JMIR Public Health and Surveillance

Impact Factor (2023): 3.5
Volume 8 (2022), Issue 12 ISSN 2369-2960 Editor in Chief: Travis Sanchez, PhD, MPH

Contents

Reviews

Original Papers

Short Paper

Review

An Assessment of the Predictive Performance of Current Machine Learning–Based Breast Cancer Risk Prediction Models: Systematic Review

Ying Gao[1], PhD; Shu Li[2], PhD; Yujing Jin[1], MM; Lengxiao Zhou[1], MM; Shaomei Sun[1], MM; Xiaoqian Xu[1], MM; Shuqian Li[1], MM; Hongxi Yang[3], PhD; Qing Zhang[1], MM; Yaogang Wang[4], MD, PhD

[1]Health Management Center, Tianjin Medical University General Hospital, Tianjin, China

[2]School of Management, Tianjin University of Traditional Chinese Medicine, Tianjin, China

[3]Department of Bioinformatics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China

[4]School of Public Health, Tianjin Medical University, Tianjin, China

Corresponding Author:
Yaogang Wang, MD, PhD
School of Public Health, Tianjin Medical University
Qixiangtai Road 22, Heping District
Tianjin, 300070
China
Phone: 86 02260361238
Fax: 86 02260361234
Email: wangyg@tmu.edu.cn

Abstract

Background: Several studies have explored the predictive performance of machine learning–based breast cancer risk prediction models and have shown controversial conclusions. Thus, the performance of the current machine learning–based breast cancer risk prediction models and their benefits and weakness need to be evaluated for the future development of feasible and efficient risk prediction models.

Objective: The aim of this review was to assess the performance and the clinical feasibility of the currently available machine learning–based breast cancer risk prediction models.

Methods: We searched for papers published until June 9, 2021, on machine learning–based breast cancer risk prediction models in PubMed, Embase, and Web of Science. Studies describing the development or validation models for predicting future breast cancer risk were included. The Prediction Model Risk of Bias Assessment Tool (PROBAST) was used to assess the risk of bias and the clinical applicability of the included studies. The pooled area under the curve (AUC) was calculated using the DerSimonian and Laird random-effects model.

Results: A total of 8 studies with 10 data sets were included. Neural network was the most common machine learning method for the development of breast cancer risk prediction models. The pooled AUC of the machine learning–based optimal risk prediction model reported in each study was 0.73 (95% CI 0.66-0.80; approximate 95% prediction interval 0.56-0.96), with a high level of heterogeneity between studies (Q=576.07, I^2=98.44%; P<.001). The results of head-to-head comparison of the performance difference between the 2 types of models trained by the same data set showed that machine learning models had a slightly higher advantage than traditional risk factor–based models in predicting future breast cancer risk. The pooled AUC of the neural network–based risk prediction model was higher than that of the nonneural network–based optimal risk prediction model (0.71 vs 0.68, respectively). Subgroup analysis showed that the incorporation of imaging features in risk models resulted in a higher pooled AUC than the nonincorporation of imaging features in risk models (0.73 vs 0.61; $P_{\text{heterogeneity}}$=.001, respectively). The PROBAST analysis indicated that many machine learning models had high risk of bias and poorly reported calibration analysis.

Conclusions: Our review shows that the current machine learning–based breast cancer risk prediction models have some technical pitfalls and that their clinical feasibility and reliability are unsatisfactory.

