in each city [9]. All young women who sell sex were referred for DREAMS services. Importantly, recruitment through RDS (but not peer outreach) was incentivized; young women who sell sex were given an incentive of US $3 for participating in the survey themselves, and an additional US $2 for each peer recruited [9]. RDS was conducted for 4 months, from April to July 2017.

Key similarities between the approaches were that both were network-based and started with a purposive sample of peers for peer outreach or seeds for RDS. However, for peer outreach, peer educators received a week of training, were supported by a network of outreach workers, and had a host of materials to support mobilization activities. Their relationship with the peers they recruited was intended to be ongoing. Recruitment through RDS was limited to 2 recruits per recruiter, and participation was incentivized [9]. Seeds received a short script of what to say when passing coupons on but not formal training and were not expected to continue their engagement with the young women who sell sex.

Unique Identifiers

Both DREAMS program beneficiaries and survey participants were assigned an alphanumeric identifier that comprised the first letter of the first name, the last 3 letters of the surname, and the date of birth to link young women who sell sex across DREAMS services. We deidentified peer outreach and Sisters point of access data, and used the alphanumeric identifier to establish if young women who sell sex had engaged with DREAMS services through the Sisters program.

Statistical Analyses

Participants included in these analyses were young women who sell sex aged 18 to 24 years, who were reached in the 2 cities where both approaches were conducted. Of note, there were no age restrictions for young women who sell sex referred for DREAMS services through peer outreach, but analyses were restricted to young women who sell sex, who were aged 18 to 24 years to match RDS data. We described the sociodemographic characteristics of young women who sell sex recruited into the RDS survey by city and compared the characteristics across the 2 cities. RDS data were RDS-II weighted, with women's responses weighted by the inverse of the reported number of young women who sell sex that they knew, that is, the number of other young women who sell sex that they could have recruited to the survey. Description of sociodemographic characteristics of peer outreach data was not possible owing to limited data routinely captured at the time of program referral.

Among young women who sell sex, who accessed DREAMS services, we described the demographic characteristics, sexual behaviors, and previous service engagement by referral approach and assessed if these differed by referral approach, using the chi-square test. Analysis was performed regardless of the referral period. For peer outreach, we restricted the comparison to the period when RDS took place. Young women who sell sex referred through both approaches were retained in the analysis since excluding them did not make any difference to the results.

Subsequently, we used logistic regression to identify sociodemographic characteristics and sexual behaviors associated with ever access of DREAMS services among young women who sell sex, who were recruited through RDS. For logistic regression analyses, seed participants were dropped. Factors associated with access of DREAMS services at $P\leq.10$ in univariable analysis were included in the multivariable regression model, adjusting for all factors associated with access of DREAMS services in the univariable analysis. Again, this was not possible for young women who sell sex, who accessed services through peer outreach owing to limited point of referral data.

Finally, we estimated the relative incremental costs of recruiting young women who sell sex using each strategy for referral to DREAMS services.

Ethics Approval

The DREAMS impact evaluation was reviewed and approved by the Medical Research Council of Zimbabwe (MRCZ/A/2085) and the London School of Hygiene and Tropical Medicine (11835). All participants were given information about the study and were asked for written informed consent for participation.

Results

Characteristics of Young Women Who Sell Sex Recruited Through RDS

Between April and July 2017, 1204 young women who sell sex were recruited to RDS surveys and referred for DREAMS services in the 2 cities. Among these young women who sell sex, the majority were aged 20 to 24 years (799/1204, 64.6%), had some or complete secondary school education (1079/1204, 88.2%), were never married (801/1204, 67.9%), and self-identified as female sex workers (801/1191, 65.0%) (Table 1). Young women who sell sex in Bulawayo were less likely to be divorced or separated (75/601, 12.2% vs 290/603, 44.8%), were less likely to self-identify as female sex workers (367/590, 60.1% vs 434/601, 69.7%), and were more likely to report having more than six alcoholic drinks in 1 night during the last 12 months (277/601, 42.5% vs 138/602, 19.9%), compared to those in Mutare. Additionally, young women who sell sex from Bulawayo were less likely to report condom use at the last sex act with a regular partner (275/476, 57.8% vs 312/431, 74.3%) or client (419/490, 85.0% vs 422/456, 92.0%) compared to young women who sell sex in Mutare. A higher proportion of young women who sell sex from Bulawayo reported condom-less sex with a regular partner (246/477, 54.0% vs 164/431, 33.8%) or client (98/491, 20.4% vs 65/458, 14.1%) in the past month compared to those from Mutare. Overall, almost a quarter (471/1204, 37.8%) of young women who sell sex reported being at risk of common mental disorders within the last week.

Table 1. Sociodemographic and sexual behavioral characteristics of young women who sell sex recruited to respondent-driven sampling (RDS) surveys in 2 Zimbabwean cities (RDS-II weighted; N=1204).

Characteristic	Bulawayo (N=601), n/N (%)	Mutare (N=603), n/N (%)	Total (N=1204), n/N (%)	*P* value
Age at recruitment (years)				.04
18-19	222/601 (38.8)	183/603 (32.0)	405/1204 (35.4)	
20-24	379/601 (61.2)	420/603 (68.0)	799/1204 (64.6)	
Highest level of education				.36
None/incomplete primary	12/601 (2.6)	29/603 (4.6)	41/1204 (3.6)	
Complete primary	49/601 (8.9)	35/603 (7.4)	84/1204 (8.1)	
Incomplete secondary	245/601 (40.4)	277/603 (43.2)	522/1204 (41.8)	
Complete secondary/higher	295/601 (48.1)	262/603 (44.8)	557/1204 (46.4)	
Marital status				<.001
Single/never married	501/601 (83.2)	300/603 (52.8)	801/1204 (67.9)	
Married/living together as if married	22/601 (4.3)	9/603 (1.8)	31/1204 (3.1)	
Divorced/separated	75/601 (12.2)	290/603 (44.8)	365/1204 (28.5)	
Widowed	3/601 (0.2)	4/603 (0.7)	7/1204 (0.5)	
Self-identification as a sex worker				.004
No	223/590 (39.9)	167/601 (30.3)	390/1191 (35.0)	
Yes	367/590 (60.1)	434/601 (69.7)	801/1191 (65.0)	
Age at start of selling sex (years)				.01
≤15	86/600 (13.2)	60/603 (9.2)	146/1203 (11.2)	
16-17	222/600 (37.2)	190/603 (31.6)	412/1203 (34.4)	
18-24	292/600 (49.6)	353/603 (59.1)	645/1203 (54.4)	
Years selling sex				.02
0-2	291/600 (52.2)	323/603 (55.9)	614/1203 (54.1)	
3-4	204/600 (32.7)	150/603 (24.8)	354/1203 (28.7)	
≥5	105/600 (15.1)	130/603 (19.3)	235/1203 (17.2)	
Number of sexual partners in the past month				.07
0-4	272/601 (50.4)	265/603 (47.8)	537/1204 (49.1)	
5-9	159/601 (24.7)	127/603 (20.6)	286/1204 (22.6)	
≥10	170/601 (24.9)	211/603 (31.6)	381/1204 (28.3)	
Number of clients in the past month				.08
0-4	313/601 (57.5)	296/603 (53.6)	609/1204 (55.5)	
5-9	131/601 (20.0)	110/603 (17.4)	241/1204 (18.7)	
≥10	157/601 (22.5)	197/603 (29.0)	354/1204 (25.8)	
Condom use at the last sex act with a regular partner				<.001
No	201/476 (42.2)	119/431 (25.7)	320/907 (34.3)	
Yes	275/476 (57.8)	312/431 (74.3)	587/907 (65.7)	
Condom-less sex with a regular partner in the past month				<.001
No	231/477 (46.0)	267/431 (66.2)	498/908 (55.7)	
Yes	246/477 (54.0)	164/431 (33.8)	410/908 (44.3)	
Condom use at the last sex act with a client				.006

Characteristic	Bulawayo (N=601), n/N (%)	Mutare (N=603), n/N (%)	Total (N=1204), n/N (%)	*P* value
No	71/490 (15.0)	34/456 (8.0)	105/946 (11.6)	
Yes	419/490 (85.0)	422/456 (92.0)	841/946 (88.4)	
Condom-less sex with a client in the past month				.03
No	393/491 (79.6)	393/458 (85.9)	786/949 (82.6)	
Yes	98/491 (20.4)	65/458 (14.1)	163/949 (17.4)	
STI[a] symptoms in the last 12 months				.17
No	470/601 (78.7)	449/603 (74.8)	919/1204 (76.7)	
Yes	131/601 (21.3)	154/603 (25.2)	285/1204 (23.3)	
Accessed DREAMS[b] services				<.001
No	277/601 (46.2)	426/603 (69.6)	703/1204 (57.9)	
Yes	324/601 (53.8)	177/603 (30.4)	501/1204 (42.1)	
Relationship with other young women who sell sex				.10
Good	450/600 (72.6)	464/601 (76.0)	914/1201 (74.3)	
Neither good nor bad	102/600 (19.3)	81/601 (14.0)	183/1201 (16.7)	
No relation	48/600 (8.1)	56/601 (10.0)	104/1201 (9.0)	
Number of close friends				.02
0	85/601 (16.2)	133/603 (22.8)	218/1204 (19.5)	
≥1	516/601 (83.8)	470/603 (77.2)	986/1204 (80.5)	
Binge drinking[c]				<.001
No alcohol in the last 12 months	153/601 (29.4)	248/602 (44.8)	401/1203 (37.1)	
Drank alcohol but no occasions of binge drinking	171/601 (28.1)	216/602 (35.3)	387/1203 (31.7)	
Yes, at least one occasion of binge drinking	277/601 (42.5)	138/602 (19.9)	415/1203 (31.2)	
Risk of CMD[d]				.73
No	366/601 (61.6)	367/603 (62.7)	733/1204 (62.2)	
Yes	235/601 (38.4)	236/603 (37.3)	471/1204 (37.8)	
Experienced any form of violence from a sexual partner				<.001
No	351/601 (60.8)	280/603 (49.2)	631/1204 (55.0)	
Yes	250/601 (39.2)	323/603 (50.8)	573/1204 (45.0)	
Experienced any form of violence from police				.26
No	576/601 (96.5)	568/602 (95.1)	1144/1203 (95.8)	
Yes	25/601 (3.5)	34/602 (4.9)	59/1203 (4.2)	

[a] STI: sexually transmitted infection.

[b] DREAMS: Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe.

[c] Had more than six alcoholic drinks in 1 night during the last 12 months.

[d] CMD: common mental disorder.

Characteristics of Young Women Who Sell Sex, Who Accessed DREAMS Services by Referral Approach

Between January 2017 and July 2018, 5386 young women who sell sex were referred to DREAMS services via peer outreach. Between April and July 2017, 1204 young women who sell sex were referred to DREAMS services via RDS. Young women who sell sex referred through RDS were over twice as likely to access DREAMS services through the Sisters program (501/1204, 41.6%) compared to young women who sell sex referred through peer outreach (930/5386, 17.3%; *P*<.001). Additionally, 45 young women who sell sex, 5% (45/930) of those who accessed DREAMS via peer outreach and 9% (45/501) of those who accessed DREAMS via RDS, were referred through both strategies. Services accessed included HIV testing, free condoms and contraception, and other services such as sexually transmitted infection treatment.

Between April and July 2017, when the RDS was ongoing, 1228 young women who sell sex were referred through peer outreach. When restricting our comparison to the period when RDS was ongoing, young women who sell sex referred through RDS remained more likely to access DREAMS services (501/1204, 41.6%) compared to young women who sell sex referred through peer outreach (212/1228, 17.3%; *P*<.001). Within this period, 3.0% (15/501) of those who accessed DREAMS via RDS and 7.1% (15/212) of those who accessed DREAMS via peer outreach were referred through both strategies.

A higher proportion of young women who sell sex accessing DREAMS services through RDS were younger (18-19 years old: 167/501, 33.3% vs 243/930, 26.1%; *P*=.004) and reported having ever been tested for HIV (441/501, 88.0% vs 661/827, 79.9%; *P*<.001) compared to young women who sell sex accessing DREAMS services through peer outreach (Table 2). Additionally, young women who sell sex accessing DREAMS services through RDS were more likely to self-report an HIV-positive status (39/439, 8.9% vs 0/661, 0.0%; *P*<.001), more likely to be aware of their HIV status (395/501, 78.8% vs 396/930, 42.6%; *P*<.001), and less likely to report no condom use at the last sex act with any partner (158/501, 31.5% vs 313/775, 40.4%; *P*=.001) compared to those accessing DREAMS services through peer outreach.

Similar proportions of young women who sell sex had completed some secondary school or higher (456/501, 91.0% vs 788/873, 90.3%; *P*=.65) and tested HIV positive as part of the RDS survey or through the Sisters program (76/499, 15.2% vs 123/745, 16.5%; *P*=.55; Table 2). Restricting to the period when RDS was ongoing did not change the results, except that the age distribution of the young women who sell sex was similar regardless of referral approach (18-19 years old: 167/501, 33.3% vs 61/212, 28.8%; *P*=.23; Table 3).

Table 2. Comparison of the characteristics of young women who sell sex, who accessed Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe (DREAMS) services through the Sisters program by referral approach.

Characteristic	Accessed DREAMS[a] through peer outreach (N=930), n/N (%)	Accessed DREAMS through RDS[b] referral (N=501), n/N (%)	*P* value
Age (years)			.004
18-19	243/930 (26.1)	167/501 (33.3)	
20-24	687/930 (73.9)	334/501 (66.7)	
Educational attainment			.65
Primary school or less	85/873 (9.7)	45/501 (9.0)	
Some secondary school or more	788/873 (90.3)	456/501 (91.0)	
City			<.001
Bulawayo	466/930 (50.1)	324/501 (64.7)	
Mutare	464/930 (49.9)	177/501 (35.3)	
Ever tested for HIV			<.001
No	166/827 (20.1)	60/501 (12.0)	
Yes	661/827 (79.9)	441/501 (88.0)	
Reported HIV status before contact with RDS or peer outreach[c]			<.001
Negative	661/661 (100.0)	400/439 (91.1)	
Positive	0/661 (0.0)	39/439 (8.9)	
HIV status at contact with RDS or peer outreach			.55
Negative	622/745 (83.5)	423/499 (84.8)	
Positive	123/745 (16.5)	76/499 (15.2)	
Aware of HIV status at contact with RDS or peer outreach[d]			<.001
No	534/930 (57.4)	106/501 (21.2)	
Yes	396/930 (42.6)	395/501 (78.8)	
No condom used at the last sex act with any partner			.001
No	462/775 (59.6)	343/501 (68.5)	
Yes	313/775 (40.4)	158/501 (31.5)	

[a] DREAMS: Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe.

[b] RDS: respondent-driven sampling.

[c] Among young women who sell sex reporting being ever tested for HIV.

[d] Proportion ever testing HIV positive or having an HIV-negative test during the past 12 months.

Table 3. Comparison of the characteristics of young women who sell sex, who were referred during the period when respondent-driven sampling was ongoing and accessed Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe (DREAMS) services through the Sisters program by referral approach.

Characteristic	Accessed DREAMS[a] through peer outreach (N=212), n/N (%)	Accessed DREAMS through RDS[b] referral (N=501), n/N (%)	*P* value
Age (years)			.23
18-19	61/212 (28.8)	167/501 (33.3)	
20-24	151/212 (71.2)	334/501 (66.7)	
Educational attainment			.16
Primary school or less	12/207 (5.8)	45/501 (9.0)	
Some secondary school or more	195/207 (94.2)	456/501 (91.0)	
City			<.001
Bulawayo	91/212 (42.9)	324/501 (64.7)	
Mutare	121/212 (57.1)	177/501 (35.3)	
Ever tested for HIV			<.001
No	49/204 (24.0)	60/501 (12.0)	
Yes	155/204 (76.0)	441/501 (88.0)	
Reported HIV status before contact with RDS or peer outreach[c]			<.001
Negative	155/155 (100.0)	400/439 (91.1)	
Positive	0/155 (0.0)	39/439 (8.9)	
HIV status at contact with RDS or peer outreach			.63
Negative	175/203 (86.2)	423/499 (84.8)	
Positive	28/203 (13.8)	76/499 (15.2)	
Aware of HIV status at contact with RDS or peer outreach[d]			<.001
No	107/212 (50.5)	106/501 (21.2)	
Yes	105/212 (49.5)	395/501 (78.8)	
No condom used at the last sex act with any partner			.01
No	118/201 (58.7)	343/501 (68.5)	
Yes	83/201 (41.3)	158/501 (31.5)	

[a] DREAMS: Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe.

[b] RDS: respondent-driven sampling.

[c] Among young women who sell sex reporting being ever tested for HIV.

[d] Proportion ever testing HIV positive or having an HIV-negative test during the past 12 months.

Factors Associated With Access of DREAMS Services in the RDS Sample

In adjusted analyses, there was strong evidence that access to DREAMS services was lower among young women who sell sex from Mutare compared to those from Bulawayo (177/603, 30.4% vs 324/601, 53.8%; adjusted OR [aOR] 0.37, 95% CI 0.28-0.50; *P*<.001; Table 4). Young women who sell sex, who started selling sex at 16 to 17 years (183/412, 45.5% vs 46/146, 30.1%; aOR 2.39, 95% CI 1.32-4.32; *P*=.004) or 18 to 24 years (271/645, 42.4% vs 46/146, 30.1%; aOR 2.42, 95% CI 1.36-4.29; *P*=.003) were more likely to access DREAMS services compared to those who started selling sex at 15 years or less. There was evidence that young women who sell sex, who reported at least one close friend were more likely to access DREAMS services compared to those who did not have a close friend (425/986, 44.0% vs 76/218, 34.3%; aOR 1.62, 95% CI 1.03-2.53; *P*=.04).

Table 4. Factors associated with ever access of Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe (DREAMS) services among young women who sell sex, who were referred through respondent-driven sampling (RDS; RDS-II weighted).

Characteristic	Total (N=1204), n (%)	YWSS[a] who accessed DREAMS[b] services via RDS[c] (N=501), n (%)	Crude OR[d] (95% CI)	*P* value	Adjusted OR (95% CI)	*P* value
Age at recruitment (years)				.75		N/A[e]
18-19	405 (35.4)	167 (41.4)	1		N/A	
20-24	799 (64.6)	334 (42.5)	1.05 (0.79-1.39)		N/A	
Highest level of education				.01		.87
Primary school or less	125 (11.7)	45 (33.5)	1		1	
Incomplete secondary school	522 (41.8)	199 (38.9)	1.26 (0.78-2.04)		1.01 (0.55-1.85)	
Complete secondary or higher	557 (46.4)	257 (47.2)	1.77 (1.10-2.85)		1.11 (0.60-2.06)	
City				<.001		<.001
Bulawayo	601 (49.9)	324 (53.8)	1		1	
Mutare	603 (50.1)	177 (30.4)	0.37 (0.28-0.50)		0.43 (0.29-0.62)	
Marital status				.001		.94
Never married	801 (67.9)	366 (46.0)	1		1	
Ever married	403 (32.1)	135 (33.8)	0.60 (0.44-0.81)		0.98 (0.66-1.47)	
Self-identification as a sex worker				.37		N/A
No	390 (35.0)	177 (44.1)	1		N/A	
Yes	801 (65.0)	317 (40.8)	0.88 (0.66-1.17)		N/A	
Age at the start of selling sex (years)				.02		.008
≤15	146 (11.2)	46 (30.1)	1		1	
16-17	412 (34.4)	183 (45.5)	1.93 (1.22-3.07)		2.39 (1.32-4.32)	
18-24	645 (54.4)	271 (42.4)	1.71 (1.10-2.64)		2.42 (1.36-4.29)	
Years selling sex				.20		N/A
0-2	614 (54.1)	275 (44.1)	1		N/A	
3-4	354 (28.7)	143 (41.9)	0.92 (0.67-1.26)		N/A	
≥5	235 (17.2)	82 (36.0)	0.72 (0.50-1.03)		N/A	
Number of sexual partners in the past month				.002		.58
0-4	537 (49.1)	227 (44.0)	1		1	
5-9	286 (22.6)	141 (48.9)	1.22 (0.87-1.71)		1.21 (0.61-2.38)	
≥10	381 (28.3)	133 (33.4)	0.64 (0.46-0.88)		0.69 (0.20-2.33)	
Number of clients in the past month				.001		.96
0-4	609 (55.5)	267 (44.1)	1		1	
5-9	241 (18.7)	114 (49.0)	1.21 (0.85-1.72)		1.09 (0.53-2.27)	
≥10	354 (25.8)	120 (32.3)	0.60 (0.43-0.83)		0.97 (0.28-3.39)	
Condom use at the last sex act with a regular partner				.03		.19
No	320 (34.3)	155 (49.2)	1		1	
Yes	587 (65.7)	236 (40.0)	0.69 (0.49-0.95)		0.74 (0.47-1.16)	
Condom-less sex with a regular partner in the past month				.03		.92
No	498 (55.7)	205 (39.3)	1		1	
Yes	410 (44.3)	186 (47.8)	1.41 (1.03-1.93)		1.02 (0.66-1.57)	
Condom use at the last sex act with a client				.52		N/A

Characteristic	Total (N=1204), n (%)	YWSS[a] who accessed DREAMS[b] services via RDS[c] (N=501), n (%)	Crude OR[d] (95% CI)	*P* value	Adjusted OR (95% CI)	*P* value
No	105 (11.6)	45 (44.7)	1		N/A	
Yes	841 (88.4)	350 (40.8)	0.85 (0.53-1.38)		N/A	
Condom-less sex with a client in the past month				.26		N/A
No	786 (82.6)	322 (40.2)	1		N/A	
Yes	163 (17.4)	74 (46.0)	1.27 (0.84-1.90)		N/A	
STI[f] symptoms in the last 12 months				.07		.96
No	919 (76.7)	404 (43.8)	1		1	
Yes	285 (23.3)	97 (36.6)	0.74 (0.53-1.02)		0.99 (0.67-1.47)	
Relationship with other YWSS				.19		N/A
Good	914 (74.3)	379 (42.1)	1		N/A	
Neither good nor bad	183 (16.7)	84 (47.1)	1.22 (0.84-1.78)		N/A	
No relation	104 (9.0)	37 (34.1)	0.71 (0.43-1.17)		N/A	
Number of close friends				.03		.04
0	218 (19.5)	76 (34.3)	1		1	
≥1	986 (80.5)	425 (44.0)	1.51 (1.05-2.17)		1.62 (1.03-2.53)	
Binge drinking[g]				.046		.44
No alcohol in the last 12 months	401 (37.1)	156 (38.6)	1		1	
Drank alcohol but no occasions of binge drinking	387 (31.7)	153 (40.1)	1.06 (0.76-1.49)		1.18 (0.78-1.77)	
Yes, at least one occasion of binge drinking	415 (31.2)	191 (48.2)	1.48 (1.06-2.06)		1.32 (0.86-2.03)	
Risk of CMD[h]				.93		N/A
No	733 (62.2)	296 (42.0)	1		N/A	
Yes	471 (37.8)	205 (42.3)	1.01 (0.77-1.34)		N/A	
Experienced any form of violence from a sexual partner				.009		.13
No	631 (55.0)	284 (46.1)	1		1	
Yes	573 (45.0)	217 (37.2)	0.69 (0.53-0.91)		0.76 (0.54-1.08)	
Experienced any form of violence from police				.95		N/A
No	1144 (95.8)	476 (42.1)	1		N/A	
Yes	59 (4.2)	24 (42.6)	1.02 (0.56-1.86)		N/A	

[a] YWSS: young women who sell sex.

[b] DREAMS: Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe.

[c] RDS: respondent-driven sampling.

[d] OR: odds ratio.

[e] N/A: not applicable.

[f] STI: sexually transmitted infection.

[g] Had more than six alcoholic drinks in 1 night during the last 12 months.

[h] CMD: common mental disorder.

Incremental Costs and Requirements of Recruitment Strategies

Peer outreach costs included the cost of formative work to select peer educators that was done in 3 days by a team of 3 program staff members in each city, cost of outreach support, and cost of peer educator incentives for 19 months (Table 5). RDS costs included the cost of formative work to select seed participants that was done in 3 days by a team of 3 research staff members in each city, cost of the RDS survey team (5 staff) who spent

50 days in Bulawayo and 60 days in Mutare, and cost of participant recruitment incentives. The incremental cost per young woman who sells sex recruited was US $7.46 for peer outreach and US $52.81 for RDS.

Table 5. Comparison of peer outreach and respondent-driven sampling requirements and incremental costs of recruiting young women who sell sex using each strategy for referral to Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe (DREAMS) services in 2 cities.

Variable	Peer outreach	RDS[a]
Characteristic		
Peer referral	Yes	Yes
Number of peers/seeds in 2 cities	18	20
Duration of recruitment (months)	19	4
Recruitment incentive	No	Yes
Recruitment costs (US$)[b]		
Cost of formative work to identify peers/seeds (per diem)	1350[c]	1350[d]
Cost of peer educator training and materials	5000	N/A[e]
Cost of the outreach support team (salaries)[f]	28,500	N/A
Cost of the RDS survey team (salaries and per diem)[g]	N/A	56,250
Cost of peer educator monthly incentives[h]	5310	N/A
Recruitment incentives[i]	N/A	5980
Total cost (US$)	40,160	63,580
Number of young women who sell sex recruited	5386	1204
Cost per young woman who sells sex recruited (US$)	7.46	52.81

[a] RDS: respondent-driven sampling.

[b] Some of the RDS costs that are research specific, such as cost of laboratory testing, have been omitted, and we have only focused on those that are recruitment specific.

[c] 3 program staff members in 2 teams at US $75 per diem for 3 days.

[d] 3 research staff members in 2 teams at US $75 per diem for 3 days.

[e] N/A: not applicable.

[f] 2 outreach workers for 19 months at US $750 salary per month.

[g] 5 research assistants for 4 months at US $750 salary per month + 5 research assistants at US $75 per diem for 110 days.

[h] 18 peer educators at US $15 monthly incentive for 19 months.

[i] 1204 participants at US $3 participant incentive + 1184 recruits at US $2 peer recruitment incentive.

Discussion

Principal Findings

In this study, we compared 2 recruitment strategies that focused on identifying and reaching young women who sell sex with DREAMS services in 2 cities in Zimbabwe. Our study suggested that peer referral, whether through RDS or peer outreach, can identify high-risk and underserved young women and refer them to services. Peer outreach was able to identify a higher proportion of young women who have never been tested for HIV and are therefore not aware of their HIV status. RDS was able to refer more young women who sell sex in a short period of time and refer younger women aged 18 to 19 compared to peer outreach. Restricting our analysis to the same period of recruitment, we found that peer outreach referred a higher proportion of young women who sell sex, who were not aware of their HIV status compared to RDS, but the ages of these women were similar regardless of referral approach. Among women who accessed DREAMS services, those referred by RDS were younger and appeared to be better engaged with services, and more women had previously tested for HIV and knew their status. By contrast, peer outreach identified more young women who sell sex, who had never been tested for HIV and were unaware of their HIV status. Although the differences in ever testing between recruitment approaches were not significant (76% vs 88%), when programs are aiming to ensure that all those who are vulnerable are reached, optimizing referral approaches by using a combination of approaches is likely important.

Both referral approaches were successful in reaching young women who sell sex at high risk of HIV, where the majority were HIV negative. In our previous study, we found that HIV prevalence and incidence rise steeply with age among this population [3,15], and thus, supporting young women who sell sex to engage effectively with prevention is critical. Reported noncondom use was high in both groups but even higher among those recruited through peer outreach.

The success of peer referral approaches in reaching high-risk hidden populations has been noted in many populations, including in West and Central Africa where the use of an enhanced (incentivized) peer outreach approach led to increased detection of new HIV-positive key population individuals who would not have been engaged otherwise [5,6]. With their enhanced peer referral approach, the authors were able to reach female sex workers who had not been effectively engaged by routine outreach approaches [5]. Our study, however, demonstrated that a higher proportion of previously unengaged young women who sell sex accessed DREAMS services through peer outreach than through RDS. This is possibly due to the level of training that peer educators received coupled with the tailored community mobilization activities and materials used. The fact that the RDS strategy referred young women who were more likely to already be engaged with HIV services could be because those young women who sell sex were more likely to be visible on the social network of young women who sell sex. Of note, RDS recruitment was done in a limited time frame with participants given a limited number of coupons with the goal of recruiting a representative sample of the network of the population of young women who sell sex, unlike peer outreach where peer educators were expected to recruit as many young women who sell sex as possible in a longer time frame. This may not only have restricted the absolute number of young women who sell sex recruited via RDS but also limited the performance of RDS in reaching less networked young women who sell sex. Nonetheless, our RDS diagnostics reported elsewhere [3] suggested that convergence was achieved in the 2 cities and our sample was likely to be representative of the network of young women who sell sex recruited.

While RDS enrolled young women who sell sex quickly, the requirement to incentivize recruitment at every stage can be costly to integrate into day-to-day programs. We showed that nonincentivized peer outreach is also able to reach high-risk young women who sell sex. Peer outreach provides at least the possibility of the process of referral being associated with longer-term support. Importantly, we found that the different approaches presented here recruited different groups of high-risk women, and the overlap in terms of those recruited was small, even when compared with peer outreach continued over many months. It seems likely that using a combination of approaches will be most effective at optimizing reach and coverage.

Consistent with other findings [16-19], our analysis suggested the importance of a comprehensive, well-coordinated, and scaled-up HIV program in reaching priority populations. Young women who sell sex from Bulawayo, where the DREAMS program was well coordinated with implementing partners working together to build high DREAMS acceptance, were more likely to access DREAMS via RDS compared to those from Mutare. On the other hand, peer outreach performed better in Mutare (a smaller town) where it referred a higher proportion of young women who sell sex than in Bulawayo. Lessons need to be constantly learned between program sites to generate opportunities for program improvement. Moreover, young women who sell sex, who reported at least one close friend were more likely to access DREAMS services compared to those who did not have a close friend, emphasizing the importance of building social cohesion among disempowered communities to optimize their uptake of HIV prevention and care [20-22].

Limitations

The limitations include the relatively limited data captured routinely at the time of program referral, limiting the possibility of comparing the characteristics of young women who sell sex, who went on to access services with those of young women who did not. Information on the refusal rate and the reason why some young women who sell sex refused to participate in DREAMS may be useful to refine existing recruitment approaches or operationalize novel approaches like starfish sampling that combines time location sampling and RDS [23]. We might not have compared like with like since our RDS, which was done for research purposes, recruited only young women who sell sex, who were 18 years or above, while referrals through peer outreach were able to include younger women who are even more vulnerable.

Conclusions

Peer outreach and RDS approaches can reach and refer high-risk but different groups of young women who sell sex for HIV services. Use of both these complementary approaches will likely improve reach.

Acknowledgments

We would like to thank the program and research teams, and the women who participated in the study. The DREAMS impact evaluation in Zimbabwe was funded by the Bill & Melinda Gates Foundation (OPP1136774; http://www.gatesfoundation.org). Foundation staff advised the study team, but did not substantively influence the study design, instruments, interpretation of data, or decision to publish.

Authors' Contributions

STC planned and conducted the analysis, and wrote the first draft; RM oversaw program implementation; BH was involved in data interpretation; SM was involved in program monitoring and evaluation; TC and PM led the data collection; and BH, JB, SF, IB, JRH, and FMC critically revised the article. All authors contributed to the writing and have read and approved the final version.

Conflicts of Interest

None declared.

References

1. Busza J, Mtetwa S, Mapfumo R, Hanisch D, Wong-Gruenwald R, Cowan F. Underage and underserved: reaching young women who sell sex in Zimbabwe. AIDS Care 2016 Mar 08;28 Suppl 2(sup2):14-20 [FREE Full text] [doi: 10.1080/09540121.2016.1176673] [Medline: 27391994]
2. Cowan FM, Chabata ST, Musemburi S, Fearon E, Davey C, Ndori-Mharadze T, et al. Strengthening the scale-up and uptake of effective interventions for sex workers for population impact in Zimbabwe. J Int AIDS Soc 2019 Jul;22 Suppl 4:e25320 [FREE Full text] [doi: 10.1002/jia2.25320] [Medline: 31328445]
3. Hensen B, Chabata ST, Floyd S, Chiyaka T, Mushati P, Busza J, et al. HIV risk among young women who sell sex by whether they identify as sex workers: analysis of respondent-driven sampling surveys, Zimbabwe, 2017. J Int AIDS Soc 2019 Dec;22(12):e25410 [FREE Full text] [doi: 10.1002/jia2.25410] [Medline: 31793748]
4. Sabin LL, Semrau K, DeSilva M, Le LTT, Beard JJ, Hamer DH, et al. Effectiveness of community outreach HIV prevention programs in Vietnam: a mixed methods evaluation. BMC Public Health 2019 Aug 16;19(1):1130 [FREE Full text] [doi: 10.1186/s12889-019-7418-5] [Medline: 31420034]
5. Olawore O, Astatke H, Lillie T, Persaud N, Lyons C, Kamali D, et al. Peer recruitment strategies for female sex workers not engaged in HIV prevention and treatment services in Côte d'Ivoire: Program data analysis. JMIR Public Health Surveill 2020 Oct 01;6(4):e18000 [FREE Full text] [doi: 10.2196/18000] [Medline: 33001039]
6. Lillie TA, Persaud NE, DiCarlo MC, Gashobotse D, Kamali DR, Cheron M, et al. Reaching the unreached: Performance of an enhanced peer outreach approach to identify new HIV cases among female sex workers and men who have sex with men in HIV programs in West and Central Africa. PLoS One 2019 Apr 3;14(4):e0213743 [FREE Full text] [doi: 10.1371/journal.pone.0213743] [Medline: 30943205]
7. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. Social Problems 1997 May;44(2):174-199. [doi: 10.2307/3096941]
8. Birdthistle I, Schaffnit SB, Kwaro D, Shahmanesh M, Ziraba A, Kabiru CW, et al. Evaluating the impact of the DREAMS partnership to reduce HIV incidence among adolescent girls and young women in four settings: a study protocol. BMC Public Health 2018 Jul 25;18(1):912 [FREE Full text] [doi: 10.1186/s12889-018-5789-7] [Medline: 30045711]
9. Hensen B, Hargreaves JR, Chiyaka T, Chabata S, Mushati P, Floyd S, et al. Evaluating the impact of DREAMS on HIV incidence among young women who sell sex: protocol for a non-randomised study in Zimbabwe. BMC Public Health 2018 Jan 31;18(1):203 [FREE Full text] [doi: 10.1186/s12889-018-5085-6] [Medline: 29386001]
10. Chabata ST, Hensen B, Chiyaka T, Mushati P, Musemburi S, Dirawo J, et al. The impact of the DREAMS partnership on HIV incidence among young women who sell sex in two Zimbabwean cities: results of a non-randomised study. BMJ Glob Health 2021 Apr 27;6(4):e003892 [FREE Full text] [doi: 10.1136/bmjgh-2020-003892] [Medline: 33906844]
11. Hargreaves JR, Mtetwa S, Davey C, Dirawo J, Chidiya S, Benedikt C, et al. Implementation and operational research: cohort analysis of program data to estimate HIV incidence and uptake of HIV-related services among female sex workers in Zimbabwe, 2009-2014. J Acquir Immune Defic Syndr 2016 May 01;72(1):e1-e8. [doi: 10.1097/QAI.0000000000000920] [Medline: 27093516]
12. The DREAMS Partnership: The rationale for the DREAMS Partnership, achievements to date, and noteworthy lessons learned along the way. U.S. President's Emergency Plan for Aids Relief (PEPFAR). 2017. URL: https://www.cgdev.org/sites/default/files/1_Deborah%20Birx_June%201_FINAL.PDF [accessed 2021-06-20]
13. Cowan F, Chabata S, Musemburi S, Fearon E, Davey C, Ndori-Mharadze T, et al. Strengthening the scale-up and uptake of effective interventions for sex workers for population impact in Zimbabwe. J Int AIDS Soc 2019 Jul;22 Suppl 4:e25320 [FREE Full text] [doi: 10.1002/jia2.25320] [Medline: 31328445]
14. Chiyaka T, Mushati P, Hensen B, Chabata S, Hargreaves JR, Floyd S, et al. Reaching young women who sell sex: Methods and results of social mapping to describe and identify young women for DREAMS impact evaluation in Zimbabwe. PLoS One 2018 Mar 15;13(3):e0194301 [FREE Full text] [doi: 10.1371/journal.pone.0194301] [Medline: 29543858]
15. Ali MS, Wit MDE, Chabata ST, Magutshwa S, Musemburi S, Dirawo J, et al. Estimation of HIV incidence from analysis of HIV prevalence patterns by age and years since starting sex work among female sex workers in Zimbabwe. AIDS 2022 Feb 16:3198. [doi: 10.1097/QAD.0000000000003198] [Medline: 35170527]
16. Bekker L, Alleyne G, Baral S, Cepeda J, Daskalakis D, Dowdy D, et al. Advancing global health and strengthening the HIV response in the era of the Sustainable Development Goals: the International AIDS Society—Lancet Commission. The Lancet 2018 Jul;392(10144):312-358. [doi: 10.1016/s0140-6736(18)31070-5]
17. Shannon K, Crago A, Baral S, Bekker L, Kerrigan D, Decker M, et al. The global response and unmet actions for HIV and sex workers. The Lancet 2018 Aug;392(10148):698-710 [FREE Full text] [doi: 10.1016/s0140-6736(18)31439-9]
18. Mwau M, Syeunda CA, Adhiambo M, Bwana P, Kithinji L, Mwende J, et al. Scale-up of Kenya's national HIV viral load program: Findings and lessons learned. PLoS One 2018 Jan 11;13(1):e0190659 [FREE Full text] [doi: 10.1371/journal.pone.0190659] [Medline: 29324811]
19. Musyoki H, Bhattacharjee P, Sabin K, Ngoksin E, Wheeler T, Dallabetta G. A decade and beyond: learnings from HIV programming with underserved and marginalized key populations in Kenya. Journal of the International AIDS Society 2021;24:e25729 [FREE Full text] [doi: 10.1002/jia2.25729]

20. Carrasco MA, Barrington C, Perez M, Donastorg Y, Kerrigan D. Social cohesion, condom use, and sexually transmitted infections among female sex workers living with HIV in the Dominican Republic. Int J STD AIDS 2019 Jan;30(1):64-71. [doi: 10.1177/0956462418792100] [Medline: 30170526]
21. Lippman S, Leslie H, Neilands T, Twine R, Grignon J, MacPhail C, et al. Context matters: Community social cohesion and health behaviors in two South African areas. Health Place 2018 Mar;50:98-104 [FREE Full text] [doi: 10.1016/j.healthplace.2017.12.009] [Medline: 29414427]
22. Kalolo A, Mazalale J, Krumeich A, Chenault M. Social cohesion, social trust, social participation and sexual behaviors of adolescents in rural Tanzania. BMC Public Health 2019 Feb 14;19(1):193 [FREE Full text] [doi: 10.1186/s12889-019-6428-7] [Medline: 30764797]
23. Raymond HF, Chen Y, McFarland W. "Starfish Sampling": a novel, hybrid approach to recruiting hidden populations. J Urban Health 2019 Feb 16;96(1):55-62 [FREE Full text] [doi: 10.1007/s11524-018-0316-9] [Medline: 30328063]

Abbreviations

aOR: adjusted odds ratio
DREAMS: Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe
RDS: respondent-driven sampling

Edited by Y Khader; submitted 21.07.21; peer-reviewed by S Weir, K Shanaube, T Lillie, H Astatke; comments to author 12.09.21; revised version received 31.03.22; accepted 11.05.22; published 27.07.22.

Please cite as:
Chabata ST, Makandwa R, Hensen B, Mushati P, Chiyaka T, Musemburi S, Busza J, Floyd S, Birdthistle I, Hargreaves JR, Cowan FM
Strategies to Identify and Reach Young Women Who Sell Sex With HIV Prevention and Care Services: Lessons Learnt From the Implementation of DREAMS Services in Two Cities in Zimbabwe
JMIR Public Health Surveill 2022;8(7):e32286
URL: https://publichealth.jmir.org/2022/7/e32286
doi:10.2196/32286
PMID:35896024

Original Paper

Bridging the Gap in End Tuberculosis Targets in the Elderly Population in Eastern China: Observational Study From 2015 to 2020

Kui Liu[1*], MD; Zhenhua Xie[2*], MD; Bo Xie[3], PhD; Songhua Chen[1], MD; Yu Zhang[1], MD; Wei Wang[1], MD; Qian Wu[1], MD; Gaofeng Cai[1*], MD; Bin Chen[1*], MD

[1]Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China
[2]Ningbo Yinzhou No 2 Hospital, Ningbo, China
[3]School of Urban Design, Wuhan University, Wuhan, China
[*]these authors contributed equally

Corresponding Author:
Gaofeng Cai, MD
Zhejiang Provincial Center for Disease Control and Prevention
3399 Binsheng Road
Binjiang District, Zhejiang Province
Hangzhou, 310000
China
Phone: 86 057187115020
Fax: 86 057187115020
Email: gfcai@cdc.zj.cn

Abstract

Background: With a progressive increase in the aging process, the challenges posed by pulmonary tuberculosis (PTB) are also increasing for the elderly population.

Objective: This study aimed to identify the epidemiological distribution of PTB among the elderly, forecast the achievement of the World Health Organization's 2025 goal in this specific group, and predict further advancement of PTB in the eastern area of China.

Methods: All notified active PTB cases aged ≥65 years from Zhejiang Province were screened and analyzed. The general epidemiological characteristics were depicted and presented using the ArcGIS software. Further prediction of PTB was performed using R and SPSS software programs.

Results: Altogether 41,431 cases aged ≥65 years were identified by the surveillance system from 2015 to 2020. After excluding extrapulmonary TB cases, we identified 39,832 PTB cases, including laboratory-confirmed (23,664, 59.41%) and clinically diagnosed (16,168, 40.59%) PTB. The notified PTB incidence indicated an evident downward trend with a reduction of 30%; however, the incidence of bacteriologically positive cases was steady at approximately 60/100,000. Based on the geographical distribution, Quzhou and Jinhua Cities had a higher PTB incidence among the elderly. The delay in PTB diagnosis was identified, and a significantly prolonged treatment course was observed in the elderly. Moreover, a 50% reduction of PTB incidence by the middle of 2024 was predicted using a linear regression model. It was found that using the exponential smoothing model would be better to predict the PTB trend in the elderly than a seasonal autoregressive integrated moving average model.

Conclusions: More comprehensive and effective interventions such as active PTB screening combined with physical checkup and succinct health education should be implemented and strengthened in the elderly. A more systematic assessment of the PTB epidemic trend in the elderly population should be considered to incorporate more predictive factors.

(JMIR Public Health Surveill 2022;8(7):e39142) doi:10.2196/39142

KEYWORDS

pulmonary tuberculosis; elderly population; prediction

Introduction

Tuberculosis (TB) ranks 13th among the leading causes of death globally and is also the top cause of death from a single infectious agent, which has been a substantial public health concern and an urgent global public health priority [1,2]. Nearly a quarter of the global population is infected with *Mycobacterium tuberculosis*, although only <10% of the people developed active TB during their lifetime, especially in low- and middle-income areas [3-5]. According to the Global Tuberculosis Report 2021 released by the World Health Organization (WHO), an estimated 9.9 million people developed illness with an incidence of 127/100,000 in 2020 [6]. Meanwhile, the latest report has demonstrated that the estimated number of newly diagnosed TB cases was 842,000 with a notified incidence of 59/100,000 in China, which still ranks second among the 30 highly burdened countries [6]. Despite being a preventable and curable disease, it has serious effects on some targeted groups such as the elderly and continues to be a major challenge.

With progressive population aging in China, the population aged >60 years had reached 264 million based on the 7th population census, among which 190 million were >65 years old, accounting for 13.5% of the total population [7]. Recently, evidence indicated a progressive increase in the notification rate of pulmonary TB (PTB) with age, particularly in East Asian and Southeast Asian countries [8]. The geriatric population is susceptible to the development of TB, which may be attributed to reactivation of the lesions and change in immunity levels [9]. With the implementation of the directly observed treatment strategy, the prevalence of PTB in China declined sharply, whereas the exploration of the change trend in the elders indicated a paucity of data [10,11]. Therefore, exploring the potential characteristics of PTB among the elderly is necessary.

Zhejiang Province, a developed area located in Eastern China, has an economic output of exceeding US $1.08 trillion. According to the Zhejiang Provincial Health Commission in 2019, the elderly population aged ≥60 years reached 11,217,200, accounting for 22.43% of the total population of the province, which was 4.53% higher than that at the national level.

This study aimed to identify the epidemiological distribution of PTB among the elderly, assess the performance of the WHO's 2025 goal in this specific group, and predict the further advancement of PTB in Eastern China. It would contribute to filling the existing policy gaps, prompting optimization of the current health policy, and constructing a new strategic health framework to realize the end TB targets.

Methods

Overview

This study was focused on Zhejiang Province, which is in the eastern region of China and comprises the following cities: Hangzhou, Ningbo, Wenzhou, Jiaxing, Huzhou, Shaoxing, Jinhua, Quzhou, Zhoushan, Taizhou, and Lishui. Its location and demographic information have been presented in previous studies [12,13]. Additionally, according to the latest data released by the Zhejiang Bureau of Statistics, there were 65.4 million permanent residents, of which 9.26 million were aged ≥65 years and accounted for 14.2% of the provincial population. The demographic data of this specific population showed an increase of 0.9% over the previous year [14].

Data Collection

All PTB cases recorded in the Zhejiang Province during 2015-2020 were selected and screened from the TB information management system (TBIMS). The TBIMS was designed and constructed as a first-generation web-based information system by the National Center for TB Control and Prevention in 2005 [15]. All users of the TBIMS including designated hospitals, communities, and the Center for Disease Control and Prevention at the provincial, city, and county levels were authorized to fulfill their duties in this specific system [16]. Patients with PTB aged ≥65 years were included, and the data pertaining to their demographics, diagnosis, laboratory outcomes, and treatment outcomes were collected. Additionally, the basic data of Zhejiang Province were approved and obtained from the Chinese Information System for Disease Control and Prevention and local statistical yearbook [12].

Definition

In this study, the PTB cases consisted of laboratory-confirmed PTB (also called bacteriologically diagnosed PTB) and clinically diagnosed PTB. The former is diagnosed by bacteriological evidence acquired by sputum smear and culture or a rapid diagnostic system, such as the GeneXpert MTB/RIF. The latter is diagnosed by chest imaging, epidemiological findings, and clinical symptoms along with other relevant testing [17]. All identifications are based on the National Diagnostic Criteria for Pulmonary Tuberculosis (WS288-2008 and WS288-2017) and Classification of Tuberculosis (WS196-2001 and WS196-2017) in China [18,19].

General Characteristics of PTB in the Elderly

According to bacteriological and clinical diagnoses, the results are presented as the following parameters: sex, age groups, different cities, occupations, source of patients, treatment classification, anti-TB treatment, drug sensitivity results, treatment outcomes, and various delays in seeking treatment. Additionally, the notified incidence and spatial distribution depict the epidemiological characteristics of PTB in the elderly.

Predictive Model for PTB in the Elderly

Based on the WHO's 2025 goal, a 50% reduction in TB morbidity and 75% decline in the absolute number of TB deaths should be achieved. PTB in Zhenjiang Province was predicted using 2 methods. In method I, the available notified incidence was used to fit and screen regression models according to the R^2 value. The linear regression model (LRM), exponential curve model, and growth curve model (GCM) were considered. These models are described in our previous study [20]. In method II, the seasonal autoregressive integrated moving average (ARIMA) model and exponential smoothing (ETS) method, 2 commonly used time series models, were compared to determine the appropriate one for predicting the trend of PTB cases in the elderly. The seasonal ARIMA and ETS models were used not only for short-term prediction but also for middle- and long-term forecasting [21,22]. The former consisted of p, d, and q and seasonal parameters p, d, and q, where p represents autoregression, d represents the degree of difference, and q represents the order of the moving average [23]. Furthermore, considering the patterns of addition, multiplication, or no patterns, the ETS model consisted of three vital parameters: error, trend, and seasonal components [24]. The formulas of the seasonal ARIMA and ETS models have been presented elsewhere, so they were omitted here [24,25]. For each method, the optimal model was selected based on the indicators of the Akaike information criterion, the corrected Akaike information criterion, and Bayesian information criterion [26-28]. Ultimately, the root mean squared error (RMSE) and mean absolute percentage error (MAPE) were used to compare the accuracy of prediction [24].

Ethics Approval

Information on all the cases was fully anonymized in the processing of data. This study was approved by the Ethics Committee of the Zhejiang Provincial Center for Disease Control and Prevention (2021-027-01). As only public health surveillance data were used, the requirement of informed consent was waived by the abovementioned ethics committee. Moreover, all details used in this research were in accordance with the Law of the Prevention and Treatment of Infectious Diseases in the People’s Republic of China.

Statistical Analysis

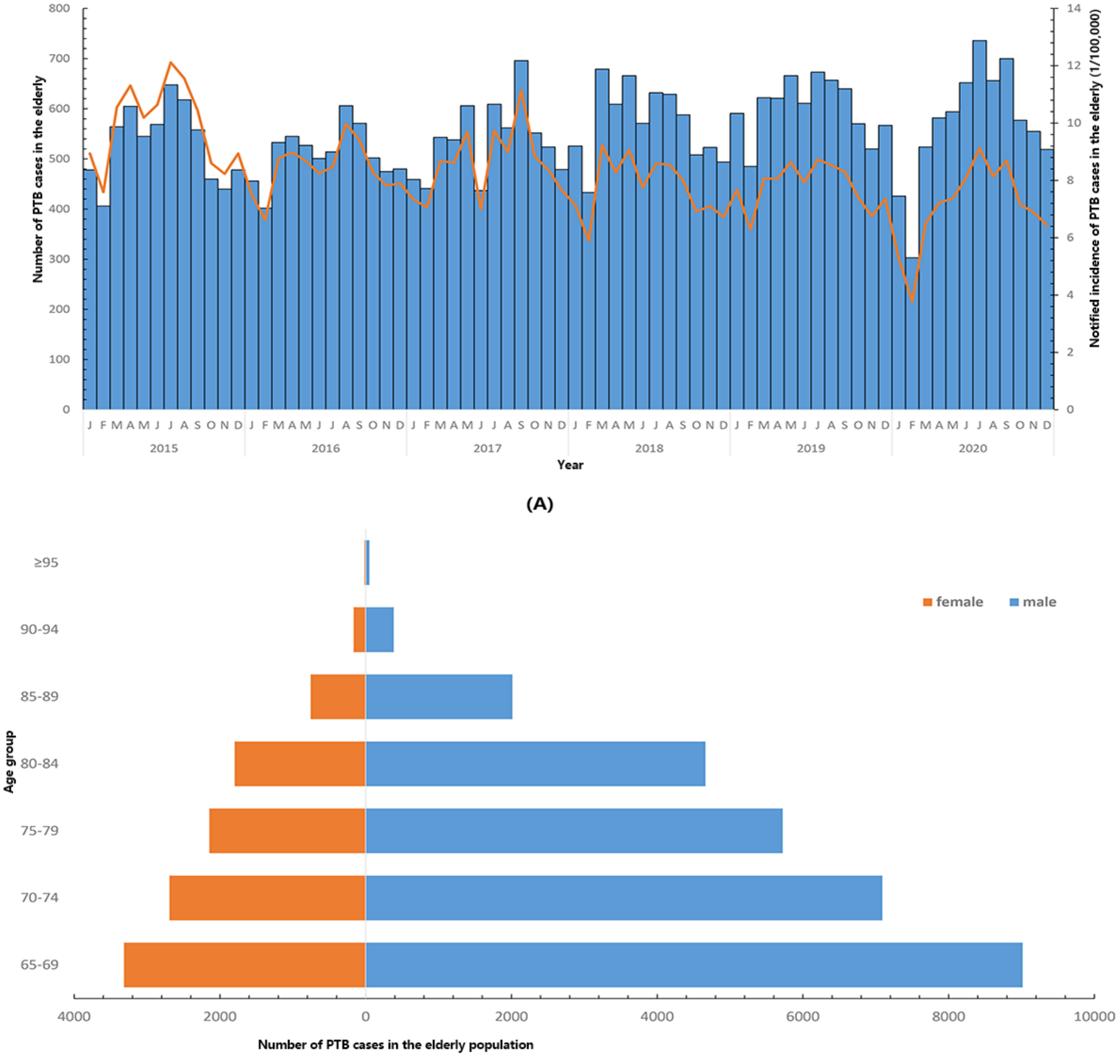
Descriptive analysis was performed using the R software (version 3.5.3, R Foundation for Statistical Computing), and the map was developed using the ArcGIS software (version 10.2, Esri). Additionally, the prediction of the PTB epidemic in 2025 was performed using SPSS Statistics 20.0 (IBM Corp), and the time series calculations involving the ETS and seasonal ARIMA models were performed using the R software (version 3.5.3). Statistical significance was set at $P<.05$.

Results

Notified Incidence and Geographical Distribution

A total of 41,431 records of people with PTB aged ≥65 years were notified in the TBIMS from 2015 to 2020. After excluding cases of extrapulmonary TB, we collected 39,832 PTB cases consisting of laboratory-confirmed PTB (23,664, 59.41%) and clinically diagnosed PTB (16,168, 40.59%), with the highest registered number in 2019 (n=7223). The notified incidence of PTB among the elderly was 119.09 per 100,000 in 2015; 100.66 per 100,000 in 2016; 103.44 per 100,000 in 2017; 93.26 per 100,000 in 2018; 93.74 per 100,000 in 2019; and 84.67 per 100,000 in 2020 (Figure 1A). The notified incidence of bacteriologically positive PTB cases among the elderly was 61.18 per 100,000 in 2015; 50.77 per 100,000 in 2016; 52.75 per 100,000 in 2017; 59.98 per 100,000 in 2018; 64.29 per 100,000 in 2019; and 58.98 per 100,000 in 2020. The former revealed an evidently declining trend, whereas the latter had a comparatively stable state. In terms of the geographical distribution of PTB in the elderly, Quzhou City had the highest incidence, whereas Zhoushan City had the lowest PTB burden. Moreover, the notified PTB incidence in the elderly illustrated an overall slow downward trend in various cities but still exceeded that in the general population (nearly 40 per 100,000 to 50 per 100,000) (Figure 2A). For bacteriologically positive cases, the notified incidence demonstrated a significant difference in various cities, where the incidence in Quzhou City was 5 times more than that in Zhoushan City (data not shown). Meanwhile, incidence of bacteriologically positive cases revealed significant differences among the 11 cities (Figure 2B). In addition, a U-trend was revealed for bacteriologically positive PTB incidence in Quzhou and Jinhua Cities, whereas a successive reduction was recorded in Jiaxing City. However, despite a relatively low notified incidence, Wenzhou City presented a steady upward trend. Additionally, nearly 9 cities demonstrated a unimodal trend for bacteriologically positive incidence in 2018-2019.

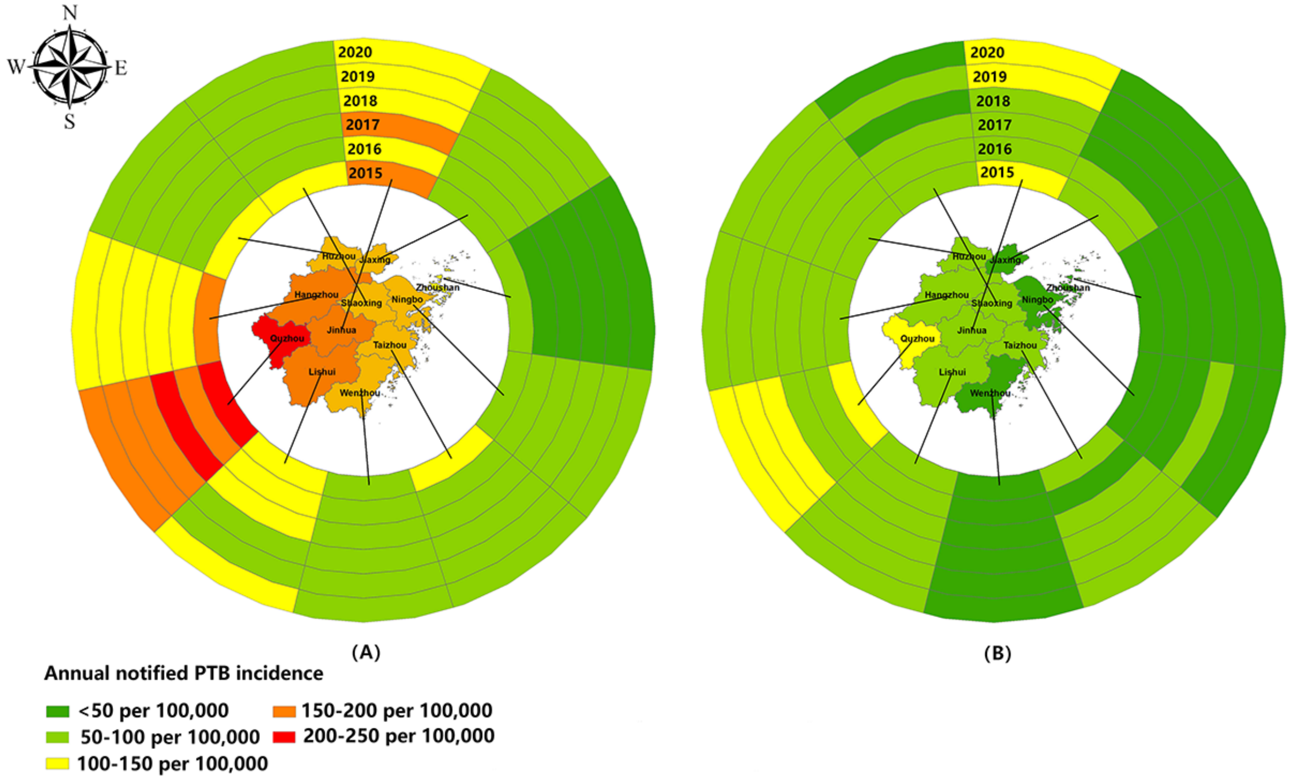
Figure 1. General epidemiological characteristics of PTB in the elderly: (A) Change trend of notified PTB incidence and recorded case number by months. (B) Distribution of notified PTB cases in different sexes by various age groups. PTB: pulmonary tuberculosis.



(A)

(B)

Figure 2. Geographical distribution of PTB cases in the elderly population during the study period: (A) Annual notified PTB incidence among different cities. (B) Annual bacteriologically positive PTB incidence among different cities. PTB: pulmonary tuberculosis.



General Characteristics of PTB in the Elderly

Of all the 39,832 geriatric PTB cases, men accounted for 72.67% (28,944 cases), whereas women accounted for 27.33% (10,888 cases). Regarding age distribution in each sex, the onset number in males was more than that in females, and a significant declining trend was observed, in which the age group of 65-69 years accounted for 30.96% (12,332) of the cases (Figure 1B). The median age was 73 years in both sexes. Regarding geographical distribution, Hangzhou, Jinhua, Taizhou, and Quzhou cities contributed to >10% among all notified PTB cases, in which the highest bacteriological (4514, 11.33%) and clinical diagnoses (3866, 9.71%) of PTB were both in the Hangzhou region. The top 3 occupations of the elderly were farming and working; being a retiree; and housekeeping, housework, and unemployment. Besides, referral and passive finding (also called actively seeking a doctor) had accounted for nearly 86% (34,284) patients among diverse sources. Most of all the elderly PTB cases (35,177, 88.31%) belonged to the initial treatment category, and 51.26% (20,418) comprised the bacteriological diagnoses. Nearly all the elderly patients could be provided anti-TB treatment, and 59.28% (23,613) received standardized treatment based on bacterial evidence. From the available information, the known drug resistance ratio was >2.7%, the drug resistance rate (rifampicin monoresistance) was 5.86% among the people aged ≥65 years, and the treatment success rate among the study population including completion of the treatment course and cure was approximately 72%. A longer delay was observed in the interval from the disease onset to the hospital visit compared to the interval from visiting the designated hospital to confirmation of PTB. Surprisingly, nearly half of the elderly cases had a treatment time >9 months (Table 1).

Table 1. Epidemiological characteristics of pulmonary tuberculosis cases among the elderly population in Eastern China from 2015 to 2020 (N=39,832).

Characteristic	Total PTB[a]	Laboratory-confirmed PTB	Clinically diagnosed PTB
Sex, n (%)			
Male	28,944 (72.67)	17,347 (43.55)	11,597 (29.11)
Female	10,888 (27.33)	6317 (15.86)	4571 (11.48)
Age group (years), n (%)			
65-69	12,332 (30.96)	6670 (16.75)	5662 (14.21)
70-74	9782 (24.56)	5809 (14.58)	3973 (9.97)
75-79	7868 (19.75)	4826 (12.12)	3042 (7.64)
80-84	6461(16.22)	4161 (10.45)	2300 (5.77)
85-89	2769 (6.95)	1783 (4.48)	986 (2.48)
90-94	551 (1.38)	380 (0.95)	171 (0.43)
≥95	69 (0.17)	35 (0.09)	34 (0.09)
City, n (%)			
Hangzhou	8380 (21.04)	4514 (11.33)	3866 (9.71)
Ningbo	3797 (9.53)	2383 (5.98)	1414 (3.55)
Wenzhou	3803 (9.55)	2120 (5.32)	1683 (4.23)
Jiaxing	2343 (5.88)	1579 (3.96)	764 (1.92)
Huzhou	2120 (5.32)	1331 (3.34)	789 (1.98)
Shaoxing	3306 (8.30)	1998 (5.02)	1308 (3.28)
Jinhua	5661 (14.21)	3676 (9.23)	1985 (4.98)
Quzhou	3994 (10.03)	2256 (5.66)	1738 (4.36)
Zhoushan	401 (1.01)	190 (0.48)	211 (0.53)
Taizhou	4021 (10.09)	2400 (6.03)	1621 (4.07)
Lishui	2006 (5.04)	1217 (3.06)	789 (1.98)
Occupation, n (%)			
Farmer and worker	29,342 (73.66)	17,696 (44.43)	11,646 (29.24)
Retiree	6128 (15.38)	3451 (8.66)	2677 (6.72)
Housekeeping, housework, and unemployment	3406 (8.55)	1957 (4.91)	1449 (3.64)
Unknown	512 (1.29)	330 (0.83)	182 (0.46)
Others	284 (0.71)	151 (0.38)	133 (0.33)
Commercial service stratum	97 (0.24)	51 (0.13)	46 (0.12)
Cadre staff	37 (0.09)	16 (0.04)	21 (0.05)
Herder	15 (0.04)	8 (0.02)	7 (0.02)
Doctor	11 (0.03)	4 (0.01)	7 (0.02)
Source of patients, n (%)			
Health examination	176 (0.44)	77 (0.19)	99 (0.25)
Contact screening	3 (0.01)	2 (0.01)	1 (0)
Passive finding	14,391 (36.13)	8174 (20.52)	6217 (15.61)
Referral	19,893 (49.94)	11,889 (29.85)	8004 (20.09)
Tracing	5205 (13.07)	3423 (8.59)	1782 (4.47)
Others	164 (0.41)	99 (0.25)	65 (0.16)
Classification of treatment, n (%)			
Initial treatment	35,177 (88.31)	20,418 (51.26)	14,759 (37.05)

Characteristic	Total PTB[a]	Laboratory-confirmed PTB	Clinically diagnosed PTB
Retreatment	4655 (11.69)	3246 (8.15)	1409 (3.54)
Anti-TB treatment, n (%)			
Yes	39,753 (99.8)	23,613 (59.28)	16,140 (40.52)
No	74 (0.19)	51 (0.13)	23 (0.06)
Results of drug sensitivity test, n (%)			
Monodrug resistance	866 (2.17)	817 (2.05)	49 (0.12)
Polydrug resistance	3 (0.01)	2 (0.01)	1 (0)
Sensitivity	17,289 (43.40)	16,422 (41.23)	867 (2.18)
Multidrug resistance	208 (0.52)	195 (0.49)	13 (0.03)
No test	21,466 (53.89)	6228 (15.64)	15,238 (38.26)
Treatment outcome, n (%)			
Completion of the treatment course[b]	13,977 (35.09)	1546 (3.88)	12,431 (31.21)
Cure[c]	14,380 (36.1)	14,335 (35.99)	45 (0.11)
Death	2854 (7.17)	2171 (5.45)	683 (1.71)
Failure	429 (1.08)	348 (0.87)	81 (0.20)
Adverse reaction	598 (1.50)	316 (0.79)	282 (0.71)
Transfer to MDR-TB[d] treatment	254 (0.64)	230 (0.58)	24 (0.06)
Others	7340 (18.43)	4718 (11.84)	2622 (6.58)
Interval between onset and visit to the designated hospital (days), n (%)			
0-14	17,315 (43.47)	10,586 (26.58)	6729 (16.89)
15-29	7944 (19.94)	4410 (11.07)	3534 (8.87)
30-44	5417 (13.60)	3074 (7.72)	2343 (5.88)
45-59	1826 (4.58)	992 (2.49)	834 (2.09)
≥60	7284 (18.29)	4580 (11.50)	2704 (6.79)
Unknown	46 (0.12)	22 (0.06)	24 (0.06)
Interval between visit to designated hospital and confirmation of PTB (days), n (%)			
0-14	34,253 (85.99)	20,198 (50.71)	14,055 (35.29)
15-29	2911 (7.31)	1746 (4.38)	1165 (2.92)
30-44	1090 (2.74)	680 (1.71)	410 (1.03)
45-59	458 (1.15)	280 (0.70)	178 (0.45)
≥60	1053 (2.64)	718 (1.80)	335 (0.84)
Unknown	67 (0.17)	42 (0.11)	25 (0.06)
Interval between confirmation of PTB and end of therapy (days), n (%)			
<180	5059 (12.7)	3537 (8.88)	1522 (3.82)
180-270	12,597 (31.63)	7644 (19.19)	4953 (12.43)
≥270	17,224 (43.24)	9148 (22.97)	8076 (20.28)

Characteristic	Total PTB[a]	Laboratory-confirmed PTB	Clinically diagnosed PTB
Unknown	4952 (12.43)	3335 (8.37)	1617 (4.06)

[a] PTB: pulmonary tuberculosis.

[b] Completion of the treatment course referred to the following situations: (1) Patients with negative etiology finished the standardized course with negative results of sputum smear/culture or no test was performed. (2) Patients with positive etiology finished the standardized course without undergoing a sputum result at the end of the treatment whereas the last sputum smear or culture yielded a negative result.

[c] Cure referred to the situation where patients with positive etiology finished the standardized course of treatment and had continuous negative results of sputum smear or culture in the last month and during the penultimate test.

[d] MDR-TB: multidrug-resistant TB.

Trend Prediction for PTB in the Elderly

Based on the WHO's 2025 goal, a reduction of 50% in TB incidence and 75% in TB absolute numbers should be achieved during the period from 2015 to 2025. All the 3 included models were considered (model with $P<.05$ for ANOVA). Considering a minor difference in R^2, GCM and LRM (Table 2) were reserved for the prediction of notified PTB incidence. Eventually, the result indicated a possible achievement of the desired TB incidence in the middle of 2024 using LRM. Besides, given a considerable decline in PTB-related deaths in 2020, no model to predict the notified PTB death number based on available data was established.

Table 2. Predictive models for pulmonary tuberculosis incidence in the elderly.

Model	*F* (*df*1,*df*2)	*P* value	R^2	Coefficients			
				b0	*P* value	b1	*P* value
LRM[a]	23.80 (1,4)	<.001	0.86	11619.97	<.001	–5.71	<.001
GCM[b]	27.69 (1,4)	<.001	0.87	119.27	<.001	–0.06	<.001

[a] LRM: linear regression model.

[b] GCM: growth curve model.

Predictive Model for PTB Cases in the Elderly

For comparing fitness to forecasting of the PTB epidemic trend among the elderly in 2021, the Holt-Winters exponential smoothing model (H-W ETS) demonstrated better predictive performance with lower RMSE and MAPE values than the seasonal ARIMA model (Table 3). The composition of the original sequence and the predictive results determined using the H-W ETS model are presented in Figure 3.

For this specific group, the authentic number of notified PTB cases was included in the 95% CI and nearly 91.7% included in the 80% CI, as observed in Table 4.

Table 3. Critical indices of the Holt-Winters exponential smoothing model and seasonal autoregressive integrated moving average model.

Model	RMSE[a]	MAPE[b]
H-W ETS[c]	44.59	6.13
Seasonal ARIMA[d]	47.17	6.15

[a] RMSE: root mean squared error.

[b] MAPE: mean absolute percentage error.

[c] H-W ETS: Holt-Winters exponential smoothing.

[d] ARIMA: autoregressive integrated moving average.

Figure 3. Composition of notified PTB cases in the elderly and prediction using the Holt-Winters exponential smoothing model. (A) Sequence composed of the seasonal effect, common trend, and random fluctuation (also called remainder). (B) Predicted number of the elderly PTB cases in 2021. ETS: exponential smoothing; PTB: pulmonary tuberculosis.