(JMIR Public Health Surveill 2022;8(12):e35750) doi:10.2196/35750

KEYWORDS

breast cancer; machine learning; risk prediction; cancer; oncology; systemic review; review; meta-analysis; cancer research; risk model

Introduction

Of all the cancers worldwide among women, breast cancer shows the highest incidence and mortality [1]. Early access to effective diagnostic and treatment services after breast cancer screening could have reduced breast cancer mortality by 25%-40% over the last several decades [2,3]. The development and implementation of risk-based breast cancer control and prevention strategies can have great potential benefits and important public health implications. Moreover, risk-based breast cancer control and prevention strategy is more effective and efficient than conventional screening based on model evaluation [4,5]. A prerequisite for the implementation of personalized risk-adapted screening intervals is accurate breast cancer risk assessment [6]. Models with high sensitivity and specificity can enable screening to target more elaborate efforts for high-risk groups while minimizing overtreatment for the rest. Currently, the US breast cancer screening guidelines use breast cancer risk assessments to inform the clinical course, thereby targeting the high-risk population by earlier detection and lesser screening harms (eg, false-positive results, overdiagnosis, overtreatment, increased patient anxiety) [7]. Nevertheless, there is no standardized approach for office-based breast cancer risk assessment worldwide.

Traditional risk factor–based models such as Gail, BRCAPRO, Breast Cancer Surveillance Consortium, Claus, and Tyrer-Cuzick models have been well-validated and used commonly in clinical practice, but these models developed by logistic regression or Cox regression or those presented as risk scoring systems have low discrimination accuracy with the area under the receiver operating characteristic curve (AUC) between 0.53 and 0.64 [8-12] and these models show bias when applied to minority populations, accompanied by great variance in terms of the patients included, methods of development, predictors, outcomes, and presentations [13-15]. Other risk prediction models that incorporated genetic risk factors were also only best suited for specific clinical scenarios and may have limited applicability in certain types of patients [16]. Recently, with the cross research between artificial intelligence and medicine, the development and validation of breast cancer risk prediction models based on machine learning algorithms have been the current research focus. Machine learning algorithms provide an alternative approach to standard prediction modelling, which may address the current limitations and improve the prediction accuracy of breast cancer susceptibility [17,18]. Mammography is the most commonly used method for breast cancer screening or early detection. Machine learning artificial intelligence models suggest that mammographic images contain risk indicators and germline genetic data that can be used to improve and strengthen the existing risk prediction models [19]. Some studies claim that machine learning–based breast cancer risk prediction models are better than regression method–based models [7,20], but 1 study reported the opposite result [21]. These controversial conclusions prompted us to review the performance and the weaknesses of machine learning–based breast cancer risk prediction models. Therefore, this systematic review and meta-analysis aims to assess the performance and clinical feasibility of the currently available machine learning–based breast cancer risk prediction models.

Methods

Study Protocol

This systematic review and meta-analysis was performed according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement [22], the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies, and the prediction model performance guidelines [23,24].

Literature Search Strategy

Papers on machine learning–based breast cancer risk prediction models were searched in PubMed, Embase, and Web of Science by using the terms “machine learning OR deep learning” AND “mammary OR breast cancer OR carcinoma OR tumor OR neoplasm” AND “risk assessment OR risk prediction” published until June 9, 2021, and limited to papers published in English. The complete search strategy is detailed in Multimedia Appendix 1. Reviews in this field and references of the original papers were also manually checked to identify whether there were any missed studies.

Inclusion and Exclusion Criteria

Studies describing development or validation models for predicting future breast cancer risk were included in our study. The inclusion criteria were as follows: (1) breast cancer risk prediction model developed using a machine learning algorithm, (2) mean follow-up period for cohort studies should be longer than 1 year, and (3) future breast cancer risk is the assessment result. The exclusion criteria were as follows: (1) review or conference or editorial or only published abstracts, (2) the original full text not available or incomplete information, and (3) studies with no AUC or C-statistic and its 95% CI. When papers included the same population, studies with larger sample size or longer follow-up periods were finally included.

Data Extraction and Study Quality

Two researchers independently collected data on the first author, publication year, geographic region, study design, study population, sample size, study period, age of participants, time point for breast cancer risk prediction, name of the risk prediction model, number of participants and cancer cases in test data set, input risk factors, development and verification methods, and AUC with its 95% CI. The Prediction Model Risk of Bias Assessment Tool (PROBAST) was used to assess the risk of bias (ROB) and the clinical applicability of the included studies [25,26]. Any discrepancies were resolved by consensus or were consulted with the corresponding author.

Statistical Analyses

The discrimination value was assessed by AUC, which measures the machining learning risk prediction model ability to distinguish the women who will and will not develop breast cancer. An AUC of 0.5 was considered as no discrimination, whereas 1.0 indicated perfect discrimination. We calculated the pooled AUC of the risk models by using DerSimonian and Laird's random-effects model [27]. A head-to-head performance comparison of the studies that developed machine learning models and those that developed traditional risk factor–based models can help us understand the performance gain of utilizing machine learning methods in the same experimental setting. The Q test and I^2 value were employed to evaluate the heterogeneity among the studies. High values in both tests (I^2>40%, a significant Q test value with P<.05) showed high levels of inconsistency and heterogeneity. We also calculated an approximate 95% prediction interval (PI) to depict the extent of between-study heterogeneity [28]. Sensitivity analysis was performed to assess the influence of each study on the pooled effects by omitting each study. The visualized asymmetry of the funnel plot and Egger regression test were assessed for the publication bias. Pooled effects were also adjusted using the Duval and Tweedie trim-and-fill method [29,30]. All statistical meta-analyses of the predictive performance were performed using the MedCalc statistical software version 20 (MedCalc Ltd).

Results

Eligible Papers and Study Characteristics

A total of 937 papers were identified, and 8 studies with 10 data sets met our inclusion criteria and they were finally included in the meta-analysis (Figure 1) [7,19-21,31-34]. The primary characteristics of the included studies are summarized in Table 1. Totally, 218,100 patients were included in this review. Most of these patients were from America and Europe; only 1 data set's participants were from Taiwan, China. Six studies [7,20,21,32-34] predicted short-term (≤5 year) breast cancer risk, while 2 studies [19,31] predicted long-term (future or lifetime) risk. The characteristics and performance of the machine learning–based breast cancer risk prediction models are summarized in Table 2. Most of the machine learning prediction models were development models; only 1 study [7] used 3 different ethnic groups for external validation. Neural network was the most common machine learning method for the development of breast cancer risk prediction models. Only 1 neural network–based model incorporated genetic risk factors [7] and 6 neural network–based models incorporated imaging features [7,20,31,32].

Figure 1. Flowchart of the study selection in this systematic review. AUC: area under the curve.