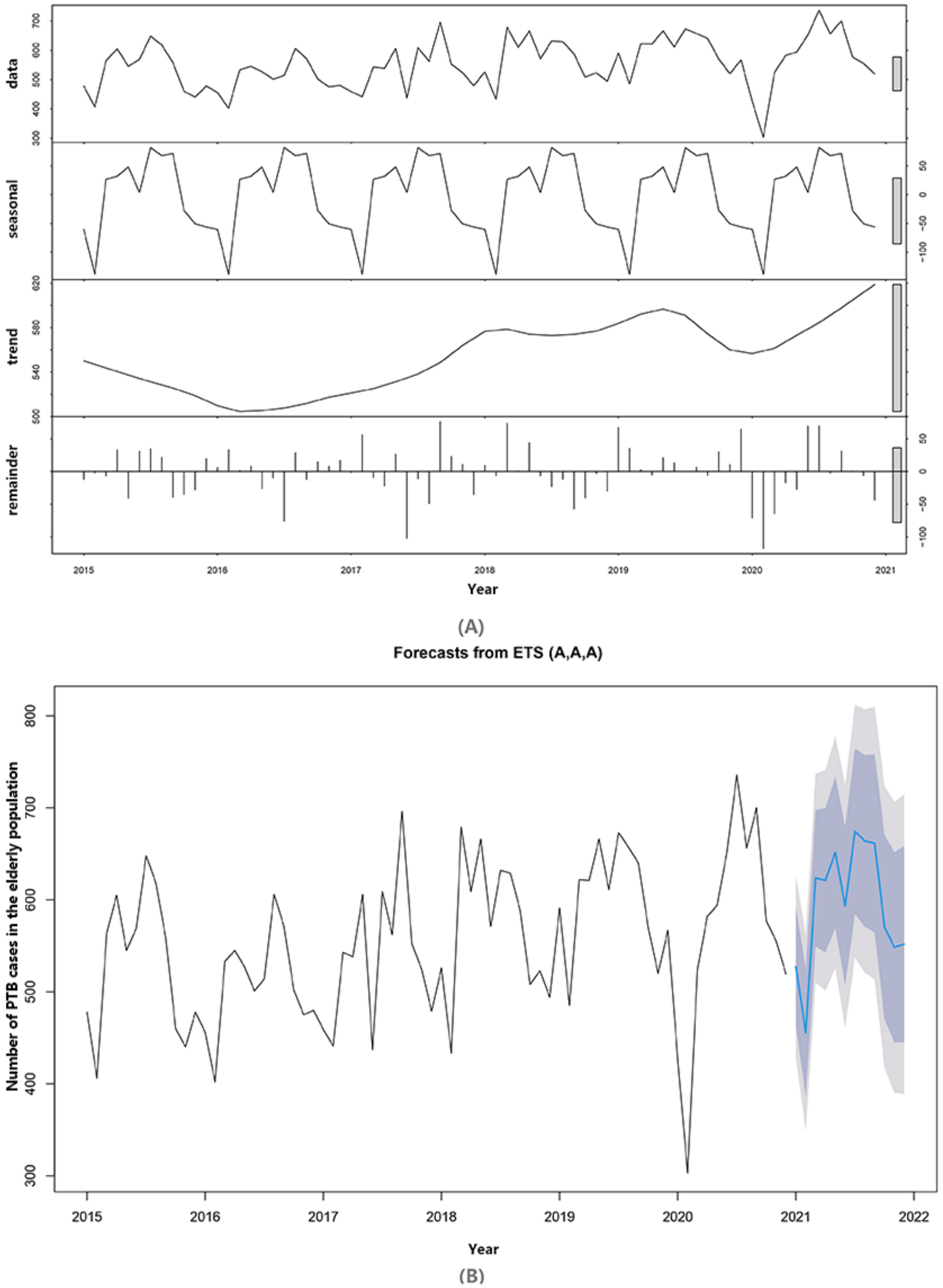


Table 4. Predicted values of notified pulmonary tuberculosis cases in the elderly using the exponential smoothing model and their actual values in 2021.

Month	Point estimation	80% CI	95% CI	Notified number
January	527	463-592	428-626	515
February	456	386-525	349-562	453
March	624	550-698	510-737	580
April	621	543-699	501-741	625
May	652	569-734	526-778	634
June	593	507-679	461-725	690
July	674	585-764	537-812	663
August	664	570-757	521-807	744
September	661	565-758	513-810	682
October	570	470-671	417-724	585
November	548	445-652	390-706	644
December	552	445-658	389-715	590

Discussion

Principal Findings

With the rising geriatric population, the increasing incidence and mortality of PTB emerged as a vital public health concern. Previous literature indicated that the increased TB incidence in older adults might be driven by the rising latent TB infections and weakened immunity [29]. Nevertheless, a comparatively developed province in Eastern China also faced a huge challenge in PTB control and prevention among the elderly. This study provides epidemiological evidence for scientific intervention and assessment of the PTB trend for this target population.

From the available epidemiological features of PTB in the elderly, the notified incidence demonstrated a significant downward trend with a reduction of 30%, whereas the bacteriologically positive incidence rate remained steady at approximately 60/100,000. The former was attributable to the deployment of the national directly observed treatment strategy combined with the internet surveillance system implemented in China since 2005, which could lower PTB transmission [10]. Moreover, increased collaboration among the designated hospitals, local Center for Disease Control and Prevention, and community health centers could improve the early identification of PTB cases, avoid delays in diagnoses causing further disease transmission, and facilitate prompt and adequate overall treatment. These possibilities may reduce PTB morbidity in the general population and target groups like the elderly. Furthermore, since the introduction of the rapid molecular testing systems for PTB such as GeneXpert MTB/RIF, the turnaround time for determining PTB has been shortened; rifampicin resistance is promptly determined, and the sensitivity and specificity are higher compared to those of classic smear microscopy, which could identify the number of etiologically positive cases despite the gradual decrease in the total incidence in this specific group [30].

From the geographical distribution, Quzhou and Jinhua Cities showed a higher PTB incidence in the elderly. In recent years, several investigations on PTB incidence have been performed in these regions. Considering the lack of PTB knowledge or the probable inability to seek care in this target group, community-based active screening for PTB in Quzhou had proved its efficacy in locally reducing the epidemic [31]. Thus, for areas with special funds available, several rounds of active PTB screening should be considered for successive identification of active PTB cases through different methods such as symptom screening, chest radiography, or molecular rapid diagnostic tests, used alone or in combination [32]. Meanwhile, for other areas, screening could be implemented in combination with the existing basic public health service project in China, which includes annual checkup for symptoms and chest radiography in the elderly [33]. Thus, strengthening the assessment through physical checkups is necessary to refer and trace elderly individuals with chest radiography abnormalities in a timely manner.

From our findings, the sex ratio in the elderly was 2.66:1 (male:female), which was slightly higher than that in the general population in Zhejiang Province [12]. Although available evidence from various countries indicated a higher susceptibility of men to PTB, the differences in age might be attributed to sociocultural roles, behaviors, and changes in the immune function among the sexes [34,35]. More PTB cases in the elderly were clustered in the age group of 65-70 years. As these people are typically free, they can form more clusters of infection. Therefore, determining active PTB cases particularly in the early stages of the disease among the elderly may be crucial for interventions and reversal of this epidemic, and additional health policies should be considered, such as inpatient isolation treatment with full reimbursement through medical insurance to reduce transmission and improve adherence to PTB treatment.

One study in Germany has reported significantly lower drug resistance rates in the elderly than the younger TB group for all TB drugs (6.5% vs 13.9%) and multidrug resistance (0.6% vs 3.1%) in 2011 [36]. Our results revealed that the drug resistance rate in the elderly was 5.86% in Zhejiang Province. Considering the reactivation of previous latent infection, the drug resistance rate in the elderly might be lower than that in the general group [4,35].

A delay in PTB diagnoses was identified, which implies the lack of knowledge on PTB's signs and symptoms in this specific group, thereby leading to a potential delay in seeking treatment. Moreover, acquiring sufficient diagnostic and treatment services for the geriatric population may be inconvenient. Therefore, succinct health education for PTB should be advocated in the elderly communities and increasing health promotions for this target group should be considered. Furthermore, a green channel in outpatient services should be provided to the elderly population to offer them a more convenient medical treatment experience.

Further, a prolonged treatment course (>9 months) was common in the elderly. Existing studies have revealed that older patients were more likely to experience drug-induced adverse reactions, such as hepatotoxicity and acute kidney injury during the treatment duration [37-39]. Furthermore, several complications and comorbidities including diabetes mellitus were prevalent in the aging group [40]. These factors might lead to the extension of standardized short-course chemotherapy.

Here, in predicting the PTB epidemic in the elderly based on the LRM, the goal of reducing PTB morbidity to 50% might be achieved by the middle of 2024. The ongoing efforts and implementations in Zhejiang Province may be effective in controlling PTB prevalence in the aging population. Due to limited data on PTB deaths and a significant decline in PTB-related deaths in 2020, we did not create a trend forecast. The significant decline might be attributed to the prevalence of the ongoing COVID-19 pandemic, which has influenced lifestyle habits such as wearing masks regularly. It would also be highly influential in preventing respiratory infectious diseases like PTB. Meanwhile, we used the optimized H-W ETS model to predict the PTB number in 2021 among the elderly, suggesting a comparatively better effect for identifying further trends. This result also implied that the H-W ETS model could provide a more elaborate atlas and consequently offer a basis for further policy developments for this specific group.

However, this study had some limitations. First, the PTB epidemiology between the general population and the elderly was not compared, which might influence other findings for these specific groups. Second, the predictive model did not include other details such as health policies, influence of the COVID-19 pandemic, and preventive methods.

Conclusions

With the global aging population and inevitable challenge of active PTB cases in the elderly, more comprehensive and effective interventions such as active PTB screening combined with physical checkup and succinct health education should be implemented in this specific group. Meanwhile, a more systematic assessment of the PTB epidemic trend in the elderly should be considered to incorporate more predictive factors.

Acknowledgments

We appreciated the local centers for disease control and prevention, community health care centers, and TB-designated hospitals in Zhejiang Province to implement TB reporting and notification during daily work. This study was supported by the National-Zhejiang Health Commission's Major S&T Project (grant WKJ-ZJ-2118) and Zhejiang Provincial Medical and Health Projects (grants 2021KY618 and 2020KY520).

Authors' Contributions

KL wrote the original draft of the manuscript and designed the methodology and visualization. ZX supervised the research and was involved in designing the methodology. BX was involved in visualization, data processing using the software, and conceptualization. SC curated the data and wrote and reviewed the manuscript. YZ supervised the research, along with being involved in reviewing and editing the manuscript, and data processing using the software. WW was involved in validation, formal analysis, and supervision. QW performed data curation, data collection, and validation. GC was involved in conceptualization and research supervision. BC acquired funding and participated in conceptualization.

Conflicts of Interest

None declared.

References

1. Gebretsadik D, Ahmed N, Kebede E, Mohammed M, Belete MA. Prevalence of tuberculosis by automated GeneXpert rifampicin assay and associated risk factors among presumptive pulmonary tuberculosis patients at Ataye District Hospital, North East Ethiopia. Infect Drug Resist 2020 May;13:1507-1516. [doi: 10.2147/idr.s248059]
2. Schrager LK, Vekemens J, Drager N, Lewinsohn DM, Olesen OF. The status of tuberculosis vaccine development. Lancet Infect Dis 2020 Mar;20(3):e28-e37. [doi: 10.1016/S1473-3099(19)30625-5] [Medline: 32014117]
3. Houben RMGJ, Dodd PJ. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. PLoS Med 2016 Oct;13(10):e1002152. [doi: 10.1371/journal.pmed.1002152]
4. Caraux-Paz P, Diamantis S, de Wazières B, Gallien S. Tuberculosis in the elderly. J Clin Med 2021 Dec;10(24):5888 [FREE Full text] [doi: 10.3390/jcm10245888] [Medline: 34945187]

5. Vynnycky E, Fine PE. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. Epidemiol Infect 1997 Oct;119(2):183-201. [doi: 10.1017/s0950268897007917] [Medline: 9363017]
6. World Health Organization. Global Tuberculosis Report 2021. Geneva: World Health Organization; 2021.
7. Haiyan L. New characteristic, trend and thinking of Chinese population-analysis based on data reported in seventh population census bulletin in 2020. Journal of Qujing Normal University 2021 Jan;40(4):97-103.
8. Yew WW, Yoshiyama T, Leung CC, Chan DP. Epidemiological, clinical and mechanistic perspectives of tuberculosis in older people. Respirology 2018 Jun;23(6):567-575. [doi: 10.1111/resp.13303] [Medline: 29607596]
9. Stead WW. Tuberculosis among elderly persons, as observed among nursing home residents. Int J Tuberc Lung Dis 1998 Sep;2(9 Suppl 1):S64-S70. [Medline: 9755968]
10. Liu S, Bi Y, Liu Y. Modeling and dynamic analysis of tuberculosis in mainland China from 1998 to 2017: the effect of DOTS strategy and further control. Theor Biol Med Model 2020 May;17(1):6 [FREE Full text] [doi: 10.1186/s12976-020-00124-9] [Medline: 32362279]
11. Zhu L, Yang Y, Guan H, Cheng S, Jin Y, Tan W, et al. Trends in drug-resistant tuberculosis after the implementation of the DOTS strategy in Shenzhen, China, 2000-2013. Int J Tuberc Lung Dis 2017 Jul 01;21(7):759-765. [doi: 10.5588/ijtld.16.0759] [Medline: 28633700]
12. Liu K, Li T, Vongpradith A, Wang F, Peng Y, Wang W, et al. Identification and prediction of tuberculosis in Eastern China: analyses from 10-year population-based notification data in Zhejiang Province, China. Sci Rep 2020 May;10(1):7425 [FREE Full text] [doi: 10.1038/s41598-020-64387-5] [Medline: 32367050]
13. Liu K, Chen S, Zhang Y, Li T, Xie B, Wang W, et al. Tuberculosis burden caused by migrant population in Eastern China: evidence from notification records in Zhejiang Province during 2013-2017. BMC Infect Dis 2022 Jan;22(1):109 [FREE Full text] [doi: 10.1186/s12879-022-07071-5] [Medline: 35100983]
14. Zhejiang's population aged 65 and above accounted for 9.26 million, accounting for 14.2%. Sina Financial Headlines. URL: https://cj.sina.com.cn/articles/view/2286092114/8842ff5204001rji1 [accessed 2022-07-14]
15. Li T, Du X, Shewade HD, Soe KT, Zhang H. What happens to migrant tuberculosis patients who are transferred out using a web-based system in China? PLoS One 2018 Nov;13(11):e0206580 [FREE Full text] [doi: 10.1371/journal.pone.0206580] [Medline: 30408131]
16. Huang F, Cheng S, Du X, Chen W, Scano F, Falzon D, et al. Electronic recording and reporting system for tuberculosis in China: experience and opportunities. J Am Med Inform Assoc 2014 Sep;21(5):938-941 [FREE Full text] [doi: 10.1136/amiajnl-2013-002001] [Medline: 24326537]
17. Jiang H, Liu M, Zhang Y, Yin J, Li Z, Zhu C, et al. Changes in incidence and epidemiological characteristics of pulmonary tuberculosis in Mainland China, 2005-2016. JAMA Netw Open 2021 Apr;4(4):e215302 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.5302] [Medline: 33835173]
18. National Health and Family Planning Commission. Classification of Tuberculosis. Beijing: National Health and Family Planning Commission; 2018.
19. Ministry of Public Health of China. China: Diagnostic Criteria for Pulmonary Tuberculosis. Beijing: People's Medical Publishing House; 2008.
20. Liu K, Huang S, Miao Z, Chen B, Jiang T, Cai G, et al. Identifying potential norovirus epidemics in China via internet surveillance. J Med Internet Res 2017 Aug;19(8):e282 [FREE Full text] [doi: 10.2196/jmir.7855] [Medline: 28790023]
21. Liu D, Li L. Application study of comprehensive forecasting model based on entropy weighting method on trend of PM2.5 concentration in Guangzhou, China. Int J Environ Res Public Health 2015 Jun;12(6):7085-7099 [FREE Full text] [doi: 10.3390/ijerph120607085] [Medline: 26110332]
22. Li Y, Han T, Wang J, Quan W, He D, Jiao R, et al. Application of ARIMA model for mid- and long-term forecasting of ozone concentration. Huan Jing Ke Xue 2021 Jul;42(7):3118-3126. [doi: 10.13227/j.hjkx.202011237] [Medline: 34212637]
23. Wang Y, Xu C, Zhang S, Wang Z, Yang L, Zhu Y, et al. Temporal trends analysis of tuberculosis morbidity in mainland China from 1997 to 2025 using a new SARIMA-NARNNX hybrid model. BMJ Open 2019 Jul;9(7):e024409 [FREE Full text] [doi: 10.1136/bmjopen-2018-024409] [Medline: 31371283]
24. Liu H, Li C, Shao Y, Zhang X, Zhai Z, Wang X, et al. Forecast of the trend in incidence of acute hemorrhagic conjunctivitis in China from 2011-2019 using the Seasonal Autoregressive Integrated Moving Average (SARIMA) and Exponential Smoothing (ETS) models. J Infect Public Health 2020 Feb;13(2):287-294 [FREE Full text] [doi: 10.1016/j.jiph.2019.12.008] [Medline: 31953020]
25. Hyndman RJ, Koehler AB, Snyder RD, Grose S. A state space framework for automatic forecasting using exponential smoothing methods. Int J Forecast 2002 Jul;18(3):439-454. [doi: 10.1016/s0169-2070(01)00110-8]
26. Wang H, Tian CW, Wang WM, Luo XM. Time-series analysis of tuberculosis from 2005 to 2017 in China. Epidemiol Infect 2018 Jun;146(8):935-939. [doi: 10.1017/S0950268818001115] [Medline: 29708082]
27. Anwar MY, Lewnard JA, Parikh S, Pitzer VE. Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence. Malar J 2016 Nov;15(1):566 [FREE Full text] [doi: 10.1186/s12936-016-1602-1] [Medline: 27876041]

28. Zeng Q, Li D, Huang G, Xia J, Wang X, Zhang Y, et al. Time series analysis of temporal trends in the pertussis incidence in Mainland China from 2005 to 2016. Sci Rep 2016 Aug;6:32367 [FREE Full text] [doi: 10.1038/srep32367] [Medline: 27577101]
29. Hochberg NS, Horsburgh CR. Prevention of tuberculosis in older adults in the United States: obstacles and opportunities. Clin Infect Dis 2013 May;56(9):1240-1247 [FREE Full text] [doi: 10.1093/cid/cit027] [Medline: 23362286]
30. Opota O, Mazza-Stalder J, Greub G, Jaton K. The rapid molecular test Xpert MTB/RIF ultra: towards improved tuberculosis diagnosis and rifampicin resistance detection. Clin Microbiol Infect 2019 Nov;25(11):1370-1376 [FREE Full text] [doi: 10.1016/j.cmi.2019.03.021] [Medline: 30928564]
31. Liu K, Peng Y, Zhou Q, Cheng J, Yu H, Tang L, et al. Assessment of active tuberculosis findings in the eastern area of China: a 3-year sequential screening study. Int J Infect Dis 2019 Nov;88:34-40 [FREE Full text] [doi: 10.1016/j.ijid.2019.07.029] [Medline: 31374346]
32. World Health Organization. WHO Consolidated Guidelines on Tuberculosis. Module 2: Screening- Systematic Screening for Tuberculosis Disease. Geneva: World Health Organization; 2021.
33. Tian CW, Zheng YM, Sun NL, Yao MJ, Bing PF, Liu L, et al. Current status of standardization of basic public health services and standardization strategy of service equalization in China. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Oct;41(10):1723-1730. [doi: 10.3760/cma.j.cn112338-20200513-00719] [Medline: 33297633]
34. Hertz D, Schneider B. Sex differences in tuberculosis. Semin Immunopathol 2019 Mar;41(2):225-237. [doi: 10.1007/s00281-018-0725-6] [Medline: 30361803]
35. Mori T, Leung CC. Tuberculosis in the global aging population. Infect Dis Clin North Am 2010 Sep;24(3):751-768. [doi: 10.1016/j.idc.2010.04.011] [Medline: 20674802]
36. Hauer B, Brodhun B, Altmann D, Fiebig L, Loddenkemper R, Haas W. Tuberculosis in the elderly in Germany. Eur Respir J 2011 Aug;38(2):467-470 [FREE Full text] [doi: 10.1183/09031936.00199910] [Medline: 21804163]
37. Schluger NW. Tuberculosis and nontuberculous mycobacterial infections in older adults. Clin Chest Med 2007 Dec;28(4):773-781 [FREE Full text] [doi: 10.1016/j.ccm.2007.07.004] [Medline: 17967293]
38. Chang C, Chen Y, Wu V, Shu C, Lee C, Wang J, et al. Acute kidney injury due to anti-tuberculosis drugs: a five-year experience in an aging population. BMC Infect Dis 2014 Jan;14:23 [FREE Full text] [doi: 10.1186/1471-2334-14-23] [Medline: 24410958]
39. Byng-Maddick R, Noursadeghi M. Does tuberculosis threaten our ageing populations? BMC Infect Dis 2016 Mar;16:119 [FREE Full text] [doi: 10.1186/s12879-016-1451-0] [Medline: 26968654]
40. Jeon CY, Murray MB. Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies. PLoS Med 2008 Jul;5(7):e152. [doi: 10.1371/journal.pmed.0050152]

Abbreviations

ARIMA: autoregressive integrated moving average
ETS: exponential smoothing
GCM: growth curve model
H-W ETS: Holt-Winters exponential smoothing
LRM: linear regression model
MAPE: mean absolute percentage error
PTB: pulmonary tuberculosis
RMSE: root mean squared error
TBIMS: TB information management system
WHO: World Health Organization

Edited by Y Khader; submitted 29.04.22; peer-reviewed by T Li, J Jiang; comments to author 11.05.22; revised version received 06.06.22; accepted 09.06.22; published 29.07.22.

Please cite as:
Liu K, Xie Z, Xie B, Chen S, Zhang Y, Wang W, Wu Q, Cai G, Chen B
Bridging the Gap in End Tuberculosis Targets in the Elderly Population in Eastern China: Observational Study From 2015 to 2020
JMIR Public Health Surveill 2022;8(7):e39142
URL: https://publichealth.jmir.org/2022/7/e39142
doi:10.2196/39142
PMID:35904857

Original Paper

Age- and Sex-Specific Association Between Vegetation Cover and Mental Health Disorders: Bayesian Spatial Study

Abu Yousuf Md Abdullah[1,2*], BSc, MSc; Jane Law[1,2*], BSc, MSc, PhD; Christopher M Perlman[2*], BSc, MSc, PhD; Zahid A Butt[2*], MBBS, MSc, PhD

[1]School of Planning, University of Waterloo, Waterloo, ON, Canada

[2]School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Abu Yousuf Md Abdullah, BSc, MSc
School of Planning
University of Waterloo
200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 1 226 978 2838
Email: aymabdul@uwaterloo.ca

Abstract

Background: Despite growing evidence that reduced vegetation cover could be a putative risk factor for mental health disorders, the age- and the sex-specific association between vegetation and mental health disorder cases in urban areas is poorly understood. However, with rapid urbanization across the globe, there is an urgent need to study this association and understand the potential impact of vegetation loss on the mental well-being of urban residents.

Objective: This study aims to analyze the spatial association between vegetation cover and the age- and sex-stratified mental health disorder cases in the neighborhoods of Toronto, Canada.

Methods: We used remote sensing to detect urban vegetation and Bayesian spatial hierarchical modeling to analyze the relationship between vegetation cover and mental health disorder cases. Specifically, an Enhanced Vegetation Index was used to detect urban vegetation, and Bayesian Poisson lognormal models were implemented to study the association between vegetation and mental health disorder cases of males and females in the 0-19, 20-44, 45-64, and ≥65 years age groups, after controlling for marginalization and unmeasured (latent) spatial and nonspatial covariates at the neighborhood level.

Results: The results suggest that even after adjusting for marginalization, there were significant age- and sex-specific effects of vegetation on the prevalence of mental health disorders in Toronto. Mental health disorders were negatively associated with the vegetation cover for males aged 0-19 years (−7.009; 95% CI −13.130 to −0.980) and for both males (−4.544; 95% CI −8.224 to −0.895) and females (−3.513; 95% CI −6.289 to −0.681) aged 20-44 years. However, for older adults in the 45-64 and ≥65 years age groups, only the marginalization covariates were significantly associated with mental health disorder cases. In addition, a substantial influence of the unmeasured (latent) and spatially structured covariates was detected in each model (relative contributions>0.7), suggesting that the variations in area-specific relative risk were mainly spatial in nature.

Conclusions: As significant and negative associations between vegetation and mental health disorder cases were found for young males and females, investments in urban greenery can help reduce the future burden of mental health disorders in Canada. The findings highlight the urgent need to understand the age-sex dynamics of the interaction between surrounding vegetation and urban dwellers and its subsequent impact on mental well-being.

(JMIR Public Health Surveill 2022;8(7):e34782) doi:10.2196/34782

KEYWORDS

mental health disorders; vegetation cover; age- and sex- specific association; Enhanced Vegetation Index; Bayesian; spatial; hierarchical modeling; marginalization; latent covariates

Introduction

The relationship between vegetation and mental health remains an issue of considerable interest. The worldwide increase in urbanization has caused some unique environmental problems [1], such as substantial loss of vegetation-covered areas [2,3], which might adversely affect the mental health of urban dwellers. However, as mental health disorders can be of different categories and types [4,5], it is quite challenging to understand how adversely vegetation loss could affect the mental health conditions of the general public. Nevertheless, with increasing evidence that urban residents are particularly susceptible to various mental health disorders [6-9], it has become urgent to understand how vegetation could affect individuals of different ages and sex groups in an urban population.

Previous studies reported that vegetation could positively affect individuals with various mental health disorders, such as affective and psychotic disorders [10,11]. For example, in forest therapies, where people with affective disorders were involved in recreational activities in the nearest suburban forest, positive outcomes were reported, such as a reduction in symptoms of the disorders [11]. In addition to its utility in therapeutic intervention, exposure to an adequately managed vegetation-covered area, where people can tangibly and regularly experience the surrounding greenness, may significantly affect the psychology of people [12,13]. Unfortunately, because of increased urbanization, which mainly promotes the growth of commercial and residential areas, little attention has been paid to managing the already declining vegetation-covered areas.

As we consider age and sex differences in the prevalence of mental health conditions [14-17], we need to consider age and sex differences in susceptibility to reduced vegetation cover. The findings from previous studies suggest that the surrounding residential vegetation or greenness could be positively associated with better mental health outcomes for males aged younger than 65 years [18] and for young to middle-aged males (aged 30-45 years) and older females (aged ≥45 years) [19]. However, contrasting results are also present, where no age- and sex-specific differences in the association between the mental health conditions and surrounding greenness could be found for any of the study participants [20]. Therefore, the findings of previous studies are inconclusive and differ considerably from one another, indicating a need for further research.

Social inequality (or marginalization) can also contribute to poor mental health outcomes in urban areas [13,21] and complicate our understanding of the relationship between reduced vegetation and mental health disorders. Along with vegetation loss, social inequality has been identified as one of the major problems of urbanization. For example, rapid urbanization in China was accompanied by the rising marginalization of disadvantaged social groups such as laid-off workers and rural migrants [22]. In addition, highly urbanized census metropolitan areas in Ontario, Canada, with higher levels of marginalization, reported higher levels of psychotic disorders [21]. Hence, proponents of the concept that there is no significant association between reduced vegetation and poor mental health outcomes argue that people living in parts of urban areas with reduced vegetation cover could already be socioeconomically disadvantaged. Hence, they argue that the level of marginalization in urban areas could be a more important determinant of mental well-being than reduced greenery in these areas. Therefore, further understanding is warranted on whether vegetation's age- and sex-specific effects on mental health could be significant after accounting for the influence of marginalization.

It is also important to note that the inconclusive findings of previous studies could be the outcomes of limitations in conventional analytical techniques. In this regard, special attention is required when selecting an appropriate vegetation measure to analyze the association between vegetation and common mental health disorders [1]. The vegetation measures that can accurately capture people's exposure to surrounding greenness and adjust for environmental perturbations in urban areas are found to be best suited for population-based mental health studies [1,13]. In addition, spatial dependencies in the observed cases because of the influence of cases from neighboring regions need to be addressed to obtain unbiased estimates of the uncertainties and magnitude of the association between vegetation and mental health indicators [23,24]. Thus, a spatial modeling approach is required to adjust for any spurious spatial effects, such as spatial autocorrelation, and to account for any unmeasured spatial covariates that could influence the distribution of the observed cases [25].

Despite the modeling complexities and challenges, there is a need to understand the dynamics of the age- and sex-specific association between vegetation cover and mental health disorders. The importance of such a study stems from several critical issues. First, the demographic conditions in an area change over time, thus changing the mental health burden and the country's mental health care costs [26]. Therefore, to understand the present and future mental health burdens, it is important to understand how a putative risk factor, such as reduced vegetation cover, can affect males and females in different age groups. Second, it is essential to understand which age and sex group could be particularly vulnerable to reduced vegetation cover for devising targeted intervention strategies [13,18,27]. Mental health promotion activities may differ based on the age and sex of the vulnerable groups [28]; therefore, it is imperative to identify the vulnerable groups to devise an effective intervention strategy. Third, as most studies attempt to understand the effect of people's exposure to vegetation cover at the population level [1,13], the age- and sex-specific effects on mental health could be diluted because of grouping mental health disorder data into a single age and sex group. This may cause any adverse impact of reduced vegetation on different age and sex groups to remain undetected, thus silently creating a mental health disorder epidemic.

This detailed study extends a previous study on Toronto neighborhoods, which demonstrated the positive effects of vegetation cover on psychotic and non-psychotic disorders of both sexes (males and females) in the >0 year age group [1]. However, as the previous study mainly focused on the methodological aspects of selecting an appropriate vegetation measure to study the relationship with mental health disorders, it could not explore the relationship between vegetation cover and age- and sex-specific mental health disorder cases. Therefore, this cross-sectional study attempted to fulfill the existing research gap and aimed to understand the effect of vegetation cover on the mental well-being of males and females in 4 different age groups in Toronto, Canada. Specifically, this study aimed to examine the spatial association between vegetation cover and the age- and the sex-specific mental health disorder cases in Toronto neighborhoods after controlling for marginalization and unmeasured spatial and nonspatial covariates.

Methods

Study Area

The City of Toronto was selected as the study area because of its high urbanization rate, which led to a substantial increase in the built environment and loss of vegetation-covered areas [29]. All analyses in this study were conducted at the neighborhood level, which are geographic units created for planning and service delivery purposes by the Social Development and Administration Division of the City of Toronto. These units were constructed by aggregating the Statistics Canada census tracts into meaningful spatial units [1,30]. In 2016, there were 140 neighborhoods in Toronto with a population of 2,731,571 [31].

Mental Health Disorder Data

Mental health disorder data were collected from the Ontario Community Health Profiles Partnership database [32]. The data, dated from April 1, 2015, to March 31, 2016 (fiscal year 2015), was generated from the study, *Enrollment, Access, Continuity, and Mental Health Gaps in Care (Institute for Clinical Evaluative Sciences Project No. 2018 0900 992 000)*. This study was supported by the Institute for Clinical Evaluative Sciences, a nonprofit corporation funded by the Ontario Ministry of Health and Long-Term Care. Further details on the project, the primary data sources, and the inclusion and exclusion criteria for generating the data set can be found in the project report [6].

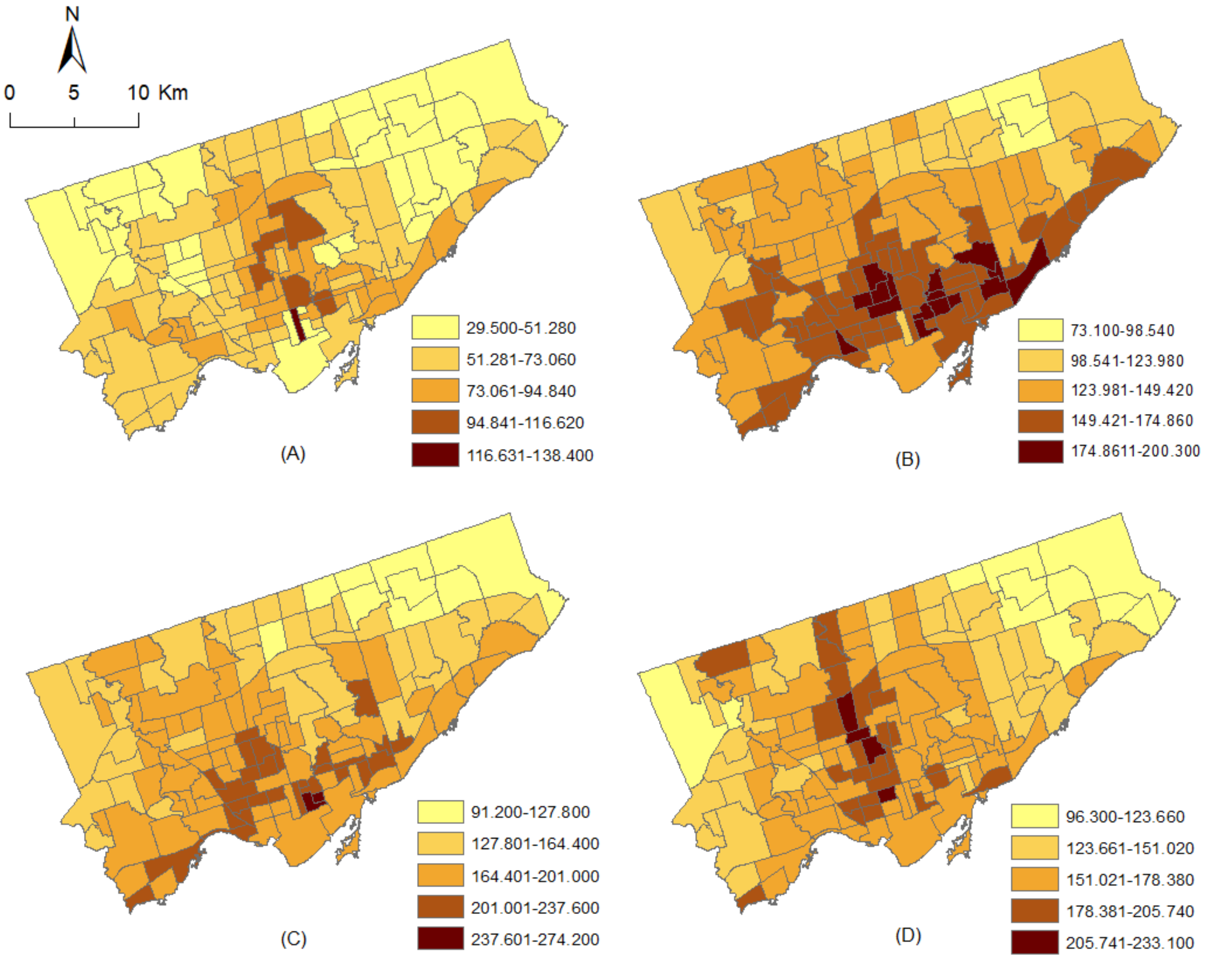
The retrieved data set comprised 4 major mental health disorders (Table 1), each created from various subcategories defined by the Ontario Health Insurance Plan codes. The combined mental health disorder category in the data set was created by adding the counts of these 4 major categories. The age- and sex-stratified observed counts for the combined mental health disorder variable were used in this study. Further details on the 4 major categories and their subcategories are provided in Table 1. This study used the combined mental health disorder data for 4 age groups: 0-19, 20-44, 45-64, and ≥65 years. Figure 1 shows the crude rates of the combined mental health disorders for males and females in the 4 age groups and the spatial distribution of the disorders in the study area.

Figure 1 shows that the mental health disorders in Toronto exhibit considerable spatial variations. For example, the rate of mental health disorders among individuals aged 20-44 years is the highest in Toronto's south-central parts. Similarly, for the remaining age groups, the rates were generally higher in the southern parts than in the other areas. Therefore, to measure the spatial autocorrelation in the retrieved data, global Moran's I tests were performed using the GeoDa software [33]. The results suggested moderate autocorrelation (Moran's I=0.491; P=.001) in the data for the age group 0-19 years and high autocorrelation (Moran's I=0.702, P=.001; Moran's I=0.717, P=.001; and Moran's I=0.696, P=.001, respectively) for the 20-44, 45-64, and ≥65 years age groups. Hence, significant spatial autocorrelation in the data necessitated selecting a spatial modeling technique for this study.

Table 1. The 4 major categories and subcategories of mental health disorders with Ontario Health Insurance Plan (OHIP) codes.

Major category and subcategory	OHIP code (subcategory)
Psychotic disorders	
Schizophrenia	295
Manic-depressive psychoses and involutional melancholia	296
Other paranoid states	297
Other psychoses	298
Non-psychotic disorders	
Hysteria, reactive depression, neurasthenia, obsessive-compulsive neurosis, and anxiety neurosis	300
Personality disorders	301
Sexual deviations	302
Psychosomatic illness	306
Adjustment reaction	309
Depressive disorder	311
Substance-use disorders	
Alcoholism	303
Drug dependence	304
Family, social, and occupational issues	
Economic problems	897
Marital difficulties	898
Parent-child problems	899
Problems with aged parents or in-laws	900
Family disruption or divorce	901
Education problems	902
Social maladjustment	904
Occupational problems	905
Legal problems	906
Other problems of social adjustment	909
Combined mental health disorders	
Psychotic, non-psychotic, substance-use and family, social and occupational issues related disorders	All codes listed above

Figure 1. The crude rates (per 1000 population) of the combined mental health disorders for both sexes in the (A) 0-19, (B) 20-44, (C) 45-64, and (D) ≥65 years age groups in the Toronto neighborhoods.



Vegetation Cover Data

The vegetation cover, specifically the density and quality (biomass vigor) of vegetation, was estimated using remote sensing techniques. For this purpose, 3 satellite images from the Landsat Operational Land Imager and Thermal Infrared Sensor or Landsat 8 were obtained from the United States Geological Survey EarthExplorer data repository [34]. The downloaded images had a spatial resolution of 30 m; contained an average cloud cover of 2.67%; and were captured between May 20, 2016, and June 14, 2016. Thus, the vegetation cover during the spring-summer seasons in 2016 was measured to be consistent with the data period of the mental health disorder data set. After radiometric and atmospheric correction, the images were mosaiced and clipped to the extent of the study area.

In this study, the Enhanced Vegetation Index (EVI) was calculated to estimate the vegetation cover in the Toronto area [35]. The EVI was calculated from the processed satellite image and using the *Raster Calculator* function in ArcMap 10.7 software [36] from Environmental Systems Research Institute. Vegetation indexes, such as EVI, are calculated by measuring the relative abundance of different electromagnetic waves (commonly known as bands) reflected from the surface of the vegetation to the sensors in the satellite. The details of the index and the formula used to calculate EVI can be found in Multimedia Appendix 1 [1,35].

The Bayesian Spatial Hierarchical Modeling

The age- and sex-stratified associations between vegetation and mental health disorders were analyzed using the Bayesian spatial hierarchical modeling technique. In the Bayesian spatial hierarchical models, the observed counts, O_i, of the combined mental health disorder in each neighborhood i (where i=1,2,…140) were assumed to follow a Poisson distribution *(*$O_i \sim Poisson(\lambda_i)$*)*. Here, λ_i denotes the expected number of mental health disorder cases in the neighborhood i.

Furthermore, the observed count of the combined mental health disorders in a neighborhood (i) was assumed to be a function of the unknown area-specific relative risk of the disorders, r_i, and the expected count, E_i. The expected counts of the combined mental health disorders were calculated separately for males and females in each age group. Specifically, the expected counts were calculated by multiplying the age- and sex-specific rates of the disorders with the residential population of each neighborhood [25]. The age- and sex-specific rates of the disorder were calculated by dividing the total number of mental health disorder cases in each age and sex group by the total number of individuals (population) in that age and sex group of the neighborhood. Thus, through the integration of E_i, the population at risk was accounted for in the Bayesian spatial hierarchical models.

In addition, the environmental and socioeconomic risk factors, such as reduced vegetation cover and marginalization, in a neighborhood could be assumed to govern the risk of developing different mental health disorders [1,8,11,25,37,38]. Therefore, the unknown area-specific relative risk in a neighborhood was modeled as a function of the vegetation cover and the marginalization factors. For this study, we mainly focused on modeling the spatial association between the age- and sex-stratified mental health disorder cases and the amount of vegetation cover in each neighborhood (X_{1i}). However, to adjust for social inequality or marginalization, we also added the 4 domains (or variables) of marginalization in the Bayesian spatial hierarchical models: material deprivation (X_{2i}), ethnic concentration (X_{3i}), residential instability (X_{4i}), and dependency (X_{5i}) covariates. All the domains were retrieved from the Ontario Marginalization Index (OMI) [39], and the weighted average score for each variable was used as a measure of the covariate. Before including the 4 OMI variables in the models, we carried out Pearson correlation and multicollinearity tests [40] and confirmed the absence of any multicollinearity among the variables. A detailed explanation for selecting these 4 socioeconomic covariates and the results of the multicollinearity tests can be found in our previous study [1]. The 4 OMI variables are described more elaborately elsewhere [39].

Hence, the observed counts of the combined mental health disorders in any neighborhood can be modeled as:

However, because of nonspatial variations in individual-level processes or variations in individual-level risks of developing mental health disorders, overdispersion could be considered a common problem for count data [25,41]. In addition, spatial autocorrelation could be present in spatial data because of unmeasured (latent) spatial covariates influencing the distribution of the observed variable (here, the mental health disorder cases). As previously mentioned in the Mental Health Disorder Data section, we found moderate to high spatial autocorrelation in our mental health disorder data set. Therefore, to adjust for the nonspatial and spatially structured unknown covariates, we integrated 2 Gaussian random-effects terms, u_i and s_i, using equation 1 to obtain the final model equation [25]:

Here, the terms u_i and s_i represent all the covariates (nonspatial and spatial, respectively) that we could not measure but could have influenced the distribution of the mental health disorders in the study area.

We also analyzed whether the nonspatial or spatially structured unmeasured covariates had significantly influenced the Bayesian models. Their relative contributions were modeled using the posterior distribution of the quantity ψ and could be defined as follows [42]:

where $SD_{spatial}$ is the empirical marginal SD of s_i and $SD_{nonspatial}$ is the empirical marginal standard deviation of u_i.

On the basis of equation 3, as ψ→1, the variations in the area-specific relative risk of mental health disorders would be mainly spatial, and as ψ→0, the variations would be nonspatial in nature.

Once the required models were run, we analyzed the relative risk of developing mental health disorders because of reduced vegetation cover in each neighborhood. The relative risk of mental health disorders for the 4 age groups of males and females in the neighborhood *i* can be expressed as follows:

Ethics Approval

No ethics approval was required for this work. All data used for the analysis and reporting of the results are in concordance to the free usage and sharing policy guidelines provided by the data providers.

Results

The Bayesian Spatial Hierarchical Modeling: The Age- and Sex-Stratified Association Between EVI and Mental Health Disorder Cases

The results of the age- and sex-stratified spatial association between vegetation cover (EVI) and mental health disorder cases at the neighborhood level are tabulated in Table 2. The results suggest that the mental health of males in the 0-19 and 20-44 years age groups is significantly affected by the surrounding vegetation cover. By contrast, the association between urban greenery (represented by EVI) and mental health was significant only among females aged 20-44 years. For males in the 0-19 and 20-44 years age groups, the magnitude of the association between vegetation and mental health disorders was $\beta_1=-7.009$ (95% CI −13.130 to −0.980) and $\beta_1=-4.544$ (95% CI −8.224 to −0.895), respectively. However, for females in the 20-44 years age group, this magnitude was $\beta_1=-3.513$ (95% CI −6.289 to −0.681). Hence, the results indicate that increased vegetation cover could potentially improve the mental health conditions of both males and females in their youth. In particular, males from the 0-19 years age group could most benefit from the presence of vegetation.

The ψ values for all the models are >0.70 and are statistically significant. As all the ψ values are approaching 1 (ψ→1), the variations in the area-specific relative risk of mental health disorders for both sexes and all age groups could be predominantly influenced by the unmeasured spatial covariates.

Table 2. The results of analyzing the spatial association between Enhanced Vegetation Index (EVI) and combined mental health disorders of males and females in 4 different age groups (years).

Summaries of the posterior means	0-19 years age group	20-44 years age group	45-64 years age group	≥65 years age group
Men				
β_0 (95% CI)	*0.550 (0.232 to 0.873)*[a]	*0.274 (0.078 to 0.472)*	0.138 (−0.056 to 0.333)	*0.221 (0.004 to 0.438)*
Vegetation cover/EVI β_1 (95% CI)	*−7.009 (−13.130 to −0.980)*	*−4.544 (−8.224 to −0.895)*	−2.920 (−6.585 to 0.721)	−3.841 (−7.913 to 0.166)
Material deprivation, β_2 (95% CI)	−0.035 (−0.097 to 0.027)	*0.095 (0.056 to 0.134)*	*0.132 (0.093 to 0.169)*	*0.053 (0.012 to 0.093)*
Ethnic concentration, β_3 (95% CI)	*−0.192 (−0.265 to −0.121)*	*−0.148 (−0.194 to −0.102)*	*−0.156 (−0.201 to −0.110)*	*−0.079 (−0.127 to −0.030)*
Residential instability, β_4 (95% CI)	−0.013 (−0.073 to 0.046)	*0.080 (0.040 to 0.119)*	*0.136 (0.098 to 0.174)*	*0.075 (0.035 to 0.116)*
Dependency, β_5 (95% CI)	*−0.131 (−0.228 to −0.034)*	*−0.074 (−0.137 to −0.012)*	*−0.109 (−0.174 to −0.044)*	−0.022 (−0.086 to 0.043)
ψ^b (95% CI)	*0.716 (0.520 to 0.899)*	*0.795 (0.647 to 0.904)*	*0.742 (0.558 to 0.886)*	*0.817 (0.688 to 0.908)*
Women				
β_0 (95% CI)	*0.325 (0.001 to 0.645)*	*0.292 (0.141 to 0.439)*	0.137 (−0.017 to 0.292)	*0.214 (0.015 to 0.413)*
Vegetation cover/EVI, β_1 (95% CI)	−1.624 (−7.632 to 4.448)	*−3.513 (−6.289 to −0.681)*	−1.376 (−4.274 to 1.533)	−2.934 (−6.660 to 0.784)
Material deprivation, β_2 (95% CI)	−0.029 (−0.089 to 0.032)	*0.096 (0.066 to 0.125)*	*0.093 (0.063 to 0.123)*	0.023 (−0.016 to 0.061)
Ethnic concentration, β_3 (95% CI)	*−0.262 (−0.334 to −0.192)*	*−0.186 (−0.221 to −0.151)*	*−0.144 (−0.180 to −0.108)*	*−0.104 (−0.150 to −0.058)*
Residential instability, β_4 (95% CI)	0.013 (−0.046 to 0.072)	*0.048 (0.019 to 0.078)*	*0.071 (0.041 to 0.101)*	*0.068 (0.030 to 0.107)*
Dependency, β_5 (95% CI)	−0.069 (−0.165 to 0.026)	*−0.068 (−0.115 to −0.021)*	*−0.064 (−0.112 to −0.016)*	0.034 (−0.025 to 0.092)
ψ^b (95% CI)	*0.738 (0.544 to 0.906)*	*0.799 (0.675 to 0.892)*	*0.779 (0.632 to 0.887)*	*0.837 (0.729 to 0.914)*

[a] The statistically significant values at a 95% CI have been italicized.

[b] The relative contribution of the spatially structured and nonstructured random effect terms.

The Variations in Relative Risks of Mental Health Disorders for Men and Women in Different Age Groups

The relative risk values of the combined mental health disorders for males and females in the 0-19, 20-44, 45-64, and ≥65 years age groups in the neighborhoods of Toronto are compared in Figure 2. The relative risks were calculated using equation 4. The results suggest similarities in relative risk values for males and females aged 0-19 and ≥65 years. However, females had higher relative risks for the 20-44 and 45-64 years age groups than males in the same age groups. All median values of relative risks were >1, implying an increased risk of developing various mental health disorders. Although males in the 20-44 and ≥65 years age groups were particularly vulnerable, females in the 20-44 and 45-64 years age groups experienced the highest risk.

The combined mental health disorder risks because of varying vegetation cover in each neighborhood after adjusting for the risks from marginalization and unmeasured covariates are shown in Figures 3 and 4. The relative risk maps were similar for males and females in all age groups, suggesting an identical risk distribution for both sexes. However, a comparison of the maps of different age groups for each sex revealed interesting distribution patterns. For example, Figures 3A and 4A for males and females in the 0-19 years age group, respectively, showed that the neighborhoods with high risk (^{r}i>1) were located in the central Toronto area. By contrast, Figures 3B and 4B show that high-risk areas for the 20-44 years age group for both sexes were located in the southern parts of the study area. Although high-risk neighborhoods in the Figures 3C and 4C for the 45-64 years age group of both males and females were localized in the southern parts, Figure 4C illustrates that the risk for females was distributed over a larger area compared with the risk for males. A good portion of the central western part of Toronto was at an elevated risk of mental health disorders for females in the 45-64 years age group. In contrast, high-risk neighborhoods in Figures 3D and 4D for males and females aged ≥65 years were found in the south to northward directions in the study area.

Figure 2. The posterior mean relative risks of combined mental health disorders for males and females in the 4 age groups (0-19, 20-44, 45-64, and ≥65 years).

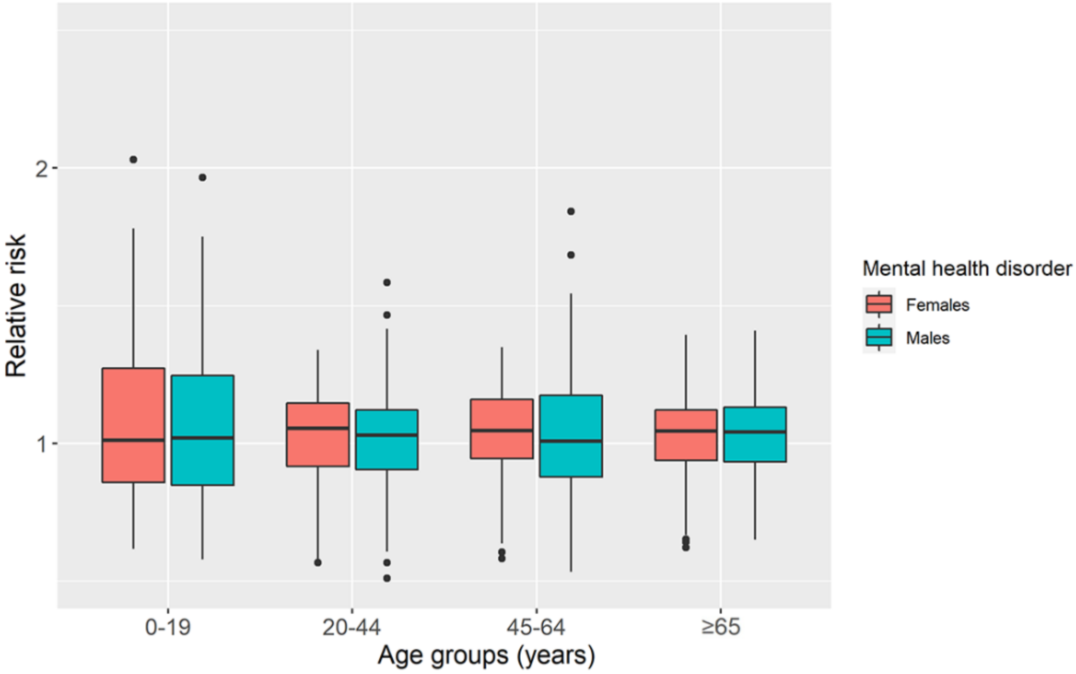


Figure 3. The posterior mean of relative risk (r_i) for males in the (A) 0-19, (B) 20-44, (C) 45-64, and (D) ≥65 years age groups.

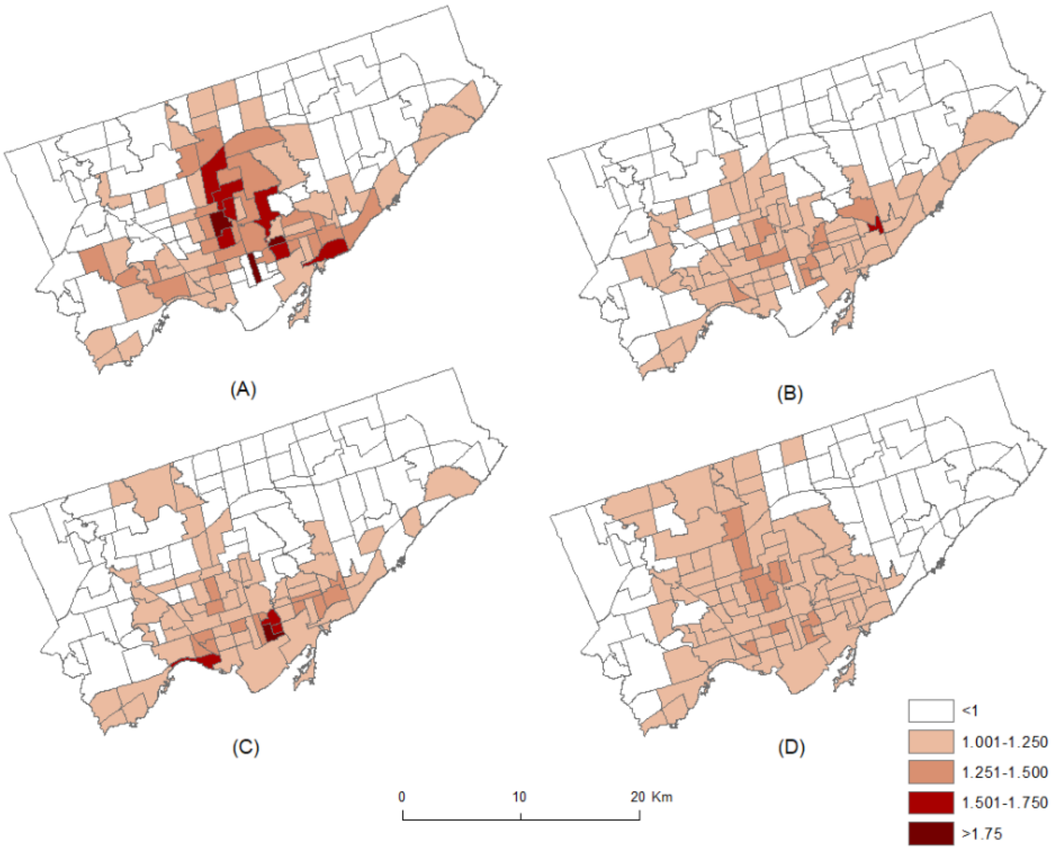
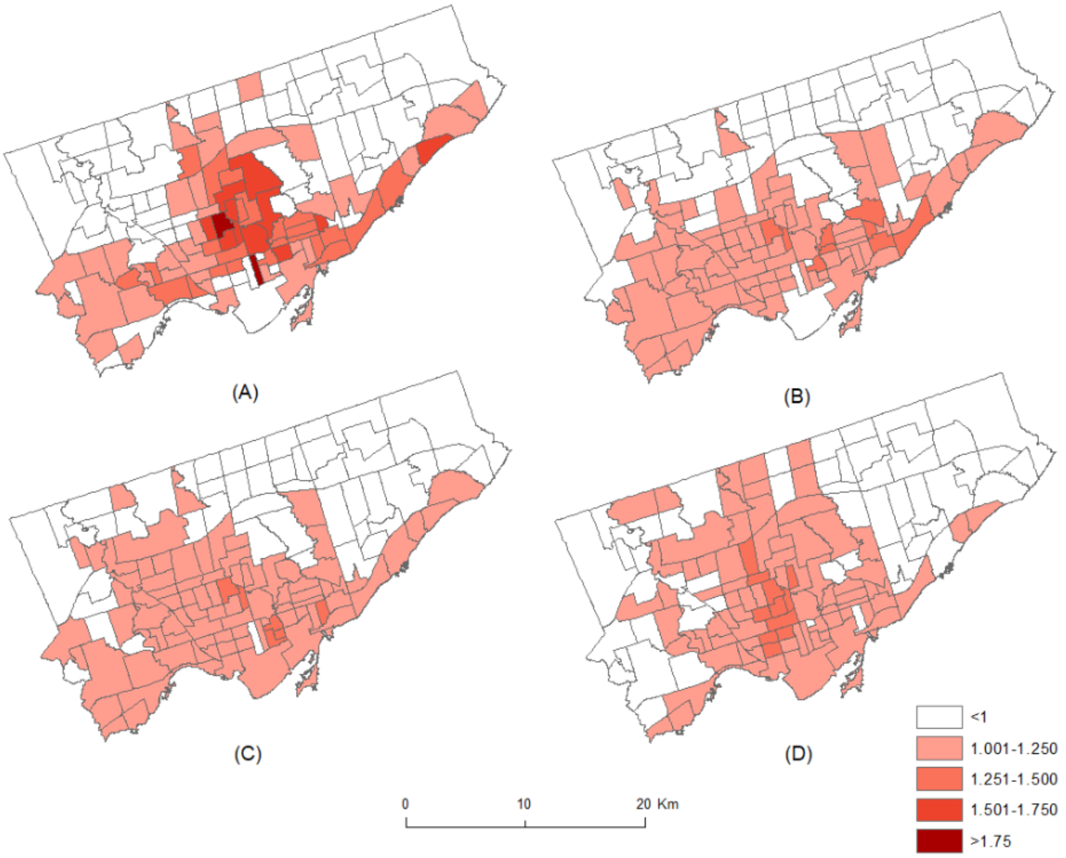


Figure 4. The posterior mean of relative risk (r_i) for females in the (A) 0-19, (B) 20-44, (C) 45-64, and (D) ≥65 years age groups.



Discussion

Principal Findings

Through the application of Bayesian spatial hierarchical modeling and remote sensing techniques, this study found that reduced vegetation cover was significantly associated with poor mental health outcomes in young males and females in urban Toronto. The relative risk maps identified hotspots of mental health disorder cases, which could be targeted for future interventions. Although the mental health disorder cases for young males (aged 0-19 and 20-44 years) showed a higher magnitude of association with reduced vegetation cover than females, the absolute values of the relative risk of mental health disorder cases were higher for adult females (aged 20-44 and 45-64 years). These findings reflect the complexities of studying the age- and sex-specific association between vegetation cover and mental health disorders. The results also highlight the importance of further research to understand the exact dynamics between the surrounding greenery and mental health disorders in urban areas.

The results suggest that vegetation was associated with mental health disorder cases in Toronto for females in the 20-44 years age group and males in the 0-19 and 20-44 years age group. However, this study could not detect any significant association between vegetation and mental health disorders for the middle-aged (45-64 years) and older adults (≥65 years). These findings are concordant with the results from previous studies on people from similar age groups but in different study settings. For example, higher surrounding greenness was found to be associated with improved neurobehavioral health in children [43]. In addition, the research suggested that improved mental health conditions such as improved attention, reduced symptoms of aggressive behavior, and also, reduced symptoms of attention deficit hyperactivity disorder were evident in children aged 6-18 years because of an increase in surrounding greenness. The associations were significantly prominent for the externalizing behaviors, especially when the greenness was within 1600 m of the children's residence. Similarly, another review study concluded that neighborhood greenness was beneficial for children's cognitive function and mental health [44].

As evidenced in our results, exposure to urban greenery could have a more prominent effect on young adults aged 20-44 years than older adults. This differential effect could be explained by reflecting on how people in different age groups engage in social cohesion and health beneficial activities. Young people are relatively more socially engaging, physically active, and more likely to maintain health-benefiting behaviors than older adults [45]. Consequently, owing to more outdoor activities than older adults, young people could be exposed to greater levels of urban greenery. Furthermore, previous research reported that the utility of greenspace lies not only as a site of physical activities but also in its restorative effect on mental health [13]. Hence, young adults may actively seek vegetation-covered areas and other forms of urban greenery because of their utility as sites for physical exercise and destressing [27,46]. Moreover, various types of neighborhood vegetation, such as trees, can lead to greater outdoor space use and increased social engagements among youth [47-49]. These social engagements, in turn, have the potential to promote social cohesion and mental well-being [13,27].

A crucial benefit of using EVI in this research to model the association between vegetation cover and the age- and sex-stratified mental health disorder cases could be discussed using findings from previous studies. Several prominent studies have discussed the importance of using vegetation measures that can capture both the quality and quantity of the surrounding greenery [1,50,51]. For example, a study conducted in Western Australia strongly emphasized the quality rather than the quantity (or number) of green spaces in reducing the psychological distress of people [50]. In concordance with these studies, vegetation indexes were recommended for use in mental health studies instead of the conventional areal-based measures of vegetation [1]. In this regard, sophisticated indexes such as EVI could be particularly suitable because they are highly sensitive to the leaf structures and photosynthetic activities in the vegetation patches. In addition, vegetation indexes such as EVI can also adjust for various disturbances such as canopy background noise and atmospheric resistance, which can greatly improve the detection of different types of vegetation in an urban setting.

The findings from this study provided critical evidence that reduced vegetation cover can impair the mental health of young people. Hence, this study demonstrated that based on the demographics (age and sex structures) in an area, investment in vegetation could help reduce the burden of mental health disorders. In addition, the results suggest a need to assess why, unlike the males in the 0-19 years age group, the mental health conditions of children and adolescent females are not affected by reduced vegetation cover. Previous studies have shown that the conservative approaches of societies toward females might cause young females to spend less time outside their homes [52,53], leading to less exposure to surrounding vegetation and thus possibly reducing the influence of vegetation on their mental health. Future studies can attempt to understand the relative exposure of young females to vegetation-covered surroundings by measuring their outdoor activities. Finally, although no association between mental health disorders in older adults (aged 45-64 and ≥65 years) and vegetation cover was found in this study, urban residents from these age groups should still be encouraged to spend time in vegetation-covered surroundings, as the present evidence suggests that such activity could lead to improved mental health outcomes [1,13,54].

Our study is a spatial cross-sectional study that demonstrates how the vegetation cover in an urban landscape could be potentially associated with mental health disorder cases in Toronto. It captures how the observed distribution of cases in the specific study period had resulted from the continuous influence of the risk factors in previous years. Various socioeconomic and environmental factors influence mental health disorder cases, with distinctive effects on males and females in different age groups [1,13,38,55]. However, these risk factors are required to function continuously in an area (continued exposure) to significantly affect the cases in a large area such as the City of Toronto. Therefore, although we could not incorporate temporal analyses owing to the unavailability of mental health spatial data in different years, the findings still reflect the outcome of changes in vegetation in the preceding years and how this change could have affected the cases in the study area. Thus, the most significant contribution of this study lies in advancing the knowledge on reduced vegetation cover and how it could be a risk factor for the mental health disorders of males and females in different age groups in an urban area. There is a paucity of conclusive evidence in the existing literature, which this research has attempted to fulfill.

Despite the strengths, several limitations should be considered when interpreting the results of this study. First, as this was an ecological study, no individual-level conclusions should be deduced from the results. The results are only applicable to groups of people at the neighborhood level in Toronto. Second, the combined mental health disorder data set, which was used as the outcome variable, was created by aggregating different subcategories of mental health disorders. Better results could be obtained if the age- and sex-specific associations were analyzed using each subcategory that created the combined data set. However, as the original data set did not contain the observed cases for each subcategory, the combined data set had to be used. Third, this study analyzed the association between mental health disorder cases and vegetation cover in the summer-spring period. In reality, there could be a seasonal effect on the association because people's activities may change during a bitterly cold winter and also because people are more prone to develop seasonal affective disorders during seasons such as fall and winter. In this regard, the 2 spatial and nonspatial random effect terms in the Bayesian models would have adjusted for any unmeasured effects in the associations. Fourth, while interpreting the results, it is important to consider that people already in better mental health conditions could be more socioeconomically advantaged and thus reside in cities rich with vegetation cover [56]. However, we considered this effect and source of potential bias during our analysis. On the basis of the findings of several prominent studies [13,21,22], we modeled this issue as a product of the socioeconomic construct in the study area. Consequently, we adjusted for social inequality using the 4 marginalization covariates, which potentially adjusted for the differential exposure to vegetation cover owing to socioeconomic status and residential location in vegetation-rich areas. Finally, the findings from this research could only be generalized to areas having similar physical and socioeconomic conditions like the City of Toronto.

Despite its limitations, this study focused on the much-debated area of mental health research and aimed to analyze in detail the age- and sex-specific association between vegetation cover and mental health disorders. This research attempted to overcome the complex spatial and nonspatial modeling constraints, which are generally overlooked in mental health studies. The findings from this study provide crucial evidence that reduced vegetation cover could disproportionately affect the mental health conditions of young people. Owing to increased urbanization across the globe, with significant loss of vegetation-covered areas, the burden of mental health disorders among youth is likely to increase. This highlights the need to reconsider urban planning strategies and develop greener and more sustainable cities for future generations.

Conclusions

Despite strong evidence of the age- and sex-specific differences in the prevalence of mental health disorders, few studies have attempted to understand the age- and sex-specific association between vegetation cover and mental health disorders in urban areas. With the rapid loss of vegetation because of urbanization and global environmental changes, there is a need to thoroughly assess the impact of reduced vegetation on males and females from different age groups. Therefore, this study analyzed the association between vegetation cover and the combined mental health disorder cases for males and females in the 0-19, 20-44, 45-64, and ≥65 years age groups in the neighborhoods of Toronto. An EVI was constructed from Landsat-8 imageries using remote sensing techniques, and the Bayesian spatial hierarchical modeling was used to model the association between vegetation and mental health disorder cases after accounting for material deprivation, ethnic concentration, residential instability, dependence, and unmeasured covariates. The spatial and nonspatial unmeasured covariates were adjusted in the models by integrating 2 Gaussian random effect terms, which also helped to adjust for the spatial autocorrelation and overdispersion in the mental health disorder data. The age- and sex-specific analyses found that the mental health of children, adolescents, and younger adults could be particularly susceptible to reduced vegetation cover. Specifically, both males and females aged 0-19 and 25-44 years were vulnerable to various mental health disorders. In contrast, no significant association was evident between vegetation and the mental health disorder cases of adults in the 45-64 and ≥65 years age groups. For the older age groups, the socioeconomic factors were found to be more significantly influential than the variations in vegetation cover. The relative risk maps identified localized hotspots of mental health disorder cases, which could be the focus of urban vegetation management practices. The study findings suggest that the young urban population could be highly susceptible to reduced vegetation cover, which may lead to deterioration of mental health conditions in the early stages of life. The results highlight the need for sustainable urban planning initiatives that prioritize the conservation and growth of urban greenery. Although the study was conducted for the residents of Toronto, Canada, similar results could be expected in areas with comparable physical and socioeconomic conditions.

Acknowledgments

The authors would like to thank all the data providers for making the data publicly available and promoting future research. The authors express gratitude to the Natural Sciences and Engineering Research Council of Canada for funding this research (grant RGPIN-2014-06359).