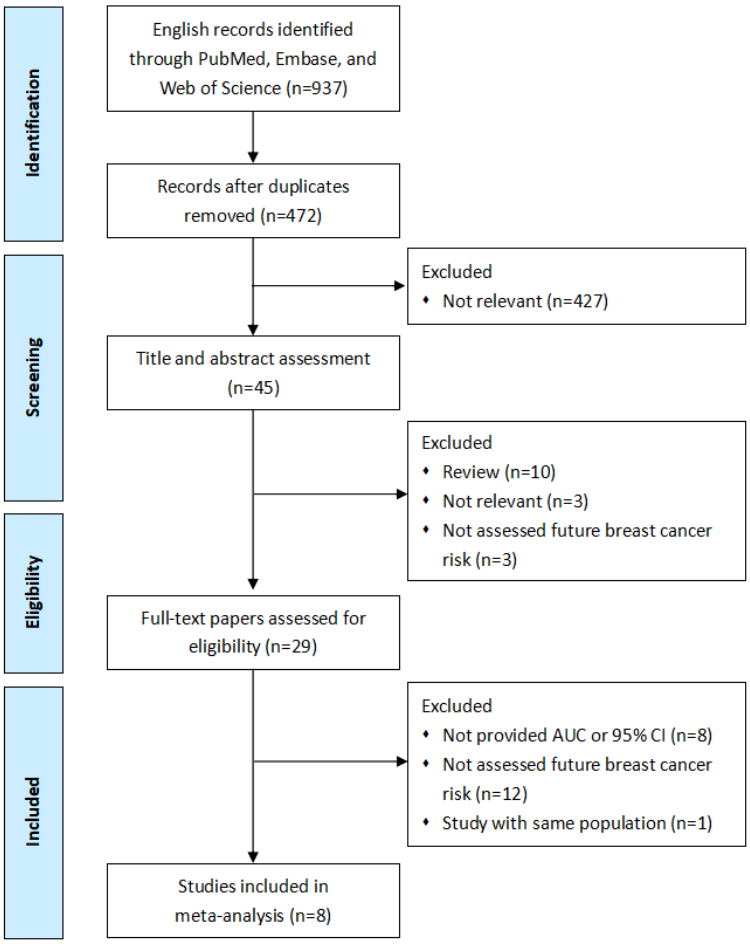


Table 1. Characteristics of the included studies on the machine learning–based breast cancer risk prediction models.

Study ID	Study design	Study population, geographic location	Sample size	Age (years)	Study period	Breast cancer risk	Participants in test data set (n)	Cancers in test data set (n)
Yala et al [7], 2021	Retrospective study	Massachusetts General Hospital, USA	70,972	40-80	2009-2016	5 years	7005	588
Yala et al [7], 2021	Retrospective study	Cohort of Screen-Aged Women, Karolinska University Hospital, Sweden	7353	40-74	2008-2016	5 years	7353	1413
Yala et al [7], 2021	Retrospective study	Chang Gung Menoral Hospital, Taiwan	13,356	40-70	2010-2011	5 years	13,356	244
Ming et al [19], 2020	Retrospective study	Oncogenetic Unit, Geneva University Hospital, Sweden	45,110	20-80	1998-2017	Lifetime	36,146	4911
Portnoi et al [20], 2019	Retrospective study	A large tertiary academic medical center, Massachusetts General Hospital, USA	1183	40-80	2011-2013	5 years	1164	96
Stark et al [21], 2019	Prospective study	Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial data set, USA	64,739	50-78	1993-2001	5 years	12,948	269
Dembrower et al [31], 2020	Retrospective study	Cohort of Screen-Aged Women, Karolinska University Hospital, Sweden	14,034	40-74	2008-2015	Future	2283	278
Arefan et al [32], 2020	Retrospective case-control study cohort	Health Insurance Portability and Accountability Act, USA	226	41-89	2013	Short-term	226	113
Tan et al [33], 2013	Retrospective study	University of Oklahoma Medical Center, USA	994	—[a]	2006	12-36 months	994	283
Saha et al [34], 2019	Retrospective study	Duke University School of Medicine, USA	133	27-76	2004-2013	2 years	133	46

[a]Not available.

Table 2. Characteristics and performance of the machine learning–based breast cancer risk prediction models.

Study ID, model name	Statistical method	Development/validation model	Model input parameters	Incorporation of imaging features	AUC[a] (95% CI)
Yala et al [7], 2021					
Tyrer-Cuzick model[b]	Logistic regression	—[c]	Age, weight, height, menarche age, given birth, menopause status, hormone replacement therapy usage, *BRCA* gene, ovarian cancer, breast biopsy, family history, hormonal factors	No	0.62 (0.59-0.66)
Radiolosit BI-RADS[d] model[e]	Logistic regression	Development model	Mammographic features	Yes	0.62 (0.60-0.65)
Image- and heatmaps model	Convolutional neural network	Development model	—	Yes	0.64 (0.60-0.68)
Imaged-only deep learning model	Convolutional neural network	Development model	Mammographic features	Yes	0.73 (0.70-0.77)
Hybrid deep learning model	Convolutional neural network	Development model	Age, weight, height, menarche age, given birth, menopause status, hormone replacement therapy usage, *BRCA* gene, ovarian cancer, breast biopsy, family history, hormonal factors	Yes	0.72 (0.69-0.76)
Mirai without risk factors model[f]	Convolutional neural network	Development model	Mammographic features	Yes	0.76 (0.73-0.79)
Mirai with risk factors model	Convolutional neural network	Development model	Age, weight, height, menarche age, given birth, menopause status, hormone replacement therapy usage, *BRCA* gene, ovarian cancer, breast biopsy, family history, hormonal factors	Yes	0.76 (0.73-0.80)
Yala et al [7], 2021					
Imaged-only deep learning model	Convolutional neural network	Validation model	Mammographic features	Yes	0.71 (0.69-0.73)
Mirai without risk factors model[f]	Convolutional neural network	Validation model	Mammographic features	Yes	0.78 (0.76-0.80)
Yala et al [7], 2021					
Imaged-only deep learning model	Convolutional neural network	Validation model	Mammographic features	Yes	0.70 (0.66-0.73)
Mirai without risk factors model[f]	Convolutional neural network	Validation model	Mammographic features	Yes	0.79 (0.75-0.82)
Ming et al [19], 2020					
BOADICEA[g] model	Logistic regression	—	Family pedigree, age, age at menarche, age at first live birth, parity, age at menopause, Ashkenazi Jewish ancestry, ovarian, prostate, pancreatic, contralateral, and lung/bronchus cancer diagnosis and age of onset, estrogen receptor status, progesterone receptor status, *HER2* status, and *BRCA/BRCA2* germline pathogenic variant	No	0.639[h]
Machine learning-Markov Chain Monte Carlo generalized linear mixed model	Markov Chain Monte Carlo	Development model	Family pedigree, age, age at menarche, age at first live birth, parity, age at menopause, Ashkenazi Jewish ancestry, ovarian, prostate, pancreatic, contralateral, and lung/bronchus cancer diagnosis and age of onset, estrogen receptor status, progesterone receptor status, *HER2* status, and *BRCA/BRCA2* germline pathogenic variant	No	0.851 (0.847-0.856)

Study ID, model name	Statistical method	Development/validation model	Model input parameters	Incorporation of imaging features	AUC[a] (95% CI)
Machine learning-adaptive boosting model[e,f]	Adaptive boosting	Development model	Family pedigree, age, age at menarche, age at first live birth, parity, age at menopause, Ashkenazi Jewish ancestry, ovarian, prostate, pancreatic, contralateral, and lung/bronchus cancer diagnosis and age of onset, estrogen receptor status, progesterone receptor status, *HER2* status, and *BRCA/BRCA2* germline pathogenic variant	No	0.889 (0.875-0.903)
Machine learning-random forest model	Random forest	Development model	Family pedigree, age, age at menarche, age at first live birth, parity, age at menopause, Ashkenazi Jewish ancestry, ovarian, prostate, pancreatic, contralateral, and lung/bronchus cancer diagnosis and age of onset, estrogen receptor status, progesterone receptor status, *HER2* status, and *BRCA/BRCA2* germline pathogenic variant	No	0.843 (0.838-0.849)
Portnoi et al [20], 2019					
Traditional risk factors logistic regression model[e]	Logistic regression	Development model	Age, weight, height, breast density, age at menarche, age at first live birth, menopause, hormone replacement therapy usage, had gene mutation, had ovarian cancer, had breast biopsy, number of first-degree relatives who have had breast cancer, race/ethnicity, history of breast cancer, and background parenchymal enhancement on magnetic resonance images	No	0.558 (0.492-0.624)
Magnetic resonance image-deep convolutional neural network model[f]	Convolutional neural network	Development model	Full-resolution magnetic resonance images	Yes	0.638 (0.577-0.699)
Tyrer-Cuzick model[b]	Logistic regression	—	Age, weight, height, breast density, age at menarche, age at first live birth, menopause, hormone replacement therapy usage, had gene mutation, had ovarian cancer, had breast biopsy, number of first-degree relatives who have had breast cancer, and race/ethnicity, and history of breast cancer	No	0.493 (0.353-0.633)
Stark et al [21], 2019					
Feed-forward artificial neural network model	Artificial neural network	Development model	Age, age at menarche, age at first live birth, number of first-degree relatives who have had breast cancer, race/ethnicity, age at menopause, an indicator of current hormone usage, number of years of hormone usage, BMI, pack years of cigarettes smoked, years of birth control usage, number of liver births, an indicator of personal prior history of cancer	No	0.608 (0.574-0.643)
Logistic regression model[e,f]	Logistic regression	Development model	Age, age at menarche, age at first live birth, number of first-degree relatives who have had breast cancer, and race/ethnicity, age at menopause, an indicator of current hormone usage, number of years of hormone usage, BMI, pack years of cigarettes smoked, years of birth control usage, number of liver births, an indicator of personal prior history of cancer	No	0.613 (0.579-0.647)
Gaussian naive Bayes model	Gaussian naive Bayes	Development model	Age, age at menarche, age at first live birth, number of first-degree relatives who have had breast cancer, and race/ethnicity, age at menopause, an indicator of current hormone usage, number of years of hormone usage, BMI, pack years of cigarettes smoked, years of birth control usage, number of liver births, an indicator of personal prior history of cancer	No	0.589 (0.555-0.623)