Authors' Contributions

All the authors have contributed equally to this research. The conceptualization was done by AYMA and JL, and they worked toward fixing the methodology of the work. AYMA ran the data analysis and compiled the results, whereas the validation of the results was confirmed through discussions among AYMA, JL, CMP, and ZAB. AYMA and JL prepared the original draft, and further writing, review, and editing were carried out together by AYMA, JL, CMP, and ZAB. The supervision was done by JL, CMP, and ZAB. JL was responsible for the project administration and fund acquisition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Calculation of the Enhanced Vegetation Index and implementation of the Bayesian Spatial Hierarchical Model.
[DOCX File , 20 KB - publichealth_v8i7e34782_app1.docx]

References

1. Abdullah AY, Law J, Butt ZA, Perlman CM. Understanding the differential impact of vegetation measures on modeling the association between vegetation and psychotic and non-psychotic disorders in Toronto, Canada. Int J Environ Res Public Health 2021 Apr 28;18(9):4713 [FREE Full text] [doi: 10.3390/ijerph18094713] [Medline: 33925179]
2. Martellozzo F, Ramankutty N, Hall RJ, Price DT, Purdy B, Friedl MA. Urbanization and the loss of prime farmland: a case study in the Calgary–Edmonton corridor of Alberta. Reg Environ Change 2014 Aug 20;15(5):881-893. [doi: 10.1007/s10113-014-0658-0]
3. Sharpe DM, Stearns F, Leitner LA, Dorney JR. Fate of natural vegetation during urban development of rural landscapes in Southeastern Wisconsin. Urban Ecol 1986 Jun;9(3-4):267-287. [doi: 10.1016/0304-4009(86)90004-5]
4. Üstün TB, Ho R. Classification of mental disorders: principles and concepts. In: Quah SR, editor. International Encyclopedia of Public Health. 2nd edition. Cambridge, MA, USA: Academic Press; 2017:51-57.
5. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. World Health Organization. 1992. URL: https://apps.who.int/iris/handle/10665/37958 [accessed 2019-07-10]

6. Glazier RH, Gozdyra P, Kim M, Bai L, Kopp A, Schultz SE, et al. Geographic Variation in Primary Care Need, Service Use and Providers in Ontario, 2015/16. Institute for Clinical Evaluative Sciences. 2018 Aug. URL: https://www.ices.on.ca/Publications/Atlases-and-Reports/2018/Geographic-Variation-in-Primary-Care [accessed 2019-07-10]
7. Fett AK, Lemmers-Jansen IL, Krabbendam L. Psychosis and urbanicity: a review of the recent literature from epidemiology to neurourbanism. Curr Opin Psychiatry 2019 May;32(3):232-241. [doi: 10.1097/YCO.0000000000000486] [Medline: 30724751]
8. Radua J, Ramella-Cravaro V, Ioannidis JP, Reichenberg A, Phiphopthatsanee N, Amir T, et al. What causes psychosis? An umbrella review of risk and protective factors. World Psychiatry 2018 Feb;17(1):49-66 [FREE Full text] [doi: 10.1002/wps.20490] [Medline: 29352556]
9. Canada's five fastest growing urban areas are all in Ontario. CBC News. 2019 Mar 29. URL: https://www.cbc.ca/news/canada/windsor/stats-can-population-census-1.5075855 [accessed 2019-09-25]
10. Chen HT, Yu CP, Lee HY. The effects of forest bathing on stress recovery: evidence from middle-aged females of Taiwan. Forests 2018 Jul 05;9(7):403. [doi: 10.3390/f9070403]
11. Bielinis E, Jaroszewska A, Łukowski A, Takayama N. The effects of a forest therapy programme on mental hospital patients with affective and psychotic disorders. Int J Environ Res Public Health 2019 Dec 23;17(1):118 [FREE Full text] [doi: 10.3390/ijerph17010118] [Medline: 31877954]
12. Takayama N, Saito H, Fujiwara A, Horiuchi M. The effect of slight thinning of managed coniferous forest on landscape appreciation and psychological restoration. Prog Earth Planet Sci 2017 Jun 28;4(1):17 [FREE Full text] [doi: 10.1186/s40645-017-0129-6]
13. Markevych I, Schoierer J, Hartig T, Chudnovsky A, Hystad P, Dzhambov AM, et al. Exploring pathways linking greenspace to health: theoretical and methodological guidance. Environ Res 2017 Oct;158:301-317. [doi: 10.1016/j.envres.2017.06.028] [Medline: 28672128]
14. Feehan M, McGee R, Williams SM. Mental health disorders from age 15 to age 18 years. J Am Acad Child Adolesc Psychiatry 1993 Nov;32(6):1118-1126. [doi: 10.1097/00004583-199311000-00003] [Medline: 8282655]
15. Morgan VA, Castle DJ, Jablensky AV. Do women express and experience psychosis differently from men? Epidemiological evidence from the Australian National Study of Low Prevalence (Psychotic) Disorders. Aust N Z J Psychiatry 2008 Jan;42(1):74-82. [doi: 10.1080/00048670701732699] [Medline: 18058447]
16. Jones PB. Adult mental health disorders and their age at onset. Br J Psychiatry Suppl 2013 Jan;54:s5-10. [doi: 10.1192/bjp.bp.112.119164] [Medline: 23288502]
17. Bangasser DA, Valentino RJ. Sex differences in stress-related psychiatric disorders: neurobiological perspectives. Front Neuroendocrinol 2014 Aug;35(3):303-319. [doi: 10.1016/j.yfrne.2014.03.008] [Medline: 24726661]
18. Dadvand P, Bartoll X, Basagaña X, Dalmau-Bueno A, Martinez D, Ambros A, et al. Green spaces and general health: roles of mental health status, social support, and physical activity. Environ Int 2016 May;91:161-167. [doi: 10.1016/j.envint.2016.02.029] [Medline: 26949869]
19. Astell-Burt T, Mitchell R, Hartig T. The association between green space and mental health varies across the lifecourse. A longitudinal study. J Epidemiol Community Health 2014 Jun;68(6):578-583. [doi: 10.1136/jech-2013-203767] [Medline: 24604596]
20. Villeneuve PJ, Ysseldyk RL, Root A, Ambrose S, DiMuzio J, Kumar N, et al. Comparing the normalized difference vegetation index with the Google street view measure of vegetation to assess associations between greenness, walkability, recreational physical activity, and health in Ottawa, Canada. Int J Environ Res Public Health 2018 Aug 10;15(8):1719 [FREE Full text] [doi: 10.3390/ijerph15081719] [Medline: 30103456]
21. Rotenberg M, Tuck A, Anderson KK, McKenzie K. The incidence of psychotic disorders and area-level marginalization in Ontario, Canada: a population-based retrospective cohort study. Can J Psychiatry 2022 Mar;67(3):216-225 [FREE Full text] [doi: 10.1177/07067437211011852] [Medline: 33896210]
22. Wu F. Urban poverty and marginalization under market transition: the case of Chinese cities. Int J Urban Regional Res 2004 Jun;28(2):401-423. [doi: 10.1111/j.0309-1317.2004.00526.x]
23. Haining RP. Spatial autocorrelation. In: Smelser NJ, Baltes PB, editors. International Encyclopedia of the Social & Behavioral Sciences. Oxford, UK: Pergamon Press; 2001:14763-14768.
24. Haining RP, Li G. Modelling Spatial and Spatial-Temporal Data: A Bayesian Approach. Boca Raton, FL, USA: CRC Press; 2020.
25. Law J, Haining R, Maheswaran R, Pearson T. Analyzing the relationship between smoking and coronary heart disease at the small area level: a Bayesian approach to spatial modeling. Geogr Anal 2006 Apr;38(2):140-159. [doi: 10.1111/j.0016-7363.2006.00680.x]
26. Richter D, Wall A, Bruen A, Whittington R. Is the global prevalence rate of adult mental illness increasing? Systematic review and meta-analysis. Acta Psychiatr Scand 2019 Nov;140(5):393-407. [doi: 10.1111/acps.13083] [Medline: 31393996]
27. Dzhambov A, Hartig T, Markevych I, Tilov B, Dimitrova D. Urban residential greenspace and mental health in youth: different approaches to testing multiple pathways yield different conclusions. Environ Res 2018 Jan;160:47-59. [doi: 10.1016/j.envres.2017.09.015] [Medline: 28961469]

28. Wetterlin FM, Mar MY, Neilson EK, Werker GR, Krausz M. eMental health experiences and expectations: a survey of youths' Web-based resource preferences in Canada. J Med Internet Res 2014 Dec 17;16(12):e293 [FREE Full text] [doi: 10.2196/jmir.3526] [Medline: 25519847]
29. Sustaining and Expanding the Urban Forest: Toronto's Strategic Forest Management Plan – 2012-2022. City of Toronto, Parks, Forestry and Recreation, Urban Forestry. 2013. URL: https://www.toronto.ca/data/parks/pdf/trees/sustaining-expanding-urban-forest-management-plan.pdf [accessed 2019-07-10]
30. Information About TCHPP Geographies – Definitions, notes and historical context. Toronto Health Profiles. URL: http://www.torontohealthprofiles.ca/a_documents/aboutTheData/0_2_Information_About_TCHPP_Geographies.pdf [accessed 2019-06-20]
31. Law J, Perlman C. Exploring geographic variation of mental health risk and service utilization of doctors and hospitals in Toronto: a shared component spatial modeling approach. Int J Environ Res Public Health 2018 Mar 26;15(4):593 [FREE Full text] [doi: 10.3390/ijerph15040593] [Medline: 29587426]
32. Ontario Community Health Profiles Partnership. URL: http://www.ontariohealthprofiles.ca/ [accessed 2020-03-15]
33. Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. In: Fischer MM, Getis A, editors. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Berlin, Germany: Springer; 2010:73-89.
34. EarthExplorer. U.S. Department of the Interior. URL: https://earthexplorer.usgs.gov/ [accessed 2019-12-20]
35. Landsat Surface Reflectance-derived Spectral Indices - Landsat Enhanced Vegetation Index. U.S. Department of the Interior. 2019 Dec 20. URL: https://www.usgs.gov/land-resources/nli/landsat/landsat-enhanced-vegetation-index?qt-science_support_page_related_con=0#qt-science_support_page_related_con [accessed 2019-12-20]
36. Arc GIS Desktop 10.7.1 quick start guide. ArcGIS Desktop. 2019. URL: https://desktop.arcgis.com/en/arcmap/10.7/get-started/setup/arcgis-desktop-quick-start-guide.htm [accessed 2019-06-20]
37. Weich S, Blanchard M, Prince M, Burton E, Erens B, Sproston K. Mental health and the built environment: cross-sectional survey of individual and contextual risk factors for depression. Br J Psychiatry 2002 May;180:428-433. [doi: 10.1192/bjp.180.5.428] [Medline: 11983640]
38. Nguyen QC, Li D, Meng HW, Kath S, Nsoesie E, Li F, et al. Building a national neighborhood dataset from geotagged Twitter data for indicators of happiness, diet, and physical activity. JMIR Public Health Surveill 2016 Oct 17;2(2):e158 [FREE Full text] [doi: 10.2196/publichealth.5869] [Medline: 27751984]
39. Matheson FI, Moloney G, van Ingen T. 2016 Ontario marginalization index: user guide. Ontario Agency for Health Protection and Promotion (Public Health Ontario). 2018. URL: https://www.publichealthontario.ca/-/media/documents/o/2017/on-marg-userguide.pdf [accessed 2019-07-10]
40. Benesty J, Chen J, Huang Y, Cohen I. Pearson correlation coefficient. In: Cohen I, Huang Y, Chen J, Benesty J, editors. Noise Reduction in Speech Processing. Heidelberg, Germany: Springer; 2009:1-4.
41. Haining R, Law J, Griffith D. Modelling small area counts in the presence of overdispersion and spatial autocorrelation. Comput Stat Data Anal 2009 Jun 15;53(8):2923-2937. [doi: 10.1016/j.csda.2008.08.014]
42. Best NG, Waller LA, Thomas A, Conlon EM, Arnold RA. Bayesian models for spatially correlated diseases and exposure data. In: Proceedings of the 6th Valencia International Meeting. 1998 Presented at: ValenciaIM '98; June 6-10, 1998; Valencia, Spain p. 131-156.
43. Lee M, Kim S, Ha M. Community greenness and neurobehavioral health in children and adolescents. Sci Total Environ 2019 Jul 01;672:381-388. [doi: 10.1016/j.scitotenv.2019.03.454] [Medline: 30959304]
44. Fong KC, Hart JE, James P. A review of epidemiologic studies on greenness and health: updated literature through 2017. Curr Environ Health Rep 2018 Mar;5(1):77-87 [FREE Full text] [doi: 10.1007/s40572-018-0179-y] [Medline: 29392643]
45. Johannsen DL, DeLany JP, Frisard MI, Welsch MA, Rowley CK, Fang X, Louisiana Healthy Aging Study. Physical activity in aging: comparison among young, aged, and nonagenarian individuals. J Appl Physiol (1985) 2008 Aug;105(2):495-501 [FREE Full text] [doi: 10.1152/japplphysiol.90450.2008] [Medline: 18556430]
46. Dzhambov AM, Markevych I, Hartig T, Tilov B, Arabadzhiev Z, Stoyanov D, et al. Multiple pathways link urban green- and bluespace to mental health in young adults. Environ Res 2018 Oct;166:223-233. [doi: 10.1016/j.envres.2018.06.004] [Medline: 29890427]
47. Kuo FE, Sullivan WC, Coley RL, Brunson L. Fertile ground for community: inner-city neighborhood common spaces. Am J Community Psychol 1998 Dec;26(6):823-851. [doi: 10.1023/a:1022294028903]
48. Coley RL, Sullivan WC, Kuo FE. Where does community grow?: the social context created by nature in urban public housing. Environ Behav 1997 Jul 1;29(4):468-494. [doi: 10.1177/001391659702900402]
49. DeJohn AD, Schulz EE, Pearson AL, Lachmar EM, Wittenborn AK. Identifying and understanding communities using Twitter to connect about depression: cross-sectional study. JMIR Ment Health 2018 Nov 05;5(4):e61 [FREE Full text] [doi: 10.2196/mental.9533] [Medline: 30401662]
50. Francis J, Wood LJ, Knuiman M, Giles-Corti B. Quality or quantity? Exploring the relationship between Public Open Space attributes and mental health in Perth, Western Australia. Soc Sci Med 2012 May;74(10):1570-1577. [doi: 10.1016/j.socscimed.2012.01.032] [Medline: 22464220]

51. Srugo SA, de Groh M, Jiang Y, Morrison HI, Hamilton HA, Villeneuve PJ. Assessing the impact of school-based greenness on mental health among adolescent students in Ontario, Canada. Int J Environ Res Public Health 2019 Nov 08;16(22):4364 [FREE Full text] [doi: 10.3390/ijerph16224364] [Medline: 31717373]
52. Beiser M. Strangers at the Gate: The 'Boat People's' First Ten Years in Canada. Toronto, Canada: University of Toronto Press; 1999.
53. Beiser M, Hou F. Mental health effects of premigration trauma and postmigration discrimination on refugee youth in Canada. J Nerv Ment Dis 2016 Jun;204(6):464-470. [doi: 10.1097/NMD.0000000000000516] [Medline: 27101023]
54. Hartig T. Three steps to understanding restorative environments as health resources. In: Thompson CW, Travlou P, editors. Open Space: People Space. London, UK: Taylor & Francis; 2007:183-200.
55. Pearce J, Cherrie M, Shortt N, Deary I, Ward Thompson C. Life course of place: a longitudinal study of mental health and place. Trans Inst Br Geogr 2018 May 02;43(4):555-572. [doi: 10.1111/tran.12246]
56. Martin CA, Warren PS, Kinzig AP. Neighborhood socioeconomic status is a useful predictor of perennial landscape vegetation in residential neighborhoods and embedded small parks of Phoenix, AZ. Landsc Urban Plan 2004 Oct 30;69(4):355-368. [doi: 10.1016/j.landurbplan.2003.10.034]

Abbreviations

EVI: Enhanced Vegetation Index
OMI: Ontario Marginalization Index

Edited by Y Khader; submitted 08.11.21; peer-reviewed by C Zhao, V Verma, RT Villarino; comments to author 16.01.22; revised version received 01.05.22; accepted 25.05.22; published 28.07.22.

Please cite as:
Abdullah AYM, Law J, Perlman CM, Butt ZA
Age- and Sex-Specific Association Between Vegetation Cover and Mental Health Disorders: Bayesian Spatial Study
JMIR Public Health Surveill 2022;8(7):e34782
URL: https://publichealth.jmir.org/2022/7/e34782
doi:10.2196/34782
PMID:35900816

Original Paper

Relationship Between Protein Intake in Each Traditional Meal and Physical Activity: Cross-sectional Study

Takae Shinto[1], MSc; Saneyuki Makino[1], MSc; Yu Tahara[1], PhD; Lyie Nitta[1], MSc; Mai Kuwahara[1], MSc; Ayako Tada[2], BA; Nanako Abe[2], BA; Mikiko Michie[2], BA; Shigenobu Shibata[1], PhD

[1]Department of Bioscience and Engineering, Waseda University, Tokyo, Japan
[2]Asken Inc, Tokyo, Japan

Corresponding Author:
Shigenobu Shibata, PhD
Department of Bioscience and Engineering
Waseda University
Wakamatsu-cho 2-2
Shinjuku-ku
Tokyo, 162-0056
Japan
Phone: 81 +81353697318
Email: shibatas@waseda.jp

Related Article:

This is a corrected version. See correction statement: https://publichealth.jmir.org/2022/8/e41460

Abstract

Background: Protein intake plays an important role in the synthesis and maintenance of skeletal muscles for the prevention of health risks. It is also widely known that physical activity influences muscle function. However, no large-scale studies have examined the relationship between daily dietary habits, especially the timing of protein intake, and daily physical activity.

Objective: The purpose of this cross-sectional study was to investigate how protein intake and composition (involving the 3 major nutrients protein, fat, and carbohydrate) in the 3 traditional meals (breakfast, lunch, and dinner) are associated with physical activity.

Methods: Using daily dietary data accumulated in the smartphone food log app “Asken” and a web-based cross-sectional survey involving Asken users (N=8458), we analyzed nutrient intake and composition, as well as daily activity levels. As very few individuals skipped breakfast (1102/19,319 responses, 5.7%), we analyzed data for 3 meals per day.

Results: Spearman rank correlation analysis revealed that breakfast and lunch protein intakes had higher positive correlations with daily physical activity among the 3 major macronutrients ($P<.001$). These findings were confirmed by multivariate logistic regression analysis with confounding factors. Moreover, participants with higher protein intake and composition at breakfast or lunch tended to exhibit significantly greater physical activity than those with higher protein intake at dinner ($P<.001$).

Conclusions: Among the 3 macronutrients, protein intake during breakfast and lunch was closely associated with daily physical activity.

(JMIR Public Health Surveill 2022;8(7):e35898) doi:10.2196/35898

KEYWORDS

protein; dietary pattern; physical activity; chrononutrition

Introduction

Maintaining and increasing muscle mass are critical for preventing various health risks, such as metabolic syndrome, diabetes, and sarcopenia [1-3]. In addition, protein intake is an effective factor for skeletal muscle synthesis and maintenance [4]. According to the Ministry of Health, Labor, and Welfare in Japan, the recommended daily protein intake values for adult Japanese males and females are 65 g and 50 g, respectively [5]. Additionally, several studies in the United States and Japan have shown that average protein intake in adults was unevenly distributed across the 3 meals, with the lowest intake at breakfast and the highest intake skewed toward dinner [6,7]. This typical pattern of protein intake has been associated with a decline in muscle function, including grip strength and muscle mass [8-10]. A cross-over study reported significantly higher 24-h muscle protein synthesis rates when protein intake was evenly distributed with adequate consumption at each meal, rather than skewed toward the evening meal [11]. In particular, breakfast is the most commonly skipped meal of the day, yet it plays an important role in health [12-14]. Indeed, recent studies have shown that breakfast protein intake effectively induced muscle hypertrophy in humans and rodents [15,16]. Therefore, it is important to consider not only the total protein intake, but also the intake and composition at each meal for regulating muscle function. In addition, physical activities, such as exercise and strength training, are effective for maintaining and increasing muscle mass [17]. However, the precise relationship between daily eating habits, especially protein intake patterns, and physical activity remains unclear. Therefore, we hypothesized that breakfast protein would have a strong positive relationship with physical activity among nutrients or intake timing. We investigated the relationship between the intake and composition of nutrients (protein, fat, and carbohydrate) in 3 meals and physical activity in approximately 8000 users of the Asken mobile health app for dietary management. Since this is a health care app generally aimed at weight maintenance and weight loss, the users are likely to be highly health conscious. In fact, the percentage of those who skip breakfast in Japan is 12% [18], while in this study, it was as low as 5.7%. In addition, most people consume snacks, but due to data availability, the analysis in this survey focused on the 3 traditional meals. To our knowledge, no previous study has investigated the dietary pattern of 3 meals and physical activity in such a large population. Therefore, this study could potentially clarify the characteristics of physical activity in the daily dietary pattern of 3 meals, including breakfast, as well as the relationship between physical activity and protein intake.

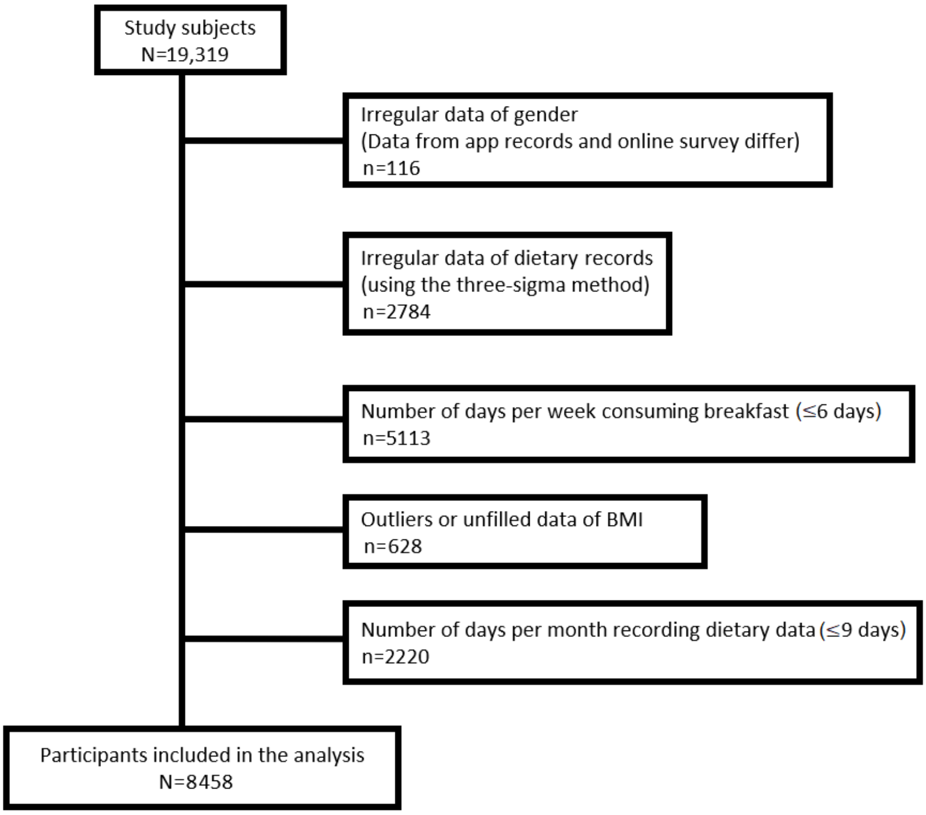
Methods

Study Participants and the Mobile Health App “Asken”

This study was approved by the Ethics Review Committee on Research with Human Subjects at Waseda University (No. 2020-046), in accordance with the guidelines of the Declaration of Helsinki. Moreover, informed consent was obtained from all individuals who participated in the study.

“Asken” is a popular food log and food coaching app that has been downloaded over 6,000,000 times [19] and has persistently ranked in the top 3 for health app categories in Japan [20]. As most users (almost 95%) use this app for body weight reduction and female individuals might be more inclined to maintain body shape than male individuals, 70% of the app users were female. The app provides feedback on dietary content based on the Dietary Intake Standards for Japanese as determined by the Ministry of Health, Labor, and Welfare. In this study, an online survey was conducted via the app, in addition to the available dietary records. The online survey was conducted in January 2021, and at the time of implementation, we announced that 200 people would be given 500 Japanese yen (approximately US $5) in a lottery. Since the app was not for clinical use and we did not ask for detailed information (eg, disease) from the participants, inclusion and exclusion criteria were not set when the online survey was conducted. The study participants included 8458 users of Asken (aged 10-80 years). As shown in Figure 1, we excluded those who gave contradictory or no responses for gender and BMI, those who did not consume breakfast daily, those whose dietary records were less than 9 days per month, and those who deviated from the normal distribution for dietary intake (using the 3-sigma method).

Figure 1. Flow of the study participants.



The self-reported food logs accumulated in the app have been deemed reliable for research purposes [21]. A recent study reported that the median correlation coefficient between daily paper-based nutrient intakes and those obtained from Asken was approximately 0.8 in adult men and women [22]. In addition, the dietary data obtained in this study were comparable with data calculated from the National Nutrition Survey in Japan (NNSJ) [23] (Multimedia Appendix 1). These results suggest that the dietary data used in this study are reliable. However, the amount of daily protein was slightly higher, while the amounts of fat and carbohydrate were slightly lower compared with the NNSJ data (Multimedia Appendix 1).

Questionnaire

The online survey included 34 items involving personal characteristics (age, gender, and prefecture of residence), lifestyle habits like diet and sleep, physical activity (International Physical Activity Questionnaire [IPAQ] short form), and personality assessment (Ten-Item Personality Inventory [TIPI]). Except for the physical activity and personality assessment, the other items were originally developed by us. In this study, “age,” “gender,” “physical activity (IPAQ short form),” and “frequency of breakfast intake” were used for analysis in the online survey. The questionnaire was designed to take approximately 15 minutes to complete.

Dietary Data

The dietary assessment was based on the average dietary records of each participant for approximately 1 month (January 2021). For each of the 3 traditional meals (breakfast, lunch, and dinner) and snacks, the intake included energy content (kcal), protein, fat, carbohydrate, sodium, potassium, cholesterol, fiber, and saturated lipids. Herein, the 3 major nutrients (ie, carbohydrate, fat, and protein) were evaluated. As timing information on snacks was lacking, data on snacks were not used for further analysis. To evaluate the compositions of the 3 macronutrients (% kcal), we converted the amount of intake (g) into the amount of energy. The calculation formulas are as follows:

Protein composition (% kcal) = protein intake (g) × 4 (kcal/g) / total energy (kcal) × 100 **(1)**

Fat composition (% kcal) = fat intake (g) × 9 (kcal/g) / total energy (kcal) × 100 **(2)**

Carbohydrate composition (% kcal) = carbohydrate intake (g) × 4 (kcal/g) / total energy (kcal) × 100 **(3)**

In addition, to emphasize the relationship between protein and physical activity, the participants were categorized into meal types according to the protein intake amount and composition in each meal. For example, in intake amount grouping, if participants consumed 20% of the daily protein intake during breakfast, 30% during lunch, and 50% during dinner, they were classified into the “dinner” group. On the other hand, for the composition type of protein intake, if participants consumed 20% of their energy during breakfast, 40% during lunch, and 30% during dinner, they were classified into the “lunch” group. The characteristics of each group are shown in Multimedia Appendix 2 and Multimedia Appendix 3. Moreover, weight, height, and frequency of recording were included in the dietary records, and these were also used in the multivariate logistic regression analysis.

Physical Activity

The IPAQ short form assessed the amount of physical activity in daily life [24]. Respondents were asked about the number of days and hours spent doing each of the following 3 types of activities during the last 7 days: vigorous-intensity activity, moderate-intensity activity, and walking. Based on IPAQ analysis guidelines [25], we calculated weekly physical activity (metabolic equivalents-minutes/week [MET-min/week]) using the following formulas:

Walking MET-min/week = 3.3 × walking min × walking days **(4)**

Moderate MET-min/week = 4.0 × moderate-intensity activity min × moderate-intensity days **(5)**

Vigorous MET-min/week = 8.0 × vigorous-intensity activity min × vigorous-intensity days **(6)**

Total physical activity MET-min/week = sum of walking + moderate + vigorous MET-min/week values **(7)**

Data cleaning was performed according to previous guidelines [25]. These guidelines describe how to analyze data from the short and long versions of the IPAQ, which were obtained through interviews (telephone or interview) and self-administered questionnaires, and describe the procedures and data processing rules for both the short and long versions. The data processing included recoding of minimum values and truncation of data to improve data comparability.

Statistical Analysis

Correlation analysis was conducted using Spearman rank correlation analysis. Moreover, multivariate logistic regression analysis was performed to investigate more detailed associations between dietary patterns and physical activity. A minimum sample size of 351 was required to have approximately 95% power to detect large effects at a significance level of .05 (G*Power, version 3.1.9.2; Universitat Kiel) in the multivariate logistic regression analysis. The objective variable, physical activity per exercise intensity, was divided into the following 2 groups: inactive (below median) and active (above median). We then divided the participants into groups according to their protein intake patterns and analyzed the amount of physical activity for each group using the Mann-Whitney *U* test. These analyses were conducted considering the following 2 factors: intake and composition. All data were analyzed using SPSS version 27 (IBM Corp), and a *P* value <.05 was considered to indicate statistical significance. Data are expressed as mean (SD) for most variables and as median and quartile (Q; 1-3) for IPAQ data.

Results

Participant Characteristics

From the obtained dietary records and questionnaires (19,319 responses), we excluded subjects for various reasons, and we eventually used the data of 8458 subjects (2321 male subjects and 6137 female subjects) in the present analysis (Figure 1). In particular, to focus on those who consume breakfast every day, we excluded about 5000 respondents, who made up the bulk of the excluded population. As this app has a gender gap among users, female individuals accounted for 73% of all participants in this analysis.

The mean participant age was 44.3 (SD 12.7) years, and the BMI was 25.0 (SD 3.7) kg/m^2 for males and 22.9 (SD 3.9) kg/m^2 for females (Table 1). The total energy intake was 2084.8 (SD 320.5) kcal for males and 1637.8 (SD 262.5) kcal for females, and the amount of energy per meal was skewed toward dinner. The daily protein intake was 85.5 (SD 17.2) g in males and 69.3 (SD 14.3) g in females, which exceeded the recommended amount of protein per day for Japanese people, as proposed by the Ministry of Health, Labor and Welfare in Japan [5]. Conversely, we observed that protein distribution was low during breakfast and skewed toward dinner, which is consistent with the findings of previous studies [6,7]. Indeed, there were significant differences in energy and protein intake comparing the 3 meals (energy: *P*<.001, protein: *P*<.001). In addition, the median amount of physical activity was 1257 (Q1-Q3 558-2346) MET-min/week. According to the Ministry of Health, Labor, and Welfare in Japan, the standard for physical activity for individuals aged 18 to 64 years is 23 MET-h/week, that is, 1380 MET-min/week of physical activity at an intensity of ≥3 MET [26]. Therefore, the distribution of the amount of physical activity in the current participants was slightly below the established standard.

Table 1. Characteristics of the participants.

Characteristic	Total (N=8458)	Male (n=2321)	Female (n=6137)
Age (years), mean (SD)	44.3 (12.7)	49.6 (11.4)	42.2 (12.5)
BMI (kg/m^2), mean (SD)	23.5 (4.0)	25.0 (3.7)	22.9 (3.9)
Total energy (kcal/d), mean (SD)	1760.5 (343.5)	2084.8 (320.5)	1637.8 (262.5)
Breakfast energy (kcal), mean (SD)	389.6 (119.7)[a]	440.8 (129.8)[a]	370.2 (109.5)[a]
Lunch energy (kcal), mean (SD)	542.2 (128.7)	630.3 (133.1)	508.8 (109.9)
Dinner energy (kcal), mean (SD)	609.1 (191.6)	780.8 (185.6)	544.2 (148.9)
Total protein intake (g/d), mean (SD)	73.8 (16.8)	85.5 (17.2)	69.3 (14.3)
Breakfast protein (g/d), mean (SD)	16.9 (6.8)[a]	19.1 (7.4)[a]	16.1 (6.4)[a]
Lunch protein (g/d), mean (SD)	22.3 (6.2)	25.2 (6.5)	21.2 (5.7)
Dinner protein (g/d), mean (SD)	28.2 (8.0)	34.1 (8.0)	25.9 (6.7)
Total fat intake (g/d), mean (SD)	53.1 (13.1)	62.7 (12.9)	49.4 (11.2)
Breakfast fat (g/d), mean (SD)	12.3 (5.2)[a]	13.9 (5.6)[a]	11.7 (4.9)[a]
Lunch fat (g/d), mean (SD)	18.4 (5.8)	21.1 (6.2)	17.3 (5.2)
Dinner fat (g/d), mean (SD)	22.4 (7.6)	27.7 (7.5)	20.4 (6.6)
Total carbohydrate intake (g/d), mean (SD)	185.9 (44.5)	219.0 (46.1)	173.4 (36.9)
Breakfast carbohydrate (g/d), mean (SD)	53.0 (18.0)[a]	59.5 (20.1)[a]	50.6 (16.5)[a]
Lunch carbohydrate (g/d), mean (SD)	69.4 (18.2)	81.0 (19.3)	65.1 (15.7)
Dinner carbohydrate (g/d), mean (SD)	63.4 (22.7)	78.5 (24.0)	57.7 (19.2)
Total PA[b] (MET[c]-min/week), median (Q1-Q3)	1257 (558-2346)	1470 (720-2612)	1164 (495-2232)
Vigorous PA (MET-min/week), median (Q1-Q3)	120 (0-960)	400 (0-1200)	0 (0-800)
Moderate PA (MET-min/week), median (Q1-Q3)	0 (0-360)	0 (0-360)	0 (0-360)
Walking (MET-min/week), median (Q1-Q3)	495 (198-990)	594 (198-1188)	495 (198-990)

[a] *P*<.001 by the Kruskal-Wallis test between breakfast, lunch, and dinner for energy and 3 macronutrients.

[b] PA: physical activity.

[c] MET: metabolic equivalents.

Relationship Between Intake of the 3 Macronutrients and Physical Activity

Based on correlation analysis, each physical activity level had the strongest positive association with breakfast or lunch protein intake among the 3 macronutrients during the 3 meals (Table 2). Total, vigorous, and moderate physical activities were positively related to breakfast protein intake, while walking was positively related to lunch protein intake (total physical activity: *P*<.001, vigorous physical activity: *P*<.001, moderate physical activity: *P*<.001, walking: *P*<.001).

Table 2. Correlation analysis (Spearman rank correlation coefficient) of intake and composition of the 3 macronutrients and International Physical Activity Questionnaire findings.

Variable	Total PA[a] (MET[b]-min/week)	Vigorous PA (MET-min/week)	Moderate PA (MET-min/week)	Walking (MET-min/week)
Intake (g)				
Breakfast protein	0.170[c]	0.191[c]	0.072[c]	0.052[c]
Breakfast fat	0.038[c]	0.021	0.019	0.033[d]
Breakfast carbohydrate	0.041[c]	0.002	0.014	0.050[c]
Lunch protein	0.142[c]	0.170[c]	0.039[c]	0.053[c]
Lunch fat	0.016	0.0211	−0.009	0.017
Lunch carbohydrate	0.023[e]	−0.001	−0.014	0.050[c]
Dinner protein	0.091[c]	0.123[c]	0.011	0.037[c]
Dinner fat	−0.014	0.010	−0.027[e]	−0.010
Dinner carbohydrate	−0.011	−0.021	−0.030[d]	0.015
Composition (% kcal)				
Breakfast protein	0.149[c]	0.196[c]	0.066[c]	0.023[e]
Breakfast fat	−0.032[d]	−0.031[d]	−0.009	−0.006
Breakfast carbohydrate	−0.047[c]	−0.086[c]	−0.022[e]	0.013
Lunch protein	0.134[c]	0.175[c]	0.065[c]	0.023[e]
Lunch fat	−0.021	−0.007	−0.004	−0.022[e]
Lunch carbohydrate	−0.047[c]	−0.083[c]	−0.020	0.007
Dinner protein	0.108[c]	0.134[c]	0.055[c]	0.028[e]
Dinner fat	−0.055[c]	−0.024[e]	−0.012	−0.052[c]
Dinner carbohydrate	−0.030[d]	−0.060[c]	−0.017	0.004

[a] PA: physical activity.

[b] MET: metabolic equivalents.

[c] *P*<.001 (Spearman rank correlation coefficient).

[d] *P*<.01 (Spearman rank correlation coefficient).

[e] *P*<.05 (Spearman rank correlation coefficient).

Multivariate logistic regression analysis revealed that vigorous and moderate physical activities were most strongly positively related with breakfast protein intake, even after adjustment for confounding factors, such as sex, age, BMI, and frequency of recording (Multimedia Appendix 4). On the other hand, although total physical activity was associated with breakfast protein intake, it showed the strongest positive relationship with lunch protein intake. Walking had a positive relationship with dinner protein intake.

Relationship Between the Composition of the 3 Macronutrients (% kcal) and Physical Activity

Correlation analysis showed that total, vigorous, and moderate physical activities were positively associated with breakfast protein intake, while walking was positively associated with dinner protein intake (total physical activity: *P*<.001, vigorous physical activity: *P*<.001, moderate physical activity: *P*<.001, walking: *P*<.010) (Table 2). On the other hand, multivariate logistic regression analysis confirmed the strongest positive association of total physical activity with breakfast, and vigorous physical activity with lunch. No association with any of the nutrients could be confirmed for moderate exercise.

Grouping by Protein Intake and Composition

The total protein intake per meal per day was similar between groups. Grouping by protein intake revealed that the breakfast and lunch groups demonstrated significantly greater physical activity (ie, total and vigorous physical activities) when compared with the dinner group (Table 3). In terms of walking, the lunch group presented significantly greater physical activity than the dinner group.

Table 3. International Physical Activity Questionnaire findings by meal type with the highest protein intake and composition from each meal.

Variable	Breakfast	Lunch	Dinner	*P* value[a]		
				B vs L[b]	B vs D[c]	L vs D[d]
Meal type with the highest protein intake						
Total PA[e] (MET[f]-min/week), median (Q1-Q3)	1440 (693-2520)	1362 (597-2672)	1200 (528-2280)	.58	<.001	.001
Vigorous PA (MET-min/week), median (Q1-Q3)	400 (0-1120)	240 (0-960)	80 (0-960)	.20	<.001	.02
Moderate PA (MET-min/week), median (Q1-Q3)	0 (0-480)	0 (0-360)	0 (0-360)	>.99	.33	.95
Walking (MET-min/week), median (Q1-Q3)	495 (198-990)	528 (198-1134)	495 (198-990)	>.99	>.99	.049
Meal type with the highest protein composition						
Total PA (MET-min/week), median (Q1-Q3)	1386 (671-2597)	1266 (528-2376)	1188 (516-2234)	.02	<.001	.34
Vigorous PA (MET-min/week), median (Q1-Q3)	400 (0-1120)	80 (0-960)	0 (0-840)	<.001	<.001	.04
Moderate PA (MET-min/week), median (Q1-Q3)	0 (0-420)	0 (0-360)	0 (0-360)	.12	.02	>.99
Walking (MET-min/week), median (Q1-Q3)	495 (198-990)	495 (198-990)	495 (198-990)	>.99	.94	>.99

[a] Mann-Whitney *U* test.

[b] Breakfast versus lunch.

[c] Breakfast versus dinner.

[d] Lunch versus dinner.

[e] PA: physical activity.

[f] MET: metabolic equivalents.

In addition, on grouping by protein composition, the breakfast group showed significantly higher total and vigorous physical activities than the lunch and dinner groups. Considering vigorous physical activity, the lunch group showed significantly greater physical activity than the dinner group. Finally, in terms of moderate physical activity, the breakfast group exhibited significantly greater physical activity than the dinner group.

Discussion

Principal Findings

In this study, we investigated the relationship of protein intake and composition during 3 meals with physical activity among adult male and female users of the mobile health app “Asken.” This study revealed that the daily amount of protein was higher than the recommended amount and the amount of physical activity was lower than the recommended amount. These findings could be attributed to the nature of the “Asken” app, which is a health care app typically designed to maintain or lose weight. In addition, because female individuals are generally more physically inactive than male individuals [27], the gender ratio of participants in this study may have influenced the results. Moreover, our results revealed that protein intake at breakfast and lunch showed the strongest positive relationship with physical activity among the 3 macronutrients consumed during the 3 meals. In addition, on grouping by protein intake patterns, those who consumed the most protein at breakfast and lunch exhibited significantly higher physical activity levels than those who consumed more protein at dinner. These results were similar for both intake and composition, indicating that not only intake but also the amount of protein in the overall diet is positively associated with the magnitude of physical activity. Several previous studies have reported that breakfast and lunch protein intakes affected the maintenance of muscle function [16,28,29]. Indeed, skeletal muscle mass increased in older individuals when protein was supplemented at breakfast and lunch [29]. A recent study has reported that in rodents, breakfast protein was effective in inducing muscle hypertrophy [16]. In addition, increased muscle mass has been shown to be negatively associated with mental health risks, such as depressed mood and stress, in addition to physical health [30,31]. Therefore, increasing protein intake at breakfast and lunch may enhance muscle mass, as well as promote physical and mental health. Indeed, among adolescent males and females, those with lower protein intake, especially those with lower milk intake at breakfast, have been reported to have lower scores for depression and anxiety symptoms [32]. Another possibility is that individuals with higher physical activity levels are more health conscious and typically consume more protein in the morning and afternoon. Several previous studies have reported a positive association between protein intake and physical activity levels [33-35]. However, these reports focused on daily protein intake, and few studies have examined intake at each meal. In addition, studies that have investigated protein intake in each diet have only investigated intake and have not confirmed the characteristics with respect to the distribution of intake [36]. Therefore, the finding that breakfast and lunch protein intakes are particularly related to physical activity is a new finding first presented in this study. Recently, 2 interesting longitudinal studies reported that high-quality protein evaluated by the "protein digestibility–corrected amino acid score" at breakfast but not lunch and dinner was positively associated with maintaining not only muscle strength but also brain cognition in Japanese older adults [37,38]. These studies have suggested future analysis of protein sources at breakfast using the "Asken" app.

Study Strengths and Limitations

In this study, we investigated the relationship between the 3 major nutrients during the 3 traditional meals and physical activity using a large data set. In addition, we examined the characteristics of physical activity by grouping individuals according to the meal with the highest protein consumption. To the best of our knowledge, no previous study has examined dietary patterns of the 3 meals and daily physical activity in such a large data set. Accordingly, the strength of this study is that it analyzed the characteristics of physical activity according to daily dietary patterns, along with the relationship between protein intake and physical activity, which is an effective factor for the regulation of muscle function that has not been addressed in previous studies.

However, this study had several limitations. First, as this survey was based on self-reports by the participants, there may be discrepancies in their actual dietary data and physical activity levels. In other words, participants may have responded by exaggerating or underestimating their lifestyles based on social expectations. Second, given the nature of the app used in this study, most participants were female, and accordingly, menstruation could not be considered. Although the survey was analyzed by eliminating the influence of gender as appropriate, it could have marginally impacted the results. Third, the app offered courses for various health purposes, such as weight maintenance, weight loss, and muscle gain. Since we recruited subjects from among all users, there may have been a mix of subjects with different objectives, which may have affected the results. Fourth, this study did not take into account stress or sleep, which could have affected the quantity and content of the meals. That is, participants may have increased consumption of foods high in fat and carbohydrates. Fifth, our results showed a small value for the correlation coefficient. This may be partly due to the large sample size, but since a similar sample size was used in a previous study [39,40], we followed the same approach in our analysis. In addition, Spearman rank correlation analysis was conducted in this study to investigate the trend of association, and more detailed relationships were confirmed by multivariate logistic regression analysis. Finally, in the present survey experiments, we could not obtain timing information on snacks. Protein content (16.9 g) and percentage energy of protein (19.1%) may participate in the increase of physical activity when a snack is taken between breakfast and lunch.

Conclusions

We analyzed the relationship of protein intake and composition during the 3 traditional meals with the amount of physical activity, as well as the characteristics of physical activity according to protein intake patterns in adult males and females. These results revealed that protein intake at breakfast and lunch had the strongest positive association with daily physical activity, and they were closely related to each other. The results suggest that breakfast may have an important role in physical activity. However, since this study focused only on protein, future studies should also consider protein sources.

Acknowledgments

All participants are gratefully acknowledged. This work was partially supported by a Grant-in-Aid for Scientific Research (A, 19H01089 for SS; C, 21K11606 for YT) from the Japan Society for the Promotion of Science, the JST-Mirai Program (JMPJM120D5 for SS), the JST-FOREST Program (JPMJFR205G for YT), and the Public Health Research Foundation (2020; for YT).

Authors' Contributions

TS, YT, MM, AT, SS, and NA designed and conducted this study. TS analyzed the data. SM, LN, and MK assisted with the statistical analysis. YT and SS advised on the interpretation of the results. The first draft of the manuscript was written by ST, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

AT and NA are full-time employees of Asken Inc. MM is a corporate officer in Asken Inc. The other authors declare no competing interests.

Multimedia Appendix 1

Comparison of average nutrient intake between the National Nutrition Survey in Japan (NNSJ) 2019 and the data of this study.
[PDF File (Adobe PDF File), 166 KB - publichealth_v8i7e35898_app1.pdf]

Multimedia Appendix 2

Characteristics of meal types with the highest protein intake from each meal.
[PDF File (Adobe PDF File), 165 KB - publichealth_v8i7e35898_app2.pdf]

Multimedia Appendix 3
Characteristics of meal types with the highest protein composition from each meal.
[PDF File (Adobe PDF File), 188 KB - publichealth_v8i7e35898_app3.pdf]

Multimedia Appendix 4
Association of the intake and composition of the 3 macronutrients in each meal with International Physical Activity Questionnaire findings.
[PDF File (Adobe PDF File), 179 KB - publichealth_v8i7e35898_app4.pdf]

References

1. Scott D, Park MS, Kim TN, Ryu JY, Hong HC, Yoo HJ, et al. Associations of low muscle mass and the metabolic syndrome in Caucasian and Asian middle-aged and older adults. J Nutr Health Aging 2016 Mar;20(3):248-255. [doi: 10.1007/s12603-015-0559-z] [Medline: 26892573]
2. Son JW, Lee SS, Kim SR, Yoo SJ, Cha BY, Son HY, et al. Low muscle mass and risk of type 2 diabetes in middle-aged and older adults: findings from the KoGES. Diabetologia 2017 May;60(5):865-872. [doi: 10.1007/s00125-016-4196-9] [Medline: 28102434]
3. Shimokata H, Ando F, Yuki A, Otsuka R. Age-related changes in skeletal muscle mass among community-dwelling Japanese: a 12-year longitudinal study. Geriatr Gerontol Int 2014 Feb;14 Suppl 1:85-92. [doi: 10.1111/ggi.12219] [Medline: 24450565]
4. Paddon-Jones D, Rasmussen BB. Dietary protein recommendations and the prevention of sarcopenia. Curr Opin Clin Nutr Metab Care 2009 Jan;12(1):86-90 [FREE Full text] [doi: 10.1097/MCO.0b013e32831cef8b] [Medline: 19057193]
5. Dietary Reference Intakes for Japanese, 2020. Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/kenkou/eiyou/syokuji_kijyun.html [accessed 2021-09-13]
6. Ishikawa-Takata K, Takimoto H. Current protein and amino acid intakes among Japanese people: Analysis of the 2012 National Health and Nutrition Survey. Geriatr Gerontol Int 2018 May;18(5):723-731. [doi: 10.1111/ggi.13239] [Medline: 29356253]
7. What We Eat in America, National Health and Nutrition Examination Survey (NHANES) 2017-2018. US Department of Agriculture. URL: https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/wweia-data-tables/ [accessed 2021-09-13]
8. Mishra S, Goldman JD, Sahyoun NR, Moshfegh AJ. Association between dietary protein intake and grip strength among adults aged 51 years and over: What We Eat in America, National Health and Nutrition Examination Survey 2011-2014. PLoS One 2018;13(1):e0191368 [FREE Full text] [doi: 10.1371/journal.pone.0191368] [Medline: 29364939]
9. Farsijani S, Payette H, Morais JA, Shatenstein B, Gaudreau P, Chevalier S. Even mealtime distribution of protein intake is associated with greater muscle strength, but not with 3-y physical function decline, in free-living older adults: the Quebec longitudinal study on Nutrition as a Determinant of Successful Aging (NuAge study). Am J Clin Nutr 2017 Jul;106(1):113-124. [doi: 10.3945/ajcn.116.146555] [Medline: 28515070]
10. Yasuda J, Asako M, Arimitsu T, Fujita S. Association of protein intake in three meals with muscle mass in healthy young subjects: a cross-sectional study. Nutrients 2019 Mar 13;11(3):612 [FREE Full text] [doi: 10.3390/nu11030612] [Medline: 30871197]
11. Mamerow MM, Mettler JA, English KL, Casperson SL, Arentson-Lantz E, Sheffield-Moore M, et al. Dietary protein distribution positively influences 24-h muscle protein synthesis in healthy adults. J Nutr 2014 Jun;144(6):876-880 [FREE Full text] [doi: 10.3945/jn.113.185280] [Medline: 24477298]
12. Lee TS, Kim JS, Hwang YJ, Park YC. Habit of eating breakfast is associated with a lower risk of hypertension. J Lifestyle Med 2016 Sep;6(2):64-67 [FREE Full text] [doi: 10.15280/jlm.2016.6.2.64] [Medline: 27924285]
13. Ballon A, Neuenschwander M, Schlesinger S. Breakfast skipping is associated with increased risk of type 2 diabetes among adults: a systematic review and meta-analysis of prospective cohort studies. J Nutr 2019 Jan 01;149(1):106-113. [doi: 10.1093/jn/nxy194] [Medline: 30418612]
14. Li Z, Xu L, Dai R, Li L, Wang H. Effects of regular breakfast habits on metabolic and cardiovascular diseases: A protocol for systematic review and meta-analysis. Medicine (Baltimore) 2021 Nov 05;100(44):e27629 [FREE Full text] [doi: 10.1097/MD.0000000000027629] [Medline: 34871228]
15. Yasuda J, Tomita T, Arimitsu T, Fujita S. Evenly distributed protein intake over 3 meals augments resistance exercise-induced muscle hypertrophy in healthy young men. J Nutr 2020 Jul 01;150(7):1845-1851 [FREE Full text] [doi: 10.1093/jn/nxaa101] [Medline: 32321161]

16. Aoyama S, Kim H, Hirooka R, Tanaka M, Shimoda T, Chijiki H, et al. Distribution of dietary protein intake in daily meals influences skeletal muscle hypertrophy via the muscle clock. Cell Rep 2021 Jul 06;36(1):109336 [FREE Full text] [doi: 10.1016/j.celrep.2021.109336] [Medline: 34233179]
17. Burd N, Tang J, Moore D, Phillips S. Exercise training and protein metabolism: influences of contraction, protein intake, and sex-based differences. J Appl Physiol (1985) 2009 May;106(5):1692-1701 [FREE Full text] [doi: 10.1152/japplphysiol.91351.2008] [Medline: 19036897]
18. National Health and Nutrition Survey 2019. Official Statistics of Japan. URL: https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450171&tstat=000001041744&cycle=7&year=20190&month=0&tclass1=000001148507 [accessed 2022-03-06]
19. Asken Inc. URL: https://www.asken.inc/eng [accessed 2022-06-23]
20. Shimpo M, Fukkoshi Y, Akamatsu R. Correlations between self-efficacy in resisting six temptations and dietary energy and macronutrient intake at each meal. Eat Behav 2014 Dec;15(4):563-566 [FREE Full text] [doi: 10.1016/j.eatbeh.2014.08.012] [Medline: 25215474]
21. Matsuzaki E, Michie M, Kawabata T. Validity of nutrient intakes derived from an internet website dish-based dietary record for self-management of weight among Japanese women. Nutrients 2017 Sep 24;9(10):1058 [FREE Full text] [doi: 10.3390/nu9101058] [Medline: 28946648]
22. Shinozaki N, Murakami K. Evaluation of the ability of diet-tracking mobile applications to estimate energy and nutrient intake in Japan. Nutrients 2020 Oct 29;12(11):3327 [FREE Full text] [doi: 10.3390/nu12113327] [Medline: 33138088]
23. The National Health and Nutrition Survey in Japan, 2019. Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/kenkou/eiyou/r1-houkoku_00002.html [accessed 2021-10-05]
24. Craig C, Marshall A, Sjostrom M, Bauman A, Booth M, Ainsworth B, et al. International physical activity questionnaire: 12-country reliability and validity. Medicine & Science in Sports & Exercise 2003;35(8):1381-1395 [FREE Full text] [doi: 10.1249/01.mss.0000078924.61453.fb]
25. IPAQ scoring protocol. The IPAQ Group. URL: https://sites.google.com/site/theipaq/scoring-protocol [accessed 2021-09-13]
26. Physical Activity Standards for Health Promotion, 2013. Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/stf/houdou/2r9852000002xple.html [accessed 2021-09-13]
27. Rosenfeld CS. Sex-dependent differences in voluntary physical activity. J Neurosci Res 2017 Jan 02;95(1-2):279-290 [FREE Full text] [doi: 10.1002/jnr.23896] [Medline: 27870424]
28. Aoyama S, Nakahata Y, Shinohara K. Chrono-nutrition has potential in preventing age-related muscle loss and dysfunction. Front Neurosci 2021;15:659883 [FREE Full text] [doi: 10.3389/fnins.2021.659883] [Medline: 33935640]
29. Norton C, Toomey C, McCormack W, Francis P, Saunders J, Kerin E, et al. Protein supplementation at breakfast and lunch for 24 weeks beyond habitual intakes increases whole-body lean tissue mass in healthy older adults. J Nutr 2016 Jan;146(1):65-69 [FREE Full text] [doi: 10.3945/jn.115.219022] [Medline: 26581685]
30. Moon J, Kong M, Kim H. Low muscle mass and depressed mood in Korean adolescents: a cross-sectional analysis of the fourth and fifth Korea national health and nutrition examination surveys. J Korean Med Sci 2018 Dec 10;33(50):e320 [FREE Full text] [doi: 10.3346/jkms.2018.33.e320] [Medline: 30534032]
31. Noh H, Park Y. Handgrip strength, dynapenia, and mental health in older Koreans. Sci Rep 2020 Mar 04;10(1):4004-4004 [FREE Full text] [doi: 10.1038/s41598-020-60835-4] [Medline: 32132581]
32. Khanna P, Aeri BT. Association of quantity and quality of protein intake with depression and anxiety symptoms among adolescent boys and girls (13–15 years) studying in public schools of Delhi. J Nutr Sci Vitaminol 2020;66(Supplement):S141-S148 [FREE Full text] [doi: 10.3177/jnsv.66.s141]
33. Takae R, Hatamoto Y, Yasukata J, Kose Y, Komiyama T, Ikenaga M, et al. Physical activity and/or high protein intake maintains fat-free mass in older people with mild disability; the Fukuoka Island City Study: a cross-sectional study. Nutrients 2019 Oct 29;11(11):2595 [FREE Full text] [doi: 10.3390/nu11112595] [Medline: 31671741]
34. An K. Physical activity level in Korean adults: the Korea National Health and Nutrition Examination Survey 2017. Epidemiol Health 2019;41:e2019047 [FREE Full text] [doi: 10.4178/epih.e2019047] [Medline: 31801321]
35. Van Walleghen EL, Orr JS, Gentile CL, Davy KP, Davy BM. Habitual physical activity differentially affects acute and short-term energy intake regulation in young and older adults. Int J Obes (Lond) 2007 Aug 6;31(8):1277-1285. [doi: 10.1038/sj.ijo.0803579] [Medline: 17342074]
36. Gaytán-González A, Ocampo-Alfaro MDJ, Torres-Naranjo F, González-Mendoza RG, Gil-Barreiro M, Arroniz-Rivera M, et al. Dietary protein intake patterns and inadequate protein intake in older adults from four countries. Nutrients 2020 Oct 16;12(10):3156 [FREE Full text] [doi: 10.3390/nu12103156] [Medline: 33081081]
37. Kinoshita K, Otsuka R, Takada M, Tsukamoto-Yasui M, Nishita Y, Tange C, et al. Low amino acid score of breakfast is associated with the incidence of cognitive impairment in older Japanese adults: a community-based longitudinal study. J Prev Alzheimers Dis 2022;9(1):151-157. [doi: 10.14283/jpad.2021.25] [Medline: 35098986]
38. Kinoshita K, Otsuka R, Nishita Y, Tange C, Tomida M, Zhang S, et al. Breakfast protein quality and muscle strength in Japanese older adults: a community-based longitudinal study. J Am Med Dir Assoc 2022 May;23(5):729-735.e2 [FREE Full text] [doi: 10.1016/j.jamda.2021.11.037] [Medline: 35007507]

39. Tahara Y, Shinto T, Inoue K, Roshanmehr F, Ito A, Michie M, et al. Changes in sleep phase and body weight of mobile health app users during COVID-19 mild lockdown in Japan. Int J Obes (Lond) 2021 Oct;45(10):2277-2280 [FREE Full text] [doi: 10.1038/s41366-021-00890-7] [Medline: 34218262]
40. Mineshita Y, Kim H, Chijiki H, Nanba T, Shinto T, Furuhashi S, et al. Screen time duration and timing: effects on obesity, physical activity, dry eyes, and learning ability in elementary school children. BMC Public Health 2021 Feb 28;21(1):422 [FREE Full text] [doi: 10.1186/s12889-021-10484-7] [Medline: 33639912]

Abbreviations

IPAQ: International Physical Activity Questionnaire
MET: metabolic equivalents
NNSJ: National Nutrition Survey in Japan

Edited by H Bradley; submitted 21.12.21; peer-reviewed by K Heinrich, DW Waqar Ali; comments to author 24.02.22; revised version received 23.03.22; accepted 25.05.22; published 12.07.22.

Please cite as:
Shinto T, Makino S, Tahara Y, Nitta L, Kuwahara M, Tada A, Abe N, Michie M, Shibata S
Relationship Between Protein Intake in Each Traditional Meal and Physical Activity: Cross-sectional Study
JMIR Public Health Surveill 2022;8(7):e35898
URL: https://publichealth.jmir.org/2022/7/e35898
doi:10.2196/35898
PMID:35819831

Corrigenda and Addenda

Correction: Reasons for COVID-19 Vaccine Hesitancy Among Chinese People Living With HIV/AIDS: Structural Equation Modeling Analysis

Yan Yao[1*], PhD; Ruiyu Chai[1*], MM; Jianzhou Yang[2], PhD; Xiangjun Zhang[3], PhD; Xiaojie Huang[4], PhD; Maohe Yu[5], PhD; Geng-feng Fu[6], PhD; Guanghua Lan[7], PhD; Ying Qiao[8], MM; Qidi Zhou[9], PhD; Shuyue Li[10], MM; Junjie Xu[11], PhD

[1]Department of Epidemiology and Biostatistics, School of Public Health, Jilin University, Changchun, China
[2]Department of Preventive Medicine, Changzhi Medical College, Changzhi, China
[3]Department of Public Health, University of Tennessee, Knoxville, TN, United States
[4]Beijing Youan Hospital, Capital Medical University, Beijing, China
[5]Tianjin Centers for Disease Control and Prevention, Tianjin, China
[6]Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China
[7]Guangxi Center for Disease Prevention and Control, Nanning, China
[8]The Second Hospital of Huhhot, Huhhot, China
[9]Department of Emergency Medicine, Peking University Shenzhen Hospital, Peking University, Shenzhen, China
[10]Changchun Maternity Hospital, Changchun, China
[11]Clinical Research Academy, Peking University Shenzhen Hospital, Peking University, Shenzhen, China
[*]these authors contributed equally

Corresponding Author:
Junjie Xu, PhD
Clinical Research Academy, Peking University Shenzhen Hospital
Peking University
No. 1120, Lianhua Road, Futian District
Shenzhen, 518036
China
Phone: 86 (755) 8392333
Email: xjjcmu@163.com

Related Article:

Correction of: https://formative.jmir.org/2022/6/e33995

(JMIR Public Health Surveill 2022;8(7):e40910) doi:10.2196/40910

In “Reasons for COVID-19 Vaccine Hesitancy Among Chinese People Living With HIV/AIDS: Structural Equation Modeling Analysis” (JMIR Public Health Surveill 2022;8(6):e33995) the authors made the following updates:

1. In the originally published article, author Yan Yao was inadvertently listed as the eleventh author. In the corrected version, author Yan Yao is listed as the first author.

2. In the originally published article, the note of equal contribution was attributed to three authors: Ruiyu Chai, Jianzhou Yang, and Xiangjun Zhang. In the corrected version, the equal contribution note is attributed to authors Yan Yao and Ruiyu Chai.

3. Accordingly, the corrected order of authors is as follows:

> *Yan Yao[1*], PhD; Ruiyu Chai[1*], MM; Jianzhou Yang[2], PhD; Xiangjun Zhang[3], PhD; Xiaojie Huang[4], PhD; Maohe Yu[5], PhD; Geng-feng Fu[6], PhD; Guanghua Lan[7], PhD; Ying Qiao[8], MM; Qidi Zhou[9], PhD; Shuyue Li[10], MM; Junjie Xu[11], PhD*
>
> *[*]these authors contributed equally.*

4. In the originally published article, author Yan Yao was erroneously listed as the corresponding author. The corresponding authorship is now correctly attributed to Junjie Xu.

5. Accordingly, the address and contact details of the corresponding author have been changed as follows:

> *Junjie Xu, Phd*
>
> *Clinical Research Academy*

> *Peking University Shenzhen Hospital*
> *Peking University*
> *1120, Lianhua Road, Futian District*
> *Shenzhen, 518036, China*
> *Phone: 86 755 83923333*
> *Email: xjjcmu@163.com.*

6. In the originally published article, the *Authors' Contributions* section was as follows:

> *RC, YY, and JX conceptualized the study and wrote, reviewed, and edited the manuscript. RC, XH, XZ, SL, and YY designed the methodology. JY, XH, MY, GF, GL, JY, and QZ curated the data. RC, JY, and YY performed the formal analysis. JY, XH, MY, GF, and GL performed project administration. XH and JY obtained the resources. QZ, SL,YY, and JX supervised the study. RC, JY, and XZ contributed equally as first authors. QZ, SL, YY, and JX contributed equally as corresponding authors. All authors have read and approved the final manuscript as submitted.*

The corrected *Authors' Contributions* is as follows:

> *RC, YY, and JX conceptualized the study and wrote, reviewed, and edited the manuscript. RC, XH, XZ, SL, and YY designed the methodology. JY, XH, MY, GF, GL, QY, and QZ curated the data. RC, JY, and YY performed the formal analysis. JY, XH, MY, GF, and GL performed project administration. XH and JY obtained the resources. QZ, SL,YY, and JX supervised the study. YY and RC contributed equally as first authors. JY, XZ, SL and JX contributed equally as corresponding authors. All authors have read and approved the final manuscript as submitted.*

The correction will appear in the online version of the paper on the JMIR Publications website on July 29, 2022 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 09.07.22; this is a non–peer-reviewed article;accepted 15.07.22; published 29.07.22.

Please cite as:
Yao Y, Chai R, Yang J, Zhang X, Huang X, Yu M, Fu GF, Lan G, Qiao Y, Zhou Q, Li S, Xu J
Correction: Reasons for COVID-19 Vaccine Hesitancy Among Chinese People Living With HIV/AIDS: Structural Equation Modeling Analysis
JMIR Public Health Surveill 2022;8(7):e40910
URL: https://publichealth.jmir.org/2022/7/e40910
doi:10.2196/40910
PMID:35905496

Short Paper

Publication and Impact of Preprints Included in the First 100 Editions of the CDC COVID-19 Science Update: Content Analysis

Jeremy Otridge[1*], BA; Cynthia L Ogden[1*], PhD; Kyle T Bernstein[1], MAS, PhD; Martha Knuth[1], MLiS; Julie Fishman[1], MPH; John T Brooks[1], MD

Centers for Disease Control and Prevention, Atlanta, GA, United States

[*]these authors contributed equally

Corresponding Author:
Jeremy Otridge, BA
Centers for Disease Control and Prevention
1600 Clifton Road
Atlanta, GA, 30329
United States
Phone: 1 2406874849
Email: jeremy.otridge@gmail.com

Abstract

Background: Preprints are publicly available manuscripts posted to various servers that have not been peer reviewed. Although preprints have existed since 1961, they have gained increased popularity during the COVID-19 pandemic due to the need for immediate, relevant information.

Objective: The aim of this study is to evaluate the publication rate and impact of preprints included in the Centers for Disease Control and Prevention (CDC) COVID-19 Science Update and assess the performance of the COVID-19 Science Update team in selecting impactful preprints.

Methods: All preprints in the first 100 editions (April 1, 2020, to July 30, 2021) of the Science Update were included in the study. Preprints that were not published were categorized as “unpublished preprints.” Preprints that were subsequently published exist in 2 versions (in a peer-reviewed journal and on the original preprint server), which were analyzed separately and referred to as “peer-reviewed preprint” and “original preprint,” respectively. Time to publish was the time interval between the date on which a preprint was first posted and the date on which it was first available as a peer-reviewed article. Impact was quantified by Altmetric Attention Score and citation count for all available manuscripts on August 6, 2021. Preprints were analyzed by publication status, publication rate, preprint server, and time to publication.

Results: Of the 275 preprints included in the CDC COVID-19 Science Update during the study period, most came from three servers: medRxiv (n=201, 73.1%), bioRxiv (n=41, 14.9%), and SSRN (n=25, 9.1%), with 8 (2.9%) coming from other sources. Additionally, 152 (55.3%) were eventually published. The median time to publish was 2.3 (IQR 1.4-3.7). When preprints posted in the last 2.3 months were excluded (to account for the time to publish), the publication rate was 67.8%. Moreover, 76 journals published at least one preprint from the CDC COVID-19 Science Update, and 18 journals published at least three. The median Altmetric Attention Score for unpublished preprints (n=123, 44.7%) was 146 (IQR 22-552) with a median citation count of 2 (IQR 0-8); for original preprints (n=152, 55.2%), these values were 212 (IQR 22-1164) and 14 (IQR 2-40), respectively; for peer-review preprints, these values were 265 (IQR 29-1896) and 19 (IQR 3-101), respectively.

Conclusions: Prior studies of COVID-19 preprints found publication rates between 5.4% and 21.1%. Preprints included in the CDC COVID-19 Science Update were published at a higher rate than overall COVID-19 preprints, and those that were ultimately published were published within months and received higher attention scores than unpublished preprints. These findings indicate that the Science Update process for selecting preprints had a high fidelity in terms of their likelihood to be published and their impact. The incorporation of high-quality preprints into the CDC COVID-19 Science Update improves this activity’s capacity to inform meaningful public health decision-making.

(JMIR Public Health Surveill 2022;8(7):e35276) doi:10.2196/35276

KEYWORDS

preprints; preprint; publishing; publish; bioRxiv; medRxiv; Centers for Disease Control and Prevention; CDC; preprint server; public health; health information; COVID-19; pandemic; publication; Altmetric attention score; Altmetric; attention score; citation count; citation; science update; decision-making

Introduction

Preprints are publicly available, non-peer reviewed articles posted to various servers such as medRXiv and bioRxiv. Preprints first appeared as information exchange groups in 1961. However, due to resistance from journals, information exchange groups closed in 1967 [1]. In 1991, an automated email server that later became arXiv, a popular modern preprint server, was established [1]. In 2019, Cold Spring Harbor Laboratory collaborated with Yale University to launch medRxiv [1]. Preprints have gained popularity and credibility during the COVID-19 pandemic due to the need for rapid, relevant information to respond to the pandemic. To help inform the public health response to COVID-19, the Centers for Disease Control and Prevention (CDC) created the COVID-19 Science Update [2].

The COVID-19 Science Update provides brief summaries of new COVID-19–related articles on topics such as health equity, vaccines, variants, natural history, and testing, among others [2]. To provide the most relevant and timely information, the COVID-19 Science Update includes both published (peer-reviewed) articles and manuscript preprints. In collaboration with the World Health Organization (WHO), the Stephen B. Thacker CDC Library does a daily systematic (exhaustive, reproducible, and defensible) search for all COVID-19–related articles, which are then cleaned and deduplicated to be included in the WHO COVID-19 Database. The full search strategy for creating this database can be found on the WHO COVID-19 Database website [3]. The resulting articles are sent to the Science Update team within the CDC COVID-19 Response, who then select peer-reviewed articles and preprints on public health priority topics in the CDC Science Agenda for COVID-19 [4] and CDC COVID-19 Response Health Equity Strategy [5].

The COVID-19 Science Update was piloted biweekly beginning in April 2020 and has been publicly available since September 2020 (previous editions were retroactively posted on the internet) [2]. In November 2020, public release became weekly, and an email subscription became available in February 2021. The objective of this analysis is to evaluate the publication rate (percent of preprints published in peer-reviewed journals) and impact (eg, Altmetric Attention Score [6] and citation count) of preprints included in the COVID-19 Science Update and to assess the performance of the COVID-19 Science Update team in selecting impactful preprints.

Methods

All preprints in the first 100 editions (April 1, 2020, to July 30, 2021) of the COVID-19 Science Update were included in this analysis. Some preprints were eventually published (categorized as “peer-reviewed preprint”) whereas others were not published (categorized as “unpublished preprint”). Time to publish was the time interval between the date on which a preprint was first posted and the date on which it was first available as a peer-reviewed article. Impact was quantified (median and interquartile range) using both Altmetric Attention Score (a weighted measure of attention received from academic, news, and social media sources) and citation count (also from Altmetric) for all available articles (peer reviewed or not) on August 6, 2021. For peer-reviewed preprints, impact metrics only measure impact accumulated following publication. To allow time for preprints to be published and to accumulate impact, we restricted comparisons to items included in the 2020 editions (through Edition 70) of the COVID-19 Science Update. Statistical comparisons (Mood median test) were performed in Minitab with $P<.05$ for statistical significance. Preprints were analyzed by publication status, publication rate, preprint server, and time to publication.

Results

Among the 1,971 articles in the COVID-19 Science Update among the analysis period, 275 (14%) were preprints (Figure 1), most of which came from 1 of the following 3 servers: medRxiv (n=201, 73.1%), bioRxiv (n=41, 14.9%), or Lancet preprints with SSRN (n=25, 9.1%), with 8 (2.9%) coming from other sources. More than half (152/275, 55.3%) were published within the analysis period (April 1, 2020, to August 6, 2021). The median time to publish was 2.3 months (IQR 1.4-3.7). When preprints posted in the last 2.3 months were excluded (to account for the time to publish), the publication rate was 67.8% (143/211). Preprints included in the COVID-19 Science Update were published in 76 different journals, and 18 journals published at least three. Moreover, 2 journals (New England Journal of Medicine and Clinical Infectious Diseases) published 11 preprints each.

Peer-reviewed articles included in the COVID-19 Science Update (n=1696) had a median Altmetric Attention Score of 365 (IQR 64-1316) and median citation count of 33 (IQR 9-111) (Table 1). For peer-reviewed preprints (n=152), median Altmetric Attention Score was 265 (IQR 29-1896) and median citation count was 19 (IQR 3-101). For unpublished preprints (n=123), Median Altmetric Attention Score was 146 (IQR 22-552) and median citation count was 2 (IQR 0-8).

To account for the differences in time that articles have been publicly available, the analytic sample was restricted to only published articles (n=1,140) and preprints (n=73, 50 peer reviewed and 23 unpublished) that were included in the 2020 editions of the COVID-19 Science Update (through edition 70; Table 1). Among the 73 preprints, 50 (68.4%) were published. The median Altmetric Attention Score for published articles (328, IQR 57-1224) was higher than that of unpublished preprints (83, IQR 9-231; P=.002). The difference in Altmetric Attention Score between peer-reviewed (260, IQR 43-1817) and unpublished preprints was not statistically significant, likely due to small sample sizes (*P*=.09). The median citation counts for peer-reviewed preprints (55, IQR 9-148) and published articles (43, IQR 15-149) were significantly higher (*P*<.001) than that for unpublished preprints (4, IQR 1-9). There was no significant difference in citation count between peer-reviewed preprints and published articles (*P*=.23).

Figure 1. Breakdown of the types of articles included in the first 100 editions (April 2020 to July 2021) of the CDC (Centers for Disease Control and Prevention) COVID-19 Science Update. The left pie displays the overall number of preprints and published articles. The right pie shows how many of the preprints in the left pie were peer reviewed or remained unpublished as of August 6, 2021.

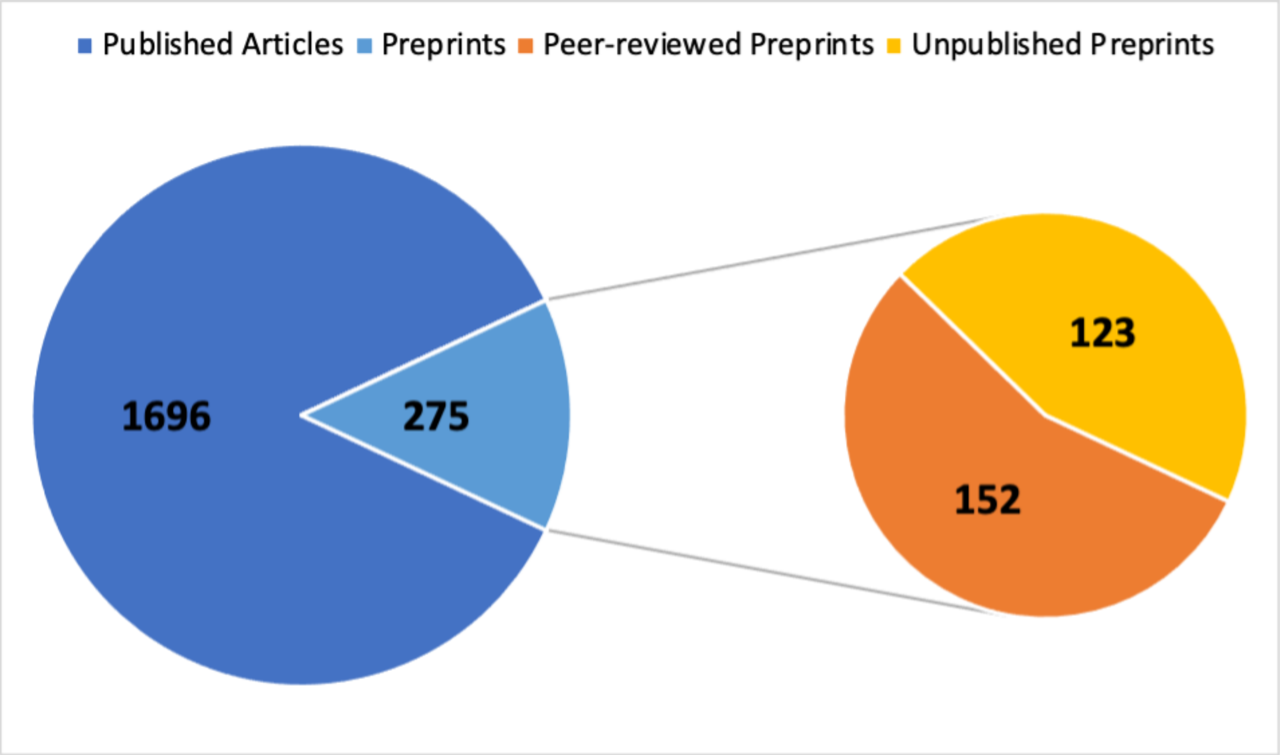


Table 1. Median Altmetric Attention Score and citation count for articles and preprints included in the first 100 editions (April 2020 to July 2021) of the CDC (Centers for Disease Control and Prevention) COVID-19 Science Update, as of August 6, 2021.

Article type	Articles, n (%)	Median Altmetric Attention Score (IQR)	Median citation count, n (IQR)
Editions 1-100			
Total	1971	324 (56-1249)	28 (7-93)
Published	1696 (86)	365 (64-1316)	33 (9-111)
Preprints	275 (14)	174 (22-899)	6 (1-29)
Peer reviewed[a]	152 (8)	265 (29-1896)	19 (3-101)
Unpublished	123 (6)	146 (22-552)	2 (0-8)
Editions 1-70 (2020), to account for time to accumulate impact			
Total	1213	312 (55-1194)	43 (14-147)
Published	1140 (94)	328 (57-1224)[b]	43 (15-149)[b]
Preprints	73 (6)	150 (25-866)	15 (4-72)
Peer reviewed[a]	50 (4)	260 (43-1817)	55 (9-148)[b]
Unpublished	23 (2)	83 (9-231)	4 (1-9)

[a] Two versions of these papers exist: the original preprint that remains on the preprint server and the peer-reviewed version that is published in a journal

[b] *P*<.05 for difference in Altmetric Attention Score or citation count from unpublished preprints. Testing was only carried out for articles in editions 1-70 to account for the differences in time that articles have had to accumulate impact.

Discussion

Prior analyses of COVID-19 preprints have found publication rates between 5.7% and 21.1% [7,8]. Preprints included in the COVID-19 Science Update were published at a higher rate than reported elsewhere [7,8], and those that were ultimately published received higher attention scores than unpublished preprints. A high Altmetric Attention Score is indicative of only the total attention a publication receives. It does not differentiate between positive and negative attention, so a high Altmetric Attention Score could equally reflect a publication of high scientific value or one that is widely refuted [6]. Preprints facilitate rapid access to information; however, they are ultimately limited by the absence of peer review and are thus subject to change and may never be published. The ability to discern high-quality impactful preprints is therefore an important tool for public health decision-making. Despite these limitations, the findings of our analysis indicate that the COVID-19 Science Update process for selecting preprints is robust, with high fidelity in terms of the likelihood of preprints to be published and to be impactful. The incorporation of high-quality preprints into the COVID-19 Science Update improves this activity’s capacity to provide timely information, which could have an impact on meaningful public health decision-making.

Acknowledgments

This project was supported in part by an appointment (JO) to the Research Participation Program at the Centers for Disease Control and Prevention (CDC) administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and CDC. The authors wish to thank the Science Update team within the Office of the Chief Medical Officer of the CDC COVID-19 Response.

Disclaimer

The findings and conclusions in this report are those of the authors and are not necessarily the official position of the Centers for Disease Control and Prevention.

Conflicts of Interest

All authors except JO, who was an ORISE fellow placed in the Office of Library Science at CDC, are US government employees (CDC).

References

1. Cobb M. The prehistory of biology preprints: A forgotten experiment from the 1960s. PLoS Biol 2017 Nov 16;15(11):e2003995 [FREE Full text] [doi: 10.1371/journal.pbio.2003995] [Medline: 29145518]
2. COVID-19 Science Update. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/library/covid19/scienceupdates.html [accessed 2022-07-11]
3. WHO COVID-19 Research Database: user guide and information. World Health Organization. URL: https://www.who.int/publications/m/item/quick-search-guide-who-covid-19-database [accessed 2021-10-14]
4. CDC Public Health Science Agenda for COVID-19. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/science/science-agenda-covid19.html [accessed 2022-07-12]
5. Health Equity: Promoting Fair Access to Health. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/index.html [accessed 2022-07-12]
6. What does Altmetric do? Altmetric. URL: https://www.altmetric.com [accessed 2021-10-19]
7. Añazco D, Nicolalde B, Espinosa I, Camacho J, Mushtaq M, Gimenez J, et al. Publication rate and citation counts for preprints released during the COVID-19 pandemic: the good, the bad and the ugly. PeerJ 2021;9:e10927 [FREE Full text] [doi: 10.7717/peerj.10927] [Medline: 33717688]
8. Fraser N, Brierley L, Dey G, Polka JK, Pálfy M, Nanni F, et al. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. PLoS Biol 2021 Apr;19(4):e3000959 [FREE Full text] [doi: 10.1371/journal.pbio.3000959] [Medline: 33798194]

Abbreviations

CDC: Centers for Disease Control and Prevention
WHO: World Health Organization

Edited by H Bradley; submitted 29.11.21; peer-reviewed by S Arnesen, D Hu, I Mircheva; comments to author 24.02.22; revised version received 01.03.22; accepted 10.05.22; published 15.07.22.

Please cite as:
Otridge J, Ogden CL, Bernstein KT, Knuth M, Fishman J, Brooks JT
Publication and Impact of Preprints Included in the First 100 Editions of the CDC COVID-19 Science Update: Content Analysis
JMIR Public Health Surveill 2022;8(7):e35276
URL: https://publichealth.jmir.org/2022/7/e35276
doi:10.2196/35276
PMID:35544426

Original Paper

Causal Modeling to Mitigate Selection Bias and Unmeasured Confounding in Internet-Based Epidemiology of COVID-19: Model Development and Validation

Nathaniel Stockham[1], MSc; Peter Washington[2], PhD; Brianna Chrisman[2], PhD; Kelley Paskov[3], MSc; Jae-Yoon Jung[4], PhD; Dennis Paul Wall[4,5], PhD

[1]Neurosciences Interdepartmental Program, Stanford University, Palo Alto, CA, United States
[2]Department of Bioengineering, Stanford University, Stanford, CA, United States
[3]Biomedical Informatics Program, Stanford University, Stanford, CA, United States
[4]Department of Biomedical Data Science, Stanford University, Stanford, CA, United States
[5]Department of Pediatrics, Stanford University, Stanford, CA, United States

Corresponding Author:
Nathaniel Stockham, MSc
Neurosciences Interdepartmental Program
Stanford University
3145 Porter Dr
Palo Alto, CA, 94304
United States
Phone: 1 2056021832
Email: stockham@stanford.edu

Abstract

Background: Selection bias and unmeasured confounding are fundamental problems in epidemiology that threaten study internal and external validity. These phenomena are particularly dangerous in internet-based public health surveillance, where traditional mitigation and adjustment methods are inapplicable, unavailable, or out of date. Recent theoretical advances in causal modeling can mitigate these threats, but these innovations have not been widely deployed in the epidemiological community.

Objective: The purpose of our paper is to demonstrate the practical utility of causal modeling to both detect unmeasured confounding and selection bias and guide model selection to minimize bias. We implemented this approach in an applied epidemiological study of the COVID-19 cumulative infection rate in the New York City (NYC) spring 2020 epidemic.

Methods: We collected primary data from Qualtrics surveys of Amazon Mechanical Turk (MTurk) crowd workers residing in New Jersey and New York State across 2 sampling periods: April 11-14 and May 8-11, 2020. The surveys queried the subjects on household health status and demographic characteristics. We constructed a set of possible causal models of household infection and survey selection mechanisms and ranked them by compatibility with the collected survey data. The most compatible causal model was then used to estimate the cumulative infection rate in each survey period.

Results: There were 527 and 513 responses collected for the 2 periods, respectively. Response demographics were highly skewed toward a younger age in both survey periods. Despite the extremely strong relationship between age and COVID-19 symptoms, we recovered minimally biased estimates of the cumulative infection rate using only primary data and the most compatible causal model, with a relative bias of +3.8% and –1.9% from the reported cumulative infection rate for the first and second survey periods, respectively.

Conclusions: We successfully recovered accurate estimates of the cumulative infection rate from an internet-based crowdsourced sample despite considerable selection bias and unmeasured confounding in the primary data. This implementation demonstrates how simple applications of structural causal modeling can be effectively used to determine falsifiable model conditions, detect selection bias and confounding factors, and minimize estimate bias through model selection in a novel epidemiological context. As the disease and social dynamics of COVID-19 continue to evolve, public health surveillance protocols must continue to adapt; the emergence of Omicron variants and shift to at-home testing as recent challenges. Rigorous and transparent methods to develop, deploy, and diagnosis adapted surveillance protocols will be critical to their success.

(JMIR Public Health Surveill 2022;8(7):e31306) doi:10.2196/31306

KEYWORDS

selection bias; COVID-19; epidemiology; causality; sensitivity analysis; public health; surveillance; method; epidemiologic research design; model; bias; development; validation; utility; implementation; sensitivity; design; research; epidemiology

Introduction

Accurate estimation of disease parameters is a fundamental problem in epidemiology. The internal and external validity of epidemiological studies is threatened by unmeasured confounding and selection bias [1,2]. There is an extensive and sophisticated literature focused on mitigating these threats by study design and poststudy statistical adjustment [3-5]. In particular, the randomization paradigm for treatment assignment and sample selection has served at the de facto standard for identifying causal effects and point estimates of disease parameters in a target population. However, even studies with perfect randomization can still suffer from unmeasured confounding and selection bias via a variety of phenomena, such as participant noncompliance, unit nonresponse, incomplete registers of the target population, and data collection failures [6]. In the past decade, there have been several advances in the theoretical treatment of these threats, particularly in the graphical causal modeling literature, where the problems of selection bias and unmeasured confounding have received a comprehensive theoretical treatment [7-9]. Although these recent methods provide a clear conceptual and mathematical framework, they have yet to be routinely deployed in the epidemiological community at large [10,11].

This gap is particularly acute in internet-based public health and surveillance. Internet-based sampling in general suffers from unknown selection mechanisms on largely unobservable and dynamic populations, making traditional adjustment methods that require external data about the target population vulnerable to model violation. Previous studies that augmented traditional surveillance mechanisms with internet-based data have proved highly successful at imputing missing or time-delayed information [12,13]. However, it is challenging to model emerging pathogens and adapt to changing internet user behavior across time and social context [14]. This fundamental difficulty was demonstrated vividly early in the COVID-19 pandemic as several highly sophisticated crowdsourced internet-based surveillance efforts were launched in response to the pandemic [15-21]. Despite explicit support by global social media and web service vendors, these early efforts yielded significantly biased estimates of key epidemiological parameters [22-26]. Internet-based epidemiology must adopt methodological approaches appropriate to the dynamic and unobservable features of internet populations.

In this work, we seek to address this gap between recent theoretical developments and the current practice of internet-based public health surveillance. We present structural causal modeling as a guide to epidemiological judgement through encoding epidemiological knowledge into models that can be tested using sample data, and we describe a general graphical method for deriving falsifiable model conditions. Importantly, this approach can be deployed using only the sampled data, whereas traditional methods for detecting confounding and selection bias require some information about the unsampled or missing data from units with partial data or external data, such as census or health care system medical records [1]. For novel and dynamic phenomena, the required external information may be unavailable, unreliable, or impractical to collect in the timespan available. Our objective is to demonstrate the practical utility of model diagnosis and selection using statistical criteria derived from structural causal models.

Methods

Structural Causal Models

Structured causal models permit the formal encoding of causal mechanisms and have been extended to formally analyze studies in the presence of selection bias and unmeasured confounding. The mathematical tool necessary for this work is d-separation on directed acyclic graphs (DAGs). Here, we briefly review d-separation notation and concepts. We can represent a probability distribution as a DAG where nodes represent variables X, Y, and Z and edges represent functional dependencies between variables. The formalism of d-separation is a mapping between the DAG of a probability distribution and the conditional independencies of that distribution; this is stated formally in the conditional independence statements of Figure 1. To state that variables X and Y are d-separated by Z is to state that X and Y are conditionally independent if conditioned on variable Z. Conversely, if X and Y are not d-separated by Z, then X and Y are conditionally dependent if conditioned on Z. A path in a DAG is a sequence of edges (regardless of direction), and every path can be decomposed into a sequence of path elements of edges, chains, forks, and colliders, as shown in Figure 1. The variables X and Y are d-separated in the DAG if all paths from X to Y in the graph are "blocked." Intuitively, a path from X to Y is blocked if no information about X can be inferred from observing Y via information transferred along that path.

The d-separation path element rules determine whether a path between X and Y is blocked. A path between X and Y can be blocked in 2 different ways by conditioning on a set of variables W. If the path contains a fork or a chain element, then it is blocked if the middle variable (Z in Figure 1) in at least 1 of the fork or chain elements is in W. If the path contains a collider element, then the path is blocked only if the middle variable is not in W. Conditioning on the middle variable of a collider element can unblock the path and make X and Y not d-separated. To illustrate this concept more explicitly, consider 2 independent binary 0,1 variables X and Y, where $Z = X + Y$. If we condition on Z such that $Z=1$, then X and Y would appear anticorrelated (nonindependent) because samples with $X=Y=0$ and $X=Y=1$ are, by definition, never observed in the subset where $Z=1$. This effect is known as *collider bias* and is a major source of selection bias in epidemiological studies. For further reading, there are several good introductions to d-separation in graphical models [6,9].

Figure 1. Conditional independence statements and d-separation rules.

Conditional independence statements:

$$\begin{aligned} &X, Y \text{ d-separated by } Z: && (X \perp\!\!\!\perp Y \mid Z) \iff P(X,Y\mid Z) = P(X\mid Z)P(Y\mid Z) \\ &X, Y \text{ not d-separated by } Z: && (X \not\perp\!\!\!\perp Y \mid Z) \iff P(X,Y\mid Z) \neq P(X\mid Z)P(Y\mid Z) \end{aligned}$$

Path element rules:

$$\begin{aligned} \text{Edge}: X \rightarrow Y &\iff (X \not\perp\!\!\!\perp Y) \\ \text{Chain}: X \rightarrow Z \rightarrow Y &\iff (X \not\perp\!\!\!\perp Y), (X \perp\!\!\!\perp Y \mid Z) && : \text{blocked by } Z \\ \text{Fork}: X \leftarrow Z \rightarrow Y &\iff (X \not\perp\!\!\!\perp Y), (X \perp\!\!\!\perp Y \mid Z) && : \text{blocked by } Z \\ \text{Collider}: X \rightarrow Z \leftarrow Y &\iff (X \perp\!\!\!\perp Y), (X \not\perp\!\!\!\perp Y \mid Z) && : \text{unblocked by } Z \end{aligned}$$

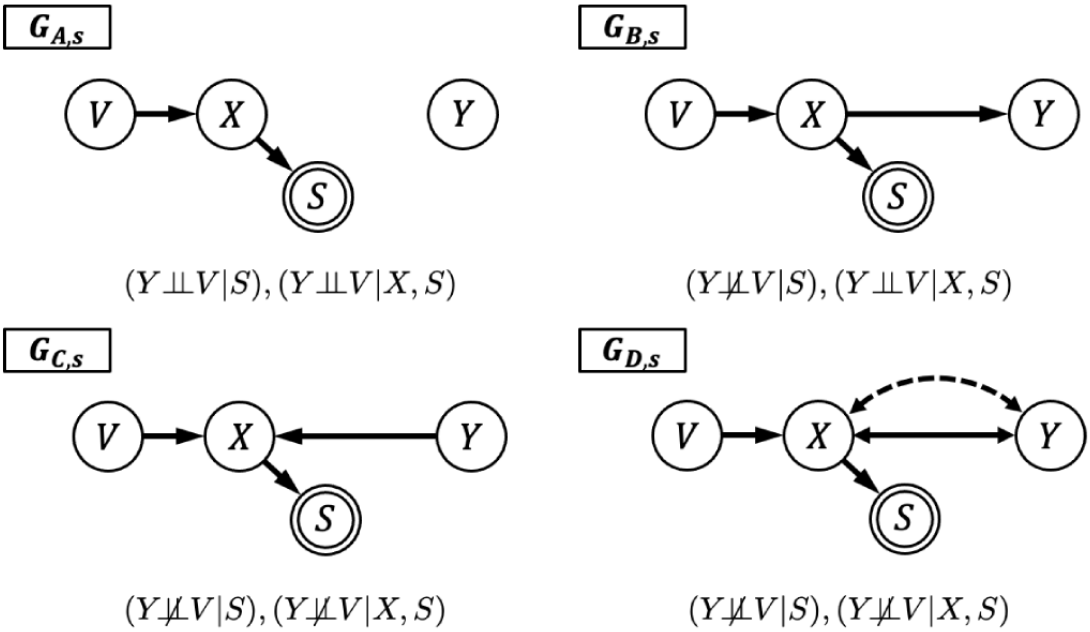
The only additional conceptual step necessary for analyzing selection bias is to add the sampling mechanism to the initial causal graph G to create the augmented causal graph G_s. The encoded sampling mechanism determines the value of the sampling indicator variable S, where S=1 if the unit was sampled and S=0 otherwise. Additionally, any mechanism that filters data after primary collection induces an additional selection bias and must also be encoded in G_s. The augmented graph G_s obeys d-separation rules, but for clarity, the sampling indicator S node is depicted in G_s with a double ring to emphasize that S=1 for all samples by definition; all d-separation statements in G_s must be evaluated conditional on S=1.

For any graph G_s, the s-recoverability condition states that for any variables Y and X in G_s, the distribution of the sample P(Y|X,S=1) is identical to the distribution of the target population P(Y|X) if and only if Y and S are d-separated by X [8]. Assuming the graph is faithful, S and Y conditioned on X are independent if and only if there is no selection bias or unmeasured confounding. It is not possible to directly test for independence between S and Y using sample data, because is no variation as S=1 by definition, but other surrogate variables in the sample data can be used to test independence of S and Y.

We demonstrate this principle in Figure 2 using the causal graphs $G_{A,s}$, $G_{B,s}$, $G_{C,s}$, and $G_{D,s}$ with outcome variable Y, instrumental variable V on variable X relative to Y, and a sample indicator variable S determined by X. Here we define an instrumental variable V on X relative to Y as a variable V that is not independent of X (V and X not d-separated), but V is independent of Y conditional on X (V and Y are d-separated by X). In graph $G_{A,s}$, the variables V and Y are trivially d-separated (d-separated without any blocking variables) and therefore independent. If there is any selection bias ($G_{B,s}$, $G_{C,s}$) or unmeasured confounding ($G_{D,s}$), then V and Y are not trivially d-separable and are not independent. For example, suppose the null hypothesis of statistical independence between V and Y is rejected in the sampled data. Then the graph $G_{A,s}$ is not compatible with the sample data, but alternative graphs $G_{C,s}$, $G_{B,s}$, and $G_{D,s}$ are compatible with nonindependent V and Y and should be considered. Any augmented causal graph G_s entails a set of conditional independencies that can be statistically tested using only the sample data. For more complex graphs, there are several software tools that will compute all the entailed independencies, of which *dagitty* is perhaps the most user friendly [26].

In this work, we focus on graphical modeling as a formalism to aid epidemiological judgment. Epidemiological knowledge tightly constrains the set of possible explanatory scenarios for a given context; the difficulty is choosing which of these scenarios is most plausible. Statistically testing the independencies implied by the causal graph encoding is a direct method to select between scenarios. We now demonstrate this approach in an applied problem of estimating the cumulative infection rate CI_P of SARS-CoV-2 in the COVID-19 New York City (NYC) spring 2020 epidemic through a prospectively collected crowdsourced internet survey.

Figure 2. Example causal graphs with selection bias and unmeasured confounding.



Recruitment

Initial crowdsourced epidemiology efforts in the COVID-19 pandemic focused on surveys collected from a variety of internet sources and target populations. Instead of recruiting via major internet platforms such as Facebook and Google, we recruited our survey participants from the Amazon Mechanical Turk (MTurk) crowdworking platform. MTurk is an internet-based labor market where a research group or business (*requesters*) can create and disseminate a human intelligence task (HIT) to a distributed human labor pool (*workers*) that can accept and complete these tasks for a known monetary reward upon satisfactory completion of the task. A HIT can range from transcribing an audio file to a personality survey, and requesters can restrict the task workers within a specific geographic area or demographic subset. All MTurk workers in the United States are adults of age 18 years or older.

We chose the MTurk population for 2 reasons. First, MTurk has been successfully used by many academic groups, including our own, across a broad array of disciplines [27-32]. Second, the demographics and health status of the MTurk worker population in the United States has been repeatedly characterized and has remained stable through time, closely matching the racial and ethnic composition of the United States but skewing toward women, a younger age, worse mental health, and lower income than the US population [28]. Any MTurk worker registered as residing in New York State or New Jersey was permitted to respond to the survey via the MTurk HIT job posting with restriction that a worker could respond only once per survey period. No other restrictions or invitation mechanisms were used in either survey period.

Human Subject Research Ethical Statement

This research was not found to be considered human subject research as the survey did not collect any personally identifying information or set of information that could be reidentifying, in compliance with MTurk's policy prohibiting any transmission of workers' personally identifiable information to requesters and Stanford University research policy GUI-H12. Research was carried out in a way that followed ethical guidelines set by the Declaration of Helsinki. All MTurk tasks are carefully reviewed before being posted, and MTurk workers are able to accept but then refuse to complete any task or any part of a task at any point in time. Furthermore, the survey task included an introduction page that informed the respondents of the purpose and content of this survey and for what purposes their response data would be used.

Overview of Survey Design

We collected primary data from the MTurk population listed as currently residing in New Jersey or New York. Data for surveys s_1 and s_2 were collected in 2 successive periods: April 11-14 and May 8-11, 2020. During this period, both New York and New Jersey were under a statewide stay-at-home order that greatly restricted travel and prohibited public gatherings [33,34]. We collected primary data from 2 survey periods to estimate the trajectory of the spring 2020 COVID-19 epidemic in NYC and assess the stability of model selection across 2 different phases of the epidemic. The context of NYC in spring 2020 was chosen because it was 1 of the first major COVID-19 epidemics. A Qualtrics survey was created for each survey run, with a reward (median completion time) for the surveys of US $1 (5 minutes) and US $1.25 (6 minutes) for s_1 and s_2, respectively. This reward is consistent with other MTurk HITs for the time required. Before accepting the task, the participant was aware of the overall survey subject (COVID-19) and the monetary reward for completion. We excluded responses from participants that were incomplete or out-of-area, as determined by geolocation information provided by Qualtrics. The included responses were split by collection period and aggregated into 3 nested geographic areas: New Jersey and New York (NJ/NY), the section of the New York City Combined Statistical Area contained within New Jersey and New York (NYC CBSA), and NYC proper.

Before answering any questions, the survey asked each participant (respondent) to privately list their 5 closest peer relationships (relations) with whom they typically socialize in person. There was large variation in the number of contacts for each person during the mandatory stay-at-home orders. Instead of asking respondents about their total number of contacts, we asked about their closest peer relationships because these are the set of persons whose current health status and household characteristics would most likely be known to the respondents. Furthermore, we only asked about 5 relations to minimize the time to complete the survey. The survey queried each respondent about the demographic, employment characteristics, and possible COVID-19 symptoms of both themselves and their relations. The survey also queried each respondent about both their household and their relations' households, including household size and whether any member had a confirmed SARS-CoV-2 infection since March 15, 2020. These questions were chosen to permit comparison of respondents and relations to known census demographic data and to estimate the cumulative number of infected households and individuals within a specified geographic area. The survey material is included in Multimedia Appendices 1 and 2.

Statistical Analysis

Estimator Definition

We defined a household-based cumulative infection rate estimator on a sample P_S for the cumulative infection rate as:

where C_p is an indicator variable for the confirmed SARS-CoV-2 infection status of person p in a population P of size N_P. We defined the household secondary attack rate (SAR_h) as the ratio of secondary household cases to the total population of exposed household members. We can write SAR_h formally as:

where H is the set of unique households in population P, indicator variable C_h=1 for if there is at least 1 SARS-CoV-2 infection in household h, and N_h the size of household h in H. Let the total population be defined as the sum of the household members $N_P=\Sigma_{h\varepsilon H}$ N_h. We can then rewrite CI_P in terms of households as:

We then defined the estimator of CI_P on a sample P_S as:

with unique households H_S. The estimator is consistent as H_S goes to H if $P(C_h, N_h|h \text{ in } H_s)=P(C_h, N_h)$ of the target population. In the special case of SAR_h=0, is an unbiased estimator of the cumulative household infection rate CI_H under the less restrictive condition of $P(C_h|S=1)=P(C_h)$.

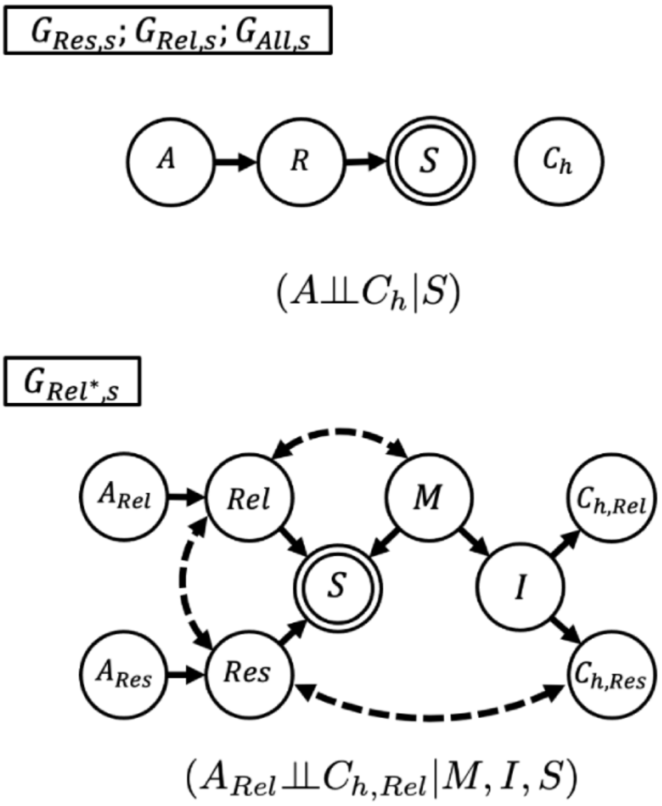
Structural Causal Models

The survey data were modeled in a causal graph encoding the variables and assumptions, as depicted in Figure 3. Every person in the population P was assigned 2 indicator variables Res_i and Rel_{ij} and the outcome variables $C_{h,Res,i}$ and $C_{h,Rel,ij}$. The variable Res_i=1 if person p_i is a respondent to the survey, and Rel_{ij}=1 if p_i in P would choose person p_j as a relation in the context of this survey. In the sample, the set of respondents is $P_{S,Res}=\{p_i|Res_i=1\}$, the set of relations is $P_{S,Rel}=\{p_j|Rel_{ij}=1, p_i \text{ in } P_{S,Res}\}$, and the total sample is $P_S=P_{S,Res} \cup P_{S,Rel}$. The outcome variable $C_{h,Res,i}$=1 if there is at least 1 confirmed SARS-CoV-2 infection in the household of respondent p_i in $P_{S,Res}$, and $C_{h,Rel,ij}$=1 if there is at least 1 confirmed infection in the household of relation p_j in $P_{S,Rel}$. For each response, there is 1 respondent and 5 relations chosen by that respondent. In the causal graphs in Figure 3, the subgraph that includes variables pertaining to relations is replicated identically for all 5 relations. Given that this was an anonymous internet survey, we assumed no information bias due to intentional misrepresentations on the part of the respondents or relations. We also assumed there was no information bias due to testing constraints at the level of the household.

We defined 4 possible causal models depicted in Figure 3, augmented with the sampling indicator variable S. All variables were conditioned on a common geographic area, which was suppressed in the graphs and notation for clarity.

The first 3 causal models $G_{Res,s}$, $G_{Rel,s}$, and $G_{All,s}$ all shared the same causal graph, as represented in the first graph in Figure 3, and differed only in terms of the data used. The first causal graph encoded that the age variable A of a person influences whether they respond as a relation/respondent in variable R, which in turn determines whether they are in the sample S. Furthermore, the household infection status C_h of the person was assumed to be unrelated to the other variables. The first causal model $G_{Res,s}$ only used data on respondents, whereas the second model $G_{Rel,s}$ only used data on relations. The third model $G_{All,s}$ combined respondents' and relations' data.

Figure 3. Alternative causal graphs.



The second graph in Figure 3 corresponds to the fourth causal model $G_{Rel^*,s}$, which encoded a possible confounding between S and C_{Res} and between S and C_{Rel}. In particular, $G_{Rel^*,s}$ modeled the case where the respondents' household status $C_{h,Res}$ is related to the relations' household status $C_{h,Rel}$ through a transmission event I between a respondent and a relation due to the recent close contact event M. The variables I and M d-separated the sampling indicator S and the outcome of interest $C_{h,Rel}$, but the infection event I was unobservable. However, if we filtered the samples so that M=0 (excluding relations with close contact events with the respondent), then I=0 for the retained samples because there can be no transmission without close contact; filtering on M=0 effectively conditioned on M=0 and I=0. Filtering on M also induced another selection bias modeled by the edge M to S, but S and $C_{h,Rel}$ remained d-separated as did $C_{h,Rel}$ and A_{Rel}. Therefore, the graph $G_{Rel^*,s}$ implied that $P(C_{h,Rel}|M=0,I=0,S=1)=P(C_h)$ and furthermore implied that $C_{h,Rel}$ and A_{Rel} are statistically independent. Effectively, this model excluded information from respondents and excluded relations that had recent in-person contact with the respondents. Practically, this model reduced overestimation of the cumulative infection rate due to possible common causes of the infection status of respondents and relations, such as when a respondent transmits an infection to a relation.

These 4 models can be compared and ranked empirically by statistical tests of the conditional independences implied by their d-separation conditions. Each causal model in Figure 3 has at least 1 independence statement that is testable using the observed data on the age and household infection status of the respondents and relations. From 1 statistical test of the independence statement, we can distinguish which causal model is compatible with the data for each survey period using only the survey sample data. This is an important methodological point, given that the current practice for model diagnosis and selection assumes strong prior knowledge about the target population on several demographic variables. No external data are required for this type of diagnostic analysis, which is a key advantage when the target population is unstudied, inaccessible, or dynamic through time.

We evaluated the causal models by statistically testing the implied independence of A; C_h in models $G_{Res,s}$, $G_{Rel,s}$, and $G_{All,s}$; and A_{Rel} and $C_{h,Rel}$ in model $G_{Rel^*,s}$ using the Fisher exact test for independence. For each model, we filtered the data, as specified in the model, median-split the age variable, and performed Fisher exact tests on the 2×2 table of the age group (A_0, A_1) by house infection status ($C_{h,0}$, $C_{h,1}$) with point test statistics shown in Table 1. Ages for respondents and relations were randomly assigned within the recorded 5-year bin across independent replications (n=1000). The median odds ratio with 95% CIs and median *P* values are reported in Table 1.

Table 1. Model selection by conditional independence tests.

Survey	Odds ratio (95% CI)	*P* value	Sample size, N
$\mathbf{G_{Res,s}}$			
s_1	0.802 (0.765-0.922)	.77	527
s_2	0.271 (0.202-0.438)	.01	513
$\mathbf{G_{Rel,s}}$			
s_1	0.955 (0.885-1.026)	.90	2635
s_2	0.634 (0.581-0.694)	.01	2565
$\mathbf{G_{All,s}}$			
s_1	0.919 (0.824-1.007)	.73	3162
s_2	0.572 (0.525-0.614)	<.001	3078
$\mathbf{G_{Rel*,s}}$			
s_1	0.977 (0.855-1.130)	.99	1340
s_2	1.472 (1.216-1.823)	.28	1104

In the general case, there will be no ground truth to compare the model against. However, in this study, we assessed the model performance directly. Due to the particular conditions of the NYC epidemic, the performance of the estimator in each causal model can be evaluated directly from the CI_P reported by the New Jersey and New York State health departments. Under the test rationing and home quarantine policies of New York and New Jersey during the spring 2020 epidemic, diagnostic real-time reverse transcription polymerase chain reaction (rRT-PCR) SARS-CoV-2 tests were restricted to individuals hospitalized with COVID-19 symptoms. Households with a member who tested positive were required by law to quarantine [35-40]. Although multiple members of a given household might have SARS-CoV-2 infections, no additional rRT-PCR tests would be performed on other household members unless they were hospitalized. Therefore, the reported CI_P is as if SAR_h=0 where no secondary household cases are reported. We aggregated confirmed SARS-CoV-2 infections as reported by the New York and New Jersey governments for each date and geographic area (NJ/NY, NYC CBSA, NYC) and calculated the CI_P relative to the American Community Survey (ACS) population for each. To evaluate the performance of each model, we computed the cumulative infection rate estimator at SAR_h=0 and calculated its relative bias from the reported CI_P for each area and period.

To demonstrate the practical epidemiological utility of this type of internet-based sampling, we calculated for different values of SAR_h using the causal model most compatible with the primary data, deriving 95% CIs by bootstrap resampling (n=1000).

Results

Demographics

In total, 527 and 513 responses met the inclusion criteria from surveys s_1 and s_2, respectively. Demographic information is summarized as frequencies for each collection period, with Pearson chi-squared tests performed to compare raw counts to demographic distributions in the 2018 ACS update of the US Census Bureau (Table 2).

Significant age skews were apparent across all survey periods, with both respondents and relations skewing significantly younger than the known population distribution, while sex distributions were not significantly different than the ACS estimate for New York and New Jersey. This large age skew made the sample highly unrepresentative of the target population, but with a correctly specified causal model, it was possible to obtain an unbiased estimate of the cumulative infection rate CI_P, as we next demonstrate through model diagnosis, selection, and evaluation.

Table 2. Demographic characteristics of survey samples.

Characteristics	Respondents, n(%)		Relations, n(%)		Combined, n(%)		ACS[a] (%)
	Survey s_1 (N=527)	Survey s_2 (N=513)	Survey s_1 (N=2635)	Survey s_2 (N=2565)	Survey s_1 (N=3162)	Survey s_2 (N=3078)	
Age (years)							
<19	2 (0.4)	5 (1.0)	264 (10.0)	356 (13.9)	266 (8.4)	361 (11.7)	22.7
19-29	184 (34.9)	192 (37.4)	620 (23.5)	575 (22.4)	804 (25.4)	767 (24.9)	15.2
30-39	170 (32.3)	163 (31.8)	531 (20.2)	525 (20.5)	701 (22.2)	688 (22.4)	13.2
40-49	89 (16.9)	77 (15)	400 (15.2)	344 (13.4)	489 (15.5)	421 (13.7)	12.9
50-59	56 (10.6)	46 (9.0)	367 (13.9)	346 (13.5)	423 (13.4)	392 (12.7)	14.1
60-69	19 (3.6)	24 (4.7)	284 (10.8)	294 (11.5)	303 (9.6)	318 (10.3)	11.2
≥70	7 (1.3)	6 (1.2)	169 (6.4)	125 (4.9)	176 (5.6)	131 (4.3)	10.6
Chi-square (*df*=6)	475	480	457	358	793	672	N/A[b]
P value	<.001	<.001	<.001	<.001	<.001	<.001	N/A
Sex[c]							
N/A	1 (0.2)	3 (0.6)	20 (0.8)	77 (3.0)	21 (0.7)	80 (2.6)	N/A
Female	267 (50.7)	286 (55.8)	1353 (51.3)	1303 (50.8)	1620 (51.2)	1589 (51.6)	51.4
Male	259 (49.1)	224 (43.7)	1262 (47.9)	1185 (46.2)	1521 (48.1)	1409 (45.8)	48.6
Chi-square (*df*=1)	0.08	4.50	0.13	0.97	0.046	3.14	N/A
P value	.77	.03	.71	.32	.83	.07	N/A
Occupation (multiple)							
Essential worker	153 (30.0)	140 (27.9)	N/A	N/A	N/A	N/A	N/A
Food service	31 (5.9)	27 (5.3)	N/A	N/A	N/A	N/A	N/A
Health care	66 (12.5)	69 (13.5)	N/A	N/A	N/A	N/A	N/A
Work from home	152 (28.8)	183 (35.7)	N/A	N/A	N/A	N/A	N/A
Not working	71 (13.5)	86 (16.8)	N/A	N/A	N/A	N/A	N/A
Other	231 (43.8)	173 (33.7)	N/A	N/A	N/A	N/A	N/A

[a] ACS: American Community Survey of the US Census Bureau.

[b] N/A: not applicable.

[c] Sex inferred from the reported gender identity for comparison with the ACS.

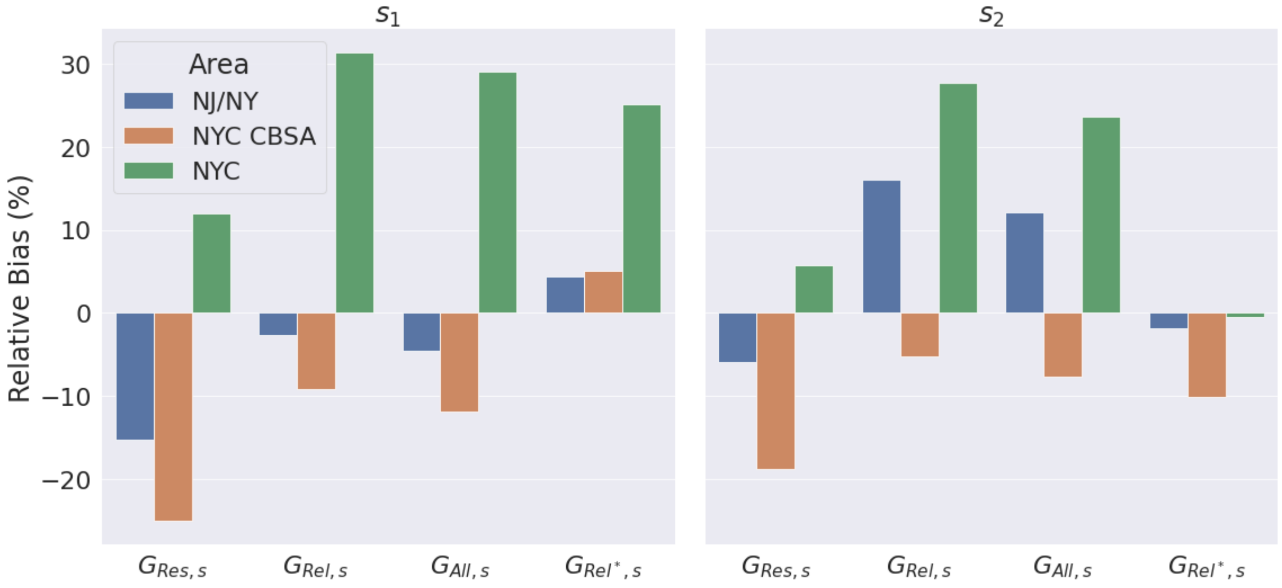
Model Diagnosis, Selection, and Evaluation

In the first survey period, no model could be rejected at nominal α=0.05, but in the second survey period, only model $G_{Rel*,s}$ could not be rejected. The most likely explanation for why the 4 models were more distinguished in the second period is that the cumulative infection rate increased through the course of the epidemic, giving greater power to detect a statistical dependence in a model, even though the total sample size was similar between periods for each model.

Model Performance

The model $G_{Rel*,s}$ recovered accurate cumulative infection rate estimates, with the lowest bias across both survey periods for the full sample (NJ/NY), as displayed in Figure 4, with the estimator variance for all models increasing as the sample size decreased with a smaller geographic area.

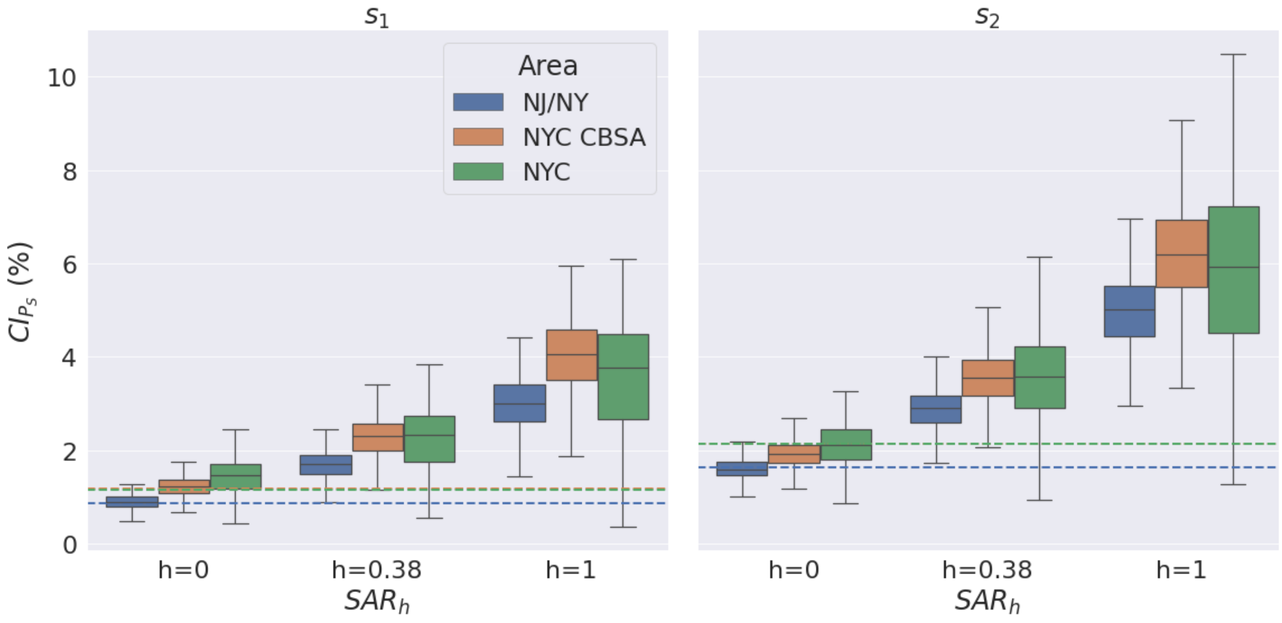
Figure 4. Relative bias of cumulative infection estimates by geographic area, model, and survey period. NJ/NY: New Jersey and New York; NYC: New York City; NYC CBSA: New York City Combined Statistical Area contained within New Jersey and New York.



Estimating the Cumulative Infection Rate from the Household Secondary Attack Rate

The model $G_{Rel*,s}$ was used to calculate the cumulative infection rate estimator for different values of SAR_h, as displayed in Figure 5. For all geographic areas and survey periods, the median estimate was approximately 1-4 times higher than the reported cumulative infection rate CI_P, with upper bounds ranging from 2.5-5 times higher.

Figure 5. Estimated cumulative infection rate by geographic area, household secondary attack rate (SARh), and survey period. Dashed lines are the reported CI_p for the survey period, color-matched to the geographic area. NJ/NY: New Jersey and New York; NYC: New York City; NYC CBSA: New York City Combined Statistical Area contained within New Jersey and New York.



Reported Symptoms Among Respondents and Relations

The number of households with at least 1 confirmed SARS-CoV-2 infection increased by 2 times, and the number of households with at least 1 member recently hospitalized for influenza-like illness (hospital ILI) increased by 1.5 times for both respondents and relations across the 2 survey periods, as shown in Table 3. Despite this, the marginal rates of common symptoms (fever, aches, anosmia, allergy) remained similar across both periods. This highlights the practical difficulties of estimating changes in CI_P by using common symptom checklists, as reported by internet surveys.

The correlation between health status indicators and symptoms remained largely the same across both periods (Figure 6). The notable exception is that the correlation between SARS-CoV-2 and hospital ILI increased from the first to the second survey period, presumably corresponding to the conversion from diagnosis to hospitalization as the epidemic progressed.

Table 3. Respondent/relation household heath status and symptoms by survey period.

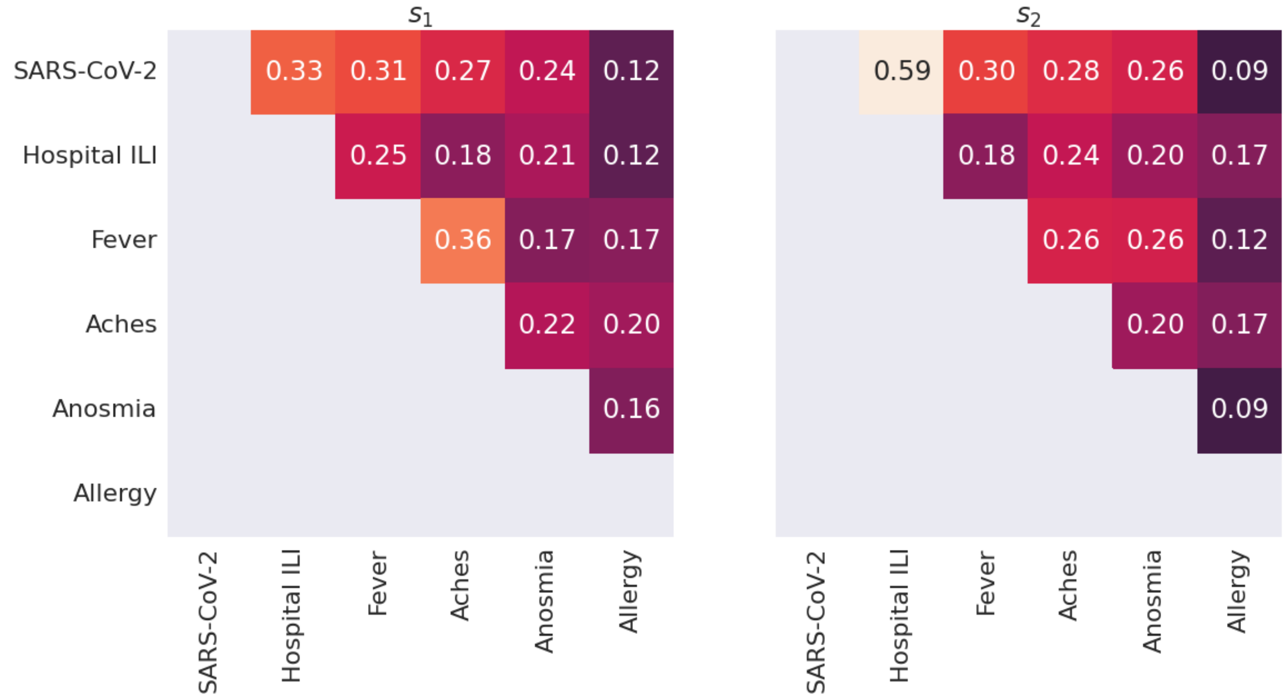
Person	Household health status, n (%)		Reported symptoms, n (%)			
	SARS-CoV-2[a]	Hospital ILI[b,c]	Fever	Aches	Anosmia	Allergy
Survey s_1						
Respondents (N=1040)	25 (2.4)	24 (2.3)	32 (3.1)	112 (10.8)	45 (4.3)	89 (8.6)
Relations (N=5200)	155 (3.0)	109 (2.1)	211 (4.1)	325 (6.3)	167 (3.2)	226 (4.4)
Survey s_2						
Respondents (N=1040)	53 (5.2)	33 (3.2)	23 (2.3)	145 (14.3)	46 (4.5)	104 (10.2)
Relations (N=5200)	295 (5.8)	169 (3.3)	154 (3.0)	337 (6.6)	180 (3.6)	192 (3.8)

[a] SARS-CoV-2: at least 1 household member had tested positive for SARS-CoV-2 infection by real-time reverse transcription polymerase chain reaction (rRT-PCR).

[b] ILI: influenza-like illness.

[c] Hospital ILI: at least 1 household member was recently hospitalized for an ILI.

Figure 6. Reported symptom correlations by survey period. SARS-CoV-2 : at least 1 household member tested positive for SARS-CoV-2 infection by rRT-PCR. Hospital ILI: at least 1 household member was recently hospitalized for an ILI. ILI: influenza-like illness; rRT-PCR: real-time reverse transcription polymerase chain reaction.



Discussion

Principle Findings

Using no external data and standard statistical independence tests, we were able to rank and reject all alternative models except the model $G_{Rel*,s}$ that yielded the lowest bias for the cumulative infection rate estimator, with a bias of less than 4% on the full sample despite the high skew toward younger ages relative to the target population. Without randomization or representativeness, this study recovered accurate estimates of a key epidemiological parameter using an internet-based crowdsourced population in a dynamic public health crisis using only a few hundred samples.

Although we primarily intend this work to demonstrate the broad utility of graphical models as an aid to epidemiologists, it is worth noting how useful internet-based epidemiology could prove in future epidemics by inspecting the estimates of for varying SAR_h (Figure 5). A major source of confusion in the early COVID-19 pandemic was diverging estimates of the cumulative infection rate. By July 27, 2020, there were 228,679 rRT-PCR cumulative confirmed infections in NYC for a reported CI_P of 2.65%. However, the actual CI_P was estimated to be 23.3% by seroprevalence studies in the July 27-August 13, 2020, period—8.8 times higher [41-44]. A similar difference would imply that the reported CI_P of 2.14% by May 8, 2020, corresponded to an actual CI_P of ~19%. Using the simple method of estimating by SAR_h=0.38, it would yield intermediate estimates implying upper bounds on the actual CI_P of 2.5 and 5 times the reported cumulative infection rate CI_P.

Limitations

The limitations of this approach are encoded directly in the set of causal graphs and entail explicit conditions where statistical tests will fail to reject incorrect models. For example, if age is a poor instrumental variable for response status, then with finite data, none of the models may be rejected by statistical tests. In contrast, if age is strongly related to the outcome variable household status C_h, but not response status, then all models could be rejected, even if there was no selection bias on C_h. More generally, if there is any relationship between a set of variables, there will be a statistically significant correlation, given sufficient data; therefore, any causal model regardless of its utility will be rejected if statistical tests are applied naively.

These inherent limitations are why we emphasize structural causal models as an aid and not a substitute for epidemiological judgment. The utility of causal modeling is the formal comparison and communication of alternative explanations of the sampled data. For example, in this study, we chose to not model information bias, instead focusing on detecting selection bias. The choice to ignore information bias is explicit in the presented causal graphs; none of them have a subgraph that models an information bias mechanism, such as rRT-PCR test constraints or inaccurate self-reporting. These causal models were constructed with these assumptions for the context and objectives of this study, and similar assumptions may not be acceptable for a different context or objective. The key point is that all these assumptions are made apparent on inspection of the causal graphs.

Conclusion

The COVID-19 pandemic is an unprecedented event, pushing the limits of the health care system worldwide. Reducing transmission via nonpharmacological interventions has been effective but requires near-real-time and accurate information across all segments of society, information that has been difficult to reliably ascertain. Given the vast divergence of cumulative infection rate estimates across early studies [43] and the consequences for undermining public trust, there is a clear use case for internet-based public health surveillance to rapidly estimate key epidemiological parameters. A major use of internet-based surveys in the COVID-19 pandemic has been estimating the rate of vaccine uptake. The Census Household Pulse and Delphi-Facebook overestimated COVID-19 vaccine uptake to May 2021 by 14% and 17%, respectively, in the United States [44], while a much smaller Axios-Ipsos online survey of a different design overestimated uptake by only 5% in the United States. Internet-based surveys are an important tool with several uses for managing a pandemic, but current methodology is hampered by an inability to successfully detect and mitigate estimate bias. However, looking beyond vaccine uptake, near-term public health interventions, and advances in treatments, COVID-19 continues to evolve along with human societies. Surveillance systems and statistical models that assume centralized reporting may not be as useful with the mass adoption of at-home tests for COVID-19; alternative approaches, such as the social network polling design used in this work, may need to be deployed. For these reasons, we hope that these recent advances in causal modeling theory are adopted by the epidemiological community for current and future epidemics.

Acknowledgments

Design, implementation, analysis, and preparation of the manuscript were performed by NS. We gratefully acknowledge conceptual input and constructive feedback from PW, BC, KP, JYJ, and DPW.

The work was supported in part by funds to DPW from the National Institutes of Health (1R01EB025025-01, 1R01LM013364-01, 1R21HD091500-01, 1R01LM013083); the National Science Foundation (Award 2014232), the Hartwell Foundation, the Bill and Melinda Gates Foundation, the Coulter Foundation, the Lucile Packard Foundation, Auxiliaries Endowment, the Islamic Development Bank (ISDB) Transform Fund, and the Weston Havens Foundation; program grants from Stanford University’s Human Centered Artificial Intelligence Program, the Precision Health and Integrated Diagnostics Center, the Beckman Center, Bio-X Center, the Predictives and Diagnostics Accelerator, Spectrum, the Spark Program in Translational Research, and MediaX; and program grants from the Wu Tsai Neurosciences Institute's Neuroscience:Translate Program. We also acknowledge generous support from David Orr, Imma Calvo, Bobby Dekesyer, and Peter Sullivan. PW would like to acknowledge support from Mr. Schroeder and the Stanford Interdisciplinary Graduate Fellowship (SIGF) as the Schroeder Family Goldman Sachs Graduate Fellow.

Data Availability

The data underlying this paper will be shared upon reasonable request to the corresponding author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey materials for survey period 1.
[PDF File (Adobe PDF File), 133 KB - publichealth_v8i7e31306_app1.pdf]

Multimedia Appendix 2
Survey materials for survey period 2.
[PDF File (Adobe PDF File), 154 KB - publichealth_v8i7e31306_app2.pdf]

References

1. Smith LH. Selection mechanisms and their consequences: understanding and addressing selection bias. Curr Epidemiol Rep 2020 Aug 09;7(4):179-189. [doi: 10.1007/s40471-020-00241-6]
2. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004 Sep;15(5):615-625. [doi: 10.1097/01.ede.0000135174.63482.43] [Medline: 15308962]
3. Pearl J. Causality. 2nd ed. New York, NY: Cambridge University Press; 2009.
4. Bareinboim E, Pearl J. Causal inference and the data-fusion problem. Proc Natl Acad Sci U S A 2016 Jul 05;113(27):7345-7352 [FREE Full text] [doi: 10.1073/pnas.1510507113] [Medline: 27382148]
5. Richardson T, Robins J. A Unification of the Counterfactual and Graphical Approaches to Causality. Working Paper. Single World Intervention Graphs (SWIGs). URL: https://csss.uw.edu/files/working-papers/2013/wp128.pdf [accessed 2022-02-26]
6. Schuessler J, Selb P. Graphical causal models for survey inference. SocArXiv 2019:26. [doi: 10.31235/osf.io/hbg3m]
7. Saadati M, Tian J. Adjustment criteria for recovering causal effects from missing data. 2019 Presented at: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD; September 16–20, 2019; Würzburg, Germany p. 561-577. [doi: 10.1007/978-3-030-46150-8_33]
8. Bareinboim E, Tian J, Pearl J. Recovering from selection bias in causal and statistical inference. 2014 Presented at: The 28th AAAI Conference on Artificial Intelligence; July 27–31, 2014; Québec City, Québec, Canada. [doi: 10.1609/aaai.v28i1.9074]
9. Hernán MA, Robins JM. Causal Inference: What If. URL: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf [accessed 2022-02-26]
10. Pearce N, Lawlor DA. Causal inference: so much more than statistics. Int J Epidemiol 2016 Dec 01;45(6):1895-1903 [FREE Full text] [doi: 10.1093/ije/dyw328] [Medline: 28204514]
11. Petersen JM, Ranker LR, Barnard-Mayers R, MacLehose RF, Fox MP. A systematic review of quantitative bias analysis applied to epidemiological research. Int J Epidemiol 2021 Nov 10;50(5):1708-1730. [doi: 10.1093/ije/dyab061] [Medline: 33880532]
12. McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data. PLoS Negl Trop Dis 2017 Jan;11(1):e0005295 [FREE Full text] [doi: 10.1371/journal.pntd.0005295] [Medline: 28085877]
13. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci U S A 2015 Nov 24;112(47):14473-14478 [FREE Full text] [doi: 10.1073/pnas.1515373112] [Medline: 26553980]
14. Aiken EL, McGough SF, Majumder MS, Wachtel G, Nguyen AT, Viboud C, et al. Real-time estimation of disease activity in emerging outbreaks using internet search information. PLoS Comput Biol 2020 Aug;16(8):e1008117 [FREE Full text] [doi: 10.1371/journal.pcbi.1008117] [Medline: 32804932]
15. Carnegie Mellon University Delphi Research Group. COVIDcast. URL: https://delphi.cmu.edu/covidcast [accessed 2022-02-26]
16. Facebook. Symptoms Map and Dashboard. URL: https://dataforgood.facebook.com/covid-survey [accessed 2022-02-26]
17. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul 11;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
18. Rossman H, Keshet A, Shilo S, Gavrieli A, Bauman T, Cohen O, et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. Nat Med 2020 May;26(5):634-638 [FREE Full text] [doi: 10.1038/s41591-020-0857-9] [Medline: 32273611]
19. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science 2020 Jun 19;368(6497):1362-1367 [FREE Full text] [doi: 10.1126/science.abc0473] [Medline: 32371477]
20. Kreuter F. Partnering with a global platform to inform research and public policy making. Surv Res Methods 2020;14(2):159-163. [doi: 10.18148/srm/2020.v14i2.7761]

21. Wang C, Han L, Stein G, Day S, Bien-Gund C, Mathews A, et al. Crowdsourcing in health and medical research: a systematic review. Infect Dis Poverty 2020 Jan 20;9(1):8 [FREE Full text] [doi: 10.1186/s40249-020-0622-9] [Medline: 31959234]
22. Post J, Class F, Kohler U. Unit nonresponse biases in estimates of SARS-CoV-2 prevalence. Surv Res Methods 2020;14(2):115-121. [doi: 10.18148/srm/2020.v14i2.7755]
23. Klingwort J, Schnell R. Critical limitations of digital epidemiology: why COVID-19 apps are useless. Surv Res Methods 2020;14(2):95-101. [doi: 10.18148/srm/2020.v14i2.7726]
24. Bendavid E, Mulaney B, Sood N, Shah S, Bromley-Dulfano R, Lai C, et al. COVID-19 antibody seroprevalence in Santa Clara County, California. Int J Epidemiol 2021 May 17;50(2):410-419 [FREE Full text] [doi: 10.1093/ije/dyab010] [Medline: 33615345]
25. Schaurer I, Weiß B. Investigating selection bias of online surveys on coronavirus-related behavioral outcomes. Surv Res Methods 2020;14(2):103-108. [doi: 10.1007/978-3-030-39903-0_301696]
26. Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol 2016 Dec 01;45(6):1887-1894. [doi: 10.1093/ije/dyw341] [Medline: 28089956]
27. Chunara R, Chhaya V, Bane S, Mekaru SR, Chan EH, Freifeld CC, et al. Online reporting for malaria surveillance using micro-monetary incentives, in urban India 2010-2011. Malar J 2012 Feb 13;11:43 [FREE Full text] [doi: 10.1186/1475-2875-11-43] [Medline: 22330227]
28. Walters K, Christakis DA, Wright DR. Are Mechanical Turk worker samples representative of health status and health behaviors in the U.S.? PLoS One 2018;13(6):e0198835 [FREE Full text] [doi: 10.1371/journal.pone.0198835] [Medline: 29879207]
29. Chan C, Holosko MJ. An overview of the use of Mechanical Turk in behavioral sciences. Res Soc Work Pract 2015 Jul 06;26(4):441-448. [doi: 10.1177/1049731515594024]
30. Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods 2012 Mar;44(1):1-23. [doi: 10.3758/s13428-011-0124-6] [Medline: 21717266]
31. Ophir Y, Sisso I, Asterhan CSC, Tikochinski R, Reichart R. The Turker blues: hidden factors behind increased depression rates among Amazon's Mechanical Turkers. Clin Psychol Sci 2019 Oct 02;8(1):65-83. [doi: 10.1177/2167702619865973]
32. Office of the Governor of New Jersey. Governor Murphy Announces Statewide Stay at Home Order, Closure of All Non-Essential Retail Businesses. URL: https://www.nj.gov/governor/news/news/562020/20200320j.shtml [accessed 2022-02-26]
33. Office of Governor of New York State. Governor Cuomo Signs the "New York State on PAUSE" Executive Order. URL: https://www.governor.ny.gov/news/governor-cuomo-signs-new-york-state-pause-executive-order [accessed 2022-02-26]
34. New York City Department of Health and Mental Hygiene. 2020 Health Alert #4 : COVID-19 Updates for New York City. URL: https://www1.nyc.gov/assets/doh/downloads/pdf/han/alert/2020/covid-19-update-030920.pdf [accessed 2022-02-26]
35. New York City Department of Health and Mental Hygiene. 2020 Health Alert #6 : COVID-19 Updates for New York City. URL: https://www1.nyc.gov/assets/doh/downloads/pdf/han/alert/2020/covid-19-03152020.pdf [accessed 2022-02-26]
36. New York City Department of Health and Mental Hygiene. 2020 Health Alert #8 : COVID-19 Update for New York City. URL: https://www1.nyc.gov/assets/doh/downloads/pdf/han/advisory/2020/covid-19-03202020.pdf [accessed 2022-02-26]
37. New York State Department of Health. Health Advisory : COVID-19 Update for Local Health Department Response Activities. URL: https://coronavirus.health.ny.gov/system/files/documents/2020/04/doh_covid19_lhdupdate_040720.pdf [accessed 2022-02-26]
38. New York State Department of Health. 2019 Novel Coronavirus (COVID-19) Interim Guidance: Protocol for COVID-19 Testing Applicable to All Health Care Providers and Local Health Departments. 2019. URL: https://coronavirus.health.ny.gov/system/files/documents/2020/04/doh_covid19_revisedtestingprotocol_042620.pdf [accessed 2022-02-26]
39. New Jersey Department of Health. Surveillance Criteria and Testing for Novel Coronavirus 2019 (COVID-19). URL: https://www.nj.gov/health/cd/documents/topics/NCOV/Surveillance [accessed 2022-02-26]
40. Bajema KL, Wiegand RE, Cuffe K, Patel SV, Iachan R, Lim T, et al. Estimated SARS-CoV-2 seroprevalence in the US as of September 2020. JAMA Intern Med 2021 Apr 01;181(4):450-460 [FREE Full text] [doi: 10.1001/jamainternmed.2020.7976] [Medline: 33231628]
41. Thompson CN, Baumgartner J, Pichardo C, Toro B, Li L, Arciuolo R, et al. COVID-19 outbreak - New York City, February 29-June 1, 2020. Morb Mortal Wkly Rep 2020 Nov 20;69(46):1725-1729 [FREE Full text] [doi: 10.15585/mmwr.mm6946a2] [Medline: 33211680]
42. Yang W, Kandula S, Huynh M, Greene SK, Van Wye G, Li W, et al. Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: a model-based analysis. Lancet Infect Dis 2021 Feb;21(2):203-212 [FREE Full text] [doi: 10.1016/S1473-3099(20)30769-6] [Medline: 33091374]
43. Meyerowitz-Katz G, Merone L. A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. Int J Infect Dis 2020 Dec;101:138-148 [FREE Full text] [doi: 10.1016/j.ijid.2020.09.1464] [Medline: 33007452]
44. Bradley VC, Kuriwaki S, Isakov M, Sejdinovic D, Meng X, Flaxman S. Unrepresentative big surveys significantly overestimated US vaccine uptake. Nature 2021 Dec;600(7890):695-700 [FREE Full text] [doi: 10.1038/s41586-021-04198-4] [Medline: 34880504]

Abbreviations

ACS: American Community Survey
DAG: directed acyclic graph
HIT: human intelligence task
ILI: influenza-like illness
MTurk: Mechanical Turk
NJ/NY: New Jersey and New York
NYC: New York City
NYC CBSA: New York City Combined Statistical Area contained within New Jersey and New York
rRT-PCR: real-time reverse transcription polymerase chain reaction

Edited by T Sanchez; submitted 16.06.21; peer-reviewed by N Hozé, M Bestek, R Poluru; comments to author 04.11.21; revised version received 22.02.22; accepted 17.05.22; published 21.07.22.

Please cite as:
Stockham N, Washington P, Chrisman B, Paskov K, Jung JY, Wall DP
Causal Modeling to Mitigate Selection Bias and Unmeasured Confounding in Internet-Based Epidemiology of COVID-19: Model Development and Validation
JMIR Public Health Surveill 2022;8(7):e31306
URL: https://publichealth.jmir.org/2022/7/e31306
doi:10.2196/31306
PMID:35605128

Original Paper

Unbalanced Risk of Pulmonary Tuberculosis in China at the Subnational Scale: Spatiotemporal Analysis

Maogui Hu[1*], PhD; Yuqing Feng[1,2*], MSc; Tao Li[3*], MSc; Yanlin Zhao[3], PhD; Jinfeng Wang[1], PhD; Chengdong Xu[1], PhD; Wei Chen[3], PhD

[1]State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Beijing, China

[2]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China

[3]National Center for Tuberculosis Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Wei Chen, PhD
National Center for Tuberculosis Control and Prevention
Chinese Center for Disease Control and Prevention
No 155 Changbai Road, Changping District
Beijing, 102206
China
Phone: 86 010 58900537
Email: chenwei@chinacdc.cn

Abstract

Background: China has one of the highest tuberculosis (TB) burdens in the world. However, the unbalanced spatial and temporal trends of TB risk at a fine level remain unclear.

Objective: We aimed to investigate the unbalanced risks of pulmonary tuberculosis (PTB) at different levels and how they evolved from both temporal and spatial aspects using PTB notification data from 2851 counties over a decade in China.

Methods: County-level notified PTB case data were collected from 2009 to 2018 in mainland China. A Bayesian hierarchical model was constructed to analyze the unbalanced spatiotemporal patterns of PTB notification rates during this period at subnational scales. The Gini coefficient was calculated to assess the inequality of the relative risk (RR) of PTB across counties.