Study ID, model name	Statistical method	Development/validation model	Model input parameters	Incorporation of imaging features	AUC[a] (95% CI)
Decision tree model	Decision tree	Development model	Age, age at menarche, age at first live birth, number of first-degree relatives who have had breast cancer, and race/ethnicity, age at menopause, an indicator of current hormone usage, number of years of hormone usage, BMI, pack years of cigarettes smoked, years of birth control usage, number of liver births, an indicator of personal prior history of cancer	No	0.508 (0.496-0.521)
Linear discriminant analysis model	Linear discriminant analysis	Development model	Age, age at menarche, age at first live birth, number of first-degree relatives who have had breast cancer, and race/ethnicity, age at menopause, an indicator of current hormone usage, number of years of hormone usage, BMI, pack years of cigarettes smoked, years of birth control usage, number of liver births, an indicator of personal prior history of cancer	No	0.613 (0.579-0.646)
Support vector machine model	Support vector machine	Development model	Age, age at menarche, age at first live birth, number of first-degree relatives who have had breast cancer, and race/ethnicity, age at menopause, an indicator of current hormone usage, number of years of hormone usage, BMI, pack years of cigarettes smoked, years of birth control usage, number of liver births, an indicator of personal prior history of cancer	No	0.518 (0.484-0.551)
Breast Cancer Risk Prediction Tool model[b]	Logistic regression	—	Age, age at menarche, age at first live birth, number of first-degree relatives who have had breast cancer, and race/ethnicity, age at menopause, an indicator of current hormone usage, number of years of hormone usage, BMI, pack years of cigarettes smoked, years of birth control usage, number of liver births, an indicator of personal prior history of cancer	No	0.563 (0.528-0.597)
Dembrower et al [31], 2020					
Deep learning risk score model	Deep neural network	Development model	Mammographic images, the age at image acquisition, exposure, tube current, breast thickness, and compression force	Yes	0.65 (0.63-0.66)
Dense area model[b,e]	Logistic regression	Development model	Mammographic features	Yes	0.58 (0.57-0.60)
Percentage density model[b]	Logistic regression	Development model	Mammographic features	Yes	0.54 (0.52-0.56)
Deep learning risk score + dense area + percentage density model[f]	Deep neural network	Development model	Mammographic images, the age at image acquisition, exposure, tube current, breast thickness, and compression force	Yes	0.66 (0.64-0.67)
Arefan et al [32], 2020					
End-to-end convolutional neural network model using GoogLeNet	Convolutional neural network	Development model	Imaging features of the whole-breast region	Yes	0.62 (0.58-0.66)
End-to-end convolutional neural network model using GoogLeNet	Convolutional neural network	Development model	Imaging features of the dense breast region only	Yes	0.67 (0.61-0.73)