Results: From 2009 to 2018, the number of notified PTB cases in mainland China decreased from 946,086 to 747,700. The average number of PTB cases in counties was 301 (SD 26) and the overall average notification rate was 60 (SD 6) per 100,000 people. There were obvious regional differences in the RRs for PTB (Gini coefficient 0.32, 95% CI 0.31-0.33). Xinjiang had the highest PTB notification rate, with a multiyear average of 155/100,000 (RR 2.3, 95% CI 1.6-2.8; *P*<.001), followed by Guizhou (117/100,000; RR 1.8, 95% CI 1.3-1.9; *P*<.001) and Tibet (108/100,000; RR 1.7, 95% CI 1.3-2.1; *P*<.001). The RR for PTB showed a steady downward trend. Gansu (local trend [LT] 0.95, 95% CI 0.93-0.96; *P*<.001) and Shanxi (LT 0.94, 95% CI 0.92-0.96; *P*<.001) experienced the fastest declines. However, the RRs for PTB in the western region (such as counties in Xinjiang, Guizhou, and Tibet) were significantly higher than those in the eastern and central regions (*P*<.001), and the decline rate of the RR for PTB was lower than the overall level (*P*<.001).

Conclusions: PTB risk showed significant regional inequality among counties in China, and western China presented a high plateau of disease burden. Improvements in economic and medical service levels are required to boost PTB case detection and eventually reduce PTB risk in the whole country.

(JMIR Public Health Surveill 2022;8(7):e36242) doi:10.2196/36242

KEYWORDS

pulmonary tuberculosis; infectious disease; pattern; notification rates; Bayesian; spatiotemporal pattern; tuberculosis; public health; China; disease burden; spatial data; regional inequality; risk; TB; unbalanced; notification data; trend; cases; incidence

Introduction

Tuberculosis (TB), an infectious disease caused by the bacterium *Mycobacterium tuberculosis*, is one of the top 10 causes of death worldwide. There were approximately 10 million new TB cases worldwide in 2019 and 1.4 million people died due to TB [1]. In the 1990s, China implemented the directly observed treatment, short-course (DOTS) approach, as recommended by the World Health Organization. Additionally, a series of major measures have been taken, including the formulation and implementation of TB prevention and control plans and the implementation of international collaborative projects for TB control. China has achieved the United Nations' 2015 Millennium Development Goals 5 years ahead of schedule and has made a significant contribution to global TB prevention and control [2-4]. However, there are still challenges in the prevention of TB in China. Specifically, in 2019, China had approximately 840,000 new TB cases, accounting for 8.4% of the global total and ranking third among the 30 countries with a high TB burden [1]. In terms of the 2030 sustainable development goals (SDGs), preventing and managing TB in China is a challenging mission.

Despite some studies on the TB risk in China at the national and provincial levels, there has been a lack of evaluation of the unbalanced risk of pulmonary tuberculosis (PTB) as well as the trend of the risk at fine scales from long-term series in China in recent years [5-12]. A county-level study of PTB risk in China can reveal variations in the severity and efficacy of PTB prevention and control in different regions (provinces, municipalities, and autonomous regions) on a fine scale.

Therefore, the aim of this study was to investigate the unbalanced risks of PTB at different levels and how they evolved in both temporal and spatial aspects using PTB notification data from 2851 counties over a decade in China. These findings could be beneficial to optimize the distribution of health resources in China for TB prevention and control.

Methods

Data Sources and Definition of PTB Notification Rate

TB is one of the 40 notifiable infectious diseases in China. To improve the timely diagnosis, treatment, and supervision of TB patients nationwide, the Chinese Ministry of Health launched the National Tuberculosis Information Management System (TBIMS) in 2005, which records detailed information on TB patients, including outpatient information, case information, treatment, and supervision, as well as information on TB planning and management. In China, all diagnosed TB (any kind) should be recorded to the TBIMS according to the national TB prevention and control guidelines; thus, all active TB is included in the system such as bacterially confirmed or clinically diagnosed TB and extrapulmonary TB. All TB prevention and control facilities across the country can use the system to record TB case information and planning activity data in real time. This can be used to increase the detection rate of TB and strengthen TB patient management, as well as to control and eventually eliminate TB.

China's administrative regions are generally divided into the following four levels: provincial, prefectural, county, and township levels. Counties are the basis of local government in the administrative division of Chinese society. There are specialized institutions for TB prevention and control at the provincial, prefectural, and county levels in China, including TB clinics, disease prevention and control centers, and TB designated hospitals.

In this study, the numbers of notified PTB cases and the populations of the 2851 county-level areas in mainland China from 2009 to 2018 were collected from the TBIMS. The annual PTB notification rate in a region was defined as the ratio of the annual number of notified PTB cases and the total population at the end of the year in the region.

Spatiotemporal Trend Modeling of the PTB Notification Rate

A Bayesian spatiotemporal model was built to analyze spatiotemporal patterns of PTB notification rates from 2009 to 2018. The spatiotemporal process of the PTB notification rate was decomposed into three components: the spatial random effect, overall time trend, and spatiotemporal interaction effect [13-15]. A Poisson regression model connected by a logarithmic function was used to model the process based on count data [16,17]. The Poisson log-normal model with the spatiotemporal effect is as follows:

$$y_{it}=\text{Poisson}(\mu_{it}) \quad \textbf{(1)}$$

$$\mu_{it}=e_{it}\theta_{it} \quad \textbf{(2)}$$

$$\log(\theta_{it})=\alpha+s_i+(b_0t^*+v_t)+b_{1i}t^*+\varepsilon_{it} \quad \textbf{(3)}$$

where y_{it} denotes the number of notified PTB cases in year t in the ith county, μ_{it} is the expected number notified PTB cases, α is the overall average level of the PTB notification rate in mainland China from 2009 to 2018, and s_i is the relative risk (RR) for PTB in the ith county during the study period; $\exp(s_i)>1$ indicates that the RR for PTB in the ith county is higher than the overall average across mainland China, while $\exp(s_i)<1$ means that the RR for PTB in the ith county is lower than the overall average. The time span relative to the middle time point (t_{mid}) of the study period is represented by $t^*=t-t_{mid}$. $b_0t^*+v_t$ describes the overall time trend of the RR, consisting of a linear trend b_0t^* and a time-random effect v_t, allowing the overall time trend to show a nonlinear change. The spatiotemporal interaction term $b_{1i}t^*$ represents the spatial variation in the time trend. b_{1i} measures the local trend decomposed from the overall time trend. If $\exp(b_{1i})>1$, the ith county has a stronger local trend (LT) than the global overall trend. Otherwise, $\exp(b_{1i})<1$ indicates that the ith county has a weaker LT than the global overall trend. The ε_{ij} item is an unstructured random effect in the model.

For each model parameter, prior distributions were allocated. A uniform distribution was assigned to the intercept term α. We selected the Besag-York-Mollie (BYM) model to reveal the overall spatial random effect s_i [18,19]. The intrinsic conditional autoregressive (ICAR) prior with a spatial adjacency matrix W was adopted for the spatial structure, where wij=1 if the ith and jth counties are neighborhoods and w_{ij}=0 otherwise. The ICAR prior implies that adjacent regions tend to have similar overall notification rates. The same BYM prior was used for the parameter b_{1i}. The temporal random effect v_t was assumed to follow a Gaussian distribution as $v_t \sim N(0, \sigma^2_v)$. Finally, ε_{ij} followed the distribution $\varepsilon_{ij} \sim N(\sigma^2_\varepsilon)$. As suggested by Gelman [20], a strictly positive half Gaussian prior $N_{+\infty}(0, 10)$ was assigned to parameters of random-effect standard deviations, including σ_v and σ_ε.

We implemented the model in WinBUGS, a software program specifically designed for Bayesian analysis [21]. Markov chain Monte Carlo (MCMC) simulations were used to obtain posterior distributions of model parameters. Specifically, with different starting values, we ran two MCMC chains for each model with 200,000 iterations. The first 150,000 iterations were burned in. Every 10th iteration for the remaining 50,000 MCMC iterations was retained to make inferences. The Gelman-Rubin statistic was used to test convergence [22]. The Gelman-Rubin statistic values for all parameters in this study were below 1.05, indicating that the model converged.

Inequality Measurement: Gini Coefficient of PTB Risk

To assess the inequality of the RR of PTB across counties, we computed the Gini coefficient using the formula:

where x is the RR, and i and j indicate counties. As shown by equation (4), the Gini coefficient can be interpreted as the average relative differences between all pairs of counties.

Results

Temporal Trend and Spatial Distribution of the PTB Notification Rate

The number of notified PTB cases in mainland China decreased from 946,086 to 747,700 between 2009 and 2018, showing a downward trend (Table 1, Figure 1). In 2017, the number of notified PTB cases was at its lowest (732,612). The average number of PTB cases in counties decreased from 346.2 to 276.8, with the lowest reported in 2017 (270.5). The average number of PTB cases in counties was 301 (SD 26). The overall average notification rate in mainland China was 60 (SD 6) per 100,000 people. The rates of PTB notifications in China’s provinces varied significantly (Figure 2), with the rate in the western region being higher than those in the eastern and central regions. Western provinces, including Xinjiang, Tibet, Guizhou, and Guangxi, had multiyear overall notification rates higher than 90/100,000. Xinjiang had the highest PTB notification rate, with a multiyear average of 155/100,000, followed by Guizhou (117/100,000) and Tibet (108/100,000). The lowest PTB notification rates were reported in Beijing and Tianjin, with multiyear overall notification rates of 23 and 22 per 100,000, respectively. During the study period, notification rates declined in most provinces, but increased in some western provinces such as Xinjiang, Yunnan, and Qinghai. The rates in Gansu and Shanxi decreased by more than 50%, while the rates in Yunnan and Qinghai increased by 10%-15%.

Table 1. Summary statistics of the number of pulmonary tuberculosis (PTB) cases in counties.

Year	Total PTB cases (N=8,176,058), n (%)	Mean (SD)	Range				
			Minimum	P25[a]	P50[b]	P75[c]	Maximum
2009	946,086 (11.57)	346.2 (276.95)	0	149	275	466	2888
2010	901,189 (11.02)	330.3 (274.56)	0	144	258	439.2	3849
2011	874,759 (10.70)	321.6 (273.46)	1	139	251.5	432	4364
2012	862,579 (10.70)	316.8 (278.50)	1	131	246	424	4125
2013	818,876 (10.02)	301.7 (267.18)	1	124	232	410	3656
2014	786,256 (9.62)	289.9 (261.73)	1	113	220	389	3704
2015	763,230 (9.33)	282.3 (263.03)	2	108.8	209.5	378.2	3351
2016	742,771 (9.08)	274.5 (259.55)	1	106	207	368.5	3274
2017	732,612 (8.96)	270.5 (263.80)	1	104	204.5	358	4007
2018	747,700 (9.14)	276.8 (343.34)	0	98	204	349	7797

[a] P25: 25th percentile.

[b] P50: 50th percentile.

[c] P75: 75th percentile.

Figure 1. Temporal trend in the number of notified pulmonary tuberculosis cases and the overall rates in mainland China from 2009 to 2018.

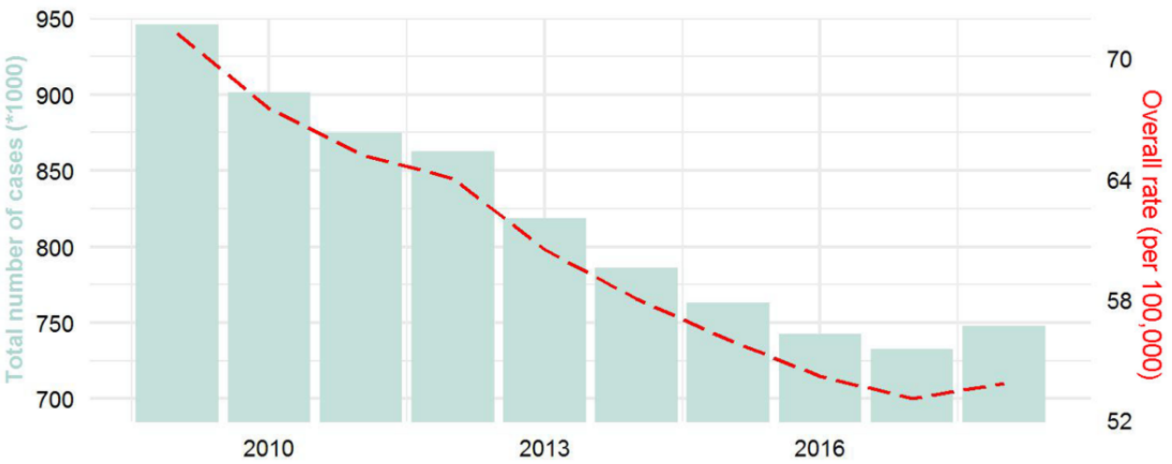
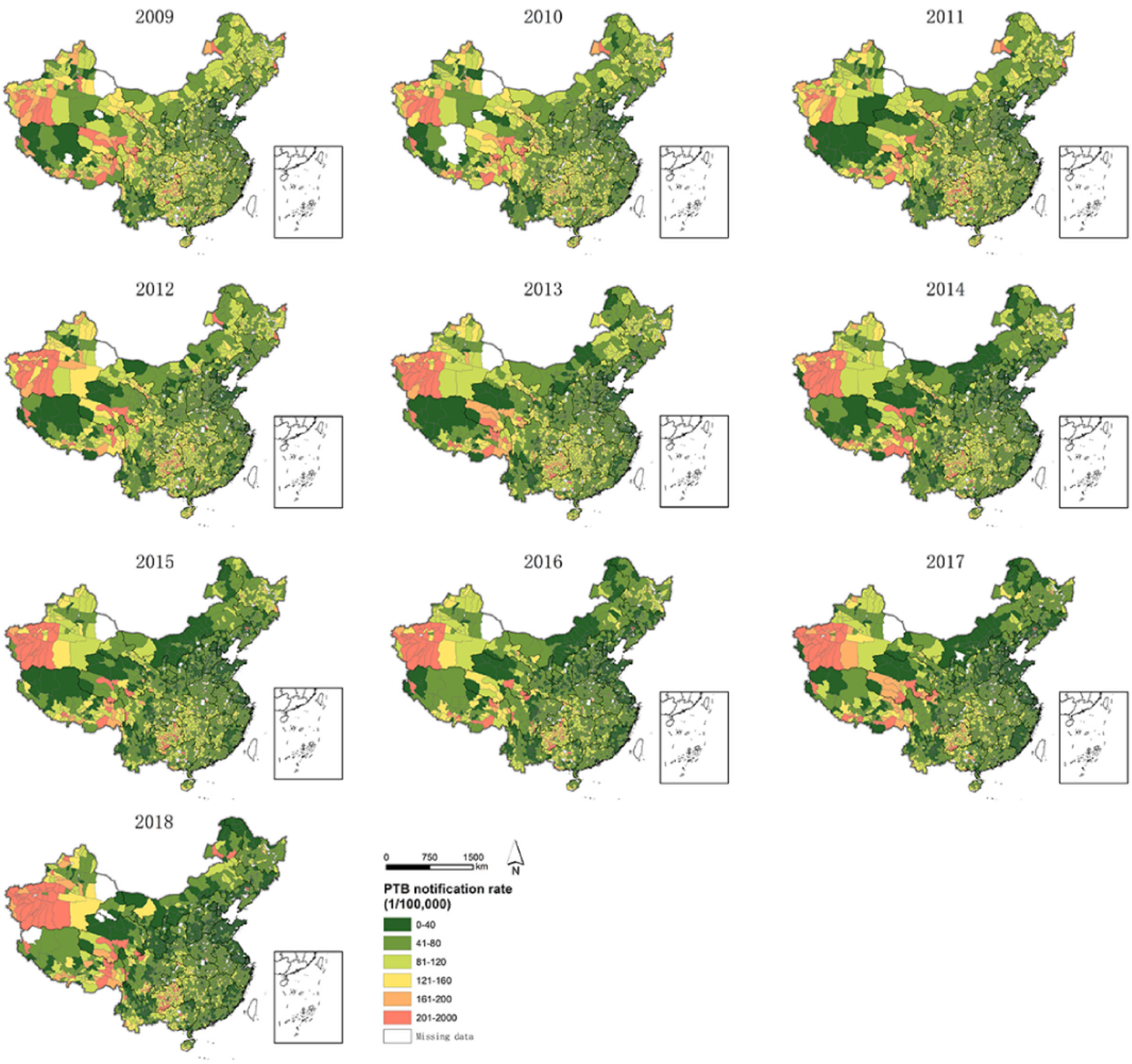


Figure 2. Spatial distributions of pulmonary tuberculosis (PTB) notification rates in counties of 31 provinces in mainland China from 2009 to 2018.

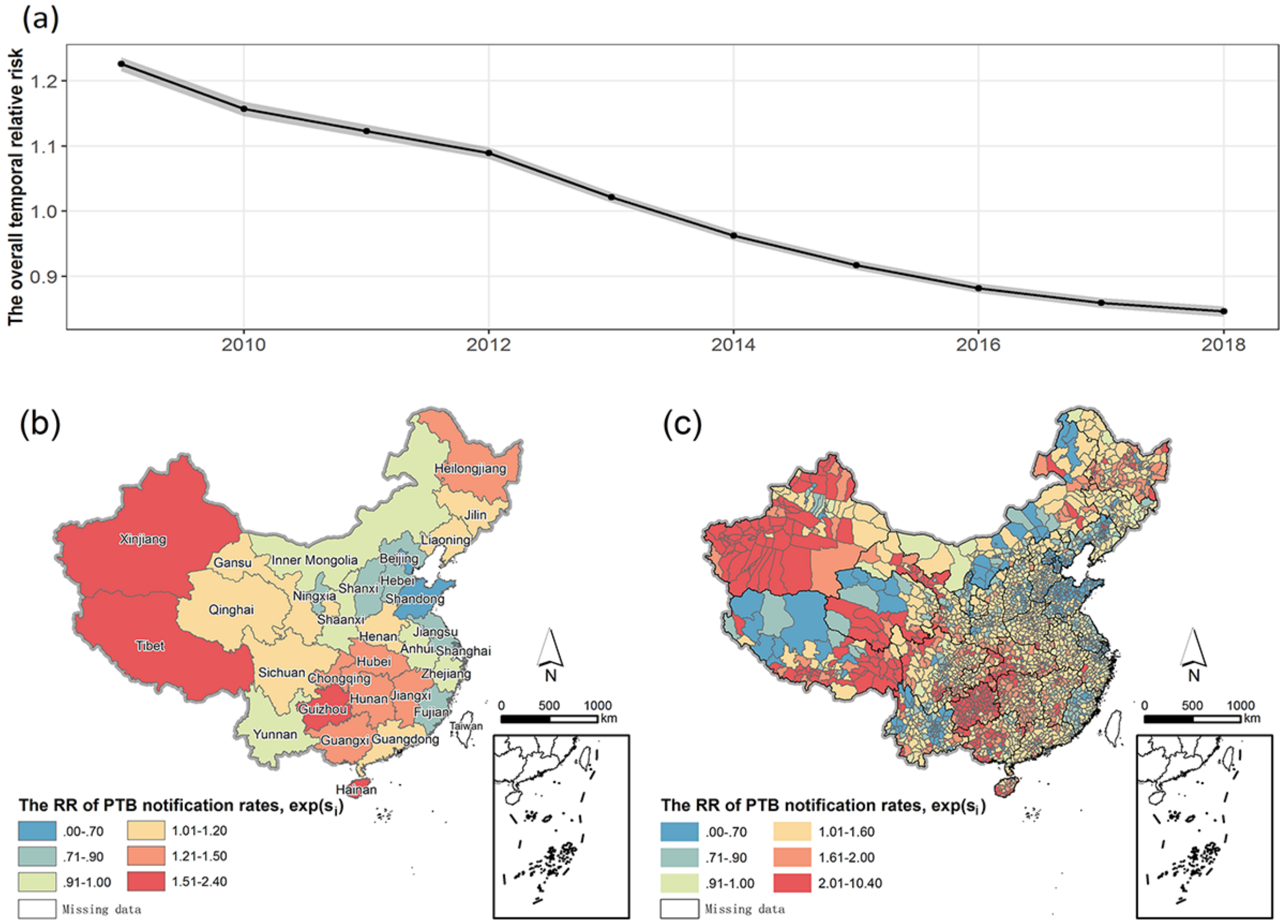


The spatial distributions of PTB notification rates within each province were also unbalanced with significant heterogeneity. The notification rates in southwestern Xinjiang, Guizhou, and southeastern Tibet were generally high in most years. The notification rate in southern Xinjiang increased dramatically. In contrast, the rates in counties in Gansu, northeastern China, central Inner Mongolia, eastern Guangxi, Shandong, Jiangsu, and Sichuan decreased significantly (Figure 2).

Spatial Relative Risk and Temporal Trend of PTB

The RR for PTB showed a steady downward trend (Figure 3a), but the rate of decline gradually slowed in later years. There were obvious regional differences in the RRs for PTB, with those in the western region being significantly higher than those in the eastern and central regions (*P*<.001). Provinces were divided into three categories according to the RR values, represented by the item exp(s_i) in the model (Figure 3b). The red and orange provinces (RR≥1.2) in Figure 3b had significantly higher RRs than the overall average. Xinjiang (RR 2.3, 95% CI 1.6-2.8; *P*<.001), Tibet (RR 1.7, 95% CI 1.3-2.1; *P*<.001), Guizhou (RR 1.8, 95% CI 1.3-1.9; *P*<.001), and Hainan (RR 1.7, 95% CI 1.3-2.1; *P*=.003) had the highest RRs, followed by Hubei, Chongqing, Hunan, Jiangxi, Guangxi, and the three northeastern provinces. The provinces indicated in blue and light blue (RR≤0.9) in Figure 3b had significantly lower RRs than the overall average. Beijing (RR 0.4, 95% CI 0.3-0.5; *P*<.001), Tianjin, Shanghai, and Shandong had the lowest RRs, followed by Hebei, Shanxi, Ningxia, Jiangsu, and Fujian. The RRs for PTB in other regions were similar to the overall average of mainland China (0.9<RR<1.2).

Figure 3. The relative risk (RR) for pulmonary tuberculosis (PTB) in mainland China from 2009 to 2018. (a) Overall temporal trend (95% CI); (b) RR at the province level; (c) RR at the county level. s_i is the RR for PTB in the ith county during the study period.



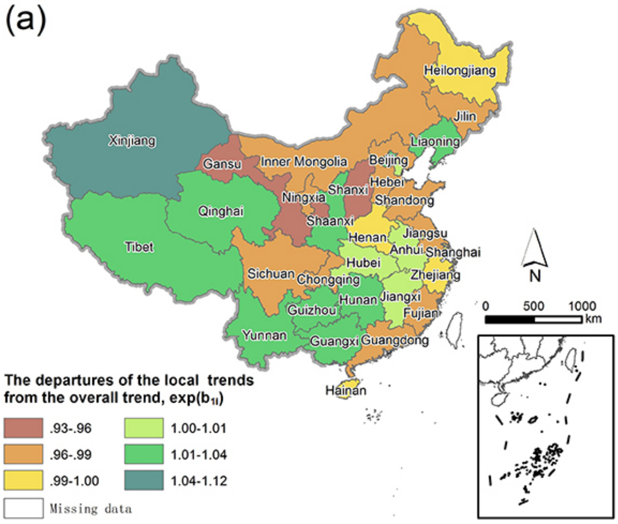
RRs were also spatially distributed heterogeneously at the county level within provinces (Figure 3c). In Xinjiang, 18% of counties had RRs<1, and they were located in the north-central region; 55% of counties had RRs>2, and most of them clustered in the south and north regions. The RRs for PTB in Tibet increased from northwest to southeast. Counties with RRs>2 (accounting for 45% of the total counties in Tibet) were distributed in the southeast. The classification based on the posterior distribution is shown in Multimedia Appendix 1.

Local Trends in the RR of PTB

In different regions of China, the decline rate of the RR varied significantly (Figure 4). The decline in the RR in the western region of China was generally lower than those in the central and eastern regions (*P*<.001). In mainland China, Xinjiang experienced the slowest RR decline, followed by Qinghai, Guizhou, Yunnan, Liaoning, and Beijing, all of which were slower than the overall trend. Gansu (LT 0.95, 95% CI 0.93-0.96; *P*<.001) and Shanxi (LT 0.94, 95% CI 0.92-0.96; *P*<.001) experienced the fastest declines, followed by Inner

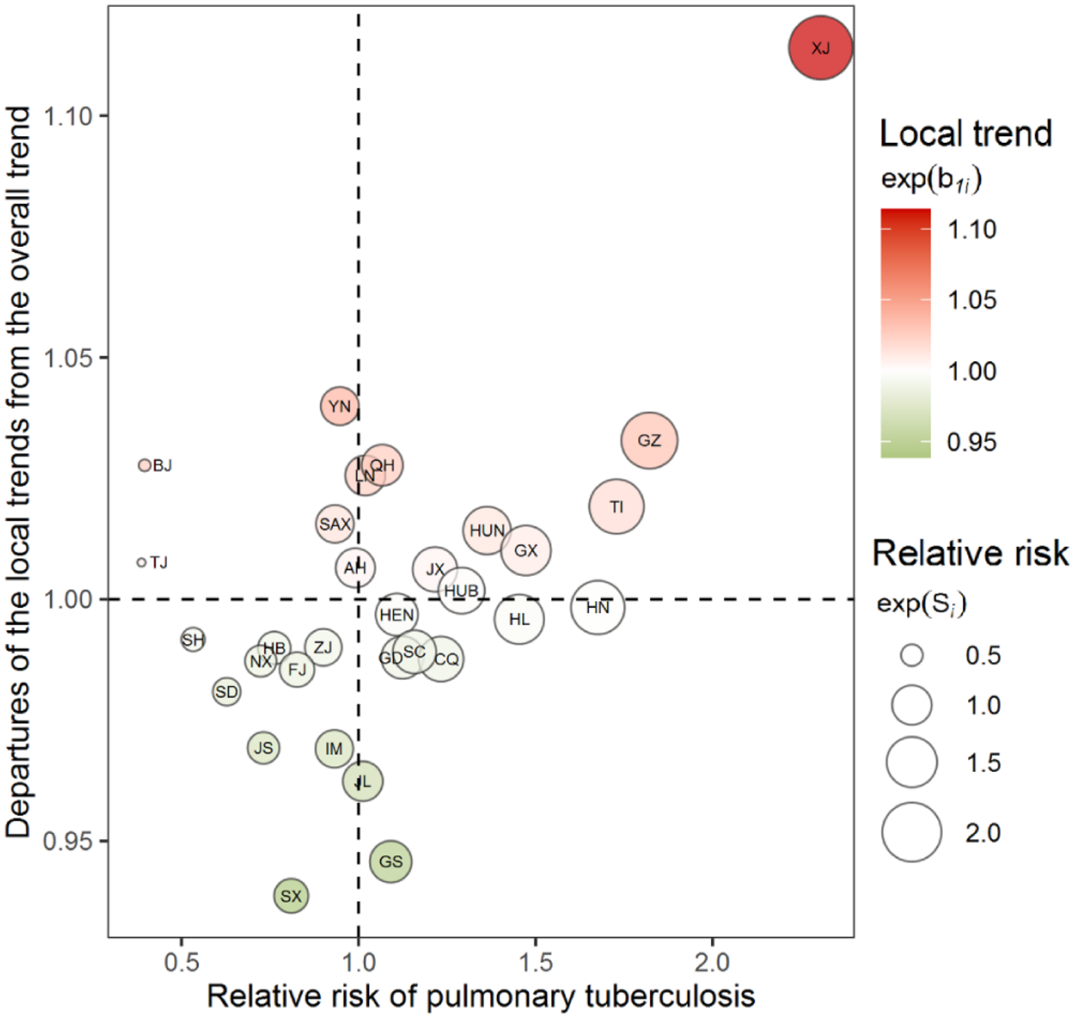
Mongolia (LT 0.97, CI 0.95-0.99, *P*<.001), Jilin (LT 0.96, 95% CI 0.95-0.98; *P*<.001), and Jiangsu (LT 0.97, 95% CI 0.95-0.99; *P*<.001), all of which were faster than the overall trend. At the county level, the proportion of counties with exp(b_{1i})>1 in Xinjiang, Guizhou, and Yunnan provinces was greater than 90%, and the proportion of counties with exp(b_{1i})<1 in Gansu and Shanxi provinces was 95%, displaying a fast decline.

Figure 4. Distribution of the departures of the local trends from the overall trend in mainland China from 2009 to 2018. (a) The departures of the local trends from the overall trend at the province level. (b) The departures of the local trends from the overall trend at the county level.



A summary of the provincial overall RRs of PTB and their departures from the country overall RR is shown in Figure 5 to compare the relative severity of PTB in different provinces. The bottom left quadrant includes provinces with lower RR and a faster decline in RR for PTB than the overall trend, such as Shanxi, Jiangsu, and Shandong. In contrast, the top right quadrant clusters provinces that had higher RRs and a slower decline in RR for PTB than the overall trend, such as Xinjiang, Guizhou, and Tibet.

Figure 5. The severity and efficacy of pulmonary tuberculosis prevention and control in different provinces. Upper left quadrant: Anhui (AH), Beijing (BJ), Shaanxi (SAX), Tianjin (TJ), Yunnan (YN). Upper right quadrant: Guangxi (GX), Guizhou (GZ), Hubei (HUB), Hunan (HUN), Jiangxi (JX), Liaoning (LN), Qinghai (QH), Tibet (TI), Xinjiang (XJ). Lower left quadrant: Fujian (FJ), Hebei (HB), Inner Mongolia (IM), Jiangsu (JS), Ningxia (NX), Shandong (SD), Shanghai (SH), Shanxi (SX), Zhejiang (ZJ). Lower right quadrant: Chongqing (CQ), Gansu (GS), Guangdong (GD), Hainan (HN), Heilongjiang (HL), Henan (HEN), Jilin (JL), Sichuan (SC).



Discussion

We used the number of PTB cases at the county level from 2009 to 2018 to analyze the risk of PTB over a decade in China. The overall average notification rate was 60 (SD 6) per 100,000 people. The RRs for PTB showed a steady downward trend. China, which has one of the world's highest TB burdens, still faces many TB-related public health challenges. Although the TB epidemic in China has decreased significantly in recent years, the imbalance in the burden of TB among different regions is severe, according to the fifth national TB prevalence sampling survey [2]. Generally, the TB prevalence in western regions is higher than those in eastern and central regions, and the TB prevalence is higher in rural areas than in urban areas. Discovering the unevenness of PTB risk between regions and the variability of its trends is important for allocating medical resources for PTB and assessing the efficiency of treatment.

China issued the "Stop Tuberculosis Action Plan (2019-2022)" to further promote the end of the TB epidemic. The plan clarifies the goals and directions of China's recent TB control program [23]. The notification rate of PTB in mainland China showed a downward trend from 2009 to 2018. This phenomenon indicated that the comprehensive implementation of the national TB prevention and control plan had a significant effect, although the declining PTB trend has gradually slowed in recent years.

The spatial distribution of TB is related to the geographical location, economy, population, and environment, and its transmission is characterized by a spatial clustering distribution [5,10,11,24]. The results showed that the RR for PTB in the western region of China was significantly higher than those in the eastern and central regions. The reason might be that patients with PTB in the western region are unable to receive a diagnosis and treatment in a timely manner due to the relatively low economic level and scarcity of resources for medical service systems and facilities. Demographic characteristics, industrial structure, and health habits may also contribute to the high risk of PTB in the western region [9]. Compared with the western region, the central and eastern regions have more developed economies and more advanced medical services [25-27].

The results show that Xinjiang has had the highest PTB notification rate in mainland China for many years [28]. Both the PTB notification case number and notification rate in Xinjiang from 2009 to 2018 gradually increased. A possible reason might be that Xinjiang has adopted a series of measures to screen cases and to control the PTB epidemic. By the end of 2010, Xinjiang had achieved 100% DOTS coverage. In 2013, Xinjiang took the lead in China to implement a new "three-in-one" TB prevention and control service model. This model transformed TB prevention and control activities managed by the local Center for Disease Control (CDC) into an integrated collaboration system involving local CDCs, designated hospitals, and primary medical and health care institutions to coordinate TB prevention and treatment. Policies including centralized isolation treatment for patients during the infectious period and home treatment for patients in the noninfectious period, such as the "centralized medication+nutritious breakfast" approach, have also been implemented in Xinjiang.

The overall PTB notification rate in Tibet showed a decreasing trend, whereas the notification rates in some counties increased. The gradual extension of the TBIMS and Infectious Disease Recording and Reporting Systems (IDRS) in Tibet in recent years is a possible reason for this. As people's health awareness increases, more people receive health examinations, and many potential PTB cases are detected through health examinations [29]. The majority of Tibet's population and cities are concentrated in the lower Yarlung Zangbo River Valley in the southeast. Therefore, the RR for PTB gradually increased in a northwest to southeast direction in Tibet. These regions with a lower altitude and a more suitable climate for human living are dense areas of PTB [30]. In addition, the higher RRs for PTB in Xinjiang and Tibet have certain relationships with attention by local governments and people's living habits [9,30,31]. Many herders have a relatively low educational level, limited economic income, and lack of access to local health services, which promote the continuous spread of PTB and influence PTB prevention and control [29].

The RR for PTB in Qinghai had a relatively stable decreasing trend in the early study period and showed a significant upward trend in 2017 and 2018 [32]. The PTB epidemic in Guizhou has always been at a high level in China, and the main factor affecting PTB transmission is the floating population [33]. The epidemic situation has greatly improved as a result of a shift in the government's focus in recent years and the strengthening of various TB intervention measures. Therefore, there is a need to strengthen health education, improve the accessibility of health services, and balance the distribution of medical and health resources to improve the rates of TB diagnosis and treatment in the western region.

A high gross domestic product per capita often indicates a high quality of life, a balanced diet, and hygiene knowledge among residents. Tianjin, Beijing, Shanghai, and Shandong were the provinces (municipalities) with the lowest RRs for PTB in mainland China. These provinces have high levels of economic development, high levels of education among the population, high levels of TB knowledge of disease prevention, and high medical standards. In contrast, less developed areas are constrained by economic and cultural factors, and the PTB epidemic is serious in these areas. Considering China's large population, high proportion of the population with latent TB infection, unbalanced regional development, aging population, increasing floating population, and heavy economic burden, it is necessary to increase the detection, diagnosis, and treatment of TB on the existing basis; optimize the TB prevention and treatment service system; and increase funding to ensure that China achieves the SDGs on schedule by 2030 [34].

There are some limitations that should be considered when interpreting the study findings. First, we did not analyze any type of TB other than PTB, which may have different temporal and spatial patterns of risk. However, considering that PTB cases account for more than 80% of total TB cases, the overall trend should be similar. Second, the specific effects of economic development and medical resources on TB were not considered in the study, which might also vary significantly in different regions.

In conclusion, the overall rate of PTB notification decreased steadily in the last decade in China. However, the notification rates between different provinces and within each province exhibited significant regional inequality. Western China presented a high plateau of the disease burden. These regions also had relatively low declining trends compared to the national overall trend. Improvements in economic and medical service levels in such regions are required to boost PTB case detection and eventually reduce PTB risk in the whole country. These findings improve our understanding of the characteristics of the PTB distribution in China, and provide evidence for the spatial strategy to contain the PTB epidemic.

Acknowledgments

This work was supported by the National Science and Technology Major Project (grant number 2017ZX10201302) and the National Natural Science Foundation of China (grant numbers 41771434, 41531179). The views expressed in this article are those of the authors and do not represent any official policy. All the authors have declared no relationships or activities that could appear to have influenced this work. The funders had no role in study design, data collection and analyses, decision to publish, or preparation of the manuscript.

Authors' Contributions

WC and MH designed the study, collected data, interpreted the findings, and commented on the manuscript. MH, YF, and TL collected data, built the model, finalized the analysis, interpreted the findings, and wrote the manuscript. YZ, JW, and CX interpreted the findings, and commented on and revised drafts of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Classification result based on exp(s_i) posterior distribution in mainland China from 2009 to 2018. (a) Classification result based on exp(s_i) posterior distribution at the province level; (b) Classification result based on exp(s_i) posterior distribution at the county level.

[PNG File , 623 KB - publichealth_v8i7e36242_app1.png]

References

1. Global Tuberculosis Report 2020. World Health Organization. 2020. URL: https://www.who.int/publications/i/item/9789240013131 [accessed 2022-05-04]
2. Technical Guidance Group of the Fifth National TB Epidemiological Survey, The Office of the Fifth National TB Epidemiological Survey. The fifth national tuberculosis epidemiological survey in 2010. Chinese J Antituberculosis 2012 Aug;34(8):485-508 [FREE Full text]
3. Notice of the General Office of the State Council on Issuing the National Tuberculosis Prevention and Control Plan during the 13th Five-year. General Office of the State Council of the People's Republic of China. URL: http://www.gov.cn/zhengce/content/2017-02/16/content_5168491.htm [accessed 2022-06-11]
4. Wang L, Zhang H, Ruan Y, Chin D, Xia Y, Cheng S, et al. Tuberculosis prevalence in China, 1990-2010; a longitudinal analysis of national survey data. Lancet 2014 Jun 14;383(9934):2057-2064. [doi: 10.1016/S0140-6736(13)62639-2] [Medline: 24650955]
5. Li Q, Liu M, Zhang Y, Wu S, Yang Y, Liu Y, et al. The spatio-temporal analysis of the incidence of tuberculosis and the associated factors in mainland China, 2009-2015. Infect Genet Evol 2019 Nov;75:103949. [doi: 10.1016/j.meegid.2019.103949] [Medline: 31279820]
6. Jiang Y, Dou X, Yan C, Wan L, Liu H, Li M, et al. Epidemiological characteristics and trends of notifiable infectious diseases in China from 1986 to 2016. J Glob Health 2020 Dec;10(2):020803. [doi: 10.7189/jogh.10.020803] [Medline: 33214900]
7. He W, Ju K, Gao Y, Zhang P, Zhang Y, Jiang Y, et al. Spatial inequality, characteristics of internal migration, and pulmonary tuberculosis in China, 2011-2017: a spatial analysis. Infect Dis Poverty 2020 Nov 19;9(1):159 [FREE Full text] [doi: 10.1186/s40249-020-00778-0] [Medline: 33213525]
8. Zuo Z, Wang M, Cui H, Wang Y, Wu J, Qi J, et al. Spatiotemporal characteristics and the epidemiology of tuberculosis in China from 2004 to 2017 by the nationwide surveillance system. BMC Public Health 2020 Aug 26;20(1):1284 [FREE Full text] [doi: 10.1186/s12889-020-09331-y] [Medline: 32843011]
9. Wang Y, Wang J, Wang W. Temporal-spatial distribution of tuberculosis in China, 2004-2016. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Apr 10;41(4):526-531. [doi: 10.3760/cma.j.cn112338-20190614-00441] [Medline: 32344476]
10. Wang Q, Guo L, Wang J, Zhang L, Zhu W, Yuan Y, et al. Spatial distribution of tuberculosis and its socioeconomic influencing factors in mainland China 2013-2016. Trop Med Int Health 2019 Sep;24(9):1104-1113. [doi: 10.1111/tmi.13289] [Medline: 31314953]
11. Liu M, Li Q, Zhang Y, Ma Y, Liu Y, Feng W, et al. Spatial and temporal clustering analysis of tuberculosis in the mainland of China at the prefecture level, 2005-2015. Infect Dis Poverty 2018 Oct;7(1):106 [FREE Full text] [doi: 10.1186/s40249-018-0490-8] [Medline: 30340513]
12. Jiang H, Zhang G, Yin J, Zhao D, Liu F, Yao Y, et al. Assessment of strategies and epidemiological characteristics of tuberculosis in Henan Province, China: observational study. JMIR Public Health Surveill 2021 Jan 22;7(1):e24830 [FREE Full text] [doi: 10.2196/24830] [Medline: 33480857]
13. Banerjee S, Carlin B, Gelfand A. Hierarchical modeling and analysis for spatial data. Boca Raton, FL: Chapman and Hall/CRC Press; 2014.
14. Cressie N, Wikle C. Statistics for spatio-temporal data. Hoboken, NJ: John Wiley & Sons; 2015.
15. Haining R. Spatial data analysis: theory and practice. Cambridge: Cambridge University Press; 2003.
16. Lawson A. Bayesian disease mapping: hierarchical modeling in spatial epidemiology, second edition. Boca Raton, FL: CRC Press; 2018.
17. McCullagh P. Generalized linear models. Milton Park, UK: Routledge; 1983.
18. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 1991 Mar 1;43(1):1-20. [doi: 10.1007/bf00116466]

19. Morris M, Wheeler-Martin K, Simpson D, Mooney S, Gelman A, DiMaggio C. Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan. Spat Spatiotemporal Epidemiol 2019 Nov 1;31:100301 [FREE Full text] [doi: 10.1016/j.sste.2019.100301] [Medline: 31677766]
20. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Anal 2006 Sep 1;1(3):515-534. [doi: 10.1214/06-BA117A]
21. Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 2000 Oct;10(4):325-337. [doi: 10.1023/A:1008929526011]
22. Gelman A, Rubin D. Inference from iterative simulation using multiple sequences. Statist Sci 1992 Nov 1;7(4):457. [doi: 10.1214/ss/1177011136]
23. Notice on issuing the action plan to stop tuberculosis. National Health Commission of the People's Republic of China. 2019. URL: http://www.nhc.gov.cn/jkj/s3589/201906/b30ae2842c5e4c9ea2f9d5557ad4b95f.shtml [accessed 2022-06-11]
24. Yang L, Li T, Chen W. Study on spatial clustering characteristics of tuberculosis in China, 2013-2018. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Nov 10;41(11):1843-1847. [doi: 10.3760/cma.j.cn112338-20200114-00029] [Medline: 33297649]
25. Yang Z, Shao S, Xu L, Yang L. Can regional development plans promote economic growth? City-level evidence from China. Socio-Economic Planning Sciences 2021 Dec:101212. [doi: 10.1016/j.seps.2021.101212]
26. Zhang Y, Wang Q, Jiang T, Wang J. Equity and efficiency of primary health care resource allocation in mainland China. Int J Equity Health 2018 Sep 12;17(1):140. [doi: 10.1186/s12939-018-0851-8] [Medline: 30208890]
27. Tian X, Geng Y, Viglia S, Bleischwitz R, Buonocore E, Ulgiati S. Regional disparities in the Chinese economy. An emergy evaluation of provincial international trade. Resour Conserv Recycl 2017 Nov;126:1-11. [doi: 10.1016/j.resconrec.2017.07.017]
28. Mijiti P, Yuehua L, Feng X, Milligan P, Merle C, Gang W, et al. Prevalence of pulmonary tuberculosis in western China in 2010-11: a population-based, cross-sectional survey. Lancet Glob Health 2016 Jul;4(7):e485-e494 [FREE Full text] [doi: 10.1016/S2214-109X(16)30074-2] [Medline: 27283762]
29. Yang G, Wang J, Zhang B, Wang T. Analysis on epidemiological characteristics of tuberculosis in Tibet Autonomous Region from 2008 to 2017. Chinese Journal of Disease Control and Prevention 2020;24(9):1106-1109. [doi: 10.16462/j.cnki.zhjbkz.2020.09.023]
30. Zhang L, Sun Y, Yang L, Li G. Prevalence of tuberculosis in Tibet. Journal of Clinical Pulmonary Medicine 2011;16(8):1268-1269 [FREE Full text]
31. Li B, Zhang X, Guo J, Wang J, Pianduo B, Wei X, et al. Prevalence of pulmonary tuberculosis in Tibet Autonomous Region, China, 2014. Int J Tuberc Lung Dis 2019 Jun 01;23(6):735-740. [doi: 10.5588/ijtld.18.0614] [Medline: 31315707]
32. Wang CC, Liu Y, Cao XZ, Jiang MX, Li EC. Epidemiological characteristics of pulmonary tuberculosis of Qinghai, 2006-2015. Modern Preventive Medicine 2017;44(3):397-400 [FREE Full text]
33. Yuan W, Chen P, He Y, Chen H, Zhang X, Song Q. Major influencing factors of tuberculosis epidemic control in Guizhou province: an analysis on surveillance data of 2006 -2017. Chinese Journal of Public Health 2019;35(2):282-285. [doi: 10.11847/zgggws1119713]
34. Wang X, Li Y, Yi F, Fu Q. Description of epidemic features and control status on tuberculosis in China, 1990-2017. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Jun 10;41(6):856-860. [doi: 10.3760/cma.j.cn112338-20190809-00587] [Medline: 32564549]

Abbreviations

BYM: Besag-York-Mollie
CDC: Center for Disease Control
DOTS: directly observed treatment, short-course
ICAR: intrinsic conditional autoregressive
IDRS: Infectious Disease Recording and Reporting Systems
LT: local trend
MCMC: Markov chain Monte Carlo
PTB: pulmonary tuberculosis
RR: relative risk
SDG: sustainable development goal
TB: tuberculosis
TBIMS: National Tuberculosis Information Management System

Edited by G Eysenbach, H Bradley; submitted 06.01.22; peer-reviewed by A Kibballi Madhukeshwar, N Mohammad Gholi Mezerji; comments to author 22.04.22; revised version received 06.05.22; accepted 10.05.22; published 01.07.22.

Please cite as:
Hu M, Feng Y, Li T, Zhao Y, Wang J, Xu C, Chen W
Unbalanced Risk of Pulmonary Tuberculosis in China at the Subnational Scale: Spatiotemporal Analysis
JMIR Public Health Surveill 2022;8(7):e36242
URL: https://publichealth.jmir.org/2022/7/e36242
doi:10.2196/36242
PMID:35776442

Original Paper

Effects of an mHealth Intervention for Pulmonary Tuberculosis Self-management Based on the Integrated Theory of Health Behavior Change: Randomized Controlled Trial

Yuhan Bao[1*], MM; Chunxiang Wang[2*], PhD; Haiping Xu[3], BS; Yongjie Lai[3], BS; Yupei Yan[4], BA; Yuanyuan Ma[5], BM; Ting Yu[5], BM; Yibo Wu[6,7], MPharm, PhD

[1]Health Clinic, Changzhou Institute of Technology, Changzhou, China
[2]School of Sciences, Changzhou Institute of Technology, Changzhou, China
[3]School of Pharmaceutical Sciences, Shandong University, Jinan, China
[4]Department of Humanities, Arts and Media, Changzhi Medical College, Changzhi, China
[5]School of Public Health, Shandong University, Jinan, China
[6]School of Public Health, Peking University, Beijing, China
[7]Key Research Base of Philosophy and Social Sciences in Shaanxi Province, Health Culture Research Center of Shaanxi, Xi'an, China
[*]these authors contributed equally

Corresponding Author:
Yibo Wu, MPharm, PhD
School of Public Health, Peking University
38 Xueyuan Road, Haidian District
Beijing, 100191
China
Phone: 86 18810169630
Email: bjmuwuyibo@outlook.com

Abstract

Background: Improving the health self-management level of patients with tuberculosis (TB) is significant for reducing drug resistance, improving the cure rate, and controlling the prevalence of TB. Mobile health (mHealth) interventions based on behavioral science theories may be promising to achieve this goal.

Objective: This study aims to explore and conduct an mHealth intervention based on the Integrated Theory of Health Behavior Change (ITHBC) in patients with pulmonary TB to increase their ability of self-care management.

Methods: A prospective randomized controlled study was conducted from May to November 2020. A total of 114 patients who were admitted consecutively to the TB clinic of Harbin Chest Hospital, China from May 2020 to August 2020 were recruited by convenience sampling. Patients were divided into the control group and intervention group, and all received a 3-month intervention. Patients in the intervention group and the control group received routine medical and nursing care in the TB clinic, including the supervision of their medications. In addition, pharmacist-assisted mHealth (WeChat) intervention based on the ITHBC theory about TB management was provided to the intervention group. The primary outcome was self-management behavior, while the secondary outcomes were TB awareness, self-efficacy, social support, and degree of satisfaction with health education. The outcomes were measured using web-based self-designed and standard questionnaires administered at baseline and at the end point of the study. Intergroup data were assessed using the Mann-Whitney *U* test, whereas intragroup data were assessed with the Wilcoxon test (for paired samples).

Results: A total of 112 patients (59 in intervention group and 53 in control group) completed the study. After the intervention, a statistically significant increase was noted in the scores of each item of self-care management behaviors compared with the scores at the baseline ($P<.001$) in the intervention group. The scores of all self-care management behaviors of the control group were lower than those of all self-care management behaviors in the intervention group (all $P<.05$), except for the item “cover your mouth and nose when coughing or sneezing” ($P=.23$) and item “wash hands properly” ($P=.60$), which had no statistically significant difference from those in the intervention group. Compared with those at baseline, TB knowledge awareness, self-efficacy, social support, and degree of satisfaction with health education in the intervention group increased significantly ($P<.001$), and the intervention group had significantly higher scores than the control group ($P<.001$).

Conclusions: mHealth intervention for TB self-management based on ITHBC could deepen the understanding of patients with TB about their disease and improve their objective initiative and self-care management behaviors, which were beneficial for promoting compliance behavior and quality of prevention and control for pulmonary TB.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2200055557; https://tinyurl.com/4ray3xnw

(JMIR Public Health Surveill 2022;8(7):e34277) doi:10.2196/34277

KEYWORDS

ITHBC; mHealth; RCT; pulmonary tuberculosis

Introduction

Tuberculosis (TB) is a communicable disease, which is one of the top 10 causes of death worldwide and the leading cause of death from a single infectious agent (ranking above AIDS). According to the estimates of the World Health Organization, in 2019, there were 9.87 million new cases of TB and 1.28 million deaths. China is one of the 30 high TB burden countries. In 2020, the number of new TB cases in China was 842,000, ranking second in the world [1]. Hence, TB is a major public health problem with high incidence and mortality worldwide.

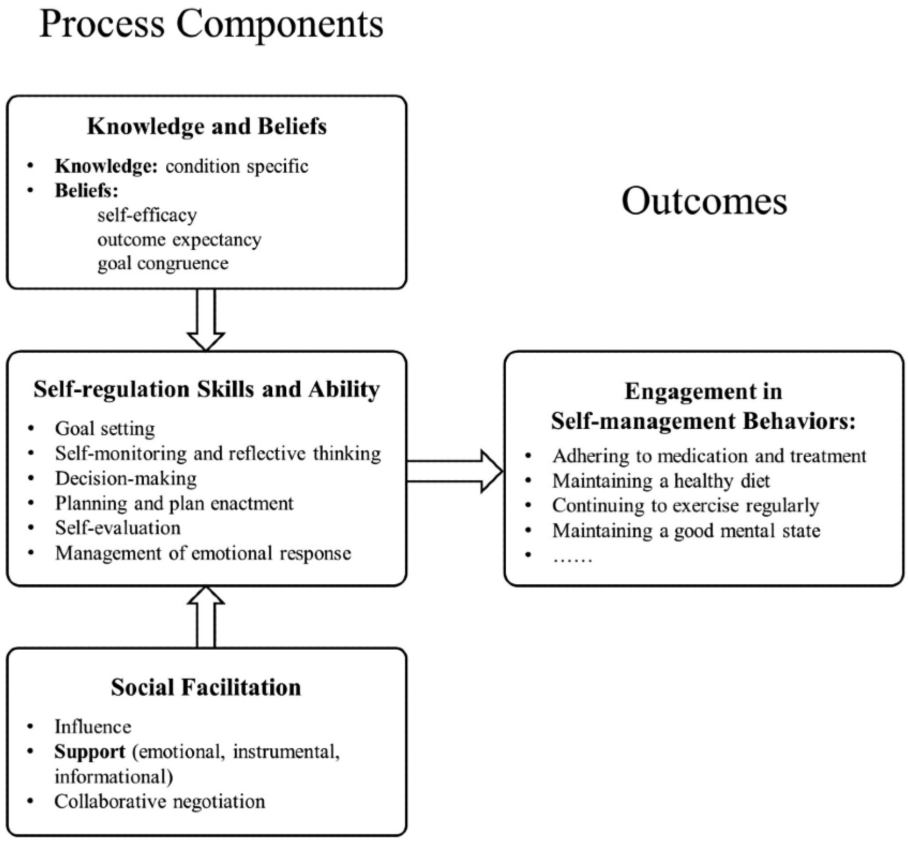
Self-management is defined as a task that patients undertake to deal with the medical, role, and emotional management of their chronic conditions [2]. For patients with TB, self-management includes adhering to medication and treatment, maintaining a healthy diet and adequate amount of exercise, keeping a good mental state, and strengthening personal capacity to solve problems [3]. Therefore, improving the self-management level of patients with TB is of great significance for controlling their illness, increasing their quality of life, improving the cure rate, and controlling the prevalence of TB [4].

Public health programs have used various interventions to improve the self-management level of patients with TB, especially with regard to their adherence to TB treatment [5]. One of the most common interventions is directly observed therapy (DOT), in which a health worker, family member, or community member observes the patient taking TB medications [6]. The fact that people can be closely monitored and the social process with peer pressure are the advantages of DOT that may improve patients’ medication adherence [7]. However, until now, DOT is still not optimal because of its inconvenience and labor-intensive practice [8,9]. Moreover, the COVID-19 pandemic has resulted in the need for “social distancing,” which has caused the suspension of DOT and an exponential use of mobile health (mHealth) approaches for patient care [10]. Further, the World Health Organization has called for maximizing remote care and support for people with TB by expanding the use of digital technologies [1]. All these factors have driven the development of mHealth that can be defined as using mobile computing and communication technologies in support of health care to supplement the traditional clinical-based care [11]. Common mHealth interventions and programs include video-based DOT and mobile phone text messaging to support treatment compliance and health education [9]. So far, the acceptability, feasibility, and efficiency of mHealth in improving patient adherence and supporting TB treatment have shown promising results [12]. This study will take advantage of mHealth and make use of web-based health interventions for patients.

Although evidence indicates that health promotion interventions based on behavioral science theories are more effective than those without theoretical models, only few interventional studies related to TB self-management have used theoretical models as guidance [13]. Therefore, we aimed to evaluate an mHealth intervention to improve the self-management level of patients with TB based on the Integrated Theory of Health Behavior Change (ITHBC). The ITHBC suggests that health behavior change can be promoted by enhancing knowledge and beliefs, strengthening social facilitation, and increasing self-regulation skills and abilities, of which the first two promote the latter [14]. Engagement in self-management behaviors is regarded as the proximal outcome influencing the distal outcome of improved health status [15]. The ITHBC summarizes the key components of health behavior change processes and provides the pivotal components for intervention development, which has great referential significance for the design of intervention measures in this study. The theoretical framework of ITHBC is shown in Figure 1.

WeChat is the most popular social media platform in China [16]. Its wide application in the daily life of every Chinese indicates that it a promising new medium for delivering health care in a cost-effective way. Accumulating evidence has robustly proved that WeChat-based mHealth is acceptable, feasible, and cost-effective in improving health outcomes in various health conditions [17,18]. Therefore, we aimed to develop an ITHBC-based mHealth intervention targeting self-management in patients with TB and to study whether this WeChat-based intervention can significantly improve the self-management level of patients with TB and the cure rate of TB and reduce the infection rate and drug resistance.

Figure 1. The theoretical framework of Integrated Theory of Health Behavior Change.



Methods

Study Design

A prospective randomized controlled trial was conducted in Harbin Chest Hospital, China from May to November 2020. Convenience sampling was performed to recruit 114 patients with pulmonary TB. They were assigned into the intervention group and control group in a 1:1 ratio by using a computer-based random number generator, and all received a 3-month intervention. WeChat groups were created by a pharmacist to provide health education for the participants. Patients in the intervention group and the control group received routine medical and nursing care in the TB clinic, including the supervision of their medications. In addition, pharmacist-assisted mHealth intervention based on the ITHBC theory about TB management was provided to the intervention group. This trial is reported in accordance with CONSORT-eHEALTH (see Multimedia Appendix 1 for the completed CONSORT-eHEALTH form V 1.6.1).

Based on the ITHBC theory, we summarized the epidemical characteristics of TB and the self-management and psychology of patients with TB through a literature review. The research team created a multidisciplinary panel comprising of a TB pharmacist, 2 nursing experts, 2 TB physicians, and a public health expert to discuss and develop the intervention program. A statistician was also included to be responsible for the random allocation of the participants and data analysis. Semistructured interviews were conducted with 4 patients randomly selected in the finally included participants to know their mastery of TB knowledge, diets, exercise habits, risk factors and awareness, personal needs and wishes, and concerns and inertia. After communication, we conducted individualized assessments to refine the intervention program. Finally, 20 participants who would not be enrolled in the final trials were randomly selected to complete the pretest to refine self-designed scales. The face and content validity of the self-designed questionnaire were determined by pretest and experts, respectively.

Sample Size

The formula for sample size calculation was as follows:

We assumed σ=1.6 and δ=1.02 with reference to a preliminary study [19] when α=1.6, 1-β=0.80, and then we got n=40 for 1 group. Considering dropouts (20%) over the course of the study, 100 patients were included (50 for control group and 50 for intervention group).

Study Participants

The participants in this study were patients who were admitted consecutively to the TB clinic of Harbin Chest Hospital from May to August 2020, and the study was conducted between May and November 2020. Each participant's intervention started as long as they were enrolled offline at the hospital by face-to-face communication. Therefore, all the patients were not recruited or started with the intervention at the same time, but all completed the 3-month intervention.

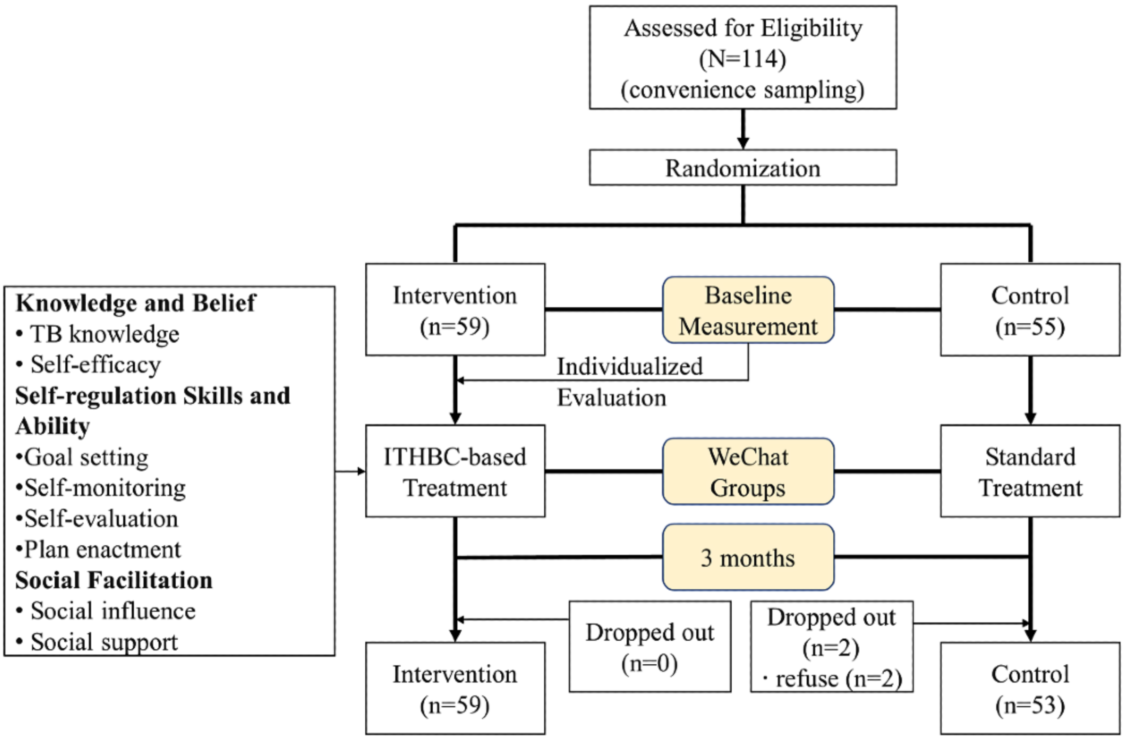
The eligibility criteria were (1) 18 years old or older, (2) experiencing newly diagnosed active pulmonary TB according to the classification of TB (WS 196-2017) issued by the People's

Republic of China State Health and Family Planning Commission [20], (3) literate and capable of using WeChat, and (4) hospitalized when enrolled in the study. The exclusion criteria were as follows: (1) experiencing drug-resistant pulmonary TB or treated according to the treatment plan of drug-resistant TB and (2) complicated with serious diseases such as AIDS, malignant tumor, and severe diseases that were newly diagnosed in the heart, brain, liver, and kidney. All patients in both the groups received the therapy in the hospital for nearly 1 month since the intervention started and then received TB outpatient treatment for 2 months. The median hospitalization time for the intervention group was 25 (IQR 18-32) days, while that for the control group was 24 (IQR 16-31) days (*P*=.35). All patients gave web-based informed consent to participate in this study (see Multimedia Appendix 2 for the informed consent). Patients whose treatment regimens were changed or lost to follow-up were terminated from this study.

Interventions

Intervention for the intervention group on the day of hospitalization included health education in WeChat groups, web-based health education lectures, and receiving health education plans as well as routine medical and nursing care such as supervision of their medications, which continued for 1 month. Then they received therapy at outpatient clinics and were provided with the same intervention as provided in the hospital for 2 months (see Multimedia Appendix 3 for the screenshots of the health education in the WeChat group). Patients in the control group received routine medical and nursing care in the TB clinic, including supervision of their medications. Meanwhile, the pharmacist also created a WeChat group for the control group to allow patients to communicate with each other but without any intervention. The content of the intervention was designed based on the ITHBC theory, and the flow chart is presented in Figure 2.

Figure 2. The flow chart of the randomized controlled trial. ITHBC: Integrated Theory of Health Behavior Change; TB: tuberculosis.



Knowledge and Belief

Information and desire to change are the prerequisites that patients engage in recommended behaviors. In our study, health education can be divided into knowledge education and belief education. Knowledge was defined as information of pulmonary TB, and belief was defined as personal perceptions about patients' health condition or health behaviors [14]. Knowledge education contained basic knowledge, diet management, hygiene routines, and medication management. Belief education was conducted in weekly meetings to increase the consciousness of crisis for disease and patient confidence for disease management. The pharmacist gave individualized health education based on baseline information. Education forms included WeChat groups and online lectures. The topics of the health education are presented in Multimedia Appendix 4, which were sent to the WeChat group in the form of articles, pictures, or videos. Patients were guided every Monday for 10 minutes in WeChat. Web-based lectures were conducted every Thursday for an hour through WeChat.

Self-regulation Skills and Ability

Patients with the desire to change health behaviors should learn self-regulation skills and the ability to apply TB knowledge to their lives. To improve self-regulation skills and ability, the pharmacist summarized patients' current behavior and issues measured at the baseline to make an individualized health education plan in a month and a short-term aim in a week during hospitalization. The plans in the latter 2 months would be made according to performance of the patients in the previous month. Patients were asked to self-monitor and record daily health behaviors and emotions. A meeting was held by the pharmacist in the WeChat group every Saturday, where patients reported and self-evaluated their own behaviors and emotions; moreover, they had exchange of views with wardmates. The pharmacist would give some timely guidance to help the patients make the next aim themselves.

Social Facilitation

Excellent social facilitation positively influences and supports patients' engagement in health behaviors. According to the ITHBC theory, social facilitation includes social influence and social support. Patients are more likely to engage in health behaviors when authorities sway their thinking and motivation, which is called social influence. In our study, social influence came from health care providers, especially pharmacists who offered the intervention, family, and significant others (eg, wardmates, friends, neighbors, colleagues, social media). Social support includes emotional, instrumental, and informational support, which were offered by health care providers, family, friends, and others. The patient's family received health education in rounds with patients, attended web-based lectures, and supervised the patients. They were encouraged to take the lead in adopting health behaviors and provide comfort and company for the patients in their family. The pharmacist provided instrumental and informational support by establishing a WeChat group and health education. Wardmates communicated with and encouraged each other in wards and in the WeChat group. During the intervention, when the country was in lockdown during the COVID-19 epidemic, patients also received much tutorial advice on preventive behaviors, as the coronavirus transmission is similar to that of pulmonary TB virus.

Outcome Measures

All the patients filled out a web-based questionnaire of demographic information at baseline (see Multimedia Appendix 5 for the questionnaire 1). The primary outcome was self-management behavior. Secondary outcomes were TB knowledge awareness, self-efficacy, social support, and degree of satisfaction with health education. The outcome data were collected by scales at baseline and at the end of the study when patients were hospitalized and at home, respectively. Clinical pharmacists used the web-based platform of Wen Juan Xing [21] to distribute the questionnaire (see Multimedia Appendix 5 for the questionnaire 2), and patients filled it while the pharmacist guided them face-to-face (in hospitalization) or telephone (after discharge). The demographic information of the patients was collected by the medical record system and the self-designed questionnaire.

Primary Outcome: Self-management Behavior

Self-management behavior was measured using a self-designed structured scale with item 4 in questionnaire 2. General self-management behaviors consisted of COVID-19-related behaviors, medication behaviors, and lifestyle habits with 13 items and a 5-level Likert scoring method. Each item was scored as 1-4 points, with 1=no intention to act, 2=intention to act, 3=having acted but stopped, 4=starting to act, and 5=having acted and considering keeping on. Higher scores indicated greater self-management behaviors. Cronbach α was .966, suggesting good internal consistency of the scale.

Secondary Outcomes

TB Knowledge Awareness

A self-designed constructed questionnaire (item 3) was administered to evaluate TB knowledge awareness in basic knowledge and hygiene routines with 5 items and a 5-level Likert scoring method. Answers were comprised of totally disagree (score=1), disagree (score=2), uncertain (score=3), agree (score=4), and totally agree (score=5). The score of each item was summed to give a total TB knowledge score, with higher score indicating a better knowledge level. The coefficient of Cronbach α was .772.

Self-efficacy

The chronic disease self-efficacy scale [22] (items 6-11) was applied to measure the self-efficacy of patients. This scale has 6 items, including symptom management, daily life management, emotional management, and disease control. A 10-point scoring test was used, and each item was rated from 1=not at all confident to 10=completely confident. The score of each item was summed to give the scale's score (range 6-60), and higher scores indicated higher self-efficacy level. The coefficient of Cronbach α was .910.

Social Support

The perceived social support scale, designed by Blumenthal et al [23] (item 2), was used to assess a patient's perception of the social support from family, friends, and significant others. The scale had 12 items and contained 3 subscales of 4 items each. Response options ranged from 1=totally dissatisfied to 7=totally satisfied. Higher scores indicated more social support. The coefficient of Cronbach α was .97.

Degree of Satisfaction With Health Education

To assess patients' satisfaction of the intervention organized by the pharmacists, a self-designed scale (item 5) based on the Unified Theory of Acceptance and Use of Technology [24] was applied to evaluate the degree of satisfaction with health education. This item had 10 subitems and consisted of 3 sections, which are performance expectancy (4 subitems), effort expectancy (4 subitems), and facilitating conditions (2 subitems). It was a 7-point scale from “totally dissatisfied” to “totally satisfied.” A higher score indicated higher satisfaction. The coefficient of Cronbach α was .963.

Quality Control

In order to ensure the quantity and homogeneity of the participants, the participants were selected in strict accordance with the criteria of inclusion and exclusion, which were checked by TB physicians. We chose newly diagnosed patients with active pulmonary TB to make sure of a consistent treatment plan and to reduce bias caused by different medication therapies. A study [25] has shown that the generation of a random sequence should be done by some independent personnel, usually a statistician, who is not going to be involved in the conduct of the randomized controlled trial. To ensure the scientific nature of the random allocation, this study had a statistician responsible for the random allocation of the participants and the data analysis. The researchers discussed the intervention measures of this study with an interdisciplinary panel in the field of TB and conducted structured interviews with the participants in order to further improve the intervention program. In addition, a pretest was conducted before formal intervention to find the problems in the design of the intervention and to optimize the intervention program. The clinical pharmacist explained the purpose and significance of this study to the participants and obtained informed consent. In the process of filling out the questionnaire, the clinical pharmacist helped to explain the questionnaire to the patients to ensure that the questionnaire was completed efficiently, but it was not instructive. Moreover, when using the data collection tool, known as Wen Juan Xing, patients could not submit the questionnaire unless they completed every question, which avoided missing data.

Statistical Analysis

Statistical analysis was implemented using SPSS Statistics 26.0 software (IBM Corp). Normality tests were applied to assess the distribution of the continuous data, and nonnormally distributed data were presented as median (IQR). The Mann-Whitney *U* test was used to compare the changes in each scale score between the 2 groups at the baseline and the end point, while the Wilcoxon test (for paired samples) was applied to compare the difference between the baseline and the end point of each scale score in each group. The demographic characteristics of the 2 groups were compared using the chi-square test. *P*<.05 (2-tailed test) was considered as statistically significant.

Ethics Approval

This study was approved by the ethics committee of Harbin Chest Hospital, China (2020-10).

Results

Characteristics of the Patients

A total of 114 patients participated in this study, and they were randomly assigned to the intervention group (n=59) and the control group (n=55). During the study period, 2 patients dropped out (2/114, 1.8%). Therefore, 112 patients were included in the statistical analysis finally. Of the 112 patients, 64 (57.1%) were males. All patients were aged 18 years and older and were mainly aged between 18 and 30 years (39/112, 34.8%), and only 7 (6.3%) were older than 60 years. Approximately half of the patients had a college degree or higher (51/112, 45.5%). Most patients (101/112, 90.2%) lived with their families or friends, and 71 (63.4%) were married. The demographic information of the patients is shown in Table 1. At baseline, approximately all patients' characteristics were similarly distributed between the intervention and the control groups.

Table 1. Demographic characteristics of the patients.

Characteristics	Intervention group (n=59), n (%)	Control group (n=53), n (%)	χ^2 (*df*)	*P* value
Gender			1.1 (1)	.30
Female	31 (53)	33 (62)		
Male	28 (47)	20 (38)		
Age group (years)			1.7 (4)	.79
18-30	21 (36)	18 (34)		
31-40	15 (25)	12 (23)		
41-50	7 (12)	11 (21)		
51-60	12 (20)	9 (17)		
>60	4 (7)	3 (6)		
Race/ethnicity			0 (1)	.93
Han	58 (98)	51 (96)		
Other ethnicities	1 (2)	2 (4)		
Religious belief			0 (1)	>.99
There is belief	4 (7)	4 (8)		
No belief	55 (93)	49 (92)		
Marital status			1.9 (2)	.40
Unmarried	18 (31)	19 (36)		
Married	40 (68)	31 (58)		
Others	1 (2)	3 (6)		
Per capita monthly income (CNY)[a]			2.8 (5)	.74
<1000	8 (14)	11 (21)		
1000-1999	8 (14)	9 (17)		
2000-2999	16 (27)	13 (25)		
3000-3999	15 (25)	13 (25)		
4000-4999	4 (7)	1 (2)		
≥5000	8 (14)	6 (11)		
Usual residence			1.4 (1)	.23
Rural	12 (20)	16 (30)		
Urban	47 (80)	37 (70)		
Education			0.7 (3)	.88
Primary school and below	3 (5)	3 (6)		
Junior middle school	14 (24)	10 (19)		
High school/technical secondary school	17 (29)	14 (26)		
College and above	25 (42)	26 (49)		
Residence status			0.6 (1)	.44
Alone	7 (12)	4 (8)		
Others	52 (88)	49 (92)		
Relatives or friends were infected by tuberculosis			0 (1)	>.99
Yes	5 (8)	5 (9)		
No	54 (92)	48 (91)		

[a] 1 CNY=US $0.14.

Primary Outcome: Self-management Behavior

The total scores of all the 13 self-management behaviors were calculated. Both groups scored greater than that at baseline (*P*<.001 vs *P*=.04, respectively), but the total score of the intervention group was higher than that of the control group (*P*<.001) (Figure 3). Table 2 shows that there was no significant difference in the scores of all self-management behaviors between the intervention group and the control group at baseline.

After the intervention, there were significant differences in the scores of the above behaviors between the 2 groups except for “covering your mouth and nose when coughing or sneezing” (*P*=.23) and “wash hands properly” (*P*=.60) (Table 3).

The scores of self-management of health behaviors in the intervention group increased significantly after the intervention (*P*<.001). The 4 health behaviors with significant increases in the scores in the control group were “stay in a good mood” (*P*=.002), “maintain a balanced diet and ensure the intake of high protein, vitamins, and minerals” (*P*=.002), “cover your nose and mouth when coughing or sneezing” (*P*<.001), and “wash hands properly” (*P*<.001). Although the score of the item “self-isolate from families in home-based treatment” was lower than that at the baseline (*P*=.01), there was no significant change in the scores of other behaviors in the control group (Table 4).

Figure 3. Mean scores of self-management behaviors for both groups before and after the intervention.

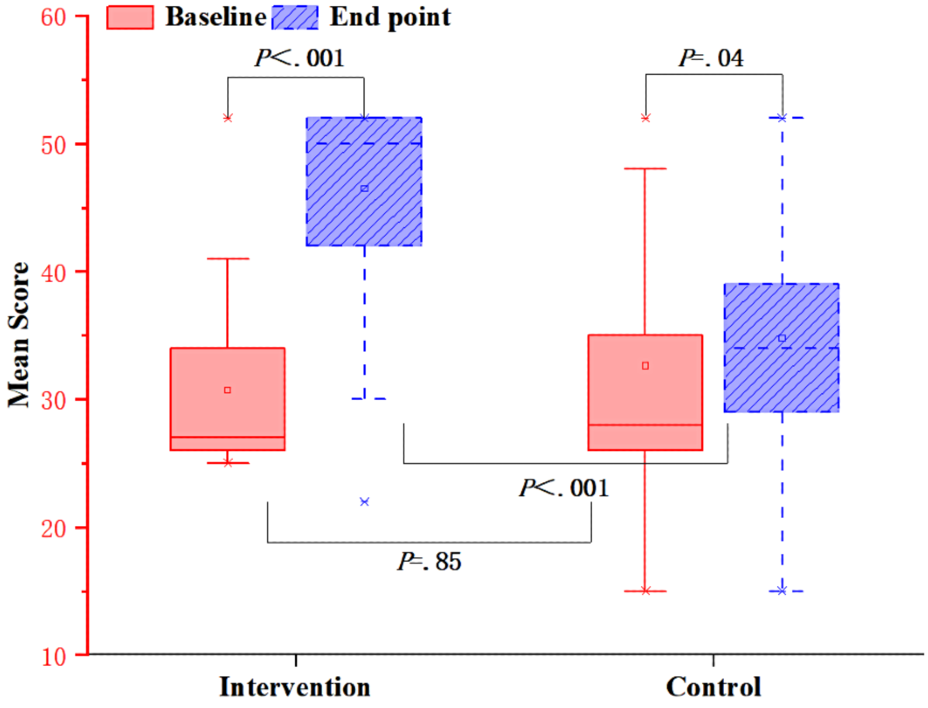


Table 2. Scores of self-management health behaviors at baseline in the intervention and control groups.

Self-management behaviors	Intervention group		Control group		*Z*	*P* value
	Median (P25-P75)	Mean	Median (P25-P75)	Mean		
(1) Self-isolate from families in home-based treatment	2.00 (1.00-2.00)	1.88	2.00 (1.00-2.00)	1.91	–0.593	.55
(2) Keep your room well-ventilated	2.00 (2.00-2.00)	2.25	2.00 (2.00-3.00)	2.49	–1.513	.13
(3) Maintain a balanced diet and ensure enough intake of high protein, vitamins, and minerals	2.00 (2.00-3.00)	2.36	2.00 (2.00-3.00)	2.47	–0.575	.57
(4) Stick to daily exercise	2.00 (2.00-2.00)	2.24	2.00 (2.00-3.00)	2.47	–1.519	.13
(5) Go to bed and wake up early, and alternate work with rest	2.00 (2.00-2.00)	2.24	2.00 (2.00-2.00)	2.36	–0.565	.57
(6) Stay in a good mood	2.00 (2.00-3.00)	2.32	2.00 (2.00-3.00)	2.51	–0.879	.38
(7) Cover your nose and mouth when coughing or sneezing	2.00 (2.00-3.00)	2.44	3.00 (2.00-3.00)	2.70	–1.794	.07
(8) Choose and wear a mask rightly	2.00 (2.00-3.00)	2.54	3.00 (2.00-4.00)	2.87	–1.723	.09
(9) Wash your hands with soap and water for at least 20 seconds regularly	2.00 (2.00-3.00)	2.61	3.00 (2.00-4.00)	2.87	–1.379	.17
(10) Adopt disinfection measures (eg, 70% alcoholic solutions or 1% sodium hypochlorite solution) in daily life	2.00 (2.00-3.00)	2.47	2.00 (2.00-3.00)	2.42	–0.799	.42
(11) Focus on your symptoms and adverse drug reactions	2.00 (2.00-3.00)	2.39	2.00 (2.00-3.00)	3.45	–0.08	.94
(12) Do not change or stop medication easily	2.00 (2.00-3.00)	2.46	2.00 (2.00-3.00)	2.55	–0.422	.67
(13) Reexamine regularly	2.00 (2.00-3.00)	2.51	2.00 (2.00-3.00)	2.55	–0.027	.98

Table 3. Scores of self-management health behaviors at the end point of the study in the intervention and control groups.

Self-management behaviors	Intervention group		Control group		*Z*	*P* value
	Median (P25-P75)	Mean	Median (P25-P75)	Mean		
(1) Self-isolate from families in home-based treatment	4.00 (2.00-4.00)	3.03	1.00 (1.00-2.00)	1.60	–5.308	<.001
(2) Keep your room well-ventilated	4.00 (4.00-4.00)	3.44	2.00 (1.00-4.00)	2.53	–3.843	<.001
(3) Maintain a balanced diet and ensure enough intake of high protein, vitamins, and minerals	4.00 (3.00-4.00)	3.47	4.00 (2.00-4.00)	3.04	–2.055	.04
(4) Stick to daily exercise	4.00 (2.00-4.00)	3.14	2.00 (1.00-4.00)	2.30	–3.096	.002
(5) Go to bed and wake up early, and alternate work with rest	4.00 (4.00-4.00)	3.53	2.00 (1.50-4.00)	2.51	–4.601	<.001
(6) Stay in a good mood	4.00 (3.00-4.00)	3.51	3.00 (2.00-4.00)	3.02	–3.246	.001
(7) Cover your nose and mouth when coughing or sneezing	4.00 (4.00-4.00)	3.75	4.00 (4.00-4.00)	3.60	–1.213	.23
(8) Choose and wear a mask rightly	4.00 (4.00-4.00)	3.80	4.00 (1.50-4.00)	3.17	–2.809	.005
(9) Wash your hands with soap and water for at least 20 seconds regularly	4.00 (4.00-4.00)	3.78	4.00 (4.00-4.00)	3.70	–0.522	.60
(10) Adopt disinfection measures (eg, 70% alcoholic solutions or 1% sodium hypochlorite solution) in daily life	4.00 (4.00-4.00)	3.44	1.00 (1.00-3.50)	2.02	–5.595	<.001
(11) Focus on your symptoms and adverse drug reactions	4.00 (4.00-4.00)	3.75	2.00 (1.00-4.00)	2.21	–6.416	<.001
(12) Do not change or stop medication easily	4.00 (4.00-4.00)	3.73	3.00 (2.00-4.00)	2.83	–4.566	<.001
(13) Reexamine regularly	4.00 (4.00-4.00)	3.61	2.00 (1.00-4.00)	2.23	–5.783	<.001

Table 4. Intragroup comparison results for the self-management health behaviors (end point–baseline).

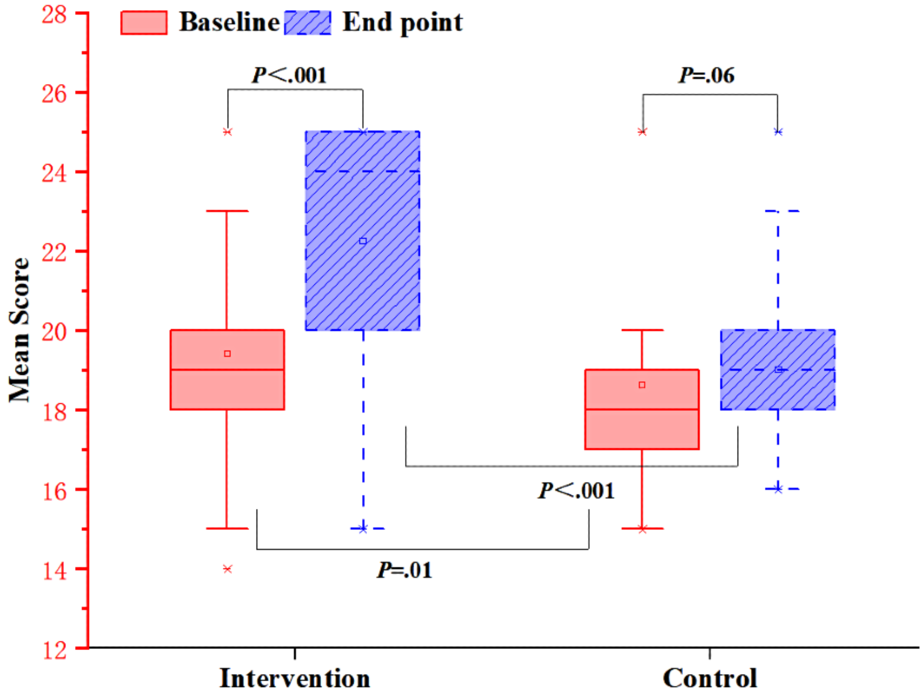
Self-management behaviors	Intervention group			Control group		
	Mean difference	*Z*	*P* value	Mean difference	*Z*	*P* value
(1) Self-isolate from families in home-based treatment	1.15	–4.828	<.001	–0.30	–2.544	.01
(2) Keep your room well-ventilated	1.19	–5.376	<.001	0.04	–0.576	.57
(3) Maintain a balanced diet and ensure enough intake of high protein, vitamins, and minerals	1.12	–5.038	<.001	0.57	–3.166	.002
(4) Stick to daily exercise	0.90	–4.568	<.001	–0.17	–0.582	.56
(5) Go to bed and wake up early, and alternate work with rest	1.29	–5.639	<.001	0.15	–1.251	.21
(6) Stay in a good mood	1.19	–5.602	<.001	0.51	–3.050	.002
(7) Cover your nose and mouth when coughing or sneezing	1.31	–5.442	<.001	0.91	–4.647	<.001
(8) Choose and wear a mask rightly	1.25	–5.323	<.001	0.3	–1.413	.16
(9) Wash your hands with soap and water for at least 20 seconds regularly	1.17	–5.324	<.001	0.83	–4.358	<.001
(10) Adopt disinfection measures (eg, 70% alcoholic solutions or 1% sodium hypochlorite solution) in daily life	0.97	–4.863	<.001	–0.40	–1.889	.06
(11) Focus on your symptoms and adverse drug reactions	1.36	–5.917	<.001	–0.25	–1.157	.25
(12) Do not change or stop medication easily	1.27	–5.648	<.001	0.28	–1.949	.051
(13) Reexamine regularly	1.10	–5.328	<.001	–0.32	–1.764	.08

Secondary Outcomes

TB Knowledge Awareness

The score of TB knowledge awareness in the intervention group increased significantly after the intervention (*P*<.001), but there was no significant differences in the control group between that at baseline and that at the end point (*P*=.06) (Figure 4).

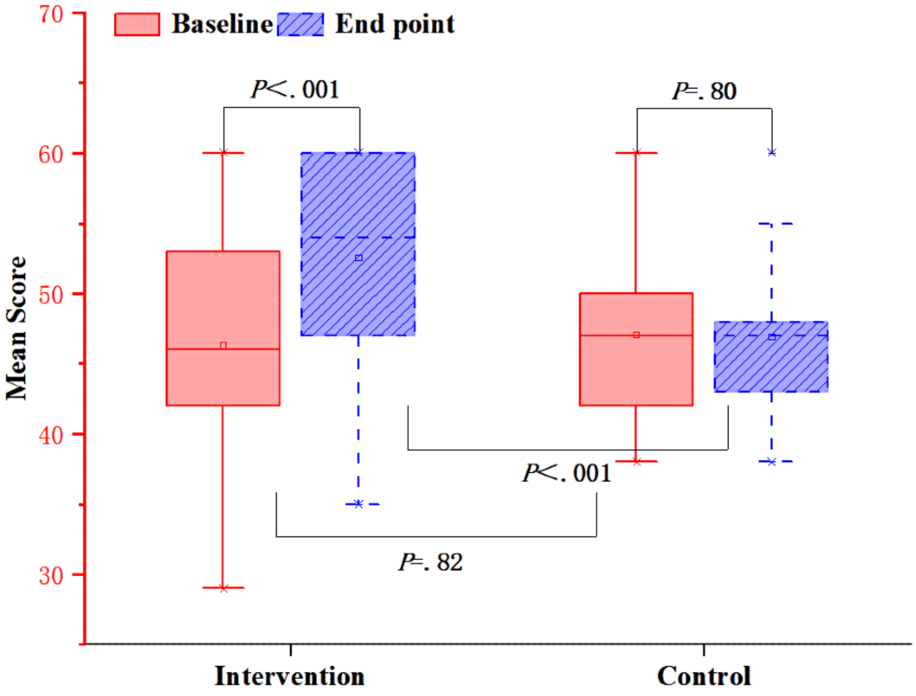
Figure 4. Mean scores of tuberculosis knowledge awareness for both groups before and after the intervention.



Self-efficacy

After the intervention, the score of self-efficacy was significantly improved in the intervention group (*P*<.001), but there was no significant change in the control group between that at baseline and that at the end point (*P*=.80). The score of the intervention group was significantly higher than that of the control group (*P*<.001) (Figure 5).

Figure 5. Mean scores of self-efficacy for both groups before and after the intervention.



Social Support

The social support scale includes 3 dimensions: family support, friend support, and other support (from significant others). Although the score of friend support in the control group was significantly higher than that in the intervention group at baseline (*P*=.008), after the intervention, the scores of the 3 dimensions were significantly improved in the intervention group (*P*<.001) and were significantly higher than those in the control group (*P*<.001). There were no significant changes in the scores of the control group before and after the intervention (Table 5 and Table 6).

Table 5. Median scores of social support scale for both groups at baseline and end point.

Social support scale	Intervention group (n=59), median (P25-P75)	Control group (n=53), median (P25-P75)	Mean difference	*Z*	*P* value
Family support					
Baseline	24 (22-24)	24 (23-24)	–0.4	–0.983	.33
End point	24 (23-27)	24 (23-24)	2.566	–5.111	<.001
Friend support					
Baseline	21 (18-24)	24 (20-24)	–1.798	–2.664	.008
End point	24 (21-25)	24 (20-24)	0.432	–0.817	.41
Other support					
Baseline	23 (21-24)	24 (23-24)	–0.11	–0.05	.96
End point	24 (22-28)	24 (23-24)	1.444	–3.569	<.001

Table 6. Intragroup comparison results for social support scale (end point–baseline).

Group	Family support			Friend support			Other support		
	Mean difference	*Z*	*P* value	Mean difference	*Z*	*P* value	Mean difference	*Z*	*P* value
Intervention (n=59)	2.136	–2.407	.02	2.475	–3.833	<.001	2.61	–3.589	<.001
Control (n=53)	–0.83	–0.255	.80	0.245	–0.809	.42	1.057	–0.115	.91

Degree of Satisfaction With Health Education

The scale for degree of satisfaction with health education evaluates performance expectancy, effort expectancy, and facilitating conditions. After the intervention, the scores of the 3 dimensions were significantly improved in the intervention group (*P*<.001) and were significantly higher than those in the control group (*P*<.001). There were no significant changes in the scores of the control group before and after the intervention (Table 7 and Table 8).

Table 7. Median scores of the degree of satisfaction with health education for both groups at baseline and end point.

Satisfaction scale	Intervention group (n=59), median (P25-P75)	Control group (n=53), median (P25-P75)	Mean difference	*Z*	*P* value
Performance expectancy					
Baseline	19 (16-22)	20 (19-22)	–0.492	–1.215	.22
End point	24 (22-28)	19 (17-24)	4.35	–5.22	<.001
Effort expectancy					
Baseline	19 (16-23)	20 (19-24)	–0.781	–1.54	.12
End point	24 (23-28)	20 (19-24)	3.848	–5.22	<.001
Facilitating conditions					
Baseline	8 (8-12)	10 (8-12)	–0.516	–1.17	.24
End point	12 (12-14)	10 (7-12)	2.781	–5.607	<.001

Table 8. Intragroup comparison results for the degree of satisfaction with health education (end point–baseline).

Group	Performance expectancy			Effort expectancy			Facilitating conditions		
	Mean difference	*Z*	*P* value	Mean difference	*Z*	*P* value	Mean difference	*Z*	*P* value
Intervention (n=59)	4.559	–5.403	<.001	4.441	–5.406	<.001	2.881	–5.591	<.001
Control (n=53)	–0.283	–0.427	.67	–0.189	–0.182	.86	–0.415	–0.988	.32

Discussion

Principal Findings

Overview

Self-management is crucial for the control of chronic diseases. ITHBC was formed on the basis of integration and absorption of many related health behavior change theory models [14], which were used for guiding the establishment and maintenance of health behaviors, thereby changing poor individual lifestyles to better health status and quality of life. In China, mobile services, especially short message service is ubiquitous, as it can deliver frequent prompts and health information to improve treatment adherence [26-29]. With the emergence of the COVID-19 pandemic and the adoption of public health measures for its containment, forms of health education were affected around the world. mHealth interventions did not have any restrictions during the COVID-19 pandemic and met the needs of the patients to the greatest extent. Moreover, the internet brings strong interactions between health care providers and patients at lower cost. We developed an mHealth intervention program based on the ITHBC theory and evaluated its effects on improving patients' self-management ability with newly diagnosed pulmonary TB. Such an intervention empowers patients to change poor health behaviors based on disease knowledge, self-efficacy, and social support [30,31].

Primary Outcomes

After calculation of the total score of 13 self-management behaviors, we found that the scores of both the groups increased, but the scores of the intervention group increased more than those of the control group. However, the total score could not easily explain the effect of our intervention; therefore, further analysis for each self-management behavior needs to be conducted. After the intervention, a statistically significant increase in the intervention group was noted in the scores of each item of self-management behavior compared with the scores at the baseline (*P*<.001). In the control group, the score of the item "self-isolate from families in home-based treatment" was lower than that at the baseline (*P*=.01), whereas scores of the item "maintain a balanced diet and ensure enough intake of high protein, vitamins, and minerals" (*P*=.002), item "stay in a good mood" (*P*=.002), item "cover your nose and mouth when coughing or sneezing" (*P*<.001), and item "wash hands properly" (*P*<.001) increased compared to the scores at the baseline. Source control is the most efficient measure to prevent infectious diseases. For item "self-isolate from families in home-based treatment," the score was lower than that of any other behaviors in the intervention and control group. Moreover, it was found that the item "no intention to adopt such a behavior (namely, self-isolation)" had the most response, which may be attributed to the fact that not each patient with pulmonary TB is infectious and that the sputum culture result is needed to determine whether to carry out self-isolation or not. However, a comparison with the scores at the baseline shows that the scores of the intervention group increased (*P*<.001) and the scores of the control group decreased (*P*=.01). Extended duration of therapy was considered to be the main reason that patients did not want to self-isolate. However, efficient health education facilitates the choice to self-isolate.

Nutritional therapy is the basis of TB treatment. Each patient in our hospital was instructed to maintain a balanced diet, which increased the scores of the related behavior in both groups, and the scores of the intervention group were higher than those of the control group (*P*=.04). A positive mentality contributes to favorable treatment, and improvement of health brings pleasure to patients. In the 2 groups, the score of the item "stay in a good mood" increased with a higher increase in the intervention group, indicating a good effect of the intervention on emotion. The scores of the item "cover your nose and mouth when coughing or sneezing" and item "wash hands properly" showed a similar increase in both the groups. Pulmonary TB and COVID-19 have a similar route of transmission and prevention [32]. It is probable that the overall increase of the above two self-management behaviors can be attributed to the increased awareness of infectious disease prevention during the COVID-19 epidemic. This is consistent with the findings of Wang et al [33], that is, more people washed their hands with soap after touching contaminated objects and covered their mouths when coughing or sneezing as precautionary strategies during the initial stage of the COVID-19 epidemic in China. Further, the adherence of the patients in the control group gradually decreased over time, thereby presenting as a discontinuation of the drug and missed doses, which is consistent with the findings of Zomahoun et al [34], whereas the adherence of the patients in the intervention group increased significantly with daily medication reminders on the WeChat group. In addition, the intervention group had a higher revisit rate than the control group (*P*=.22) through review of medical records and consultation in WeChat, which indicated a positive effect of ITHBC-based mHealth intervention on patients' adherence.

Secondary Outcomes

Compared with those at baseline, TB knowledge awareness, self-efficacy, social support, and degree of satisfaction with health education of the intervention group all increased significantly (*P*<.001), and the scores of the intervention group were higher than those of the control group (*P*<.001), thereby indicating that mHealth interventions for TB self-management based on ITHBC can improve behavior belief, self-efficacy, and social support of patients with TB. However, this finding is inconsistent with the results of most studies, which showed that single health education may result in increased patient resistance to change or engage in appropriate disease management strategies [35]. Knowledge is the foundation to change behavior. Researchers have confirmed the role of health education in promoting behavior beliefs, for example, a health education program significantly improved the health beliefs of participants with a history of opisthorchiasis [36]. Effective educational interventions are those that aim to modify patients' behaviors rather than simply providing information. It is worth noting that before the intervention, the average score of the item "not spitting, covering your nose and mouth when coughing or sneezing, and wearing a mask can reduce the spread of TB" was 4.3, which was higher than the average score of 3.5-3.7 in other items. This showed that patients had a good understanding of how to reduce the spread of TB with or without health education.

Self-efficacy reflects the confidence of patients for disease control, and good self-efficacy helps change health behaviors. Health education might have helped patients to perceive risks and expect outcomes, thereby promoting the establishment of good beliefs. This study set different self-management goals for each patient according to different contexts through review at the baseline. This individualized education provided encouragement and helped to build confidence in defeating the disease. Moreover, volitional self-efficacy may be related to social support and follow-up. Some studies involved in other chronic diseases proved that partner support and home visits were useful ways to support patients [37,38]. ITHBC-based mHealth intervention pays more attention to the effect of social support on self-management behaviors. According to previous researches, patients often failed to continue treatment for the lack of family and community support [39,40]. In our study, the WeChat groups were used as tools to build partnerships among patients and between patients and medical staff, thereby providing positive social support and ultimately improving the self-efficacy of patients with TB. The intervention content was developed and optimized by multidisciplinary medical experts, and we gave it a pretest based on the needs of patients. Satisfaction with health education was explored from content, structure, and form, and it was shown that the score of each item in the intervention group was significantly higher than the score of the control group after the 3-month intervention. This illustrated that ITHBC-based mHealth allows researchers to provide a scientific, efficient, convenient, and easy way to understand health education and give timely responses for health confusion.

In the 3-month intervention, 2 patients in the control group dropped out, and 112 patients fulfilled the criterion of the smallest sample size in the study. Schulz and Grimes [41] suggest that losses to follow-up less than 5% usually have little impact, whereas losses greater than 20% raise serious flags about study validity. White and Thompson [42] advocated the use of mean imputation and missing-indicator method as a solution for practical purposes. The data in this study were missing at random, and we finally used mean substitution to fill and calculate the outcome of the control group. For results that had no significant differences from the conclusion, it could be concluded that lost visits exerted little effect on the conclusion and the conclusion was reliable.

It is not uncommon that some participants do not receive the intervention allocated by the randomization process. This study could not show whether the participants had actually adopted the recommended behaviors. Researchers have confirmed the gold standard of reporting is "intention-to-treat" analysis and according to the intervention that they actually received (per-protocol analysis) can rarely lead to differing results [25]. Therefore, in this study, we collected the outcomes of all the participants randomly assigned to the intervention group, even if some of the participants may not have adopted the recommended behaviors.

Strengths and Limitations

This study had 2 strengths. First, we applied the ITHBC theory to study the self-management of patients with TB for the first time. Second, mHealth interventions did not face any restrictions during the COVID-19 pandemic and met the needs of the patients to the greatest extent. Moreover, healthy people are exposed to less risks owing to the application of mHealth [43,44].

This study had the following limitations. First, the intervention was only conducted in 1 hospital, which may restrict the general applicability of our results, and future studies should be conducted with larger sample sizes to confirm our results. Second, scores of TB knowledge awareness and friend support in both groups had significant differences at the baseline (P=.01), which may be due to the small sample size. Third, the short duration of observation (3 months) cannot prove the long-term efficacy. We plan to conduct studies with a long-term intervention (6 or 12 months) to focus on assessing the patient's clinical efficacy and quality of life. Fourth, the nonrandom sampling method could result in some selection bias.

Conclusion

In conclusion, ITHBC-based mHealth intervention may be a new promising therapeutic strategy for the management of TB for improving patients' subjective initiative and self-management behaviors, which are beneficial for promoting compliance and the quality of prevention and control for pulmonary TB. In addition, mHealth provides an effective solution for outpatients without nursing care.

Acknowledgments

The authors thank the multidisciplinary panel comprised of pharmacists, nursing experts, physicians, and public health experts in tuberculosis care in Harbin Chest Hospital, in particular, Haiou Lin, Xiuping Li, Changlong Fan, Lianzhi Wang, Jing Zhang, Yingying Li, and Hui Jiang. They advised and cooperated during the development and implementation of the intervention. Furthermore, we thank Danyang Li and Lu Zhang for advising during the intervention development and training.

YB was affiliated with the Pharmacy Department of Harbin Chest Hospital at the time of the trial and is currently affiliated with the Health Clinic at Changzhou Institute of Technology.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1312 KB - publichealth_v8i7e34277_app1.pdf]

Multimedia Appendix 2
Informed consent.
[PDF File (Adobe PDF File), 94 KB - publichealth_v8i7e34277_app2.pdf]

Multimedia Appendix 3
Screenshots of the health education in WeChat groups.
[PDF File (Adobe PDF File), 325 KB - publichealth_v8i7e34277_app3.pdf]

Multimedia Appendix 4
Topics of health education in the Knowledge and Belief section.
[PDF File (Adobe PDF File), 67 KB - publichealth_v8i7e34277_app4.pdf]

Multimedia Appendix 5
Questionnaire content.
[PDF File (Adobe PDF File), 167 KB - publichealth_v8i7e34277_app5.pdf]

References

1. WHO team. Global Tuberculosis Report. Geneva: World Health Organization; 2021.
2. Packer TL, Fracini A, Audulv, Alizadeh N, van Gaal BG, Warner G, et al. What we know about the purpose, theoretical foundation, scope and dimensionality of existing self-management measurement tools: A scoping review. Patient Educ Couns 2018 Apr;101(4):579-595. [doi: 10.1016/j.pec.2017.10.014] [Medline: 29239734]
3. Chen Y, Tan D, Xu Y, Wang B, Li X, Cai X, et al. Effects of a HAPA-based multicomponent intervention to improve self-management precursors of older adults with tuberculosis: A community-based randomised controlled trial. Patient Educ Couns 2020 Feb;103(2):328-335. [doi: 10.1016/j.pec.2019.09.007] [Medline: 31543353]

4. Harandi TF, Mahmoodi Z, Ghavidel N, Sharifipour Z. Factors affecting self-management in Iranian tuberculosis patients: A path analysis model. Can J Respir Ther 2021;57:73-78 [FREE Full text] [doi: 10.29390/cjrt-2021-009] [Medline: 34212074]
5. Alipanah N, Jarlsberg L, Miller C, Linh NN, Falzon D, Jaramillo E, et al. Adherence interventions and outcomes of tuberculosis treatment: A systematic review and meta-analysis of trials and observational studies. PLoS Med 2018 Jul;15(7):e1002595 [FREE Full text] [doi: 10.1371/journal.pmed.1002595] [Medline: 29969463]
6. World Health Organization. An expanded DOTS framework for effective tuberculosis control. Int J Tuberc Lung Dis 2002 May;6(5):378-388. [Medline: 12019913]
7. Volmink J, Garner P. Directly observed therapy for treating tuberculosis. Cochrane Database Syst Rev 2007 Oct 17;17(4):1-5. [doi: 10.1002/14651858.CD003343.pub3] [Medline: 17943789]
8. Guo P, Qiao W, Sun Y, Liu F, Wang C. Telemedicine Technologies and Tuberculosis Management: A Randomized Controlled Trial. Telemed J E Health 2020 Sep;26(9):1150-1156. [doi: 10.1089/tmj.2019.0190] [Medline: 31794684]
9. Latif AI, Sjattar EL, Erika KA. Models and benefits of mobile health application to support patient with tuberculosis: A literature review. Enfermería Clínica 2020 Mar;30:163-167. [doi: 10.1016/j.enfcli.2019.07.069]
10. Fekadu G, Jiang X, Yao J, You JH. Cost-effectiveness of video-observed therapy for ambulatory management of active tuberculosis during the COVID-19 pandemic in a high-income country. Int J Infect Dis 2021 Dec;113:271-278 [FREE Full text] [doi: 10.1016/j.ijid.2021.10.029] [Medline: 34688946]
11. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010 Oct 06;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
12. Keutzer L, Wicha SG, Simonsson US. Mobile Health Apps for Improvement of Tuberculosis Treatment: Descriptive Review. JMIR Mhealth Uhealth 2020 Apr 21;8(4):e17246 [FREE Full text] [doi: 10.2196/17246] [Medline: 32314977]
13. Tola HH, Shojaeizadeh D, Tol A, Garmaroudi G, Yekaninejad MS, Kebede A, et al. Psychological and Educational Intervention to Improve Tuberculosis Treatment Adherence in Ethiopia Based on Health Belief Model: A Cluster Randomized Control Trial. PLoS One 2016;11(5):e0155147 [FREE Full text] [doi: 10.1371/journal.pone.0155147] [Medline: 27167378]
14. Ryan P. Integrated Theory of Health Behavior Change: background and intervention development. Clin Nurse Spec 2009;23(3):161-172. [doi: 10.1097/nur.0b013e3181a42373] [Medline: 19395894]
15. Ryan P, Maierle D, Csuka ME, Thomson A, Szabo A. Computer-based intervention to enhance self-management of calcium and vitamin D intake in women. West J Nurs Res 2013 Sep;35(8):986-1010 [FREE Full text] [doi: 10.1177/0193945913483369] [Medline: 23539320]
16. Montag C, Becker B, Gan C. The Multipurpose Application WeChat: A Review on Recent Research. Front Psychol 2018;9:2247 [FREE Full text] [doi: 10.3389/fpsyg.2018.02247] [Medline: 30618894]
17. Li X, Li T, Chen J, Xie Y, An X, Lv Y, et al. A WeChat-Based Self-Management Intervention for Community Middle-Aged and Elderly Adults with Hypertension in Guangzhou, China: A Cluster-Randomized Controlled Trial. IJERPH 2019 Oct 23;16(21):4058. [doi: 10.3390/ijerph16214058] [Medline: 31652688]
18. Guo Y, Xu Z, Qiao J, Hong YA, Zhang H, Zeng C, et al. Development and Feasibility Testing of an mHealth (Text Message and WeChat) Intervention to Improve the Medication Adherence and Quality of Life of People Living with HIV in China: Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e10274. [doi: 10.2196/10274]
19. Gao X. Investigation on medication adherence among tuberculosis patients in consolidation period and intervention effect (in Chinese). University of South China 2013(2):17 [FREE Full text] [doi: 10.12677/qrb.2014.11004]
20. People's Republic of China state health and Family Planning Commission. Tuberculosis classification (WS196-2017). Electronic Journal of Emerging Infectious Diseases 2018;3(03):191-192. [doi: 10.19871/j.cnki.xfcrbzz.2018.03.018]
21. Wen Juan Xing. URL: https://www.wjx.cn/ [accessed 2022-07-04]
22. Self-efficacy for managing chronic disease 6-item scale. Stanford Patient Education Research Center. 2013. URL: https://www.unmc.edu/centric/_documents/SEMCD6ISinfo.pdf [accessed 2022-07-04]
23. Blumenthal JA, Burg MM, Barefoot J, Williams RB, Haney T, Zimet G. Social support, type A behavior, and coronary artery disease. Psychosom Med 1987;49(4):331-340. [doi: 10.1097/00006842-198707000-00002] [Medline: 3615762]
24. Venkatesh, Morris, Davis, Davis. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
25. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand 2018 Apr;97(4):380-387 [FREE Full text] [doi: 10.1111/aogs.13309] [Medline: 29377058]
26. Shetty AS, Chamukuttan S, Nanditha A, Raj RKC, Ramachandran A. Reinforcement of adherence to prescription recommendations in Asian Indian diabetes patients using short message service (SMS)--a pilot study. J Assoc Physicians India 2011 Nov;59:711-714. [Medline: 22616337]
27. Vervloet M, van Dijk L, Santen-Reestman J, van Vlijmen B, van Wingerden P, Bouvy M, et al. SMS reminders improve adherence to oral medication in type 2 diabetes patients who are real time electronically monitored. Int J Med Inform 2012 Sep;81(9):594-604. [doi: 10.1016/j.ijmedinf.2012.05.005] [Medline: 22652012]

28. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
29. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
30. Lorig KR, Holman HR. Self-management education: History, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/s15324796abm2601_01] [Medline: 12867348]
31. Baskerville R, Ricci-Cabello I, Roberts N, Farmer A. Impact of accelerometer and pedometer use on physical activity and glycaemic control in people with Type 2 diabetes: a systematic review and meta-analysis. Diabet Med 2017 May;34(5):612-620. [doi: 10.1111/dme.13331] [Medline: 28173623]
32. Chopra K, Arora V, Singh S. COVID 19 and tuberculosis. Indian J Tuberc 2020 Apr;67(2):149-151 [FREE Full text] [doi: 10.1016/j.ijtb.2020.06.001] [Medline: 32553303]
33. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. IJERPH 2020 Mar 06;17(5):1729. [doi: 10.3390/ijerph17051729] [Medline: 32155789]
34. Zomahoun HTV, Guénette L, Grégoire JP, Lauzier S, Lawani AM, Ferdynus C, et al. Effectiveness of motivational interviewing interventions on medication adherence in adults with chronic diseases: a systematic review and meta-analysis. Int J Epidemiol 2017 Apr 01;46(2):589-602. [doi: 10.1093/ije/dyw273] [Medline: 27864410]
35. Stott NC, Pill RM. 'Advise yes, dictate no'. Patients' views on health promotion in the consultation. Fam Pract 1990 Jun;7(2):125-131. [doi: 10.1093/fampra/7.2.125] [Medline: 2369980]
36. Kaewpitoon SJ, Thanapatto S, Nuathong W, Rujirakul R, Wakkuwattapong P, Norkaew J, et al. Effectiveness of a Health Educational Program Based on Self-Efficacy and Social Support for Preventing Liver Fluke Infection in Rural People of Surin Province, Thailand. Asian Pac J Cancer Prev 2016;17(3):1111-1114 [FREE Full text] [doi: 10.7314/apjcp.2016.17.3.1111] [Medline: 27039733]
37. Fisher EB, Boothroyd RI, Elstad EA, Hays L, Henes A, Maslow GR, et al. Peer support of complex health behaviors in prevention and disease management with special reference to diabetes: systematic reviews. Clin Diabetes Endocrinol 2017;3:4 [FREE Full text] [doi: 10.1186/s40842-017-0042-3] [Medline: 28702258]
38. Marshall J, Birriel PC, Baker E, Olson L, Agu N, Estefan LF. Widening the scope of social support: the Florida maternal, infant, and early childhood home visiting program. Infant Ment Health J 2018 Sep;39(5):595-607. [doi: 10.1002/imhj.21737] [Medline: 30074249]
39. Gebremariam MK, Bjune GA, Frich JC. Barriers and facilitators of adherence to TB treatment in patients on concomitant TB and HIV treatment: a qualitative study. BMC Public Health 2010 Oct 28;10:651 [FREE Full text] [doi: 10.1186/1471-2458-10-651] [Medline: 21029405]
40. Sagbakken M, Frich JC, Bjune G. Barriers and enablers in the management of tuberculosis treatment in Addis Ababa, Ethiopia: a qualitative study. BMC Public Health 2008 Jan 11;8:11 [FREE Full text] [doi: 10.1186/1471-2458-8-11] [Medline: 18186946]
41. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. The Lancet 2002 Mar;359(9308):781-785. [doi: 10.1016/s0140-6736(02)07882-0]
42. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. Stat Med 2005 Apr 15;24(7):993-1007. [doi: 10.1002/sim.1981] [Medline: 15570623]
43. Horna-Campos OJ, Sánchez-Pérez HJ, Sánchez I, Bedoya A, Martín M. Public transportation and pulmonary tuberculosis, Lima, Peru. Emerg Infect Dis 2007 Oct;13(10):1491-1493 [FREE Full text] [doi: 10.3201/eid1310.060793] [Medline: 18257992]
44. Zamudio C, Krapp F, Choi HW, Shah L, Ciampi A, Gotuzzo E, et al. Public transportation and tuberculosis transmission in a high incidence setting. PLoS One 2015;10(2):e0115230 [FREE Full text] [doi: 10.1371/journal.pone.0115230] [Medline: 25706530]

Abbreviations

DOT: directly observed therapy
ITHBC: Integrated Theory of Health Behavior Change
mHealth: mobile health
TB: tuberculosis

Edited by T Sanchez; submitted 14.10.21; peer-reviewed by E Baker, J Qiao, H Mehdizadeh, S Mukherjee; comments to author 06.01.22; revised version received 26.02.22; accepted 26.04.22; published 14.07.22.

Please cite as:
Bao Y, Wang C, Xu H, Lai Y, Yan Y, Ma Y, Yu T, Wu Y
Effects of an mHealth Intervention for Pulmonary Tuberculosis Self-management Based on the Integrated Theory of Health Behavior Change: Randomized Controlled Trial
JMIR Public Health Surveill 2022;8(7):e34277
URL: https://publichealth.jmir.org/2022/7/e34277
doi:10.2196/34277
PMID:35834302

Original Paper

Papers Please - Predictive Factors of National and International Attitudes Toward Immunity and Vaccination Passports: Online Representative Surveys

Paul M Garrett[1], BS, PhD; Joshua P White[1], BSc, LLM; Simon Dennis[1,2], PhD; Stephan Lewandowsky[3,4], PhD; Cheng-Ta Yang[5,6], PhD; Yasmina Okan[7], PhD; Andrew Perfors[1], PhD; Daniel R Little[1], PhD; Anastasia Kozyreva[8], PhD; Philipp Lorenz-Spreen[8], PhD; Takashi Kusumi[9], PhD; Yoshihisa Kashima[1], PhD

[1]Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, Australia
[2]Unforgettable Research Services, Melbourne, Australia
[3]School of Psychological Science, The University of Bristol, Bristol, United Kingdom
[4]School of Psychological Science, The University of Western Australia, Perth, Australia
[5]Department of Psychology, National Cheng Kung University, Tainan City, Taiwan
[6]Graduate Institute of Mind, Brain and Consciousness, Taipei Medical University, Taipei, Taiwan
[7]Centre for Decision Research, Leeds University Business School, University of Leeds, Leeds, United Kingdom
[8]Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany
[9]Graduate School of Education, Kyoto University, Kyoto, Japan

Corresponding Author:
Paul M Garrett, BS, PhD
Melbourne School of Psychological Sciences
The University of Melbourne
12th Floor Redmond Barry Building, Parkville Campus
Melbourne, 3010
Australia
Phone: 61 8344 6377 ext 03
Email: paul.garrett@unimelb.edu.au

Abstract

Background: In response to the COVID-19 pandemic, countries are introducing digital passports that allow citizens to return to normal activities if they were previously infected with (immunity passport) or vaccinated against (vaccination passport) SARS-CoV-2. To be effective, policy decision-makers must know whether these passports will be widely accepted by the public and under what conditions. This study focuses attention on immunity passports, as these may prove useful in countries both with and without an existing COVID-19 vaccination program; however, our general findings also extend to vaccination passports.

Objective: We aimed to assess attitudes toward the introduction of immunity passports in six countries, and determine what social, personal, and contextual factors predicted their support.

Methods: We collected 13,678 participants through online representative sampling across six countries—Australia, Japan, Taiwan, Germany, Spain, and the United Kingdom—during April to May of the 2020 COVID-19 pandemic, and assessed attitudes and support for the introduction of immunity passports.

Results: Immunity passport support was moderate to low, being the highest in Germany (775/1507 participants, 51.43%) and the United Kingdom (759/1484, 51.15%); followed by Taiwan (2841/5989, 47.44%), Australia (963/2086, 46.16%), and Spain (693/1491, 46.48%); and was the lowest in Japan (241/1081, 22.94%). Bayesian generalized linear mixed effects modeling was used to assess predictive factors for immunity passport support across countries. International results showed neoliberal worldviews (odds ratio [OR] 1.17, 95% CI 1.13-1.22), personal concern (OR 1.07, 95% CI 1.00-1.16), perceived virus severity (OR 1.07, 95% CI 1.01-1.14), the fairness of immunity passports (OR 2.51, 95% CI 2.36-2.66), liking immunity passports (OR 2.77, 95% CI 2.61-2.94), and a willingness to become infected to gain an immunity passport (OR 1.6, 95% CI 1.51-1.68) were all predictive factors of immunity passport support. By contrast, gender (woman; OR 0.9, 95% CI 0.82-0.98), immunity passport concern (OR 0.61, 95% CI 0.57-0.65), and risk of harm to society (OR 0.71, 95% CI 0.67-0.76) predicted a decrease in support for immunity passports. Minor differences in predictive factors were found between countries and results were modeled separately to provide national accounts of these data.

Conclusions: Our research suggests that support for immunity passports is predicted by the personal benefits and societal risks they confer. These findings generalized across six countries and may also prove informative for the introduction of vaccination passports, helping policymakers to introduce effective COVID-19 passport policies in these six countries and around the world.

(JMIR Public Health Surveill 2022;8(7):e32969) doi:10.2196/32969

KEYWORDS

COVID-19; immunity passport; vaccination passport; cross-cultural; health policy; digital certificates; SARS-CoV-2; vaccine; policy; international

Introduction

The SARS-CoV-2 virus responsible for COVID-19 has infected more than 360 million individuals worldwide and resulted in more than 5.6 million deaths [1]. As the virus continues to spread, countries seek ways to restart their economies and allow citizens to move freely without reigniting the pandemic. Vaccines are the foremost tool in combating the virus, and countries are introducing "vaccination passports" to allow low-risk individuals to travel, work, and gather under lowered restrictions [2,3]. However, there remains a stark gap between international vaccination programs, with many, predominantly poorer, countries lacking vaccines and still waiting to administer their first dose [4]. Additionally, it is unclear how effective current vaccines will be against newly emerging virus variants [5,6]. In countries where vaccines are limited, or where virus variants outpace vaccine effectiveness, immunity passports may be used.

Immunity passports identify previously infected and now recovered individuals by testing for SARS-CoV-2 antibodies [7]. Like vaccinated individuals, recovered individuals are thought to have a lower likelihood of contracting, spreading, and experiencing the most severe symptoms of the virus [7]. A recent World Health Organization report [8] suggests that recovered individuals develop antibodies within 4 weeks following infection, that immune responses remain robust for 6-8 months, and that, due to the manner by which vaccines target a specific spike protein, naturally acquired antibodies may be more robust to emerging virus variants (vaccines are effective against current variants of concern [eg, Delta and Omicron]). As such, immunity passports may prove useful in the fight against COVID-19, especially when used in conjunction with vaccination passports. Indeed, the European Union [9] has proposed exactly this with their new "green card," a digital certificate that will act as both a vaccination and immunity passport. For simplicity, we refer to these vaccination and immunity passports collectively as "immunization passports."

Immunization passports may allow economies to rapidly bounce back, with individuals perceiving crowded shops and workplaces as safer if others are recovered or vaccinated [10]. Similarly, businesses may require proof of immunization to enter their premises or use their services [11], and countries may require proof of immunization to cross their borders [2]. For example, the International Air Transport Association has developed the "Travel Pass" app [12] to store a COVID-19 vaccination record on the user's phone, such that data can be shared with governments and transport authorities before accessing flights and crossing a country's border.

Additional privacy measures may accompany these apps, as is the case with South Korea's "Green Pass" [3], a vaccination certificate that uses blockchain technology to make passes both shareable and tamper-proof [13]. Australia uses an alternative method, issuing international passes as QR codes protected with visible digital seals (nonconstrained) and administered only after one's data have been verified by the federal Australian Passport Office [14]. These immunity passport apps and QR codes are a technological extension of existing vaccination requirements such as the physical "yellow card" that accompanies yellow fever vaccination, which is necessary to enter many countries in Africa and Central and South America [15].

The potential introduction of immunization passports carries a host of scientific, legal, and ethical questions such as: Are recovered and vaccinated individuals immune to new virus variants [8]? Will these passports become a legal requirement, and how will people who cannot risk becoming infected or cannot get vaccinated be impacted? Will individuals try to become infected if doing so confers additional freedoms? [16] Each of these questions is critical to national health policies and has been a source of recent debates between privacy advocates and politicians in Britain [17], and the cause of public protests in France [18]. World governments and health policy decision-makers need scientifically informed answers to two key questions: Will people around the world accept and support the use of immunization passports? And if so, why?

We narrow the scope of our investigation to the introduction and acceptance of immunity passports—instances where an individual has been infected and recovered—in six countries around the world, as immunity passports may yet prove relevant to countries both with and without vaccination programs. Of course, these findings may also prove insightful and may extend to the conditions necessary for vaccination passport acceptance. Key to the current investigation is understanding what societal, personal, and contextual factors influence immunization passport acceptance.

Societal factors may shape one's attitude toward whether immunization passports will benefit the community at large, thereby influencing passport acceptance [19]. Health policy acceptance may improve with a sense of communal (rather than individualistic) responsibility for the public's well-being [20].

Similarly, acceptance may improve or diminish with perceptions of shared societal experiences such as stay-at-home "lockdowns" [21] and the perceived effectiveness of government COVID-19 policies (eg, COVID-19 vaccine uptake improves with perceived government effectiveness and trust in government) [22].

Personal experiences may also affect one's attitude toward using an immunization passport. For example, having had or known someone who has had COVID-19 may incentivize one toward the use of immunity passports [19]. Strong neoliberal worldviews—a belief that the free market is fair and sensitive to the social and financial needs of the people—and a desire to return to normal economic activities may also affect passport acceptance [23]. Similarly, higher education may prove important to shaping one's opinions regarding the equality and necessity of immunity passports, just as it has with vaccinations [24].

Finally, immunity passport acceptance may depend on contextual factors regarding the state of the pandemic such as COVID-19 cases, deaths, and vaccine progress, which may change country to country and across time. In developing an understanding of what factors influence immunization passport acceptance, we may consider (1) acceptance while attempting to control for the contextual influences of each country (an international model), and (2) acceptance dependent on each country (national models). The former informs us of the necessary conditions for immunity passport acceptance across countries, allowing our findings to potentially generalize beyond our sample of six countries. By contrast, the latter assesses acceptance within each sampled country and may show how it varies as a function of each country's individual context and culture.

In summary, the objective of this study was to identify which societal, personal, and contextual factors predict the uptake of immunity passports across six countries. International modeling was used to provide generalizable findings, while national modeling was used to look for factors that deviate from international interpretations. Bayesian statistics were used to provide evidence toward or against factors that predict the uptake of immunity passports. This work is intended to provide clear scientific findings for medical and health researchers and for policymakers.

Methods

Design

We surveyed attitudes toward immunity passports in six countries with different experiences during the COVID-19 pandemic: Australia, Germany, the United Kingdom, Spain, Japan, and Taiwan. Using Bayesian linear mixed models, we aimed to determine which factors—societal, personal, and contextual issues related to COVID-19—influenced immunity passport acceptance. We examined our data in two ways. First, we attempted to control for the idiosyncratic effects of each country on immunity passport acceptance (using random effects in our modeling) to create a generalized framework for immunity passport acceptance. Second, we assessed acceptance within each country to consider cultural and contextual differences.

Ethics Considerations

All participants read a plain-language statement describing the online survey, the research question—to understand what factors contribute to the uptake of COVID-19 tracing technologies and immunity passports—and the study's benefits, risks, and data protections, before providing informed consent. Participants were informed that data collection would occur through password-protected accounts, be transferred through encrypted networks, and be held on secure password-protected servers; that no identifying information would be published or released; and that anonymized data would be made available through the Open Science Framework (OSF). Australian participants were reimbursed with gift cards or points programs per their agreement with Dynata, Spanish and German participants per their agreement with Lucid, Japanese participants per their agreement with Cross Marketing, and Taiwanese participants per their agreement with Gosurvey. UK participants were reimbursed a flat rate of £0.85 (~US $1.07) per 10-minute survey. Ethics approval was obtained for data collection in Australia and Japan from the University of Melbourne (approval 1955555), in the United Kingdom from the University of Bristol (approval 103344), in Germany from the Max Planck Institute for Human Development (approval L2020-4), in Spain from the University of Leeds (approval 103402), and in Taiwan from the National Cheng Kung University (approval 108-072).

Participants

Table 1 displays demographic information for each country and sample. We sampled 13,678 participants across six countries to determine their attitudes toward and acceptance of immunity passports. Each country collected between one and four nationally representative online samples. Samples were stratified by age, gender, and, where possible, state or province, based on the country's most recent census data. Participants were aged 18 years or older, and completed a 10- to 15-minute online survey for which they were financially reimbursed. Representative samples were obtained using third-party recruitment services and assigned unique identifiers upon entering the survey to ensure anonymity. Further country-specific details are provided in Multimedia Appendix 1. Data collection was completed as part of a wider international collaboration examining the acceptability of mobile tracking technologies to address the COVID-19 pandemic [23,25-28]. However, previous publications did not address the uptake of immunity passports.

Table 1. Demographic information relevant to each sample within each country.

Characteristic	Australia		Germany	Japan	Spain	Taiwan				United Kingdom
	Sample 1	Sample 2				Sample 1	Sample 2	Sample 3	Sample 4	
Participants, n	1514	578	1514	1081	1505	1500	1500	1500	1500	1486
Age (years), mean (SD)	48 (17)	48 (17)	47 (16)	46 (17)	48 (16)	40 (12)	40 (12)	40 (12)	41 (12)	46 (16)
Gender, %										
Man	50	48	49	49	48	48	47	48	50	48
Woman	49	51	50	51	52	52	53	52	50	51
Other	<1	<1	<1	0	<1	0	< 1	<1	0	<1
Prefer not to say	<1	0	0	0	0	0	0	<1	<1	<1
Education, %										
Less than high school	9	11	14	3	10	1	1	1	1	16
High school graduate	37	40	63	39	42	12	14	13	13	17
University graduate	54	49	23	58	47	87	86	86	86	67

Procedure

As the pandemic evolved, survey designs were updated with each sample; however, the key design elements assessed in this study remained unchanged (see Figure A1 in Multimedia Appendix 1). Survey questions were designed to address primary factors of the health belief model: illness severity (harm) and sensitivity (concern), policy benefits and barriers, self-efficacy, and calls to action [29,30].

Each participant provided informed consent and demographic information before using a Likert scale to report on their perceptions and impact of the COVID-19 pandemic. Participants then read one of three hypothetical scenarios describing a different type of mobile phone COVID-19 contact-tracing system—telecommunication tracking, a government app, or the Apple/Google exposure notification system—that would alert the user if they had contact with an infected individual, before completing a comprehension check and answering questions about these scenarios (for methods and results on these items by country, see [23,25-27]; note that these studies do not model immunity passport items). Finally, participants read a description of immunity passports before responding to items examining their attitudes toward immunity passports and their neoliberal worldviews. The survey concluded with a study debrief statement (Table 2).

Before responding to the immunity passport items, each participant read the following description:

> *An ‘immunity passport’ indicates that you have had a disease [or vaccination] and that you have the antibodies for the virus causing that disease. Having the antibodies implies that you are now immune and therefore unable to spread the virus to other people. Thus, if an antibody test indicates that you have had the disease, you could be allocated an ‘immunity passport’ which would subsequently allow you to move around freely. Immunity passports have been proposed as a potential step towards lifting movement restrictions during the COVID-19 pandemic.*

Upon survey completion, data were augmented with country-specific information. Data included national indices such as the World Bank’s Perceived Government Effectiveness Scale (scale 0-100, with higher values indicating greater effectiveness) [31], and the individuality subscale from the Hofstede Index of Collectivism (scale 0-100, with higher values indicating a more individualistic, less collectivist culture) [32]. Data also included COVID-19 cumulative cases and deaths [1], mask usage (binary variable: true or false) [33], stay-at-home “lockdown” usage (binary: true or false) [33], and mobile tracking technology usage (eg, COVIDSafe in Australia or the CORONA-WARN-App in Germany) [23,25-28,33]. News articles used to determine national policy metrics (eg, mask usage and lockdowns) are available through the OSF [33].

Table 2. COVID-19 perceived risk and impact, immunity passport, and worldview items. IP: immunity passport; WV: world view.

Item	Question	Label
Perception 1	How severe do you think the novel coronavirus (COVID-19) will be for the general population?	General harm
Perception 2	How harmful would it be for your health if you were to become infected with COVID-19?	Personal harm
Perception 3	How concerned are you that you might become infected with COVID-19?	Concern self
Perception 4	How concerned are you that somebody you know might become infected with COVID-19?	Concern others
Impact 1	Have you ever tested positive for COVID-19?	Positive self
Impact 2	Has somebody you know ever tested positive for COVID-19?	Positive other
Impact 3	Have you temporarily or permanently lost your job as a consequence of the COVID-19 pandemic?	Job loss
Passport 1	Would you support a government proposal to introduce “immunity passports” for the novel coronavirus (COVID-19)?	IP Support 1st
Passport 2	How concerned are you about the idea of introducing an “immunity passport” for COVID-19?	IP Concern
Passport 3	How much would you like to be allocated an “immunity passport” for COVID-19?	IP Like
Passport 4	To what extent do you believe an “immunity passport” for COVID-19 could harm the social fabric of your country?	IP Harm
Passport 5	To what extent do you believe that it is fair for people with “immunity passports” to return to work, while those without a passport cannot?	IP Fair
Passport 6	To what extent would you consider purposefully infecting yourself with COVID-19 to get an “immunity passport”?	IP Self-infect
Passport 7	Would you support a government proposal to introduce “immunity passports” for COVID-19?	IP Support 2nd
Worldview 1	An economic system based on free markets unrestrained by government interference automatically works best to meet human needs	WV Economy
Worldview 2	The free-market system may be efficient for resource allocation, but it is limited in its capacity to promote social justice [reverse-scored item]	WV Freemarket
Worldview 3	The government should interfere with the lives of its citizens as little as possible	WV Small Gov

Data Analysis and Reporting

Overview

Anonymized data and analysis codes for this study are available through the OSF [33]. Participants were excluded from analyses for missing a response to the immunity passport support item or for not completing the survey (removed n=790; details in Multimedia Appendix 1). The reported analyses are based on Bayesian methods and credible intervals to determine effects in the data. Bayesian methods sample a posterior distribution of plausible values (the probability that, given our data, the true population mean is “x”) by weighing the likelihood of a given observation against its prior probability of occurring in the sample. Under parametric assumptions, posterior distributions act to constrain the effect of outliers in the tails of the sampled data, allowing the highest region of data density—credible regions of the data distribution—to inform our decisions. Practically, this means that instead of testing a threshold of significance (ie, *P* value or Bayes factor), we may instead compare the 95% credible regions of the data distributions and determine whether or not they overlap.

Immunity Passport Perceptions

Bayesian ordinal probit regressions were used to compare Likert-scale responses using the *MCMCoprobit* and *HPDinterval* functions in the R packages *MCMCpack* [34] and *Coda* [35], respectively. This method compares Likert items by assuming there are latent normally distributed continuous variables underlying the ordinal responses. These latent variables are then segmented into ordinal Likert responses by the number of response options minus one as thresholds. To set the location of the underlying latent variable, the lowest threshold parameter is fixed at zero [36] and all other thresholds are estimated. Country-level data were modeled together [37] and individual samples within countries were not modeled. This approach allowed us to directly compare attitudes to immunity passport items across countries and poses more reasonable assumptions than directly comparing the mean or raw distribution of the Likert scales [37]. This analysis was completed for immunity passport items, as presented in the main text, and for COVID-19 perception and worldview items (Multimedia Appendix 2).

International Modeling

Bayesian generalized linear mixed effects modeling was used to assess what factors did or did not predict support for immunity passports. Demographics, perceptions, and impact of COVID-19; COVID-19 cases and deaths by country; neoliberal worldviews; and immunity passport items were treated as additive and independent predictor variables of immunity passport support. Random intercept effects were included to account for dependencies introduced in the data by each country. Likert ratings were treated as numeric data and noncategorical variables were scaled within each country to have a mean of 0 and SD of 1.

Posterior distributions of model parameters were estimated using Hamiltonian Markov Chain Monte Carlo No-U-turn Sampling implemented in Stan via the R package *brms* [38,39]. Four chains each with 2000 iterations and 1000 burn-ins were used. Noninformative priors were set for the intercept and random effect SD parameters (both Cauchy distributions centered on 0 and a scale parameter of 2.5), and fixed effects were estimated from weakly informative priors with a Laplacian distribution centered on 0 and a scale parameter of 1. Practically, this means that factors able to overcome this strong prior bias toward zero (ie, no effect) are meaningful.

Models reported in the main text assess passport support after answering immunity passport questions. Outcome variables were reduced to a binary response set: "support yes" ("moderate," "a lot," or "extreme" Likert-scale items) and "support no" ("none," "a bit," or "some"). Sample order (present in only two counties) and gender "Other" or "Prefer not to say" were removed as small samples led to unstable model fits. The remaining factors had adequate responses for stable model fits. Models predicting passport support prior to answering questions about immunity passports, and models of international attitudes using the full range of ordinal response options are included in Multimedia Appendix 2. All relevant results were comparable to the model presented in the main text.

National Modeling

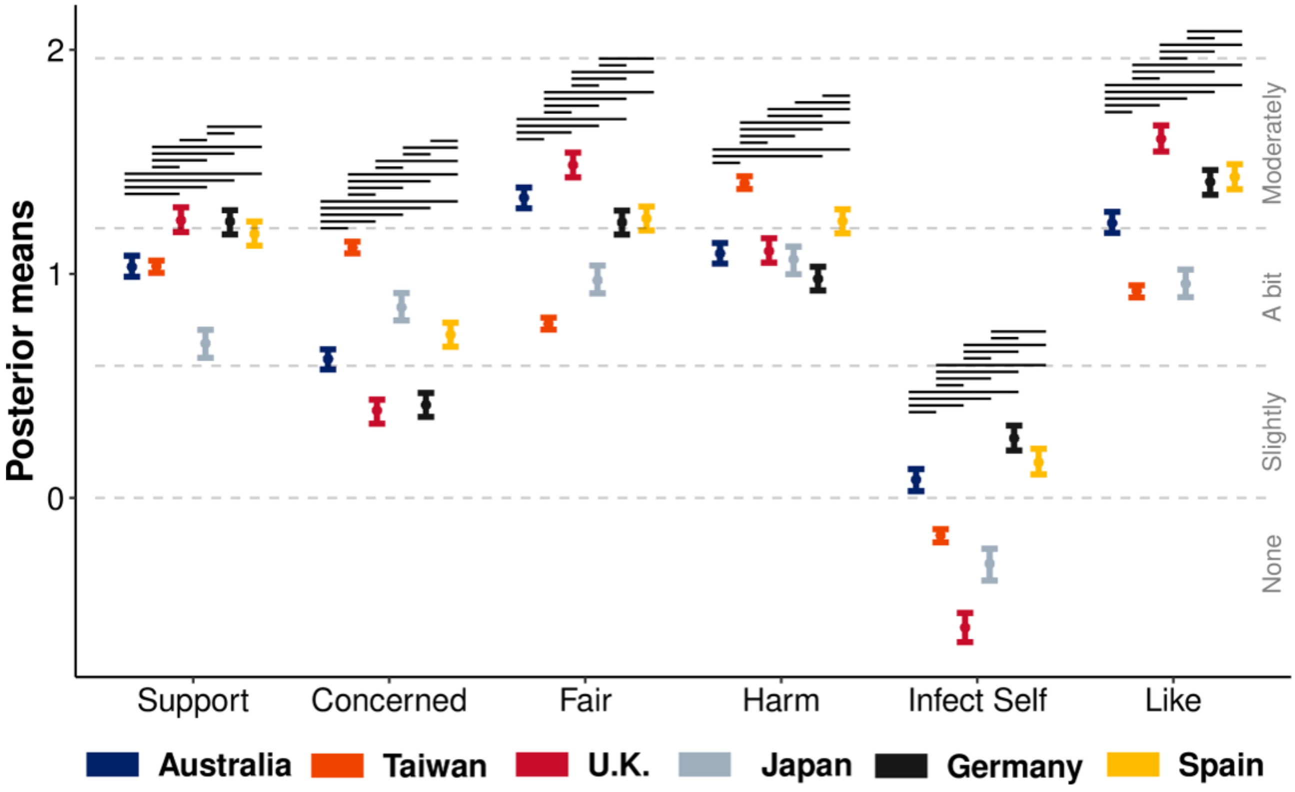
National modeling replicated the model procedures described above; however, only factors within each country were assessed. This allows for cultural and contextual variation to be observed at a national level, which may be informative for readers, researchers, and policymakers in those countries. Modeling for each country is reported separately in Multimedia Appendix 3, and a summary of the primary differences to the international model is presented in the text.

Results

Immunity Passport Perceptions

Figure 1 displays the mean ordinal regression posterior distributions and associated Likert-style responses for immunity passport perceptions across the six countries. Mean immunity passport support scores based on binary classifications (support: yes=["moderate," "a lot," or "extreme"], no=["none," "a bit," or "some"]) showed that support was the highest in Germany (775/1507, 51.43%) and the United Kingdom (759/1484, 51.15%); followed by Taiwan (2841/5989, 47.44%), Australia (963/2086, 46.16%), and Spain (693/1491, 46.79%); and the lowest in Japan (241/1081, 22.29%). All countries display little to no inclination for infecting one's self to gain an immunity passport, and although most countries are only "a bit" concerned by the introduction of immunity passports, they are generally deemed as posing a moderate risk of harm to society. As these were secondary analyses to the main focus of this paper, we include a full description of the COVID-19 impact variables and worldview items in Multimedia Appendix 3.

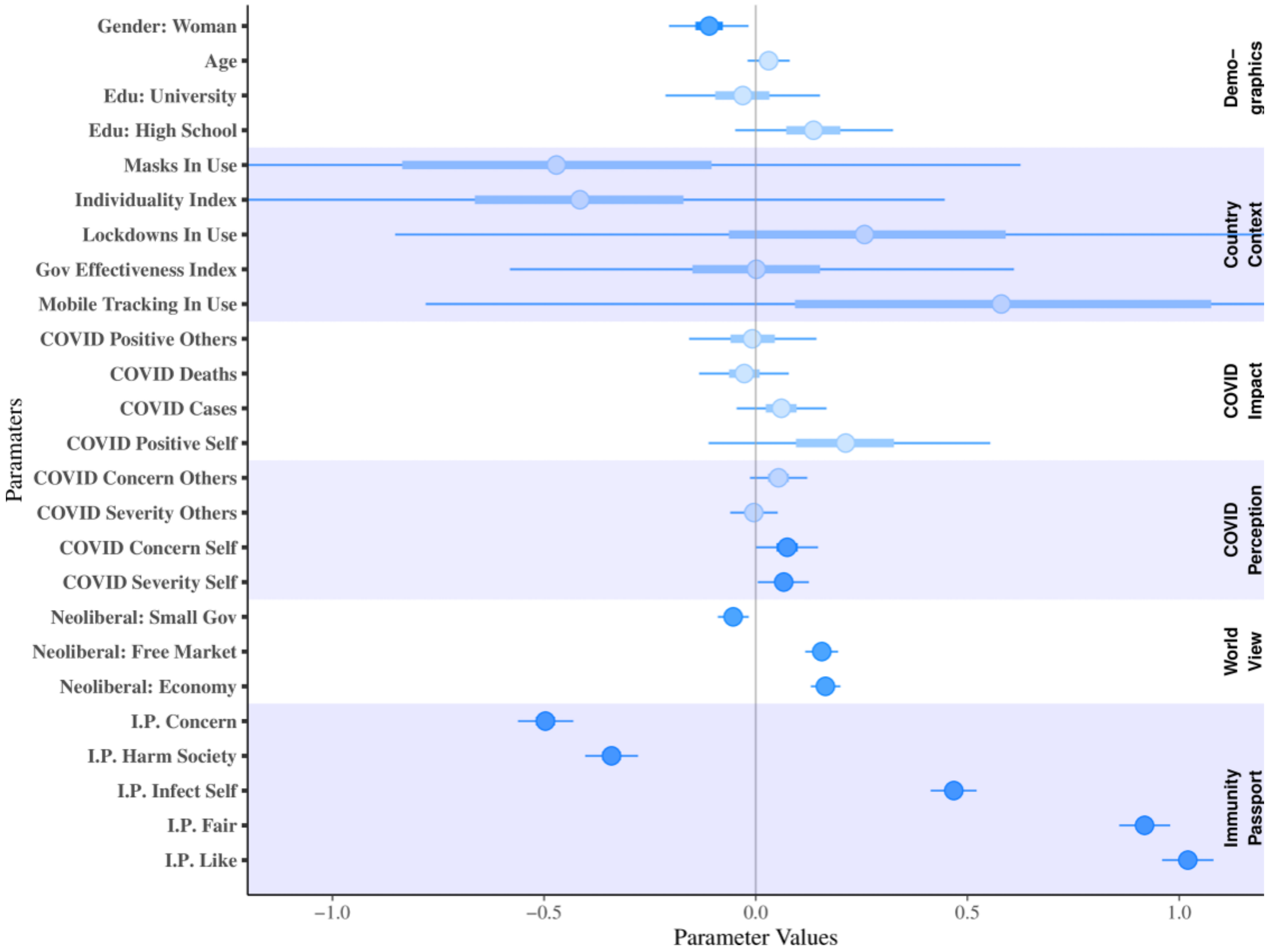
Figure 1. Ordinal regression mean posterior distributions (left axis; vertical error bars) and latent Likert-scale ratings (right axis; dotted horizontal lines) for immunity passport perceptions in Australia, Taiwan, the United Kingdom, Japan, Germany, and Spain. Error bars display the 95% highest posterior density interval. Dotted lines indicate Likert-scale categories, and nonoverlapping intervals (ie, effects) between countries are denoted by black horizontal lines within each item.