Study ID, model name	Statistical method	Development/validation model	Model input parameters	Incorporation of imaging features	AUC[a] (95% CI)
GoogLeNet combining a linear discriminant analysis model	Linear discriminant analysis	Development model	Imaging features of the whole-breast region	Yes	0.64 (0.58-0.70)
GoogLeNet combining a linear discriminant analysis model[e,f]	Linear discriminant analysis	Development model	Imaging features of the dense breast region only	Yes	0.72 (0.67-0.76)
Area-based percentage breast density model[b]	Logistic regression	Development model	Percentage breast density	Yes	0.54 (0.49-0.59)
Tan et al [33], 2013					
Support vector machine classification model[e,f]	Support vector machine classification	Validation model	Age, family history, breast density, mean pixel value difference, mean value of short run emphasis; maximum value of short run emphasis, standard deviation of the r-axis cumulative projection histogram, standard deviation of the y-axis cumulative projection histogram, median of the x-axis cumulative projection histogram, mean pixel value, mean value of short run low gray-level emphasis, and median of the x-axis cumulative projection histogram	Yes	0.725 (0.689-0.759)
Saha et al [34], 2019					
Mean reader scores model[b]	Logistic regression	Development model	—	Yes	0.59 (0.49-0.70)
Median reader scores model[b]	Logistic regression	Development model	—	Yes	0.60 (0.51-0.69)
Machine learning model 1	Machine learning logistic regression	Development model	Magnetic resonance image background parenchymal enhancement features were based on the fibroglandular tissue mask on the fat saturated sequence	Yes	0.63 (0.52-0.73)
Machine learning model 2[e,f]	Machine learning logistic regression	Development model	Magnetic resonance image background parenchymal enhancement features were based on the fibroglandular tissue segmentation using the non–fat-saturated sequence	Yes	0.70 (0.60-0.79)

[a]AUC: area under the curve.

[b]Traditional risk factor–based optimal breast cancer risk prediction model.

[c]Not available.

[d]BI-RADS: Breast Imaging-Reporting And Data System.

[e]Nonneural network–based optimal breast cancer risk prediction model.

[f]Machine learning–based optimal breast cancer risk prediction model.

[g]BOADICEA: Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm.

[h]95% CI not available.

Study Quality

PROBAST was used to assess the quality of the included studies in terms of both ROB and clinical applicability. All 8 studies demonstrated a low applicability risk; only 1 of the papers had low ROB [7], indicating that most machine learning models have technical pitfalls (Table 3). The other 7 studies that had high ROB were mostly in the domain of analysis, with several reasons as follows: (1) no information was provided on how the continuous/categorical predictors handle or they were handled unreasonably, (2) complexities in the data were not assessed in the final analysis, (3) model calibration was not assessed or lack of standardized evaluation of model calibration, (4) the calculation formulae of the predictors and their weights were not reported in the final model, and (5) insufficient number

of participants was used to develop the models. The details are shown in Multimedia Appendix 2. Only 3 neural network–based models were developed by bootstrap and cross-validation to evaluate the discrimination ability of the prediction model, whereas other machine learning models and regression models were developed by using random split or nonrandom split.

Table 3. Presentation of the Prediction Model Risk of Bias Assessment Tool results of the included studies.

Study	Risk of bias				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	Risk of bias	Applicability
Yala et al [7], 2021	LR[a]	LR	LR	HR[b]	LR	LR	LR	LR	LR
Ming et al [19], 2020	LR	HR	LR	HR	LR	LR	LR	HR	LR
Portnoi et al [20], 2019	LR	LR	LR	HR	LR	LR	LR	HR	LR
Stark et al [21], 2019	LR	LR	LR	HR	LR	LR	LR	HR	LR
Dembrower et al [31], 2020	LR	LR	LR	HR	LR	LR	LR	HR	LR
Arefan et al [32], 2020	LR	LR	LR	HR	LR	LR	LR	HR	LR
Tan et al [33], 2013	LR	LR	LR	HR	LR	LR	LR	HR	LR
Saha et al [34], 2019	LR	LR	LR	HR	LR	LR	LR	HR	LR

[a]LR: low risk.