International Modeling

Figure 2 displays the posterior estimates of the Bayesian generalized linear mixed effects model of immunity passport support using demographics, COVID-19 perceptions and impact, country-specific indices (eg, mask usage, government effectiveness), worldview, and attitudes to immunity passports as additive factors, with a random intercept for each country. Error bars display the 95% highest density interval. The global intercept had a mean of –1.67 (95% CI –3.14 to 0.32). Country intercept means were ordered from the lowest to the highest as Japan, Spain, Australia, United Kingdom, Germany, and Taiwan; credible intervals were the lowest for Japan (mean –0.66, 95% CI –2.10 to 0.57) and the highest for Taiwan (mean 0.61, 95% CI –0.79 to 2.31); and intervals for all countries extended over the zero midpoints, indicating no effect. As posterior mean estimates are rather opaque, we provide an explanation of the international model variables in terms of their odds ratios.

Figure 2. Bayesian generalized linear mixed effects model of immunity passport support (post immunity passport questions) across countries. Positive parameters display immunity passport support; negative values display a decrease in support. Bars represent 50% of the parameter distribution centered on the parameter mean and tails display the 95% highest density interval. Opaque variables show instances where the posterior interval does not overlap zero. IP: immunity passport.



Predictive variables of immunity passport acceptance—those where the 95% highest density interval did not cross zero—included increased COVID-19 concern, perceived virus severity to one’s self, worldview (believing the free market works best and that it is limited in its ability to support social justice), and immunity passport items (liking and thinking immunity passports are fair, and being willing to self-infect to receive an immunity passport). Personally liking the idea of immunity passports was the strongest predictor variable, with an odds ratio of 2.8; that is, a 1-SD increase in “liking” immunity passports corresponded to a 2.8-factor SD increase in the odds of supporting their introduction. This may seem rather tautological, but shows that positive attitudes toward immunity passports are the strongest predictor of their acceptance.

Predictive variables against the introduction of immunity passports included gender (identifying as a woman), worldview (supporting minimal government interference), and immunity passport risk items (concern and risk of harm to society). Immunity passport concern was the most predictive item against the acceptance of immunity passports, with a 1-SD increase therein corresponding to a 0.61-factor increase in the odds of supporting the introduction of immunity passports (equivalent to a 1.65-factor increase in the odds of not supporting the introduction of immunity passports).

National Modeling

Table 3 summarizes the factors that met our criteria for an effect—credible intervals that did not overlap zero—in the international and national models. Parameters are displayed as odds ratios—the degree to which each parameter increases the odds of immunity passport support—and indicate whether they increase or decrease the likelihood of immunity passport support. An odds ratio of 1 indicates no effect, less than 1 indicates a negative relationship, and an odds ratio greater than 1 indicates a positive relationship between parameters. Three notable differences were observed between national and international parameters: gender and COVID-19 severity-self were only identified in the international model, and COVID-19 concern for others was only identified in the national model for Japan.

Table 3. Odds ratios for international and national model parameters that did/did not support immunity passport acceptance.[a]

Model	Gender	COVID-19 perceptions			Worldview items			Immunity passport items				
	Woman[b]	Concern self[c]	Severity self[c]	Concern others[c]	Small government[b]	Free market[c]	Economy[c]	Concern[b]	Harm[b]	Infect self[c]	Fair[c]	Like[c]
International model	0.9	1.07	1.07	—[d]	0.98	1.17	1.17	0.61	0.71	1.6	2.51	2.77
Australia	—	—	—	—	0.88	—	1.31	0.64	—	2.03	3.42	3.71
Germany	—	—	—	—	0.9	1.14	1.23>	0.76	0.73	1.58	2.36	3.67
Japan	—	—	—	1.48	—	1.25	—	0.63	—	2.01	1.82	2.56
Spain	—	—	—	—	0.85	1.28	1.12	0.54	—	1.9	3.19	3.29
Taiwan	—	1.14[c]	—	—	—	1.16	1.17	0.61	0.63	1.38	2.27	2.05
United Kingdom	—	—	—	—	—	—	—	0.57	0.53	—	3.16	5.47

[a] Ratios represent the multiplicative increase each coefficient confers to immunity passport support. Displayed parameters are those with credible intervals that did not cross zero.

[b] Column variables that decreases the likelihood of immunity passport support.

[c] Column variables that increases the likelihood of immunity passport support.

[d] Not applicable.

Discussion

Principal Findings

The introduction of immunity passports received moderate support across the six sampled countries, except for Japan. International modeling showed that immunity passport acceptance was primarily driven by perceived personal risks (COVID-19 concern and severity, willingness to self-infect) and benefits (liking immunity passports and believing they are fair), and societal factors (neoliberal worldviews). Acceptance was not influenced by contextual factors such as COVID-19 cases and fatalities, or mask, lockdown, or tracing technology usage. National modeling displayed little variation from international results, suggesting that our international findings may prove inferential to the global community.

International Modeling

International modeling identified several predictive factors of passport support, including worldview, COVID-19 concern for one's self, and perceived virus severity to one's self; however, critical variables were those directly assessing attitudes toward immunity passports. Desiring a passport, perceiving passports as fair, and being willing to infect one's self to gain an immunity passport were all positively associated with immunity passport support. Although immunity passport perceptions displayed greater concern for others than for one's self, concern and perceived virus severity toward others were not predictive of passport support. These findings highlight that immunity passport support hinges upon personal benefits. Similar findings have been observed for vaccine uptake [40] and for mobile health technologies that emphasize patient self-efficacy [41].

Immunity passport support also improved with neoliberal worldviews, specifically seeing the free market as fair and as working best if unrestrained by government interference. By contrast, limiting government interference was negatively predictive of immunity passport support, along with gender, immunity passport concern, and perceived risk of harm to society. These parameters code societal factors that influence one's judgment on immunity passport acceptance. Additionally, we posit that these worldview items may serve as a proxy for correlated attitudes such as political worldviews. This may prove important in countries prone to political tribalism (eg, the United States) [42] where bipartisan support would be needed when promoting immunity passports, not from a legislative standpoint but rather from the view of gaining public support and the "social licence to operate" [43,44].

Contextual factors such as COVID-19 policy decisions (ie, wearing masks, home lockdowns, and the introduction of mobile tracking technologies) and country-specific indices (ie, COVID-19 cases and deaths, individualism, government effectiveness) were not predictive of immunity passport support. This reinforces our theory that attitudes toward the uptake of immunity passports are driven primarily by personal risks and benefits, and to a lesser extent, societal factors.

National Modeling

National modeling revealed minor differences to the international model. Some countries emphasized concern for others (eg, Japan) or concern for one's self (eg, Taiwan), or differed by their lack of a predictive variable when compared to the international model. For example, immunity passport support increased with the likelihood of infecting one's self in every country except the United Kingdom. By contrast, some factors were consistently predictive across countries (eg, seeing immunity passports as being “fair”). Understanding international and national variance is key to this study; no single country stands as a monolith from which understanding or predictions may be extrapolated. Attitudes modeled across countries provide insights (eg, the predictive qualities of gender and the severity of COVID-19 to one’s self) otherwise lost at national levels. By contrast, national accounts provide a nuanced view of attitudes that allow policymakers to consider how immunity passports would be perceived within a single country relative to the global community.

Limitations

The current investigation was primarily limited by our sampling options. Representative online sampling was performed in all countries; however, being online samples, they may be biased toward technological solutions for large-scale problems. Further, samples were not representative for education, with respondents in each country skewing toward being more educated than their respective populations.

We were also severely limited by public perceptions at the time of this investigation. In April-May of 2020, international vaccine rollouts were yet to begin and the focus was on nonpharmaceutical methods for virus suppression. Attitudes may have since shifted as media begin to report on governments seeking to introduce vaccination and/or immunity passports, and the risks and benefits these documents provide. This discussion will only become more heated as corporations such as airlines begin limiting services based on whether individuals have been vaccinated or have recently recovered, and as the long-term side effects of COVID-19 become apparent. Discussions will also evolve as counties reconsider what being “fully vaccinated” entails (eg, one, two, or several booster shots) and what vaccines are deemed suitable for entry to a country. Additionally, public discourse will evolve as people experience the usability of immunization passport technology, as a key barrier for mobile health technology uptake [45].

Finally, a key limitation of our study was our inability to directly assess attitudes to specific security and privacy-preserving digital passport techniques such as blockchain technologies employed by South Korea [3,13] or the Visual Digital Seals (QR code) technology employed by Australia [14]. This omission was due to the survey being conducted before these technologies were in use. Regardless, concern over immunity passports remained a key factor in our modeling, and may be inclusive of privacy and security issues as these are established barriers for the adoption of other nonpharmaceutical COVID-19 interventions (eg, COVID-19 contact-tracing apps) [25,46,47]. Decision-makers should address and minimize these concerns among potential users.

Conclusion

Governments and corporations are now introducing immunity and vaccination passports to quickly return society and the economy to normal, while encouraging the public to get vaccinated to protect themselves and their loved ones. However, the introduction of these passports will only work if the public supports their use. Policymakers can take from our findings several clear conclusions on how to effectively introduce immunization passports. Passport acceptance will benefit from highlighting the societal benefits (shorter lockdowns, a return to normal work and activities, improved community health) and personal health benefits conferred by these passports, and by addressing and minimizing the societal risks (eg, creating “vaccinated” vs “unvaccinated” social classes) and personal risks (privacy and anonymity) posed by their introduction. To a lesser extent, acceptance would also benefit by framing immunization passports as benefiting the economy and workforce (ie, neoliberal worldview). Finally, we note that internationally, women were less accepting of immunization passports; however, this trend was not observed within individual countries. Hopefully, by successfully accounting for these factors in policy decisions regarding immunity passports, governments and businesses may avoid public backlash when members of the public are prompted: “Papers please?”

Acknowledgments

Data collection in Australia was supported by anonymous philanthropic funding to the Peter Doherty Institute for Infection and Immunity to SD. Data collection in Germany was funded by the planning grant of the Volkswagen Foundation to RH, SL, and SH (Initiative “Artificial Intelligence and the Society of the Future”). SL was supported by a Research Award from the Humboldt Foundation in Germany while this research was conducted. SL also received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement 964728 (JITSUVAX). Data collection in Taiwan was supported by the National Cheng Kung University and Ministry of Science and Technology (MOST 110-2321-B-006-004 and MOST 108-2321-B-006-022-MY2) to CTY. Data collection in the United Kingdom was supported by the Elizabeth Blackwell Institute, University of Bristol, with funding from the University’s Alumni and Friends fund. Data collection in Spain was supported by funding from Leeds University Business School, University of Leeds to YO, who was in part supported by a Population Research Fellowship awarded by Cancer Research UK (C57775/A22182). Data collection in Japan was supported by funding from the School of Psychological Sciences, the University of Melbourne, and provided to YK.

Authors' Contributions

The overarching project was led by SD and SL, and coordinated with the input of all named authors. Data collection in Australia was handled by PMG, JPW, and SD; by CTY in Taiwan; by AK and PLS in Germany; by TK in Japan; by SL in the United Kingdom; and by YO in Spain. Survey translations for Germany were handled by AK and PLS, for Spain by YO, for Taiwan by

CTY, and for Japan by TK. Data collation and analysis were completed by PMG. Manuscript writing and revisions were led by PMG and contributed to by all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Methodological differences between countries and where to find additional methodology details, code, and data for each country.
[DOCX File , 178 KB - publichealth_v8i7e32969_app1.docx]

Multimedia Appendix 2
Alternative ordinal and binomial modeling accounts of the immunity passport data.
[DOCX File , 179 KB - publichealth_v8i7e32969_app2.docx]

Multimedia Appendix 3
Immunity passport modeling completed for each country separately.
[DOCX File , 47 KB - publichealth_v8i7e32969_app3.docx]

References

1. WHO Coronavirus (COVID-19) dashboard. World Health Organization. 2021. URL: https://covid19.who.int [accessed 2022-01-28]
2. Kelleher SR. Europe update: These 12 countries are now using E.U.’s digital Covid certificate, with another 16 on deck. Forbes. 2021. URL: https://www.forbes.com/sites/suzannerowankelleher/2021/06/13/europe-digital-covid-certificate-vaccine-passport/?sh=79f5600c2251 [accessed 2021-07-10]
3. Cha S. South Korea to issue blockchain-protected digital 'vaccine passports'. Reuters. 2021 Apr. URL: https://www.reuters.com/article/us-health-coronavirus-southkorea-idUSKBN2BO43W [accessed 2021-07-10]
4. Guarascio F. Exclusive: WHO vaccine scheme risks failure, leaving poor countries with no COVID shots until 2024. Reuters. 2020 Dec 16. URL: https://www.reuters.com/business/healthcare-pharmaceuticals/exclusive-who-vaccine-scheme-risks-failure-leaving-poor-countries-no-covid-shots-2020-12-16 [accessed 2021-08-12]
5. Mallapaty S, Callaway E. What scientists do and don't know about the Oxford-AstraZeneca COVID vaccine. Nature 2021 Apr;592(7852):15-17. [doi: 10.1038/d41586-021-00785-7] [Medline: 33762708]
6. Kim JH, Marks F, Clemens JD. Looking beyond COVID-19 vaccine phase 3 trials. Nat Med 2021 Feb 19;27(2):205-211. [doi: 10.1038/s41591-021-01230-y] [Medline: 33469205]
7. Liew CH, Flaherty GT. Immunity passports to travel during the COVID-19 pandemic: controversies and public health risks. J Public Health 2021 Apr 12;43(1):e135-e136 [FREE Full text] [doi: 10.1093/pubmed/fdaa125] [Medline: 32756915]
8. COVID-19 natural immunity. World Health Organization. 2021 May 10. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Natural_immunity-2021.1 [accessed 2021-06-02]
9. Coronavirus: Commission proposes a Digital Green Certificate. European Commission. 2021. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1181 [accessed 2021-06-06]
10. Brown RCH, Savulescu J, Williams B, Wilkinson D. Passport to freedom? Immunity passports for COVID-19. J Med Ethics 2020 Oct;46(10):652-659 [FREE Full text] [doi: 10.1136/medethics-2020-106365] [Medline: 32817362]
11. Dye C, Mills MC. COVID-19 vaccination passports. Science 2021 Mar 19;371(6535):1184. [doi: 10.1126/science.abi5245] [Medline: 33737463]
12. IATA travel pass initiative. International Air Transport Association. URL: https://www.iata.org/en/programs/passenger/travel-pass [accessed 2021-07-10]
13. Tsoi KKF, Sung JJY, Lee HWY, Yiu KKL, Fung H, Wong SYS. The way forward after COVID-19 vaccination: vaccine passports with blockchain to protect personal privacy. BMJ Innov 2021 Apr 07;7(2):337-341. [doi: 10.1136/bmjinnov-2021-000661]
14. How the International COVID-19 Vaccination Certificate works. Australian Government Department of Foreign Affairs and Trade. 2022. URL: https://www.passports.gov.au/how-international-covid-19-vaccination-certificate-works [accessed 2022-01-25]
15. Gelb A, Mukherjee A. A COVID vaccine certificate building on lessons from digital id for the digital yellow card. Center for Global Development. 2021. URL: https://www.cgdev.org/publication/covid-vaccine-certificate-building-lessons-digital-id-digital-yellow-card [accessed 2022-01-25]
16. Voo TC, Clapham H, Tam CC. Ethical implementation of immunity passports during the COVID-19 pandemic. J Infect Dis 2020 Aug 04;222(5):715-718 [FREE Full text] [doi: 10.1093/infdis/jiaa352] [Medline: 32582943]
17. Davies G. Vaccine passports are being trialed in the UK: Why are they so controversial? ABC News. 2021 Apr 12. URL: https://abcnews.go.com/International/vaccine-passports-trialed-uk-controversial/story?id=76976981 [accessed 2021-06-12]

18. Wilford G. French mob destroys portrait of macron in protest against vaccine passports. The Times. 2021. URL: https://www.thetimes.co.uk/article/french-mob-destroys-portrait-of-macron-in-protest-against-vaccine-passports-q9tl9x88d [accessed 2021-06-12]
19. Tsai C. Personal risk and societal obligation amidst COVID-19. JAMA 2020 Apr 28;323(16):1555-1556. [doi: 10.1001/jama.2020.5450] [Medline: 32242889]
20. Seifu Estifanos A, Alemu G, Negussie S, Ero D, Mengistu Y, Addissie A, et al. 'I exist because of we': shielding as a communal ethic of maintaining social bonds during the COVID-19 response in Ethiopia. BMJ Glob Health 2020 Jul;5(7):e003204 [FREE Full text] [doi: 10.1136/bmjgh-2020-003204] [Medline: 32709704]
21. Zhou J, Yang S, Xiao C, Chen F. Examination of community sentiment dynamics due to COVID-19 pandemic: a case study from a state in Australia. SN Comput Sci 2021;2(3):201 [FREE Full text] [doi: 10.1007/s42979-021-00596-7] [Medline: 33851137]
22. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
23. Lewandowsky S, Dennis S, Perfors A, Kashima Y, White JP, Garrett P, et al. Public acceptance of privacy-encroaching policies to address the COVID-19 pandemic in the United Kingdom. PLoS One 2021;16(1):e0245740 [FREE Full text] [doi: 10.1371/journal.pone.0245740] [Medline: 33481877]
24. Khubchandani J, Sharma S, Price JH, Wiblishauser MJ, Sharma M, Webb FJ. COVID-19 vaccination hesitancy in the United States: a rapid national assessment. J Community Health 2021 Apr;46(2):270-277 [FREE Full text] [doi: 10.1007/s10900-020-00958-x] [Medline: 33389421]
25. Garrett PM, White JP, Lewandowsky S, Kashima Y, Perfors A, Little DR, et al. The acceptability and uptake of smartphone tracking for COVID-19 in Australia. PLoS One 2021;16(1):e0244827 [FREE Full text] [doi: 10.1371/journal.pone.0244827] [Medline: 33481841]
26. Garrett P, Wang YW, White JP, Kashima Y, Dennis S, Yang CT. High acceptance of COVID-19 tracing technologies in Taiwan: a nationally representative survey analysis. Int J Environ Res Public Health 2022 Mar 11;19(6):3323 [FREE Full text] [doi: 10.3390/ijerph19063323] [Medline: 35329008]
27. Kozyreva A, Lorenz-Spreen P, Lewandowsky S, Garrett PM, Herzog SM, Pachur T, et al. Psychological factors shaping public responses to COVID-19 digital contact tracing technologies in Germany. Sci Rep 2021 Sep 21;11(1):18716. [doi: 10.1038/s41598-021-98249-5] [Medline: 34548550]
28. Garrett PM, Wang Y, White JP, Hsieh S, Strong C, Lee Y, et al. Young adults view smartphone tracking technologies for COVID-19 as acceptable: the case of Taiwan. Int J Environ Res Public Health 2021 Feb 02;18(3):1332 [FREE Full text] [doi: 10.3390/ijerph18031332] [Medline: 33540628]
29. Hochbaum GM. Public participation in medical screening programs: A socio-psychological study. Washington, DC: Public Health Service; 1958.
30. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
31. Worldwide governance indicators. World Bank. 2020. URL: http://info.worldbank.org/governance/wgi [accessed 2021-06-08]
32. Country comparison. Hofstede Insights. URL: https://www.hofstede-insights.com/country-comparison [accessed 2021-06-08]
33. Garrett P. Predictive factors for the uptake of national and international COVID-19 immunity and vaccination passports. Open Science Foundation. 2021 Aug 07. URL: https://osf.io/pwb9x [accessed 2022-01-02]
34. Martin AD, Quinn KM, Park JH. Mcmcpack: Markov Chain Monte Carlo. J Stat Soft 2011;42(9):1-21. [doi: 10.18637/jss.v042.i09]
35. Plummer M, Best N, Cowles K, Vines K. Coda: convergence diagnosis and output analysis for MCMC. R News 2006;6(1):7-11.
36. Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. J Am Stat Assoc 1993 Jun;88(422):669. [doi: 10.2307/2290350]
37. Bürkner P, Vuorre M. Ordinal regression models in psychology: a tutorial. Adv Methods Pract Psychol Sc 2019 Feb 25;2(1):77-101. [doi: 10.1177/2515245918823199]
38. Bürkner P. brms: An R package for Bayesian multilevel models using Stan. J Stat Soft 2017;80(1):1-28. [doi: 10.18637/jss.v080.i01]
39. Bürkner P. Advanced Bayesian multilevel modeling with the R Package brms. The R Journal 2018;10(1):395. [doi: 10.32614/RJ-2018-017]
40. Ashworth M, Thunström L, Cherry TL, Newbold SC, Finnoff DC. Emphasize personal health benefits to boost COVID-19 vaccination rates. Proc Natl Acad Sci U S A 2021 Aug 10;118(32):e2108225118 [FREE Full text] [doi: 10.1073/pnas.2108225118] [Medline: 34315824]
41. Salgado T, Tavares J, Oliveira T. Drivers of mobile health acceptance and use from the patient perspective: survey study and quantitative model development. JMIR Mhealth Uhealth 2020 Jul 09;8(7):e17588 [FREE Full text] [doi: 10.2196/17588] [Medline: 32673249]

42. Helmuth B, Gouhier TC, Scyphers S, Mocarski J. Trust, tribalism and tweets: has political polarization made science a “wedge issue”? Clim Chang Responses 2016 May 30;3(1):11895. [doi: 10.1186/s40665-016-0018-z] [Medline: 31417176]
43. Moffat K, Lacey J, Zhang A, Leipold S. The social licence to operate: a critical review. Forestry 2015 Nov 22;89(5):477-488. [doi: 10.1093/forestry/cpv044]
44. White JP, Dennis S, Tomko M, Bell J, Winter S. Paths to social licence for tracking-data analytics in university research and services. PLoS One 2021;16(5):e0251964 [FREE Full text] [doi: 10.1371/journal.pone.0251964] [Medline: 34019592]
45. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
46. Thomas R, Michaleff ZA, Greenwood H, Abukmail E, Glasziou P. Concerns and misconceptions about the Australian government's COVIDSafe App: cross-sectional survey study. JMIR Public Health Surveill 2020 Nov 04;6(4):e23081 [FREE Full text] [doi: 10.2196/23081] [Medline: 33048826]
47. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 contact tracing apps: predicted uptake in the Netherlands based on a discrete choice experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]

Abbreviations

OSF: Open Science Framework

Edited by T Sanchez, A Mavragani; submitted 17.08.21; peer-reviewed by C Jacob, P Williams; comments to author 23.01.22; revised version received 28.01.22; accepted 01.04.22; published 15.07.22.

Please cite as:
Garrett PM, White JP, Dennis S, Lewandowsky S, Yang CT, Okan Y, Perfors A, Little DR, Kozyreva A, Lorenz-Spreen P, Kusumi T, Kashima Y
Papers Please - Predictive Factors of National and International Attitudes Toward Immunity and Vaccination Passports: Online Representative Surveys
JMIR Public Health Surveill 2022;8(7):e32969
URL: https://publichealth.jmir.org/2022/7/e32969
doi:10.2196/32969
PMID:35377317

Original Paper

Explaining the Varying Patterns of COVID-19 Deaths Across the United States: 2-Stage Time Series Clustering Framework

Fadel M Megahed[1], PhD; L Allison Jones-Farmer[1], PhD; Yinjiao Ma[2], MPH; Steven E Rigdon[2], PhD

[1]Farmer School of Business, Miami University, Oxford, OH, United States

[2]Department of Epidemiology and Biostatistics, College for Public Health and Social Justice, Saint Louis University, St Louis, MO, United States

Corresponding Author:
Steven E Rigdon, PhD
Department of Epidemiology and Biostatistics
College for Public Health and Social Justice
Saint Louis University
3545 Lafayette Ave
St Louis, MO, 63104
United States
Phone: 1 3149772781
Email: steve.rigdon@slu.edu

Abstract

Background: Socially vulnerable communities are at increased risk for adverse health outcomes during a pandemic. Although this association has been established for H1N1, Middle East respiratory syndrome (MERS), and COVID-19 outbreaks, understanding the factors influencing the outbreak pattern for different communities remains limited.

Objective: Our 3 objectives are to determine how many distinct clusters of time series there are for COVID-19 deaths in 3108 contiguous counties in the United States, how the clusters are geographically distributed, and what factors influence the probability of cluster membership.

Methods: We proposed a 2-stage data analytic framework that can account for different levels of temporal aggregation for the pandemic outcomes and community-level predictors. Specifically, we used time-series clustering to identify clusters with similar outcome patterns for the 3108 contiguous US counties. Multinomial logistic regression was used to explain the relationship between community-level predictors and cluster assignment. We analyzed county-level confirmed COVID-19 deaths from Sunday, March 1, 2020, to Saturday, February 27, 2021.

Results: Four distinct patterns of deaths were observed across the contiguous US counties. The multinomial regression model correctly classified 1904 (61.25%) of the counties' outbreak patterns/clusters.

Conclusions: Our results provide evidence that county-level patterns of COVID-19 deaths are different and can be explained in part by social and political predictors.

(JMIR Public Health Surveill 2022;8(7):e32164) doi:10.2196/32164

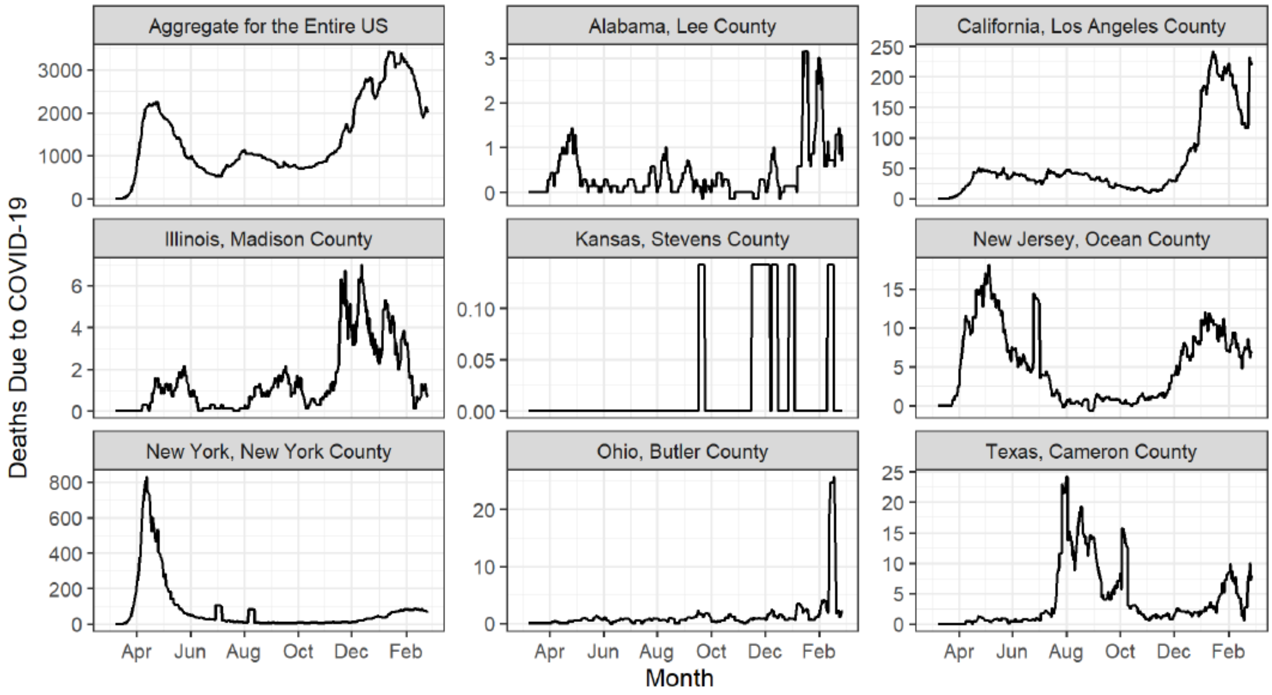
KEYWORDS

explanatory modeling; multinomial regression; SARS-CoV-2; COVID-19; socioeconomic analyses; time series analysis

Introduction

A geographically, politically, and socioeconomically diverse nation, the United States consists of 50 states, 48 of which are contiguous. When considering the COVID-19 pandemic in different regions throughout the United States, different patterns of outcomes emerge. Based on data obtained from the open source COVID-19 data hub [1], Figure 1 shows the national 7-day moving average of deaths as well as the various patterns that arise among 8 example counties from Sunday, March 1, 2020, to Saturday, February 27, 2021. For example, New York, NY, experienced a large first wave of deaths, followed by a relatively low death count through the remainder of the study. Nearby Ocean County, NJ, a populous county near the New Jersey shore had a large first wave of deaths, followed by a second wave beginning in late 2020. In contrast, Butler County, OH, a populous midwestern county, showed low death counts until late in the study period. None of these patterns mimics the overall pattern for the aggregate death counts in the United States.

Figure 1. Time series profiles of the 7-day moving average of new COVID-19 deaths for the entire United States and 8 sample counties.



Based on data from March 01, 2020 - Feb 27, 2021

Early in the COVID-19 pandemic, the county-level population mortality and case fatality rates were significantly different among the US regions [2]. Explanations for regional differences in health outcomes related to COVID-19 may be the structure of the government and policy making within the United States as it relates to the social vulnerability of the population. In the United States, each state consists of county governments that set health and economic policies for local communities. The counties within the states vary in terms of population size, demographics, access to health care, housing, and transportation. Some have noted that the regional differences in COVID-19 policies, compliance, and subsequent outcomes could be due to political differences across the regions. Goldwitzer et al [3] showed Republican-leaning counties displayed less physical distancing compared to Democratic-leaning counties and a subsequent increase in COVID-19 cases and deaths. Another study showed Democratic governors were 50% more likely to implement stay-at-home orders [4], which have been associated with increased physical distancing and reduction in COVID-19 cases and deaths [5].

Here, we investigate the regional patterns in deaths attributed to COVID-19. The phenomenon of differing national and regional patterns within the United States was illustrated for confirmed COVID-19 cases in Megahed et al [6]. In addition, a report by the *Financial Times* [7] argued, “Across the world, public health data are gathered at a very local level before aggregation into regional and national figures.... While useful as a summary, local distinctions get lost, painting a misleading image of whole countries being affected uniformly.” In this study, we investigated the various patterns of COVID-19 deaths across 3108 contiguous counties in the United States. We also sought to determine what factors relate to the pattern of deaths. Specifically, we posed 3 questions:

- How many distinct clusters of counties in the United States exhibit similar time series patterns in the deaths due to COVID-19?
- How are these clusters geographically distributed across the United States?
- Are certain geographic, political, government, and social vulnerability variables associated with the patterns of COVID-19 related deaths?

To address the first question, we performed a cluster analysis on the time series of the 3108 US counties. We provided maps to show the geographic distribution of the clusters. To address the third question, we applied a multinomial logistic regression analysis using geographic, political, and social vulnerability data to explain the patterns of deaths due to COVID-19 over time.

Methods

This study was conducted in 3 stages: (1) data gathering and preprocessing, (2) time series clustering, and (3) modeling and cluster validation.

Data

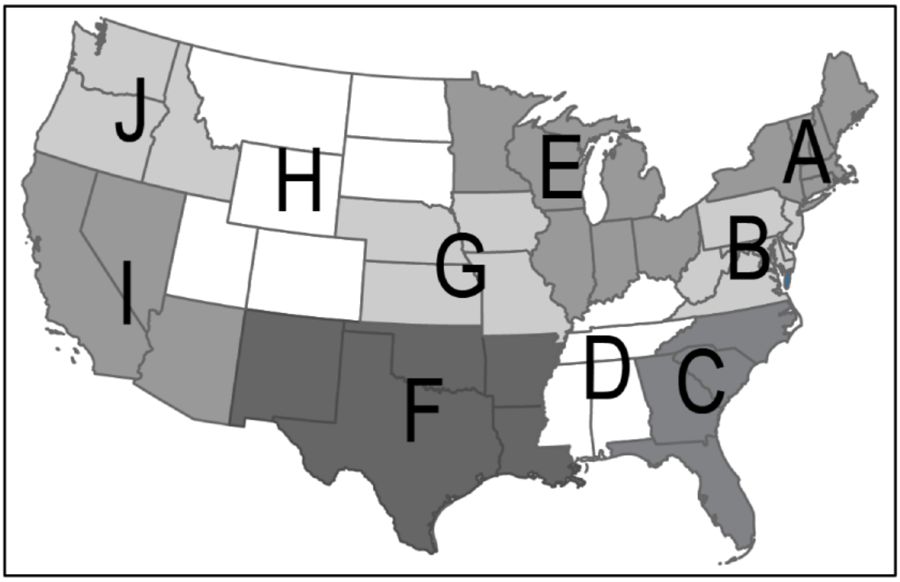
The open source COVID-19 data hub [1] was used to extract county-level time series data related to confirmed COVID-19 deaths from Sunday, March 1, 2020, to Saturday, February 27, 2021. Data were extracted from 3108 counties in the 48 contiguous US states and were completely anonymous. This data set was used to compute the daily confirmed deaths related to COVID-19 by county and contained the sole data used to inform the time series cluster analysis.

To develop the explanatory model describing the clusters, the following additional variables were gathered: region, governor's party affiliation, government response. the Centers for Disease Control and Prevention’s (CDC) social vulnerability index (SVI), and population density.

Region

The CDC produces a 10-region Framework for Chronic Disease Prevention and Health Promotion [8]. Figure 2 shows the 10 regions used in our explanatory model. The CDC's National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP) developed these regions to promote consistency in technical assistance and communications for chronic disease prevention [8].

Figure 2. The 10 CDC regions. CDC: Centers for Disease Control and Prevention.



Governor's Party Affiliation

The political party affiliation of each US state governor (within the 48 contiguous US states) at the start of the pandemic (March 2020) was determined. Since the District of Columbia does not have a governor, the political party of the mayor (Democrat) was used. The party affiliation of the governor was used as this affects the political actions and policies taken, often in the form of executive orders from the governor, during the pandemic [4].

Government Response

The overall government response index (at the US state level) from the Blavatnik School of Government [9] was downloaded on March 16, 2021. The index considers containment and closure indicators, such as school and workplace closings; economic response, such as income support and debt relief; and health systems, such as testing policies, contact tracing, and investment in vaccines. Higher values of the government response index indicate a stronger government response related to the pandemic. This index changed over the time of the study period. To capture the index over the majority of the study period, we summarized the index using the median value over the study period. Details of the methodology used to compute the index can be found at Oxford University COVID-19 Tracker Github [10].

The Social Vulnerability Index

The CDC’s SVI is computed by the CDC's Agency for Toxic and Disease Registry's Geospatial Research, Analysis, and Services Program [11]. The SVI provides the relative vulnerability of each US county based on US Census data and is ranked on 15 social factors, including unemployment, minority status, and disability. Note that the SVI data from the CDC returned results for 3107 counties, with no data on Rio Arriba County, New Mexico, and hence this county was excluded from our explanatory analysis. The SVI data were grouped into the following 4 themes:

- SVI theme 1: socioeconomic
- SVI theme 2: household composition and disability
- SVI theme 3: minority status and language
- SVI theme 4: housing and transportation

Our study included each of the 4 SVI themes. To construct the SVI for each theme, the percentile rank for each variable across the counties was computed. These were summed across the themes and then ranked within each domain. The SVIs ranged from 0 to 1, with higher values of SVIs for a particular theme indicating a higher level of social vulnerability. For more details on the SVI, see Flanagan et al [12].

Population Density

The population density in each county was computed based on the land area in square miles and the 2014-2018 American Community Survey (ACS) population estimates in each county. Both land area and population estimate variables were obtained from the CDC’s SVI 2018 data set [11]. Due to right-skewness in this variable, the natural logarithm of population density was used in the analysis.

Time Series Clustering

Time series cluster analysis was based solely on the daily confirmed deaths related to COVID-19 by county. The goal was to separate counties into groups (clusters) that show similar time series patterns. There are 3 important decisions that affect the cluster solution: (1) the scaling of the data, (2) the measure of distance between the clusters, and (3) the clustering algorithm. Liao [13] gives an overview of time series clustering methods.

For this study, the daily confirmed deaths related to COVID-19 by county were smoothed using a 7-day moving average to account for weekly patterns due to reporting. Moreover, the 7-day moving averages were rescaled so that all values fell between 0 and 1 to focus on the pattern of the progression of the deaths rather than the magnitude of the death counts. The magnitude of the death counts in each county depends on many factors, such as county size, population density, and region. The scaled 7-day moving average for county i at time t is

where $MA7_{i,t}$ is the 7-day moving average of deaths related to COVID-19 for county i at time t. The maximum in the denominator is taken over all time, $0 \leq t \leq T$. The outer maximum function in Equation (1) is used to account for reporting adjustments that occur with negative death counts on some days.

For illustration, suppose that county *i* recorded deaths only on days 7, 8, and 9, when, respectively, 7, 21, and 14 deaths occurred. On all other days, no deaths were recorded. For clarity, this sequence of death counts, the calculations of the 7-day moving averages ($MA7_{i,t}$), and the scaled moving averages () for the first 17 days are shown in Table 1.

This method of scaling the 7-day moving averages ensured that we evaluated the shape of the death profile for each county across time.

Many metrics can be used to measure the distance between time series, including Euclidean distance, dynamic time warping [14], and the Pearson correlation coefficient. An elastic measure, such as dynamic time warping, is commonly used with time series clustering [13] because it aligns or *warps* the time series so that the distance between them is minimized. Elastic measures such as this do not preserve the timing of the outbreak and deaths in a meaningful way. For this reason, we used the Euclidean distance to measure the distance between the time series clusters. In our case, the Euclidean distance between 2 death profiles of length *T* was

There are numerous clustering algorithms that have been suggested for time series clustering [13,15]. We used *k*-means clustering for this analysis. A heuristic-based method of clustering, *k*-means clustering partitions *n* objects into $k \leq n$ mutually exclusive clusters and each cluster is represented by the most centrally located object in the cluster. One limitation of the *k*-means clustering approach is that the number of clusters must be determined a priori in order to obtain a solution. It is common practice in exploratory research to evaluate cluster solutions for several sizes of *k* and select the *best* based on measures of cluster validity or homogeneity [16]. The R package *NbClust* [17] can be used to compute up to 30 cluster validity indices for cluster solutions of several sizes, *k*. This approach provides a systematic, data-driven method for selecting the optimal number of clusters in a data set without capitalizing on a single validity measure. For this analysis, *k*-means clustering was used to find the cluster solutions and the *NBClust* package was used to determine the optimal number of clusters to retain.

Table 1. Example calculation of the scaled 7-day moving averages ().

Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Deaths	0	0	0	0	0	0	7	21	14	0	0	0	0	0	0	0	0
$MA7_{i,t}$	N/A[a]	N/A	N/A	N/A	N/A	N/A	1	4	6	6	6	6	6	5	2	0	0
	N/A	N/A	N/A	N/A	N/A	N/A	1/6	4/6	1	1	1	1	1	5/6	2/6	0	0

[a] N/A: not applicable.

Explanatory Modeling

The time series clustering method described before resulted in mutually exclusive clusters of time series profiles containing counties with similar patterns in the daily deaths related to COVID-19. To further validate the cluster solution and to explain the differences in the progression of daily deaths across the counties, a multinomial regression analysis [18] was fit using the explanatory variables described in the Data section. The *multinom* function from the R package *nnet* [19] was used for this analysis.

Model performance was evaluated in terms of the ability to meaningfully interpret the model coefficients and by evaluating the in-sample classification performance. Specifically, the model predicted cluster was compared to the cluster as determined by the time series cluster solution for each county. The in-sample classification performance was measured by sensitivity, specificity, and balanced accuracy:

Sensitivity = TP/(TP + FN),

where TP and FN are the number of true-positive and false-negative predictions, respectively,

Specificity = TN/(TN + FP),

where TN and FP are the number of true-negative and false-positive prediction, respectively, and

Balanced accuracy = (Sensitivity + Specificity)/2.

Results

Number of Distinct Clusters

To address our first research question regarding the number of distinct clusters, we used time series cluster analysis of the scaled 7-day moving average of daily deaths due to COVID-19. Figure 3 shows the scaled time series of the daily deaths due to COVID-19 for 9 randomly selected contiguous counties in the United States during the study period. We evaluated 2≤*k*≤51 time series cluster solutions using 23 cluster validity indices [17]. Of the 23 validity indices, 7 (30.4%) preferred a 4-cluster solution. The second-most preferred cluster solution was a 2-cluster solution, which was preferred by 6 (26.1%) of the 23 indices. Using a majority rule of the validity indices, we retained a 4-cluster solution.

Figure 4 shows the geographic distribution of the 4-cluster solution across the United States. Cluster C1 is primarily concentrated in the Upper Midwest and mountain states, as well as in Ohio, Central Kentucky, Virginia, and Maine. Cluster C2 is located along the coast in the Northeast and in some of the larger US cities, such as Chicago, Detroit, Seattle, and New Orleans. Cluster C3 is scattered throughout much of the United States but particularly in Missouri, Illinois, and the states surrounding the Great Lakes. Cluster C4 occurs across the United States but shows concentrations in California, East Texas, the Southwest, and the Southeast. For an interactive color version of this map, please see Section 3.3.3 in Megahed et al [20].

Figure 5 shows the 25th, 50th, and 75th percentiles of the time series profiles for the counties within each cluster and provides insight into the shape of the cluster patterns. From Figure 5, it is clear that counties in cluster C1 experienced a low number of deaths due to COVID-19 throughout the study period. Counties clustering in C2 experienced early death counts beginning in April 2020, but the death counts tapered off in early summer. These counties maintained low death counts throughout the late summer and early fall, until rising again in November 2020. In C3, counties experienced few COVID-19 deaths until October 2020, when they saw a rapid rise in deaths. The death counts in C3 began dropping in December 2020, which continued through March 2021. The fourth cluster, C4, showed a small increase in deaths in late summer, followed by a steady rise throughout the fall and a higher peak in early 2021.

Figure 3. Time series profiles of the scaled 7-day moving average of new COVID-19 deaths for 9 sample counties.

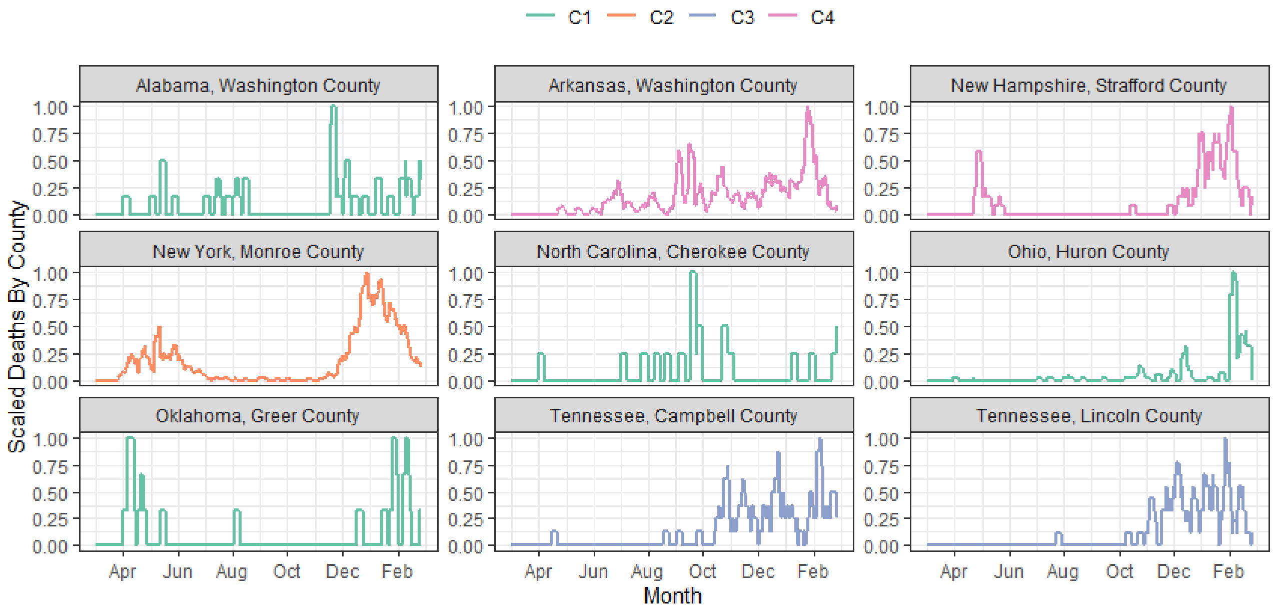


Figure 4. Map of 4 scaled time series profile clusters of COVID-19 deaths by county in contiguous US counties.

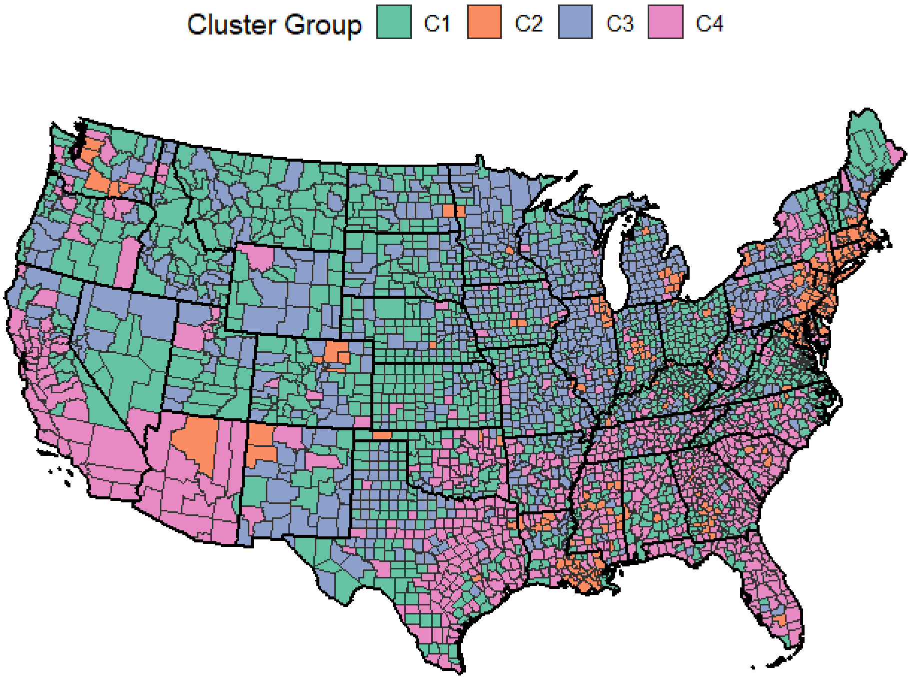
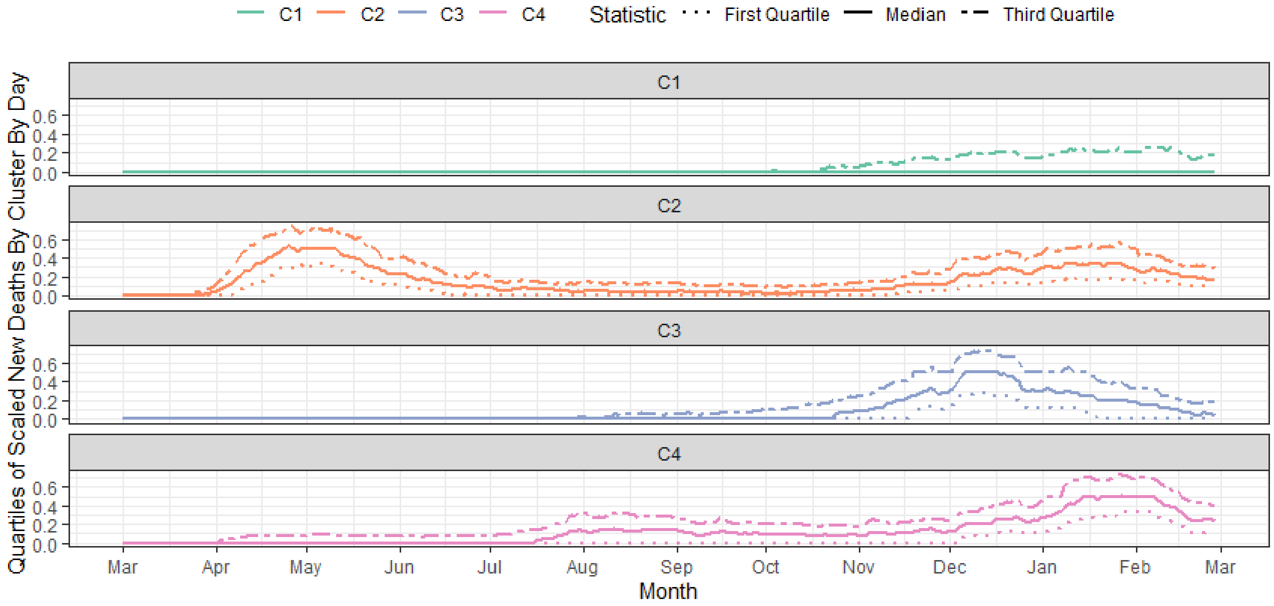


Figure 5. A summary plot, where the median scaled time series profile for each cluster is depicted using the solid bold line. The first and third quartiles are shown by dotted and 2-dash lines, respectively.



Explaining the Clusters

To address the second research question regarding factors that relate to the patterns of COVID-19–related deaths, we used an explanatory multinomial regression analysis to validate our cluster solution. Table 2 provides a summary of the explanatory study variables for each cluster.

Table 3 gives the coefficients from the multinomial logistic regression analysis. The dependent variable was cluster. The baseline category for the analysis was C1, the cluster of counties with few deaths related to COVID-19. The coefficients showed the linear change in the natural log of the odds ratio (OR) of a county classified in a corresponding cluster (eg, C2, C3, or C4) versus the baseline cluster (C1). From Table 3, it is clear that several geographic, political, government, and social vulnerability variables are associated with the patterns in COVID-19–related deaths.

Table 2. A summary of how the predictor variables were distributed per cluster. For each numeric variable, we report the mean (SD). For categorical variables, we report the distribution of each subcategory across the 4 clusters. The row summation of percentages for a subcategory may deviate slightly from 100% due to rounding errors.

Variables	C1 (N=1261)	C2 (N=226)	C3 (N=827)	C4[a] (N=794)
Continuous variables, mean (SD)				
Theme 1: socioeconomic	0.48 (0.30)	0.44 (0.31)	0.45 (0.27)	0.61 (0.26)
Theme 2: household composition and disability	0.50 (0.28)	0.37 (0.31)	0.49 (0.28)	0.56 (0.29)
Theme 3: minority status and language	0.41 (0.28)	0.71 (0.22)	0.43 (0.27)	0.65 (0.24)
Theme 4: housing and transportation	0.42 (0.29)	0.60 (0.28)	0.49 (0.26)	0.60 (0.27)
Log(population density)	3.01 (1.71)	5.86 (1.81)	3.73 (1.31)	4.60 (1.29)
Government response index median	47.09 (8.45)	52.87 (9.13)	47.24 (8.25)	48.13 (7.65)
Categorical variables, n (%)				
Governor’s party (Democratic)	579 (45.9)	142 (62.8)	428 (51.8)	202 (25.4)
Governor’s party (Republican)	682 (54.1)	84 (37.2)	399 (48.2)	591 (74.4)
Region A	41 (3.3)	43 (19.0)	21 (2.5)	24 (3.0)
Region B	131 (10.4)	63 (27.9)	62 (7.5)	48 (6.0)
Region C	101 (8.0)	19 (8.4)	13 (1.6)	239 (30.1)
Region D	140 (11.1)	20 (8.8)	51 (6.2)	153 (19.3)
Region E	188 (14.9)	30 (13.3)	283 (34.2)	23 (2.9)
Region F	154 (12.2)	31 (13.7)	116 (14.0)	201 (25.3)
Region G	236 (18.7)	7 (3.1)	144 (17.4)	25 (3.1)
Region H	187 (14.8)	7 (3.1)	88 (10.6)	9 (1.1)
Region I	22 (1.7)	1 (0.4)	14 (1.7)	53 (6.7)
Region J	61 (4.8)	5 (2.2)	35 (4.2)	18 (2.3)

[a] Rio Arriba County, New Mexico, assigned to C4 based on the time series clustering was not modeled using the multinomial logistic regression, since we could not obtain values for its predictor variables. Hence, the reported mean (SDs) and n (%) for C4 exclude this county.

Table 3. Results of multinomial logistic regression for clusters C2, C3, and C4. We used C1 as the reference cluster since it contained the largest number of counties.

Variables	C2		C3		C4	
	β (SE)	OR[a] (95% CI)	β (SE)	OR (95% CI)	β (SE)	OR (95% CI)
Theme 1: socioeconomic	0.419 (0.592)	1.52 (0.48-4.85)	–0.356 (0.286)	0.70 (0.40-1.23)	–0.018 (0.376)	0.98 (0.47-2.05)
Theme 2: household composition and disability	–0.245 (0.432)	0.78 (0.34-1.83)	0.392 (0.223)	1.48 (0.96-2.29)	0.638 (0.267)	1.89 (1.12-3.19)
Theme 3: minority status and language	3.661 (0.469)	38.90 (15.51-97.54)	0.004 (0.222)	1.00 (0.65-1.55)	1.162 (0.268)	3.20 (1.89-5.40)
Theme 4: housing and transportation	0.557 (0.428)	1.75 (0.75-4.04)	1.086 (0.227)	2.96 (1.90-4.62)	0.599 (0.270)	1.82 (1.07-3.09)
Log(population density)	1.009 (0.078)	2.74 (2.35-3.20)	0.417 (0.043)	1.52 (1.39-1.65)	0.959 (0.057)	2.61 (2.33-2.92)
Governor’s party (Republican)	–0.101 (0.233)	0.90 (0.57-1.43)	–0.323 (0.122)	0.72 (0.57-0.92)	1.093 (0.173)	2.98 (2.13-4.19)
Region B	–1.879 (0.464)	0.15 (0.06-0.38)	–0.509 (0.354)	0.60 (0.30-1.20)	–1.108 (0.395)	0.33 (0.15-0.72)
Region C	–2.621 (0.496)	0.07 (0.03-0.19)	–1.673 (0.437)	0.19 (0.08-0.44)	0.502 (0.376)	1.65 (0.79-3.45)
Region D	–1.717 (0.537)	0.18 (0.06-0.51)	–0.574 (0.369)	0.56 (0.27-1.16)	0.242 (0.401)	1.27 (0.58-2.80)
Region E	–1.941 (0.461)	0.14 (0.06-0.35)	0.884 (0.324)	2.42 (1.28-4.57)	–1.925 (0.403)	0.15 (0.07-0.32)
Region F	–1.520 (0.522)	0.22 (0.08-0.61)	0.629 (0.367)	1.88 (0.91-3.85)	0.814 (0.444)	2.26 (0.95-5.39)
Region G	–2.886 (0.647)	0.06 (0.02-0.20)	0.363 (0.361)	1.44 (0.71-2.92)	–1.536 (0.444)	0.22 (0.09-0.51)
Region H	–2.221 (0.681)	0.11 (0.03-0.41)	0.374 (0.396)	1.45 (0.67-3.16)	–1.329 (0.570)	0.26 (0.09-0.81)
Region I	–3.509 (1.117)	0.03 (0.00-0.27)	0.657 (0.479)	1.93 (0.75-4.93)	2.139 (0.476)	8.49 (3.34-21.58)
Region J	–2.527 (0.666)	0.08 (0.02-0.29)	0.228 (0.396)	1.26 (0.58-2.73)	–0.213 (0.480)	0.81 (0.32-2.07)
Government response	–0.028 (0.018)	0.97 (0.94-1.01)	–0.030 (0.009)	0.97 (0.95-0.99)	–0.020 (0.012)	0.98 (0.96-1.00)
Constant	–5.171 (1.292)	0.01 (0.00-0.07)	–1.308 (0.684)	0.35 (0.09-1.35)	–5.115 (0.934)	0.01 (0.00-0.04)

[a] OR: odds ratio.

We found that the clusters can be roughly described as follows:

- C1: low death rates throughout much of the pandemic; found mostly in Upper Midwest and mountain states
- C2: high death rates in spring 2020, with another spike in December 2020/January 2021; found mostly in the northeast and other large cities
- C3: low death rates until fall 2020, followed by a peak in December 2020; spread throughout the United States with concentrations in Central Midwest and Great Lakes
- C4: steady death rates from late summer through December 2020, followed by a peak in January; spread throughout the United States with concentrations in California, the Southwest, and the Southeast

“SVI theme 3: minority status and language” was significantly associated with clustering in C2 versus C1, yielding an OR of 38.90. Counties with high levels of SVI theme 3 were strongly associated with membership in C2 compared to C1. All CDC regions (B-J) showed a significant, negative association with C2 versus C1, indicating that being located outside region A (the Northeast, baseline category for region) is associated with lower odds of clustering in C2 versus C1. This is consistent with our initial finding from the map in Figure 4, which showed that the counties in C2 are primarily located in the Northeast.

The variable with the strongest positive association to C3, relative to C1, was “SVI theme 4: housing and transportation.” Population density was also significant and positively related to C3. The governor's party was significant and negatively associated with C3, indicating that counties in states with Republican governors are associated with lower odds of clustering in C3 than in C1. The government response was also significant and negatively related to membership in C3, but the effect was small. Among the regions, the coefficient for region C (North Carolina, South Carolina, Georgia, and Florida) was significant and negative; thus, counties in these states are associated with lower odds of being classified in C3 than in C1. In contrast, the coefficient for region E was significant and positive, which suggests that counties in Minnesota, Wisconsin, Illinois, Indiana, Michigan, and Ohio are associated with higher odds of clustering in C3.

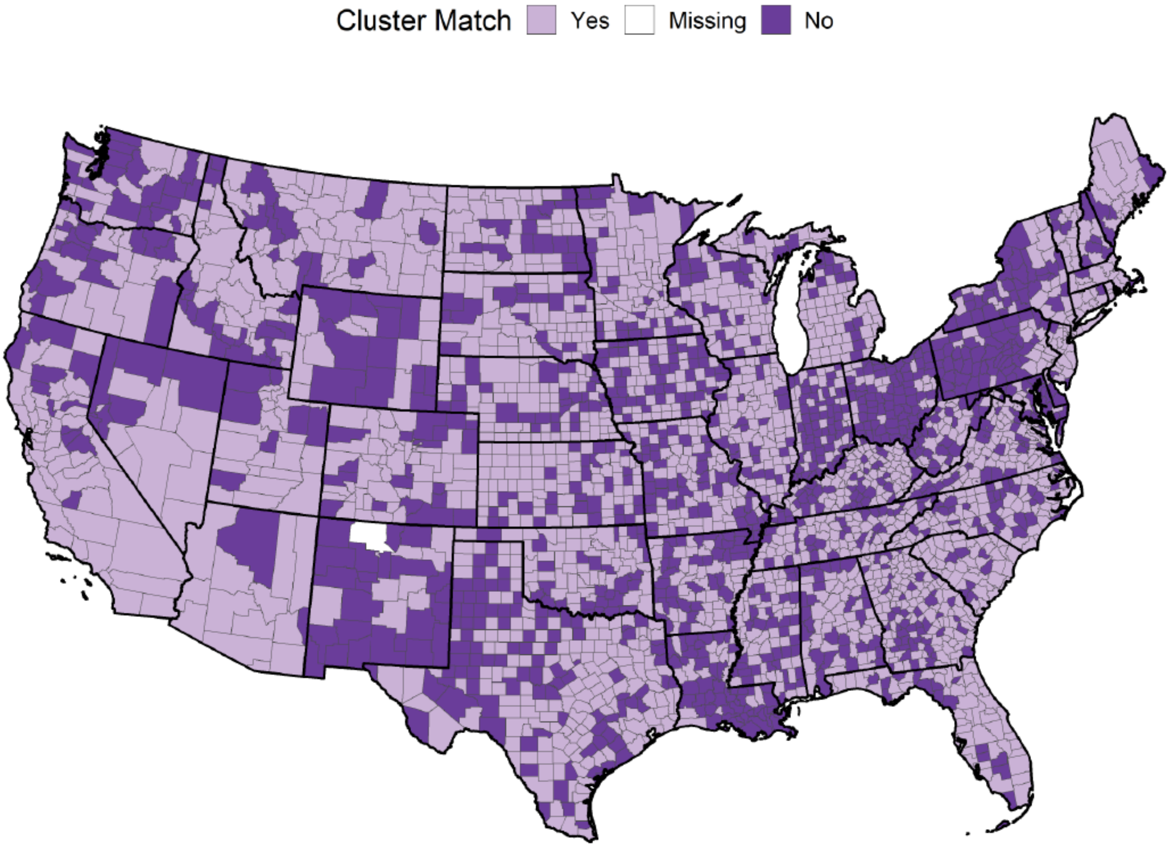
“SVI theme 1: socioeconomic” was not significant for membership in any of clusters C2-C4; however, 3 of the SVIs (household composition and disability, minority status and language, and housing and transportation) were significant and positively associated with membership in C4. In addition, counties located in states with Republican governors were also associated with higher odds of classification in C4 relative to C1. Among the CDC regions, regions I (California, Nevada, and Arizona) and F (New Mexico, Texas, Oklahoma, and Louisiana) had positive coefficients. Regions B, E, G, and H had significantly negative coefficients. The logarithm of population density was also a significant predictor for classification in C2, C3, and C4, relative to C1, which indicates that a low population density is associated with clustering in C1.

Overall, the multinomial regression model correctly classified 1904 (61.25%) of the 3108 counties into 1 of 4 clusters. Table 4 gives the in-sample predictive performance of the multinomial regression model broken down by cluster. The balanced accuracy was similar for all 4 clusters, ranging from 0.63 to 0.80. A more nuanced view of the performance can be seen from sensitivity and specificity. The model performed well in correctly classifying counties in cluster C4 (sensitivity=0.74), which shows a sustained emergence in deaths beginning in late summer 2020. The model also performed well in classifying counties in cluster C1 (sensitivity=0.71), counties with few deaths. However, it had only moderate ability to correctly classify counties into clusters C2 and C3 (sensitivity=0.42 and 0.39, respectively). Note that the sensitivity performance for clusters C2 and C3 exceeded the expected sensitivity of 0.25 that would be obtained from random allocation among 4 classes in a balanced or imbalanced multiclass classification problem (see Megahed et al [21] for more details). In terms of specificity, the model performed well at identifying which counties are not in clusters C1-C4, with specificity values ranging from 0.71 to 0.98. Figure 6 shows the distribution of the accuracy of the multinomial logistic model in predicting cluster membership. Counties that were correctly predicted from the model are indicated in a light color, while those that were incorrectly predicted are indicated in a dark color. The model provides some insight into the patterns across the United States, but additional data are needed to more accurately classify counties in terms of the pattern of death rates due to COVID-19. For an interactive version of this map, please see Section 4.2.4 in Megahed et al [20].

Table 4. The predictive performance of the multinomial regression model for each cluster.

Cluster	Balanced accuracy	Sensitivity	Specificity
C1	0.71	0.71	0.71
C2	0.70	0.42	0.98
C3	0.63	0.39	0.88
C4	0.80	0.74	0.86

Figure 6. Map of the prediction accuracy of the multinomial logistic model describing the time series cluster solution. Counties in a light color (labeled “Yes”) were correctly classified by the model. Counties in a dark color (labeled “No”) were incorrectly classified. Rio Arriba County, New Mexico (in white), was not classified due to missing data.



Discussion

Principal Findings

This research provides a framework for understanding the pattern of COVID-19–related deaths across the United States. Using time series clustering with county-level data on the occurrence of COVID-19–related deaths, we observed 4 distinct patterns from March 1, 2020, to February 27, 2021. The second stage of our analysis revealed that these patterns can be partially explained by region as well as social and political predictors.

Our findings add to the literature on the relationship between COVID-19 outcomes and vulnerable populations [22-24]. The largest number of counties in the United States experienced few deaths during the study period (cluster C1). These counties were, on average, at or below the median of all measures of social vulnerability. With lower population densities, and spread throughout the United States, C1 counties served as our model baseline.

The county-level COVID-19 death data were extracted using the COVID19 R package [1], which extracted confirmed deaths from a GitHub repository [25]. The cross-sectional data set containing the predictors used in the multinomial regression was compiled by the authors from disparate sources and is available in Megahed [26]. R statistical software version 4.0.4 was used for all processing and analysis of data. A reproducible workflow of our analysis is made available using R Markdown and is hosted in Megahed et al [20], following the best practices of Jalali et al [27] in reporting and documenting analyses for COVID-19.

Cluster C3 (low death rates until fall 2020, peaking in December 2020) had the second largest number of counties. C3 counties are spread across much of the country but have concentrations in the Great Lakes and Central Midwest regions. Interestingly, few incidences of C3 occur in the Southeastern United States and along the eastern seaboard from Washington, DC, to Massachusetts. Like C1, counties in C3 had SVI measures below the median, on average. These counties experienced a single late wave in COVID-19 deaths beginning in late October 2020 that declined by the end of the study period. There were a few distinguishing features between counties being classified in C3 versus C1: a higher population density, Democratic state leadership, location outside the Southeast, location in the Great Lakes region, and higher vulnerability in the SVI housing and transportation theme. This index indicates a higher incidence of multiunit housing, mobile homes, crowding, lack of vehicles, or group living situations.

The 226 counties that are clustered in C2 (high death rates in spring 2020 and December 2020/January 2021) are mostly populous counties in the Northeast, Washington, southeast Louisiana (including New Orleans), and the Four Corners region of Arizona and New Mexico. C2 counties experienced an early outbreak of deaths, followed by a second wave beginning in November 2020 but few deaths in summer 2020. These counties showed a strong relationship with the SVI minority and language theme, indicating a large percentage of residents who are minority or nonnative English speakers.

Cluster C4 (steady death rates beginning late summer, peaking in January) is located throughout the United States, with concentrations in the Southeast and Southwest. The counties in C4 showed a steady incidence of deaths beginning in late summer 2020 that continued through the study period. C4 counties were, on average, above the median on all SVI themes, and 3 of the 4 themes were significant in classifying counties in C4 versus C1. Specifically, the themes related to household and disability, minority and language, and housing and transportation all showed a positive association with this sustained pattern of COVID-19–related deaths. The majority (n=591, 74.4%) of these counties are located in Republican-led states.

Limitations

The local patterns in COVID-19–related deaths suggest that local-level factors, including geographic, demographic, and social vulnerability characteristics, are related to adverse outcomes from COVID-19. There are several limitations to this research. These include the observational nature of the study, which was conducted as the pandemic continues to emerge. The retrospective, secondary use of data makes it impossible to infer causation from our model. Outbreaks and adverse outcomes changed over time as local and national governments adopted new policies and vaccines to react to the emerging pandemic. Further, the government response index is available only at the state level and is constant across all counties within a state. Using a state-level predictor to explain cluster membership at the county level could lead to an ecological fallacy.

Conclusion

Despite limitations, this exploratory study revealed new insights into the most severe outcome of the COVID-19 pandemic. The identification of 4 distinct patterns of death incidences in 3108 US counties provides evidence of the differences in the realization of severe outcomes from the pandemic. The United States is a demographically and politically diverse nation, and it is important to understand the differences in pandemic-related outcomes across communities. By examining the relationship between county-level predictors and membership in the 4 cluster patterns, we showed that there are important demographic, political, and socioeconomic differences related to death patterns across the United States.

Acknowledgments

Our data acquisition and computations were supported in part by the Ohio Supercomputer Center (Grant PZS1007).

Conflicts of Interest

None declared.

References

1. Guidotti E, Ardia D. COVID-19 data hub. JOSS 2020 Jul;5(51):2376. [doi: 10.21105/joss.02376]
2. Ahmed R, Williamson M, Hamid MA, Ashraf N. United States county-level COVID-19 death rates and case fatality rates vary by region and urban status. Healthcare (Basel) 2020 Sep 09;8(3):330. [doi: 10.3390/healthcare8030330] [Medline: 32917009]
3. Gollwitzer A, Martel C, Brady WJ, Pärnamets P, Freedman IG, Knowles ED, et al. Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. Nat Hum Behav 2020 Nov;4(11):1186-1197. [doi: 10.1038/s41562-020-00977-7] [Medline: 33139897]
4. Baccini L, Brodeur A. Explaining governors' response to the COVID-19 pandemic in the United States. Am Polit Res 2020 Dec 01;49(2):215-220. [doi: 10.1177/1532673x20973453]
5. Le N, Le A, Brooks J, Khetpal S, Liauw D, Izurieta R, et al. Impact of Government-Imposed Social Distancing Measures on COVID-19 Morbidity and Mortality around the World. URL: http://www.who.int/bulletin/online_first/20-262659.pdf [accessed 2022-06-07]
6. Megahed F, Allison JL, Rigdon S. A retrospective cluster analysis of COVID-19 cases by county. bioRxiv Preprint posted online November 12, 2020. [doi: 10.1101/2020.11.12.379537]
7. Covid-19: The global crisis — in Data. URL: https://ig.ft.com/coronavirus-global-data/ [accessed 2022-06-07]
8. Centers for Disease Control and Prevention (CDC). National Center for Chronic Disease Prevention and Health Promotion Regions. URL: https://www.cdc.gov/coordinatedchronic/docs/nccdphp-regions-map.pdf [accessed 2022-06-07]
9. Hale T, Atav T, Hallas L, Kira B, Phillips T, Petherick A, et al. Variation in US states responses to COVID-19. BSG Working Paper Series. URL: https://www.bsg.ox.ac.uk/sites/default/files/2020-08/BSG-WP-2020-034.pdf [accessed 2022-06-07]
10. Oxford COVID-19 Government Tracker Methodology. URL: https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/index_methodology.md [accessed 2022-06-07]
11. Agency for Toxic Substances and Disease Registry. CDC Social Vulnerability Index. URL: https://www.atsdr.cdc.gov/placeandhealth/svi/index.html [accessed 2022-06-07]
12. Flanagan B, Gregory E, Hallisey E. A Social Vulnerability Index for Disaster Management. URL: https://www.degruyter.com/document/doi/10.2202/1547-7355.1792/html [accessed 2022-06-07]
13. Warren Liao T. Clustering of time series data—a survey. Pattern Recognit 2005 Nov;38(11):1857-1874. [doi: 10.1016/j.patcog.2005.01.025]
14. Bellman R. Adaptive Control Processes: A Guided Tour, Volume 2045 in the series Princeton Legacy Library. Princeton, NJ: Princeton University Press; 2015.
15. Aghabozorgi S, Seyed Shirkhorshidi A, Ying Wah T. Time-series clustering – a decade review. Inf Syst 2015 Oct;53:16-38. [doi: 10.1016/j.is.2015.04.007]
16. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. J Stat Softw 2014;61(6):1-36. [doi: 10.18637/jss.v061.i06]
17. Charrad M, Ghazzali N, Boiteau V, Niknafs A. Package 'NbClust'. URL: https://cran.r-project.org/web/packages/NbClust/NbClust.pdf [accessed 2022-06-07]
18. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression, 3rd Ed. Hoboken, NJ: John Wiley & Sons; 2013.
19. Ripley B, Venables W. Package 'nnet'. URL: https://cran.r-project.org/web/packages/nnet/nnet.pdf [accessed 2022-06-07]
20. Megahed FM, Jones-Farmer A, Rigdon S. A Reproducible GitHub Page for a Two-Stage Modeling Framework for Analyzing COVID-19 Deaths by County. URL: https://fmegahed.github.io/covid_deaths.html [accessed 2022-06-07]
21. Megahed FM, Jones-Farmer A, Rigdon S. A Numerical Study to Examine the Obtained Sensitivity for Arbitrary and Proportional Guessing Scenarios for a 4-Class Classification Problem. URL: https://fmegahed.github.io/covid_deaths.html [accessed 2022-06-07]
22. Chen JT, Krieger N. Revealing the unequal burden of COVID-19 by income, race/ethnicity, and household crowding: US county versus zip code analyses. J Public Health Manag Pract 2021;27 Suppl 1, COVID-19 and Public Health: Looking Back, Moving Forward:S43-S56. [doi: 10.1097/PHH.0000000000001263] [Medline: 32956299]
23. Stokes AC, Lundberg DJ, Elo IT, Hempstead K, Bor J, Preston SH. COVID-19 and excess mortality in the United States: a county-level analysis. PLoS Med 2021 May;18(5):e1003571 [FREE Full text] [doi: 10.1371/journal.pmed.1003571] [Medline: 34014945]
24. Khanijahani A. Racial, ethnic, and socioeconomic disparities in confirmed COVID-19 cases and deaths in the United States: a county-level analysis as of November 2020. Ethn Health 2021 Jan;26(1):22-35. [doi: 10.1080/13557858.2020.1853067] [Medline: 33334160]
25. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2022-06-07]

26. Megahed FM. Covid-19 Deaths. URL: https://github.com/fmegahed/covid19-deaths/tree/master/Data/Output [accessed 2022-06-07]
27. Jalali MS, DiGennaro C, Sridhar D. Transparency assessment of COVID-19 models. Lancet Global Health 2020 Dec;8(12):e1459-e1460 [FREE Full text] [doi: 10.1016/S2214-109X(20)30447-2] [Medline: 33125915]

Abbreviations

CDC: Centers for Disease Control and Prevention
OR: odds ratio
SVI: social vulnerability index

Edited by T Sanchez, A Mavragani; submitted 28.07.21; peer-reviewed by A Couture, D Snider; comments to author 22.01.22; revised version received 19.02.22; accepted 26.04.22; published 19.07.22.

Please cite as:
Megahed FM, Jones-Farmer LA, Ma Y, Rigdon SE
Explaining the Varying Patterns of COVID-19 Deaths Across the United States: 2-Stage Time Series Clustering Framework
JMIR Public Health Surveill 2022;8(7):e32164
URL: https://publichealth.jmir.org/2022/7/e32164
doi:10.2196/32164
PMID:35476722

Review

Experiences and Challenges of Implementing Universal Health Coverage With China's National Basic Public Health Service Program: Literature Review, Regression Analysis, and Insider Interviews

Guixia Fang[1*], MD; Diling Yang[1*], BA; Li Wang[1], PhD; Zhihao Wang[2,3], MA; Yuanyuan Liang[4], MA; Jinxia Yang[1], PhD

[1]School of Health Service Management, Anhui Medical University, Hefei, China
[2]The First Affiliated Hospital, University of Science and Technology of China, Hefei, China
[3]Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China
[4]Anhui Provincial Center for Disease Control and Prevention, Hefei, China
[*]these authors contributed equally

Corresponding Author:
Jinxia Yang, PhD
School of Health Service Management
Anhui Medical University
81 Meishan Road
Hefei, 230032
China
Phone: 86 15345699071
Email: yangjinxiaamu@163.com

Abstract

Background: Public health service is an important component and pathway to achieve universal health coverage (UHC), a major direction goal of many countries. China's National Basic Public Health Service Program (the Program) is highly consistent with this direction.

Objective: The aim of this study was to analyze the key experience and challenges of the Program so as to present China's approach to UHC, help other countries understand and learn from China's experience, and promote UHC across the world.

Methods: A literature review was performed across five main electronic databases and other sources. Some data were obtained from the Department of Primary Health, National Health Commission, China. Data obtained included the financing share of the national/provincial/prefectural government among the total investment of the program in 32 provinces in 2016, their respective per capita funding levels, and some indicators related to program implementation from 2009 to 2016. The Joinpoint regression model was adopted to test the time trend of changes in program implementation indicators. Face-to-face individual interviews and group discussions were conducted with 48 key insiders.

Results: The program provided full life cycle service to the whole population with an equitable and affordable financing system, enhanced the capability and quality of the health workforce, and facilitated integration of the public health service delivery system. Meanwhile, there were also some shortcomings, including lack of selection and an exit mechanism of service items, inadequate system integration, shortage of qualified professionals, limited role played by actors outside the health sector, and a large gap between the subsidy standard and the actual service cost. The Joinpoint regression analysis demonstrated that 13 indicators related to program implementation showed a significant upward trend ($P<.05$) from 2009 to 2016, with average annual percent change values above 10% for 6 indicators and below 6% for 7 indicators. Three indicators (coverage of health records, electronic health records, and health management among the elderly) rose rapidly with annual percent change values above 30% between 2009 and 2011, but rose slowly or remained stable between 2011 and 2016. In 2016, the subsidy standard per capita in the eastern, central, and western regions was equivalent to US $7.43, $7.15, and $6.57, respectively, of which the national-level subsidy accounted for 25.50%, 60.57%, and 79.52%, respectively.

Conclusions: The Program has made a significant contribution to China's efforts in achieving UHC. The Program focuses on a key population and provides full life cycle services for the whole population. The financing system completely supported by the government makes the services more equitable and affordable. However, there are a few challenges to implementing the Program in China, especially to increase the public investment, optimize service items, enhance quality of the services, and evaluate the health outcomes.

(JMIR Public Health Surveill 2022;8(7):e31289) doi:10.2196/31289

KEYWORDS

universal health coverage; basic public health service; China; experience; challenge

Introduction

Countries across the world are making great efforts to achieve universal health coverage (UHC) [1]. In April 2018, at the 70th anniversary of its founding, the World Health Organization (WHO) announced that the tagline for the 69th World Health Day was “Universal health coverage: everyone, everywhere.” To this end, public health is indispensable [2]. The Declaration of Alma-Ata adopted in 1978 defined primary care as essential health care that is universally accessible to all individuals and families in a community [3], highlighting the significance of primary care for UHC [4]. In 2018, the Declaration of Astana reiterated the importance of health promotion and disease prevention, and the role of primary care as a bridge between UHC and public health service in the current context [5]. In China, as a major component of primary care, public health service once made a remarkable contribution to reducing infectious diseases and maternal and pediatric diseases, and promoted the development of primary care globally [6]. China made a commitment to prioritizing UHC in its recent strategic agenda [7], and aligned public health service delivery with efforts toward UHC by the National Basic Public Health Service Program (hereafter referred to as “the Program”) was launched in 2009 [8].

According to the WHO, five key elements are needed to achieve UHC: (1) an efficient, well-run health system that can meet major health needs by providing people-centered health care services to all, particularly pregnant women, children, and people living with HIV/AIDS, tuberculosis, malaria, or noncommunicable diseases (NCDs); (2) a strong health financing system ensuring equitable access to affordable health care services; (3) access to essential medicines, equipment, and technologies for diagnosis and treatment; (4) well-trained and motivated health workers; and (5) recognition that all sectors, inter alia education, transportation, agriculture, and urban construction, have their own roles to play for the health of the population [9]. The Program provides basic public health services to all residents free of charge to address major health challenges among urban and rural citizens, especially children, pregnant women, the elderly, and those with NCDs. The existing service package covers 55 items in 14 categories (see below for details). It is fair to say that the Program has made a significant contribution to China’s efforts toward UHC.

First, the Program targets issues that are highly consistent with the UHC concept. The former covers many priorities of the latter, including vaccination, reporting of and response to communicable diseases and emergency public health events, case management of hypertension and type 2 diabetes, maternal and child health, and health management of the elderly [10]. Second, these services are delivered to all Chinese citizens free of charge, which ensures their affordability and equity [11]. Third, the Program helps to build up the expertise of primary health workers, as they are trained and equipped with adequate technical skills to deliver the services with good quality and high efficiency [12]. The Program also requires the ability to communicate with the population in the catchment, process information, and identify problems. Such nontechnical skills of health workers are also enhanced, as they provide health education, health record management, and follow-up visits. As a result, a more versatile and qualified health workforce is taking shape. Fourth, implementation of the Program contributes to some other building blocks of UHC. For example, health record management requires development of an information system [13]; maternal and child health management and physical checkups require necessary equipment; and health promotion involves collaboration with the sports sector, urban planning institutions, civil affair authorities, and others [14].

Currently, national and international academic communities study UHC from different perspectives. Urquieta-Salomón and Villarreal [15] analyzed the coverage of health insurance and access to preventive medical care, and concluded that interventions on prenatal care and NCDs prevention were greatly inadequate. Atun [16] summarized the experience in achieving UHC by reviewing health insurance reform. Zheng et al [17] elaborated upon China’s health care service delivery system with a focus on the measures for UHC and their outcomes. Shi [18] studied how to achieve UHC in China in terms of concept, policy, and strategy, with little attention paid to prevention and public health. Shan et al [19] and Jiang and Ma [20] identified potential challenges in China’s health insurance on its journey to UHC. Chi et al [21] performed a gray relational analysis to evaluate to what extent UHC has been achieved in China, Germany, Britain, Cuba, and Thailand; the results showed that China was the least close to achieving UHC, and that equity in financing and benefits was a major challenge. Liang and Langenbrunner [22] analyzed the development path of UHC in China from an external perspective, pointing out that there is still a long way to go before China can truly achieve UHC, and put forward suggestions toward achieving this goal, including strengthening disease risk protection, improving health service equity, and improving health system performance. The UHC Monitoring Report released by the WHO in 2019 pointed out that most countries are underinvesting in primary health care and recommended that countries increase spending on primary health care to increase UHC [23].

To date, there have been few studies investigating the relationship between the Program and UHC. However, implementing the Program has become an important pathway and means for China to achieve UHC. Thus, it is necessary and important to analyze its key experience and challenges so as to present China's approach to UHC, help other countries understand and learn from this experience, and ultimately promote UHC worldwide.

Methods

Literature Search

Both the English and Chinese literature was searched on the PubMed, Web of Science, CNKI (China National Knowledge Infrastructure), Wanfang Data, and CQVIP databases with the key words "universal health coverage" and "basic public health service." Related policy documents and data were also collected from the websites of the National Health Commission (NHC), WHO, and World Bank Database, among others. After eliminating duplicate records, the titles and abstracts of articles with close relevance were thoroughly reviewed by subrelated topics such as "basic public health service projects," "project financing system," "health human resource allocation," and "integration status of public health service system," and articles with high relevance to these topics were reviewed in full. At the same time, the third edition of the National Basic Public Health Service Specification published on the official website, along with the policy documents, statistical yearbooks, and service guidelines issued by China from 2009 to 2020 to coordinate the implementation of basic public health service projects were collected. The differences in the types, projects, and contents of the services in the three versions of the National Basic Public Health Service Specification were obtained, and the subsidy standards for per capita basic public health services in each year were obtained from the policy documents. These data played a complementary and cross-verification role with the subsequent quantitative data analysis and qualitative interview data, which was conducive to synthesizing all aspects of the information and for performing in-depth exploration and analysis. The specific search strategies in each database are shown in Table 1.

Table 1. Literature review search strategies.

Database	Search strings
PubMed	(1) "universal health coverage"[Title/Abstract] AND ("China"[MeSH[a] Terms] OR "China"[All Fields] OR "China s"[All Fields] OR "Chinas"[All Fields]); (2) "basic public health service"[Title/Abstract]
Web of Science	(1) universal health coverage (subject) and China (subject); (2) basic public health service and China (theme)
CNKI[b]	(1) (Theme: Universal Health Coverage (precise)) AND (Full text: China (precise)); (2) (Theme: "Basic Public Health Services" (Precise)) AND (Theme: Experience + Problem + Challenge (Precise))
CQVIP	(1) Title or keyword=Universal Health Coverage AND any field=China; (2) Title or keyword=Basic public health service AND title or keyword=Experience OR problem OR challenge
Wanfang	(1) Title or keyword: ("Universal Health Coverage") and All: (China); (2) Title or keyword: ("essential public health services") and Title or keyword: (Experience or problem or challenge)
Policy document	Basic public health service

[a] MeSH: Medical Subject Heading.

[b] CNKI: China National Knowledge Infrastructure.

Statistical Analysis

Entrusted by the Department of Primary Health, NHC, we started to evaluate the results of the Program in 2018 and obtained some data, including the financing from the national government as a share of the total investment in the Program in 32 provinces in 2016, their respective per capita funding levels, and some indicators of the Program from 2009 to 2016, such as coverage of health records, coverage of electronic health records (EHRs), utilization rate of health records, amount of printed health education materials distributed per 10,000 people, number of participants of health education activities per 10,000 people, share of children with a vaccination record book, share of children receiving standard Expanded Program on Immunization (EPI) vaccination, rate of newborn home visits, coverage of health management among children aged 0-6 years, rate of registration in early pregnancy, coverage of postnatal visits, coverage of health management among the elderly, percentage of patients with hypertension under standardized management, percentage of patients with type 2 diabetes under standardized management, and coverage of health management by traditional Chinese medicine (TCM). The data were analyzed in Microsoft Office Excel 2010 for Windows (Microsoft Corp). The data were initially collected by the provincial health commission and then reported to the NHC (as China's top-level basic health management authority) before finally reaching us. With an official check by the above-mentioned authorities, the accuracy and reliability of all the data are highly guaranteed.

To test the changes of time trend in the basic public health service project, the Joinpoint regression program version 4.9.0.0 (National Cancer Institute) was adopted to analyze and determine the year of significant changes. The Joinpoint regression model was adopted to evaluate the average annual percentage change (AAPC) and annual percentage change (APC) with the corresponding 95% CIs in each time trend segment.

Interviews With Key Insiders

The interviewees included policymakers, frontline workers, and other relevant personnel who have been engaged in China's public health work for many years. Face-to-face individual interviews and group discussions were conducted. The 48 interviewees included one representative from each of the three national-level institutions, namely the Department of Primary

Health of the NHC, National Center for Disease Control (CDC), and the National Disease Prevention and Control Bureau of the NHC; three leaders and three staff from the Primary Health Division of the provincial-level health commission who were responsible for the Program; three from the provincial-level CDC; three county hospital presidents and three public health workers in county hospitals; three township health center (THC) directors and three public health workers in THCs; nine village doctors; three directors of community health care centers (CHCs), which are in urban areas, and three public health workers in CHCs; and nine doctors from CHC stations. Provincial-level personnel and grassroots personnel came from Liaoning, Anhui, and Chongqing, representing China's eastern, central, and western regions, respectively. All of the interviewees have good knowledge about the practices related to UHC and basic public health service in China.

The full process was recorded, and transcripts were noted for the classification and analysis by topics.

The main interview questions were as follows:

1. Is the local basic public health service information platform connected with other systems?

2. What agencies/departments have you cooperated with or contacted while carrying out basic health service projects locally?

3. How well do you think residents are accepting basic public health services?

4. What role do you consider the implementing basic public services play to advance universal health coverage?

5. What problems do you think exist in the implementation of basic public health service projects?

Ethics Considerations

Verbal informed consent was received in all interviews. This study was approved by the Biomedical Ethics Committee of Anhui Medical University (code 20150082).

Results

Achievements of the Program

The Program Provides a Full Life Cycle Service to the Whole Population

After the Program was launched in 2009, the National Health and Family Planning Commission issued three editions of the National Basic Public Service Specifications in 2009, 2011, and 2017, respectively. The service package expanded from 41 items in 9 categories in 2009 to 55 items in 14 categories in 2017. Regardless of the specific changes to the service items, the Program has always covered the entire life cycle and the whole population, especially pregnant women, children, the elderly, and those with NCDs or tuberculosis (Table 2).

Table 2. Changes in the service categories among the three editions of the National Basic Public Service Specifications.

Edition (year)	Categories of the service	Changes
First edition (2009)	10 categories: (1) health record management for urban and rural residents; (2) health education; (3) health management of children aged 0-6 years; (4) maternal health management; (5) elderly health management; (6) vaccination; (7) reporting of and response to communicable diseases; (8) management of patients with hypertension; (9) management of patients with type 2 diabetes; and (10) management of patients with severe mental illness	Not applicable
Second edition (2011)	11 categories: (1) health record management for urban and rural residents; (2) health education; (3) vaccination; (4) health management of children aged 0-6 years; (5) maternal health management; (6) elderly health management; (7) management of patients with hypertension; (8) management of patients with type 2 diabetes; (9) management of patients with severe mental illness; (10) reporting of and response to communicable diseases and public health emergencies; and (11) health inspection and coordination	"Reporting of and response to communicable diseases" was changed to "reporting of and response to communicable diseases and public health emergencies," and "health inspection and coordination" was added
Third edition (2017)	14 categories: (1) health record management for urban and rural residents; (2) health education; (3) vaccination; (4) health management of children aged 0-6 years; (5) maternal health management; (6) elderly health management; (7) management of people with NCDs[a] (including patients with hypertension and patients with type 2 diabetes); (8) management of patients with severe mental illness; (9) management of patients with TB[b]; (10) health management by TCM[c]; (11) reporting of and response to communicable diseases and public health emergencies; (12) health inspection and coordination; (13) free contraceptives; and (14) health promotion	"Management of patients with hypertension" and "management of patients with type 2 diabetes" were merged into "management of people with NCDs"; new categories include "management of patients with TB," "free contraceptives," "health management by TCM," and "health promotion"

[a] NCD: noncommunicable disease.

[b] TB: tuberculosis.

[c] TCM: traditional Chinese medicine.

Some indicators showed that the coverage of the Program had gradually expanded to the whole population. In addition to the rising share of citizens with health records and higher utilization rate of those records, the share of children with a vaccination record book, share of children receiving standard EPI vaccination, and coverage of maternal and child health management were all above 90%. Coverage of elderly health management was approximately 70%. The share of patients with NCDs under management, along with the control rate of blood pressure and glucose, was on the rise. Since 2012, health management by TCM has developed rapidly. All of these facts indicated that the Program provided UHC services to hundreds of millions of people (Table 3).

The results of Joinpoint regression analysis are presented in Table 4. Thirteen indicators relating to the implementation of the program showed a significant upward trend ($P<.05$) from 2009 to 2016, including coverage of EHRs, coverage of health management among the elderly, and glucose control rate among the managed patients with type 2 diabetes, whereas there was no significant change in the share of children with a vaccination record book, share of children receiving standard EPI vaccination, and percentage of patients with hypertension under standardized management (all $P>.05$). The AAPC values of six indicators, including coverage of health records, coverage of EHRs, utilization rate of health records, coverage of health management among the elderly, coverage of health management by TCM among the elderly, and coverage of health management by TCM among children aged 0-36 months, were all above 10%, and those of the other seven indicators, including the rate of newborn home visits, coverage of health management among children aged 0-6 years, and rate of registration in early pregnancy, were all below 6%. The coverage of health records, EHRs, and health management among children aged 0-6 years rose rapidly between 2009 and 2011, but rose slowly between 2011 and 2016. The coverage of health management among the elderly increased significantly between 2009 and 2011, but remained stable between 2011 and 2016, and the glucose control rate among the managed patients with type 2 diabetes increased significantly between 2009 and 2013, but remained stable between 2013 and 2016.

Table 3. Representative indicators on implementation of the Program 2009-2016.[a]

Indicators	2009	2010	2011	2012	2013	2014	2015	2016
Coverage of health records, %	33.38	49.12	64.31	73.33	79.81	82.35	84.97	88.56
Coverage of electronic health records, %	17.64	28.40	58.99	69.62	76.36	78.01	81.65	85.46
Utilization rate of health records, %	14.22	14.80	17.66	22.34	29.95	46.98	51.4	55.11
Hard copies of print health education materials distributed per 10,000 people, n	1402.04	1613.33	5404.75	4501.36	4205.84	4845.88	4819.35	4408.12
Participants of health education activities per 10,000 people, n	387.20	570.87	1030.72	1487.17	1473.60	1001.92	1044.09	938.24
Share of children with a vaccination record book, %	99.73	98.91	99.46	99.71	99.63	99.71	99.85	99.97
Share of children receiving standard EPI[b] vaccination, %	98.19	98.93	99.32	99.01	99.29	99.33	99.44	99.14
Rate of newborn home visits, %	81.80	81.55	87.32	89.66	90.39	92.44	93.83	94.77
Coverage of health management among children aged 0-6 years, %	73.41	80.47	85.33	88.64	89.87	90.85	91.6	91.82
Rate of registration in early pregnancy, %	78.62	80.62	83.01	86.76	87.53	89.20	91.57	91.42
Coverage of postnatal visits, %	87.33	85.27	89.11	90.83	91.83	92.95	94.30	94.41
Coverage of health management among the elderly, %	35.17	50.59	65.21	71.80	70.03	72.47	73.29	70.74
Share of patients with hypertension under standardized management, %	68.96	70.77	68.29	70.56	69.28	74.67	70.03	71.64
Blood pressure control rate among the managed patients with hypertension, %	40.68	50.52	48.92	49.96	56.51	59.23	60.96	62.40
Share of patients with type 2 diabetes under standardized management, %	69.58	71.19	68.88	69.54	71.48	75.25	73.01	74.58
Glucose control rate among the managed patients with type 2 diabetes, %	40.03	40.18	45.88	48.54	56.16	55.63	54.25	57.93
Coverage of health management by TCM[c] among the elderly, %	13.85	16.42	18.05	16.86	26.13	37.87	46.98	53.23
Coverage of health management by TCM among children aged 0-36 months, %	13.79	15.54	17.16	18.14	27.15	41.87	48.27	55.69

[a] Source: National Survey on Primary Health Facilities 2016 by the Department of Primary Health, National Health Commission, China.

[b] EPI: Expanded Program on Immunization.

[c] TCM: traditional Chinese medicine.

Table 4. Joinpoint regression analysis of trends for representative indicators on the implementation of the Program, 2009-2016.

Indicators	AAPC[a], % (95% CI)	*P* value	Trend[b] 1			Trend 2		
			Period	APC[c], % (95% CI)	*P* value	Period	APC, % (95% CI)	*P* value
Coverage of health records, (%)	15.0 (11.7 to 18.4)	<.001	2009-2011	42.1 (21.0 to 66.8)	.006	2011-2016	5.7 (3.8 to 7.5)	.002
Coverage of electronic health records (%)	28.1 (21.5 to 35.0)	<.001	2009-2011	102.9 (51.1 to 172.4)	.005	2011-2016	6.5 (4.6 to 8.5)	.002
Utilization rate of health records (%)	23.7 (15.6 to 32.3)	<.001	2009-2016	23.7 (15.6 to 32.3)	<.001	N/A[d]	N/A	N/A
Share of children with a vaccination record book (%)	0.1 (–0.0 to 0.2)	.09	2009-2016	0.1 (–0.0 to 0.2)	.09	N/A	N/A	N/A
Share of children receiving standard EPI[e] vaccination (%)	0.1 (–0.0 to 0.3)	.09	2009-2011	0.5 (–0.3 to 1.3)	.15	2011-2016	0.0 (–0.2 to 0.2)	.99
Rate of newborn home visits (%)	2.3 (1.6 to 2.9)	<.001	2009-2016	2.3 (1.6 to 2.9)	<.001	N/A	N/A	N/A
Coverage of health management among children aged 0-6 years (%)	3.3 (2.9 to 3.8)	<.001	2009-2011	8.6 (6.3 to 10.9)	.001	2011-2016	1.3 (0.9 to 1.7)	.002
Rate of registration in early pregnancy (%)	2.3 (1.8 to 2.8)	<.001	2009-2016	2.3 (1.8 to 2.8)	<.001	N/A	N/A	N/A
Coverage of postnatal visit (%)	1.4 (0.9 to 1.9)	<.001	2009-2016	1.4 (0.9 to 1.9)	<.001	N/A	N/A	N/A
Coverage of health management among the elderly (%)	10.7 (7.4 to 14.1)	<.001	2009-2011	38.2 (17.2 to 62.9)	.008	2011-2016	1.3 (–0.8 to 3.4)	.15
Share of patients with hypertension under standardized management (%)	0.6 (–0.5 to 1.6)	.22	2009-2016	0.6 (–0.5 to 1.6)	.22	N/A[d]	N/A	N/A
Blood pressure control rate among the managed patients with hypertension (%)	5.3 (3.4 to 7.3)	<.001	2009-2016	5.3 (3.4 to 7.3)	<.001	N/A	N/A	N/A
Share of patients with type 2 diabetes under standardized management (%)	1.1 (0.3 to 1.9)	.02	2009-2016	1.1 (0.3 to 1.9)	.02	N/A	N/A	N/A
Glucose control rate among the managed patients with type 2 diabetes (%)	5.8 (3.0 to 8.6)	<.001	2009-2013	9.2 (3.3 to 15.4)	.02	2013-2016	1.4 (–5.3 to 8.5)	.57
Coverage of health management by TCM[f] among the elderly (%)	24.1 (18.5 to 30.1)	<.001	2009-2016	24.1 (18.5 to 30.1)	<.001	N/A	N/A	N/A
Coverage of health management by TCM among children aged 0-36 months (%)	25.1 (18.8 to 31.8)	<.001	2009-2016	25.1 (18.8 to 31.8)	<.001	N/A	N/A	N/A

[a] Trend: two temporal trend segments (Trend 1 and Trend 2) that join up at one join point recognized by the Joinpoint regression model.

[b] AAPC: average annual percent change.

[c] APC: annual percent change.

[d] N/A: not applicable.

[e] EPI: Expanded Program on Immunization.

[f] TCM: traditional Chinese medicine.

The Financing Mechanism of the Program Ensures Equity and Affordability

Since 2009, China has strived for the gradual equalization of basic public health services [24]. To this end, the Program is completely financed by governments at all levels, including the national, provincial, and prefecture levels and below. Due to regional economic disparity, government fiscal strength among provinces varies greatly. To ensure the equitable access to basic public health services, the national government requires that actual per capita financing for the Program should not be lower than the national standard. For example, in 2016, the state required that the per capita subsidy standard for basic public health service projects should not be less than the equivalent of US $6.51 (1 Yuan Renminbi=US $0.15 in 2016) and the result was that the per capita subsidy standard in the eastern, central, and western regions was US $7.43, $7.15, and $6.57, respectively, of which the national-level subsidy accounted for 25.50%, 60.57%, and 79.52%, respectively (see Table 5 for details). This means that according to the actual situation of economic development, the central government will give more financial support to the underdeveloped provinces in the west through transfer payments. The total public investment in the Program increased from US $2.4 billion in 2009 to US $9.6 billion in 2016 [25]. The per capita standard rose from US $2.20 in 2009 [26] to US $11.16 in 2020 [27], and will definitely continue to grow. Such a financing mechanism recognizes regional gaps and guarantees equity and affordability, therefore helping to provide better access to health services [28].

Table 5. Fiscal investment in the Program by regions for 2016.

Financing standard	Eastern region (Yuan[a]), average (% share)	Central region (Yuan), average (% share)	Western region (Yuan), average (% share)
Per capita standard	51.32 (100.00)	49.38 (100.00)	45.42 (100.00)
By national budget	13.09 (25.50)	29.91 (60.57)	36.12 (79.52)
By provincial budget	8.98 (17.50)	8.51 (17.23)	5.36 (11.80)
By prefectural budget and below	29.25 (57.00)	10.96 (22.20)	3.94 (8.68)

[a] 1 Chinese Yuan Renminbi=US $0.1447 in 2016.

Implementation of the Program Enhances Capability and Quality of the Health Workforce

Primary health workers are the mainstay to deliver basic public health services and a major contributor to UHC. To achieve the goals, competent departments provide them with targeted training to improve their capability and quality. For instance, health authorities introduced a package of policies such as the Program of Strengthening Rural Healthcare Service Delivery Network to enhance their management and technical skills through training [29]. In addition, since the new round of health care reform was launched in 2009, the central government has asked THCs to design well-structured and targeted training programs to update village doctors' knowledge and skills in public health services [30].

Local governments organize various training on the policies, technical protocols, knowledge, and skills related to the Program for those who manage or implement it, including public health managers, professionals, and village doctors. In some places, they also implement contests on the Programs and exams on the technical protocols of the services. In other provinces, ear-marked transfer from the national level is used for capacity building of CHC health workers. Other activities include a contracted-based education program of public health doctors or village doctors, which requires trainees to work at a primary level for a certain period after graduation, further education for the village doctors who had only received secondary vocational training, and training to upgrade village doctors into licensed assistant doctors [31].

In general, the Program improves the health workers' capability and quality in three areas. First, they are more capable of providing integrated care. The Program is established to deliver integrated care, including both preventive and curative interventions. This requires the services of primary health workers to shift from a disease-centered to a health-centered model [32]. Second, as the services covered by the Program are all delivered at the primary level, primary health workers have direct contact with local residents and need to maintain a good relationship with them [33]. More attention is given to communication skills, patience, and empathy. Third, they are better at identifying problems during service delivery; managing, analyzing, and utilizing information; and diagnosing and addressing regional health issues [34]. Some studies found that the theoretical knowledge and practical skills of the trained primary health workers had improved greatly [35-38].

The Program Facilitates the Integration of a Public Health Service Delivery System

Previously, China's public health service delivery system was fragmented, composed of many specialist public health institutions responsible for health education, disease prevention and control, endemic and occupational disease prevention, and maternal and child health care, respectively. Primary health care institutions were not included in the system. However, the Program was mainly implemented by primary institutions, along with participation by these specialist institutions through technical assistance, technical training, and development of protocols. Various players in the system work together. They complement, collaborate with, and support each other, which gradually integrate the public health service delivery system [39]. This is a significant step forward to achieve integrated continuous care as well as UHC.

Shortcomings of the Program

No Well-Established Selection and Exit Mechanism for Service Items, and the Number of Items Keeps Rising

First, the service package of the Program cannot meet the actual needs of the general public. This is largely associated with the fact that service items are set based on how much public financing is available [40]. UHC aims to provide people-centered, quality services to satisfy the needs of the population. However, with higher health literacy, the needs are growing and becoming more diversified. Moreover, public health priorities are continuously changing over time, which requires timely updates of the service package [41].

Since the Program covers a large number of services and a huge target population, primary health workers can only manage to hit the set target in service volume but must compromise on quality. Indeed, the services sometimes fail to live up to the standards set in the Basic Public Health Service Specification (3rd edition) [42]. For example, some follow-ups fall short of requirement of the Specification. In some places, providers may only adopt the service as a matter of formality. While service volume and items increase continuously, the conflict between the actual needs of citizens and the service capacity of the primary level is not taken into consideration adequately. For instance, two new items, health management by TCM and management of patients with severe mental illness, were included in the package recently; however, primary institutions do not have the capability of delivering these services. Without sufficient knowledge and skills in TCM, village doctors have difficulties in carrying out TCM interventions. As a result, the pressure of health workers mounts on the one hand and citizens are not satisfied with the services on the other hand, which negatively affects the result of the Program.

Poor Integration of the Service Delivery System Lowers the Efficiency

Second, poor integration among the service items lowers the efficiency of the system. There are many service items in the package but little integration among them. For example, health records for the whole population are isolated from maternal, child, and elderly health management, and NCD cases are managed separately according to conditions [43]. All of these aspects result in unnecessary duplication in terms of effort and systems. Moreover, basic public health services and essential medical care are not well connected [44]. Taking NCD management as an example, services by general practitioners (GPs), specialists, and public health workers are so poorly integrated that a closed-end service network is absent, which has a negative influence on the quality of NCD management.

Furthermore, the information technology (IT) system is fragmented and poorly integrated [45]. Without a single top-level framework, IT systems are not compatible and connected with each other, which cannot meet the needs of the Program. There are nationwide systems for vaccination, management of patients with severe mental illness, and communicable disease reporting, as well as province-wide systems for maternal and child care. For the remaining items, there are few province-wide or prefecture-level information systems. In most cases, it is the counties/districts that authorize IT companies to develop the systems needed, which wastes substantial human and financial resources. Li et al [6] showed that the IT systems of primary health institutions with in-house EHRs were developed by over 80 IT providers independently with little connectivity and interoperability. Poor integration among these systems hinders information sharing across institutions and regions, which impacts the result of the Program [46]. For instance, although health records cover the majority of the population, many of them are not sufficiently active [47]. Moreover, the isolation between health record systems and the systems of health institutions hampers the integration of preventive and medical care.

Shortage of Well-Trained Health Professionals

The third challenge is insufficient well-trained primary health workers, who are the main implementers of the Program. According to the Statistic Bulletin on China’s Health Sector 2018, health workers in THCs and CHCs only account for 32.2% of the national total [48]. GPs, representing 8.49% of all doctors, represent the profession that faces the greatest shortage. At the end of 2018, the number of GPs per 10,000 people was only 2.22, which is still lower than the upper-end of target for 2020 of 2–3 GPs per 10,000 citizens [49].

The increasing workload on health workers also highlights the understaffed issue in health institutions. As service items and the population served are both on the rise, it is common that one primary health worker takes on the responsibilities of several. A survey among primary health workers in three provinces showed that 27.1% considered that after implementation of the Program, the workload increased to an unbearable level [50]. Shi et al [51] indicated that 4307 out of 10,626 (40.53%) primary care physicians were extremely tired. Li et al [52] found that 627 of 1221 (51.35%) village doctors in three provinces were not satisfied with the situation. Sun et al [53] found that 1244 of 3212 (38.73%) CHC health workers in five provinces wanted to quit the job. Meanwhile, given the restriction in the current recruitment system and lack of supportive measures, it is very hard to effectively replenish the primary health workforce. Moreover, primary health workers are often less knowledgeable with a lower education level and professional title [54]. Among staff of CHCs and THCs, those with an associate degree and below account for 71.9% and 91.3%, respectively. Village doctors are even more poorly educated [55]; 21% of primary health professionals are not licensed physicians or licensed assistant physicians [6]. Lack of human resources and inadequate service capability have greatly impacted implementation of the Program.

Government Agencies Outside the Health Sector Are Rarely Engaged and Interdepartment Collaboration is Weak

Fourth, China’s basic public health service delivery system remains an isolated island. The services are mainly delivered by health institutions. Local governments usually do not take the lead and deliver due responsibility in the Program implementation, because the basic public health service is often considered to be the mandate of the health sector only, and thus has little to do with the government as a whole [56]. Therefore, other government agencies are poorly motivated to engage in this area. However, many of them, such as public security, civil affairs, neighborhood committees, and health insurers, should play important roles in information sharing and promotion campaigns, while primary health institutions function as the service providers. In practice, it is always the latter that propose collaboration with other agencies. In some cases, the public security department might even refuse to share information of the population in the catchment under the excuse of confidentiality. This is, to a large extent, because departments cannot reach a consensus on their cooperation in the absence of incentives or institutionalized operable binding mechanisms for coordination [57].

Per Capita Public Financing Varies Among Regions and is Lower Than the Cost

Fifth, public financing varies among regions, which may lead to inequity in health outcomes. In most provinces, underdeveloped provinces in particular, public financing for the Program follows the national standard and 80% of the financing is derived from transfer payments by the central government. However, in developed provinces, especially in the eastern region, the per capita financing is more than the national standard and service packages are more generous than the national package, because provincial and municipal governments allocate more investment to the Program. In 2016, 24 out of the 32 provinces (autonomous regions and municipalities directly under the central government) followed the national financing standard (US $6.51 per capita), and in the other 8 provinces, the financing was higher than the national standard (eg, US $19.68 per capita in Beijing and US $11.14 per capita in Shanghai) [56]. Unsurprisingly, in the areas where the per capita standard is much higher than the national standard, local residents enjoy better services. Health inequity derived from the regional economic gap is a stark fact, which is inconsistent with UHC [58].

In the exiting financing mechanism, a national standard guarantees basic equity among regions. However, in practice, provinces and regions with a strong economy have more financial and material resources available to finance the Program, while underdeveloped regions cannot afford the Program, even though their needs in disease prevention and control are more urgent. This results in a regressive effect, and widens the gap in public health service between the rich and poor regions [59].

In addition, there is an increasingly large gap between the financing standard and the actual service cost, which threatens the sustainability of the Program. With the growing demand for health, Chinese people, especially those living in cities and economically developed areas, want not only a longer life but also a healthier one, and their demand for basic public health service is increasing. For example, the elderly want their physical checkups to be as comprehensive as those performed in large hospitals, covering more complicated items such as computed tomography scans. Patients with hypertension or diabetes would like to receive imported drugs, which also comes with a higher cost. For instance, a complex package of lab tests will cost health institutions US $15. In practice, the Program was designed to provide universally accessible basic public health services. With a high-cost package in place, local financing is stressful, and primary health institutions have to finance public health services by the income from the medical service to reach the set target. Moreover, due to the geographical environment, customs, and other factors in different regions, the service delivery cost varies greatly.

Discussion

Achievements and Shortcomings of Program Implementation

The comprehensive promotion of the Program and the rapid development of the universal medical insurance system are two important measures that have jointly affected UHC in the New Medical Reform, which has attracted worldwide attention since 2009. These measures serve to build up a UHC system integrating preventive and curative care. Since its implementation, the Program has played a positive role in providing life cycle public health services for the whole population; enhancing equity, accessibility, and affordability of the public health service; improving the quality of the primary health workforce; and integrating the public health service delivery system. However, at the same time, there are still some challenges to overcome, such as a rigid item inclusion and exclusion mechanism that is irresponsive to new developments, overemphasis on prevention but neglect of medical care, backward and fragmented information systems, insufficient human resources, and inadequate government functioning.

In the Program, five categories of services, including health record management for urban and rural residents, elderly health management, management of patients with hypertension, management of patients with type 2 diabetes, and health management by TCM, are considered to be developing from scratch in China. Among the 10 indicators relating to these services, except for the percentage of patients with hypertension under standardized management, the remaining 9 indicators have increased significantly, especially the 6 indicators relating to health record management for urban and rural residents, health management by TCM, and elderly health management, which increased at a relatively higher rate. Two categories of services (health management of children aged 0-6 years and maternal health management) had been implemented for a longer period of time before the Program was established and were thus relatively mature; consequently, all 4 indicators relating to these services rose relatively slowly. The vaccination service has been carried out in China for decades, and it is so widely accepted that the two indicators relating to this type of service have remained at about 99% after the implementation of the Program, with no room for growth.

Five indicators, including coverage of health records, coverage of EHRs, health management among children aged 0-6 years, health management among the elderly, and glucose control rate among the managed patients with type 2 diabetes, appeared in 2011 or 2013, rose rapidly before the connection point, and then the increase slowed down or tended to be stable after the connection point, which may be related to the increase both in the number of service items and in the workload of grassroots personnel after 2011 [60]. According to the requirements of the National Basic Public Health Service Specification, the standardized management of patients with hypertension and type 2 diabetes should be a follow-up 4 times a year. Between 2009 and 2016, the standardized management rate of these two types of patients was between 68% and 76%; the percentage of patients with hypertension under standardized management did not increase significantly and the percentage of patients with type 2 diabetes under standardized management rose slowly, which may be related to the low compliance of the two types of patients with follow-up services and population flow [61]. In addition, with the popularity of sphygmomanometers and blood glucose meters, coupled with the limited service capacity of grassroots medical personnel, simple follow-up is no longer attractive to patients with type 2 diabetes or hypertension [62].

International Experience With UHC

Experience from other countries demonstrates that a health system with universal coverage is the foundation of the health sector. Since 2015, advocated by the WHO and United Nations (UN) Assembly, UHC has become a global goal. To this end, countries adopt different measures based on their own context and development stage but also face some common challenges.

The United Kingdom is a frontrunner in UHC; its National Health Service (NHS) provides high-quality health care services to British citizens under the principle of universal access, free care, and needs-based service. This is largely achieved owing to its strong funding system. The NHS budget accounts for 9.6% of the gross domestic product (GDP), whereas China's health expenditure only accounts for up to 5.6% of its GDP. In addition, 82% of NHS expenditure is funded by the general taxation revenue, only 2% is funded by individual copayment, and the rest is funded by national insurance and donations. This system maximizes the access to health services. However, it shares the same challenges as faced by the Program. For example, each GP in the NHS is responsible for 1500 patients, and the income of physicians has nothing to do with their workload, which leads to low efficiency of the health service delivery system and very long waiting lists [63]. In China, the implementation of the Program also brings about an extremely heavy workload to primary health workers but lacks effective incentives. Germany has also achieved a high level of UHC, where legislation of the health insurance system plays a critical role. The Statutory Health Insurance is the mainstay of Germany's health insurance system, covering 90% of the population. Such high coverage is partly attributed to the arrangement that the premium covers spouses and children who have no income [64]. The system is operated by a third party selected through a competitive process, which contributes to gains in service quality and efficiency. As a low- and middle-income country, Cubans have good health status, which may be greatly associated with the country's system of free health care for all. The system guarantees the equity of health insurance and benefits. Thailand's "30 Baht Plan" expands the health insurance system to the uncovered, which accounts for approximately 30% of the population, so that they have access to health care. As a result, Thailand has officially achieved UHC [65]. The Thai health insurance system is pooling at a national level with better capacity of risk resistance. In contrast, the three major insurance schemes (which are integrated with each other) in China are mostly pooling at the prefectural levels, and even at the county level in some cases. Although the above institutional arrangements in these countries are not specific to the public health service, service items related to prevention, health management, and rehabilitation are all covered by the free service package or the benefit package of the health insurance system.

China's Experience With UHC

Countries such as the United Kingdom, Germany, Cuba, Thailand, and others achieve UHC by following a slightly different approach from that adopted by China. Most of these countries do not consider a public health review service as an independent system, but rather cover this service in the health insurance system. By contrast, China defines a basic public health service package with universal coverage, which only includes preventive care and health management [66]. This approach draws attention to prevention and public health, which had been neglected for a long time; facilitates a strong organizational system, logistic safeguard, and political will for its financing and service delivery; and covers the whole population within a short time frame. However, this system also introduces new fragmentation between medical and preventive care. In some cases, the original treatment-centered model shifted to the other extreme: prevention is overemphasized and treatment is neglected. In terms of financing, the public health system is completely independent from the health insurance system; the latter only covers treatment, while the former only focuses on prevention. Such fragmentation is not conducive to the integration of the two and can delay progress in achieving UHC [67].

Suggestions to Promote UHC in China

Besides financing, the service delivery system for prevention is also relatively independent from that used for curative care, which has become the bottleneck for quality gains at the primary level and the health status of the population [66]. The health care under UHC requires a shift to an integrated service delivery system featured by primary care as its core, better coordination and cooperation among service providers, as well as continuity and integration of services [67]. Better integration can boost the efficiency of the system. South Africa has made impressive progress in primary care largely owing to its focus on integrated care [68]. Erondu et al [2] also appealed that UHC requires a more integrated health system, including public health.

Fang et al [69] argued that without efficiency gains in the health care system, some measures such as strengthening primary care and increasing investment can only bring moderate improvements in financial protection, one of the key objectives of UHC [70].

From a broader perspective, the basic public health service needs to connect with other health-related factors. "Health in All Policies" (HiAP) has become an important guideline for China's efforts in the health sector in recent years [71]. HiAP is also described as a necessary component of primary care [72]. The Program has started to recognize the negative impact of unhealthy lifestyles, but it is not sufficient to only change individuals' lifestyle. Social and environmental factors bearing on health should receive more attention. For example, people can have a healthy lifestyle and dietary structure, but cannot avoid the negative impacts due to inhaling polluted air during outdoor exercises or the intake of antibiotics, steroids, and pesticide residues in the diet [71]. The Program should seize the opportunity to work with more partners, especially players in nonhealth sectors, to address policies and structural factors rather than only focusing on individual factors [73].

The Program, with ear-marked funding and specific governance arrangement, has been implemented in full swing very quickly, with some compromise on service quality to a certain extent [66]. At the initial stage, the performance indicators were mainly related to volume, such as the number of records on follow-up visits and coverage of health records, rather than incentives for better quality. The latest study indicates that quality of care, and not only accessibility, is the key determinant to improve population health and reduce the disease burden [74]. For instance, South Africa, where remarkable progress has been made in UHC, takes quality as the core of its efforts [71].

The quality of the basic public health service is also closely related to the existing financing methods and overall health system in China. On the one hand, despite continuous growth of public investment, the financing standard is still too low to satisfy the public expectation of the public health service. Due to regional economic disparity and a decentralized fiscal system, local governments' expenditure on public health varies greatly, leading to a quality gap among regions. The Outline of the Planning for National Healthcare Service Delivery System (2015-2020) clearly points out that in western China, low-quality health resources and insufficient service capacity at the primary level affect the equity and efficiency of health service [75]. On the other hand, China's health system remains hospital-centered, and patients tend to go to the higher-level providers for health services. From 2005 to 2015, the proportion of outpatient care provided at the primary level declined by 7% [76], although "strong primary providers" had been continuously emphasized since the launch of the reform in 2009. This is a problem worth pondering. When patients are "voting with their feet," high-quality human resources are siphoned from the primary level, which aggravates the shortage of primary health workers, especially those of high quality. In 2010, only 5.6% of THC doctors had a 5-year or above medical education. This proportion increased very slowly in the following years, reaching 10% in 2017 [77]. As the capacity of health professionals undoubtedly influences service quality, it is hard to assure the quality of basic public health services delivered at the primary level [78]. At present, China is trying to change the medical education system and education model. The health workforce is expected to improve in the future. However, it remains a significant challenge to ensure the availability of adequate, high-quality health professionals at the primary level [36].

The UN Sustainable Development Goals require all countries to achieve UHC by 2030, focusing on two priorities in particular: universal coverage of essential health services and financial protection [70]. China has made substantial progress toward the first priority, largely due to implementation of the Program and the rapid expansion of health insurance system. However, it still has a long way to go in the second area because of gaps in health financing, service quality, and coordination within the system.

Conclusion

Based on a review of the implementation of basic public health service projects in China, this study systematically analyzed the role and challenges of the implementation of basic public health services in China, a country with a population of 1.4 billion, to achieve UHC. This experience can also serve as a reference for other countries. The services in the package focus on key populations and provide full life cycle services for the whole population. The financing system completely supported by the government makes the services more equitable and affordable. However, health system reform is complicated and full of daunting challenges, especially for a populous, developing country such as China. In the future, the Program should aim to first balance the relationship between curative and preventive care, and adjust its incentive mechanism. Meanwhile, a continuous increase in public investment is the key for both the health insurance system and to achieve equal access to basic public health services. A policy opportunity is to facilitate the integration of the Program and clinical care, which means a "family doctor team" made up of clinical physicians and public health workers that manages the health of the population in the catchment. Under this model, the continuity and quality of NCD management can be improved and the Program can directly contribute to better health outcomes. At present, many far-sighted experts are working on restructuring China's service delivery system, and public health should be an important component. The restructuring should follow the principle of HiAP, and aim for a people-centered, community-based, and coordinated service delivery system, which can provide continuous, accountable, and integrated care to satisfy the needs of the population.

Acknowledgments

Funding for work on this paper was provided by a grant from the National Health Commission, China ("Research on Improving the Equalization of National Basic Public Health Services"), which contributed substantially to the study design and the collection and analysis of data; and a grant by the National Natural Science Foundation of China ("A Study on Strategies of Equalization of Basic Public Health Based on Mechanism Design Theory" 71503010), which contributed to the study design for some of the content of this paper. The views expressed in the paper are the sole responsibility of the authors.

Authors' Contributions

FG and YD contributed equally. FG and WL developed the structure of the paper. FG and YD analyzed the data and drafted the manuscript. YJ designed the study, discussed the methodology, and contributed to drafting the paper. LY and WZ contributed to the production of some content of the draft. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Sachs JD. Achieving universal health coverage in low-income settings. Lancet 2012 Sep 08;380(9845):944-947. [doi: 10.1016/S0140-6736(12)61149-0] [Medline: 22959391]
2. Erondu NA, Martin J, Marten R, Ooms G, Yates R, Heymann DL. Building the case for embedding global health security into universal health coverage: a proposal for a unified health system that includes public health. Lancet 2018 Oct 20;392(10156):1482-1486. [doi: 10.1016/S0140-6736(18)32332-8] [Medline: 30343862]
3. Passmore R. The declaration of Alma-Ata and the future of primary care. Lancet 1979 Nov 10;2(8150):1005-1008. [doi: 10.1016/s0140-6736(79)92572-8] [Medline: 91728]
4. Beattie A, Yates R, Noble DJ. Accelerating progress towards universal health coverage in Asia and Pacific: improving the future for women and children. BMJ Glob Health 2016;1(Suppl 2):i12-i18 [FREE Full text] [doi: 10.1136/bmjgh-2016-000190] [Medline: 28588989]
5. Declaration of Astana. World Health Organization. URL: https://www.who.int/publications/i/item/WHO-HIS-SDS-2018.61 [accessed 2020-03-05]
6. Li X, Lu J, Hu S, Cheng KK, De Maeseneer J, Meng Q, et al. The primary health-care system in China. Lancet 2017 Dec 09;390(10112):2584-2594. [doi: 10.1016/S0140-6736(17)33109-4] [Medline: 29231837]
7. Liu Q. World Health Assembly announces health goals, China says it will prioritize the development of universal health coverage and China says it will give priority to universal health coverage. Guangming Daily. 2018 May 23. URL: http://epaper.gmw.cn/gmrb/html/2018-05/23/nw.D110000gmrb_20180523_2-12.htm [accessed 2020-07-23]
8. Chen Z. Launch of the health-care reform plan in China. Lancet 2009 Apr 18;373(9672):1322-1324. [doi: 10.1016/S0140-6736(09)60753-4] [Medline: 19376436]
9. Universal Health Coverage. World Health Organization. URL: http://www.who.int/health_financing/universal_coverage_definition/en/ [accessed 2020-10-30]
10. National standards for basic public health services, third edition. National Health and Family Planning Commission. 2017. URL: http://www.nhc.gov.cn/ewebeditor/uploadfile/2017/04/20170417104506514.pdf [accessed 2020-07-23]
11. Liu X, Wang Z, Zhang H, Meng Q. Measuring and evaluating progress towards Universal Health Coverage in China. J Glob Health 2021 May 01;11:08005. [doi: 10.7189/jogh.11.08005] [Medline: 33981413]
12. Kunjumen T, Okech M, Diallo K, Mcquide P, Zapata T, Campbell J. Global experiences in health workforce policy, planning and management using the Workload Indicators of Staffing Need (WISN) method, and way forward. Hum Resour Health 2022 Jan 28;19(Suppl 1):152 [FREE Full text] [doi: 10.1186/s12960-021-00695-9] [Medline: 35090485]
13. Huang M, Wang J, Nicholas S, Maitland E, Guo Z. Development, status quo, and challenges to China's health informatization during COVID-19: evaluation and recommendations. J Med Internet Res 2021 Jun 17;23(6):e27345 [FREE Full text] [doi: 10.2196/27345] [Medline: 34061761]
14. Liu ZY, Xiao Y, Zhao K, Liu AZ. Implementation progress and effect of National Essential Public Health Services Program in China. Chinese J Public Health 2019;35(6):657-664 [FREE Full text] [doi: 10.11847/zgggws1121468]
15. Urquieta-Salomón JE, Villarreal HJ. Evolution of health coverage in Mexico: evidence of progress and challenges in the Mexican health system. Health Policy Plan 2016 Feb;31(1):28-36. [doi: 10.1093/heapol/czv015] [Medline: 25823751]
16. Atun R. Transforming Turkey's health system--lessons for universal coverage. N Engl J Med 2015 Oct;373(14):1285-1289. [doi: 10.1056/NEJMp1410433] [Medline: 26422719]
17. Zheng Y, Li L, Zhang L. Universal health coverage: China's experience in health care system construction and its implications for Ethiopia. Medical Higher Vocat Educ Modern Nurs 2016;33(8):1-4. [doi: 10.3969/j.ssn.1-1137.2016.01.001]
18. Shi G. Concepts, policies and strategies: how can we achieve universal health coverage in our country. Health Econ Res 2013;10:6-8. [doi: 10.14055/j.cnki.33-1056/f.2013.10.002]
19. Shan L, Wu Q, Liu C, Li Y, Cui Y, Liang Z, et al. Perceived challenges to achieving universal health coverage: a cross-sectional survey of social health insurance managers/administrators in China. BMJ Open 2017 Jun 02;7(5):e014425 [FREE Full text] [doi: 10.1136/bmjopen-2016-014425] [Medline: 28576890]
20. Jiang CH, Ma JD. Analyzing the role of overall basic medical insurance in the process of universal health coverage. Chinese Health Service Management 2015;32(141):108-110 [FREE Full text]
21. Chi Y, Wu QH, Hao YH. International comparisons of the extent to which universal health coverage is achieved. Chinese Health Resources 2016;19:379-366. [doi: 10.13688/j.cnki.chr.2016.16011]

22. Liang LL, Langenbrunner J. The long march to universal health coverage: outside looking in. Chinese J Health Policy 2013;6(02):1-3. [doi: 10.3969/j.issn.1674-2982.2013.02.001]
23. World Health Organization. Improving universal health coverage requires countries to increase spending on primary health care. Chinese J Health Policy 2019;12(9):40 [FREE Full text]
24. On promoting the gradual equalization of basic public health services. The State Council of the People's Republic of China. URL: http://www.gov.cn/ztzl/ygzt/content_1661065.htm [accessed 2020-10-30]
25. Zheng JH. China Statistical Yearbook 2017. Beijing: China Statistics Press; 2017.
26. Notice of the General Office of the State Council on Printing and Distributing the Work Arrangements for the Five Key Reforms of the Medical and Health System in 2009. General Office of the State Council. URL: http://www.nhc.gov.cn/bgt/s9511/200907/d0fa1a47084a4108bf5b84f4a0f08a76.shtml [accessed 2021-07-23]
27. A circular on the implementation of the Basic Public Health Service Project by 2020. National Health Commission. URL: http://www.nhc.gov.cn/jws/s7874/202006/619506aa0fd14721b7e5711d389c323f.shtml [accessed 2021-06-16]
28. Reeves A, Gourtsoyannis Y, Basu S, McCoy D, McKee M, Stuckler D. Financing universal health coverage--effects of alternative tax structures on public health systems: cross-national modelling in 89 low-income and middle-income countries. Lancet 2015 Jul 18;386(9990):274-280 [FREE Full text] [doi: 10.1016/S0140-6736(15)60574-8] [Medline: 25982041]
29. Zhang Y, Chen M, Shi XQ. Challenges of basic public health services provided by village doctors in Guizhou, China. Asia Pac J Public Health 2015 Mar;27(2 Suppl):69S-76S. [doi: 10.1177/1010539514568708] [Medline: 25673280]
30. Li T, Lei T, Xie Z, Zhang T. Determinants of basic public health services provision by village doctors in China: using non-communicable diseases management as an example. BMC Health Serv Res 2016 Feb 04;16:42 [FREE Full text] [doi: 10.1186/s12913-016-1276-y] [Medline: 26846921]
31. Zhang R, Chen Y, Liu S, Liang S, Wang G, Li L, et al. Progress of equalizing basic public health services in Southwest China--- health education delivery in primary healthcare sectors. BMC Health Serv Res 2020 Mar 24;20(1):247 [FREE Full text] [doi: 10.1186/s12913-020-05120-w] [Medline: 32209085]
32. Qin JM. Progress in basic public health service projects in China. Chinese J Public Health 2017;33:1289-1297. [doi: 10.11847/zgggws2017-33-09-01]
33. Li WL. Development path of rural doctors' excellence under the background of village revitalization. Chinese Rural Health Service Administration 2021 Aug;41(08):554-557. [doi: 10.19955/j.cnki.1005-5916.2021.08.005]
34. Kanchanachitra C, Lindelow M, Johnston T, Hanvoravongchai P, Lorenzo FM, Huong NL, et al. Human resources for health in southeast Asia: shortages, distributional challenges, and international trade in health services. Lancet 2011 Feb 26;377(9767):769-781. [doi: 10.1016/S0140-6736(10)62035-1] [Medline: 21269674]
35. Li SY, Sun YY. Analysis on the training short-term effect of training program on disease prevention and control knowledge and skills for grassroots public health personnel. Chinese Primary Health Care 2015;29:101-103. [doi: 10.3969/j.issn.1001-568X.2015.06.0024]
36. Li J, Li J, Huang Y, Thornicroft G. Mental health training program for community mental health staff in Guangzhou, China: effects on knowledge of mental illness and stigma. Int J Ment Health Syst 2014;8(1):49 [FREE Full text] [doi: 10.1186/1752-4458-8-49] [Medline: 25520750]
37. Bai YJ, Xiao FG, Yun CH. Study on training effect of public health personnel in township health centers. China Higher Medical Education 2015;1:30-31. [doi: 10.3969/j.issn.1002-1701.2015.01.016]
38. Zhang ZX. Application and effect evaluation of blended training on public health for primary healthcare workers in rural areasD, Huazhong University of Science and Technology, 2016. Huazhong University of Science and Technology. URL: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=1016088792.nh&DbName=CDFD2017 [accessed 2020-06-17]
39. Wang K, Mao A, Meng YL, Yang YJ, Dong P, Qiu WQ. Development history, current situation, problems and strategies of public health system construction in China. Chinese J Public Health 2019;35:801-805. [doi: 10.11847/zgggws1123311]
40. Fan WX. Implementation status and mechanism optimization of basic public health service projects. Macroeconomic Manag 2017;11:61-65. [doi: 10.19709/j.cnki.11-3199/f.2017.11.013]
41. Yip WC, Hsiao WC, Chen W, Hu S, Ma J, Maynard A. Early appraisal of China's huge and complex health-care reforms. Lancet 2012 Mar 03;379(9818):833-842. [doi: 10.1016/S0140-6736(11)61880-1] [Medline: 22386036]
42. Tian CW, Zheng YM, Sun NL, Yao MJ, Bing PF, Liu L, et al. Current status of standardization of basic public health services and standardization strategy of service equalization in China. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Oct 10;41(10):1723-1730. [doi: 10.3760/cma.j.cn112338-20200513-00719] [Medline: 33297633]
43. Chen S, Guo L, Wang Z, Mao W, Ge Y, Ying X, et al. Current situation and progress toward the 2030 health-related Sustainable Development Goals in China: a systematic analysis. PLoS Med 2019 Nov;16(11):e1002975 [FREE Full text] [doi: 10.1371/journal.pmed.1002975] [Medline: 31743352]
44. Ren XC, Niu YZ, Meng LG. Study on the path of synergy between public health services and basic medical services. Med Philosophy 2021;42(19):65-69. [doi: 10.12014/j.issn.1002-0772.2021.19.14]
45. Qin PP, Chen Q, Xie LQ, Chen QK, Hu HP. The development situation of national health informatization in China since New Medical Reform. J Med Informat 2019;7:7-11. [doi: 10.3969/j.issn.1673-6036.2019.07.002]

46. D'Agostino M, Marti M, Otero P, Doane D, Brooks I, Garcia Saiso S, et al. Toward a holistic definition for Information Systems for Health in the age of digital interdependence. Rev Panam Salud Publica 2021;45:e143 [FREE Full text] [doi: 10.26633/RPSP.2021.143] [Medline: 34840555]
47. Wang CK, Zhu Y, Wu SY, Tang XJ. Retrospective analysis of the national primary health informatization development during the 13th Five-Year Plan in China. Chinese J Health Informat Manag 2021;18(03):319-323. [doi: 10.3969/j.issn.1672-5166.2021.03.05]
48. National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/guihuaxxs/s10748/201905/9b8d52727cf346049de8acce25ffcbd0.shtml [accessed 2020-06-15]
49. Tam Y, Leung JYY, Ni MY, Ip DKM, Leung GM. Training sufficient and adequate general practitioners for universal health coverage in China. BMJ 2018 Jul 26;362:k3128. [doi: 10.1136/bmj.k3128] [Medline: 30049679]
50. Wang L, Wang Z, Ma Q, Fang G, Yang J. The development and reform of public health in China from 1949 to 2019. Global Health 2019 Jul 02;15(1):45 [FREE Full text] [doi: 10.1186/s12992-019-0486-6] [Medline: 31266514]
51. Shi L, Song K, Rane S, Sun X, Li H, Meng Q. Factors associated with job satisfaction by Chinese primary care providers. Prim Health Care Res Dev 2014 Jan;15(1):46-57. [doi: 10.1017/S1463423612000692] [Medline: 23388523]
52. Li T, Lei T, Sun F, Xie Z. Determinants of village doctors' job satisfaction under China's health sector reform: a cross-sectional mixed methods study. Int J Equity Health 2017 Apr 18;16(1):64 [FREE Full text] [doi: 10.1186/s12939-017-0560-8] [Medline: 28420396]
53. Sun Y, Luo Z, Fang P. Factors influencing the turnover intention of Chinese community health service workers based on the investigation results of five provinces. J Community Health 2013 Dec;38(6):1058-1066. [doi: 10.1007/s10900-013-9714-9] [Medline: 23775034]
54. Wei W, Wang X, Luo Y, Feng Z, Wang S. Evaluation of the implementation effect of national basic public health service project. Chinese Primary Health Care 2013;27(5):1-2 [FREE Full text]
55. National HPC. China Health and Family Planning Statistical Yearbook. In: National Health and Family Planning Commission. Beijing: China Union Medical College Press; 2017.
56. Liu WQ, Yang JX, Wang ZH, Xie PP. Experience, problems and challenges in implementation of basic public health services in China: a brief discussion. Chinese J Public Health 2020;36(12):1677-1681. [doi: 10.11847/zgggws1127929]
57. Zhang YC, Qin JM. Challenges and strategies of non-communicable disease prevention and control from the perspective of health in all policies: a survey in typical health cities in China. Chinese J Public Health 2014;7:65-69. [doi: 10.3969/j.issn.1674-2982.2014.01.014]
58. Gwatkin DR, Ergo A. Universal health coverage: friend or foe of health equity? Lancet 2011 Jun 25;377(9784):2160-2161. [doi: 10.1016/S0140-6736(10)62058-2] [Medline: 21084113]
59. Pu C. Research on the implementation strategies of equalization of basic public health service: a case study of Chongqing. Soft Science 2010;24:73-77 [FREE Full text]
60. Feng JX, Lin YM, Li L, Wu M. A study on primary health human factors influencing the health management of patients with hypertension and type 2 diabetes. Pract Prevent Med 2021;28(02):246-249. [doi: 10.3969/j.issn.1006-3110.2021.02.032]
61. Zhou DH, Feng ZC, He SS, Sun X, Ma C, Lv B, et al. Equity of the essential public health service in rural china: evidence from a nationwide survey of hypertensive patients. Pak J Med Sci 2013 Jul;29(4):1012-1017 [FREE Full text] [doi: 10.12669/pjms.294.3773] [Medline: 24353678]
62. Pan EC, Zhang Q, Li Y, Yv H, Chen SH, Wang Y. Investigation on the health management of hypertension and diabetes in basic public health service projects carried out by grass-roots medical personnel. China Gen Med 2014;17(28):3316-3320. [doi: 10.3969/j.issn.1007-9572.2014.28]
63. Hu XZ, Gu Q, Liang F. Comparative study on forecasting methods of major medical equipments in China: taking CT as an example. Chinese Health Resources 2013;16:117-119 [FREE Full text]
64. Busse R, Blümel M, Knieps F, Bärnighausen T. Statutory health insurance in Germany: a health system shaped by 135 years of solidarity, self-governance, and competition. Lancet 2017 Aug 26;390(10097):882-897. [doi: 10.1016/S0140-6736(17)31280-1] [Medline: 28684025]
65. Hughes D, Leethongdee S. Universal coverage in the land of smiles: lessons from Thailand's 30 Baht health reforms. Health Aff 2007;26(4):999-1008. [doi: 10.1377/hlthaff.26.4.999] [Medline: 17630443]
66. Yuan B, Balabanova D, Gao J, Tang S, Guo Y. Strengthening public health services to achieve universal health coverage in China. BMJ 2019 Jun 21;365:l2358 [FREE Full text] [doi: 10.1136/bmj.l2358] [Medline: 31227480]
67. Meng Q, Mills A, Wang L, Han Q. What can we learn from China's health system reform? BMJ 2019 Jun 19;365:l2349 [FREE Full text] [doi: 10.1136/bmj.l2349] [Medline: 31217222]
68. Matsoso MP, Hunter JR, Brijlal V. Embedding quality at the core of universal health coverage in South Africa. Lancet Glob Health 2018 Nov;6(11):e1153-e1154 [FREE Full text] [doi: 10.1016/S2214-109X(18)30323-1] [Medline: 30196092]
69. Fang H, Eggleston K, Hanson K, Wu M. Enhancing financial protection under China's social health insurance to achieve universal health coverage. BMJ 2019 Jun 21;365:l2378 [FREE Full text] [doi: 10.1136/bmj.l2378] [Medline: 31227485]
70. Sustainable development goals: ensure healthy lives and promote well-being for all at all ages. United Nations. URL: https://sustainabledevelopment.un.org/sdg3 [accessed 2020-05-05]

71. Yang JX. New: integrating health into all policies. URL: http://health.people.com.cn/n1/2016/1103/c398004-28830661.html [accessed 2020-07-16]
72. van Weel C, Kassai R, Qidwai W, Kumar R, Bala K, Prasad Gupta P, et al. Primary healthcare policy implementation in South Asia. BMJ Glob Health 2016;1(2):e000057 [FREE Full text] [doi: 10.1136/bmjgh-2016-000057] [Medline: 28588938]
73. Yang JX. Integrate health into all policies. Health People. 2016 Nov 03. URL: http://health.people.com.cn/n1/2016/1103/c398004-28830661.html [accessed 2020-05-07]
74. Rubinstein A, Barani M, Lopez AS. Quality first for effective universal health coverage in low-income and middle-income countries. Lancet Glob Health 2018 Nov;6(11):e1142-e1143 [FREE Full text] [doi: 10.1016/S2214-109X(18)30447-9] [Medline: 30322645]
75. The General Office of the State Council issued the outline of the national medical and health service system. The State Council of the People's Republic of China. URL: http://www.gov.cn/zhengce/content/2015-03/30/content_9560.htm [accessed 2020-05-07]
76. Zhang L, Cheng G, Song S, Yuan B, Zhu W, He L, et al. Efficiency performance of China's health care delivery system. Int J Health Plann Manage 2017 Jul;32(3):254-263. [doi: 10.1002/hpm.2425] [Medline: 28589685]
77. National Health Commission. China Health Statistical Yearbook 2018. Beijing: China Union Medical College Press; 2018.
78. The Lancet. China's health-care reform: an independent evaluation. Lancet 2019 Sep 28;394(10204):1113. [doi: 10.1016/S0140-6736(19)32210-X] [Medline: 31571585]

Abbreviations

AAPC: average annual percent change
APC: annual percent change
CDC: Center for Disease Prevention and Control
CHC: community health center
CNKI: China National Knowledge Infrastructure
EHR: electronic health recorded
EPI: Expanded Program on Immunization
GDP: gross domestic product
GP: general practitioner
HiAP: Health in All Policies
IT: information technology
NCD: noncommunicable disease
NHC: National Health Commission
NHS: National Health Service
TCM: traditional Chinese medicine
THC: township health center
UHC: universal health coverage
UN: United Nations
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.06.21; peer-reviewed by N Reynoso-Noverón, V Minichiello; comments to author 08.11.21; revised version received 05.02.22; accepted 27.04.22; published 22.07.22.

Please cite as:
Fang G, Yang D, Wang L, Wang Z, Liang Y, Yang J
Experiences and Challenges of Implementing Universal Health Coverage With China's National Basic Public Health Service Program: Literature Review, Regression Analysis, and Insider Interviews
JMIR Public Health Surveill 2022;8(7):e31289
URL: https://publichealth.jmir.org/2022/7/e31289
doi:10.2196/31289
PMID:35867386

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org