[b]HR: high risk.

Predictive Performance

The pooled AUC of the machine learning–based optimal breast cancer risk prediction model reported in each included study was 0.73 (95% CI 0.66-0.80; approximate 95% PI 0.56-0.96), with a high level of heterogeneity between studies (Q=576.07, I^2=98.44%; P<.001) (Figure 2). We also performed metaregression, and the results showed that the heterogeneity remains high and essentially unchanged. Sensitivity analysis showed that the pooled AUC and 95% CI were not significantly altered before and after the omission of each data set, with a range of 0.72 (95% CI 0.67-0.76; approximate 95% PI 0.60-0.85) to 0.75 (95% CI 0.68-0.82; approximate 95% PI 0.57-0.98) (Multimedia Appendix 3). The results of head-to-head comparison of the performance difference in both types of models trained by the same data set showed that the pooled AUC of machine learning prediction models (0.69, 95% CI 0.63-0.74; approximate 95% PI 0.57-0.83; Figure 3A) was higher than that of the traditional risk factor–based models, with the range from 0.56 (95% CI 0.55-0.58; approximate 95% PI 0.51-0.62) to 0.58 (95% CI 0.57-0.59; approximate 95% PI 0.51-0.62) (all $P_{\text{heterogeneity}}$<.001) (Figures 3B-3E).

Figure 2. Forest plot of the pooled area under the curve of the machine learning–based optimal breast cancer risk prediction model [7,19-21,31-34]. AUC: area under the curve.

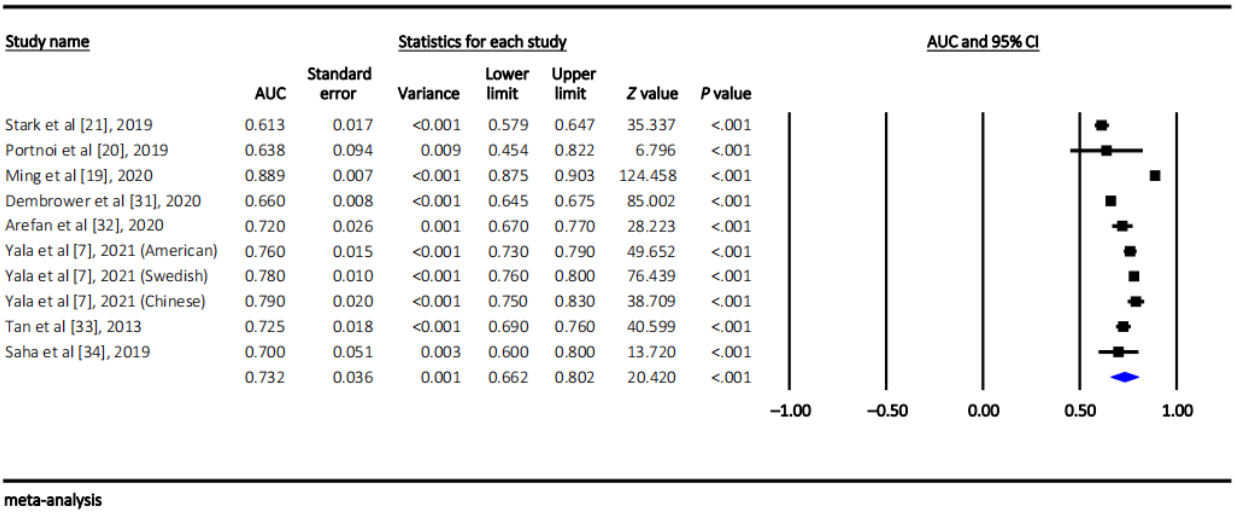
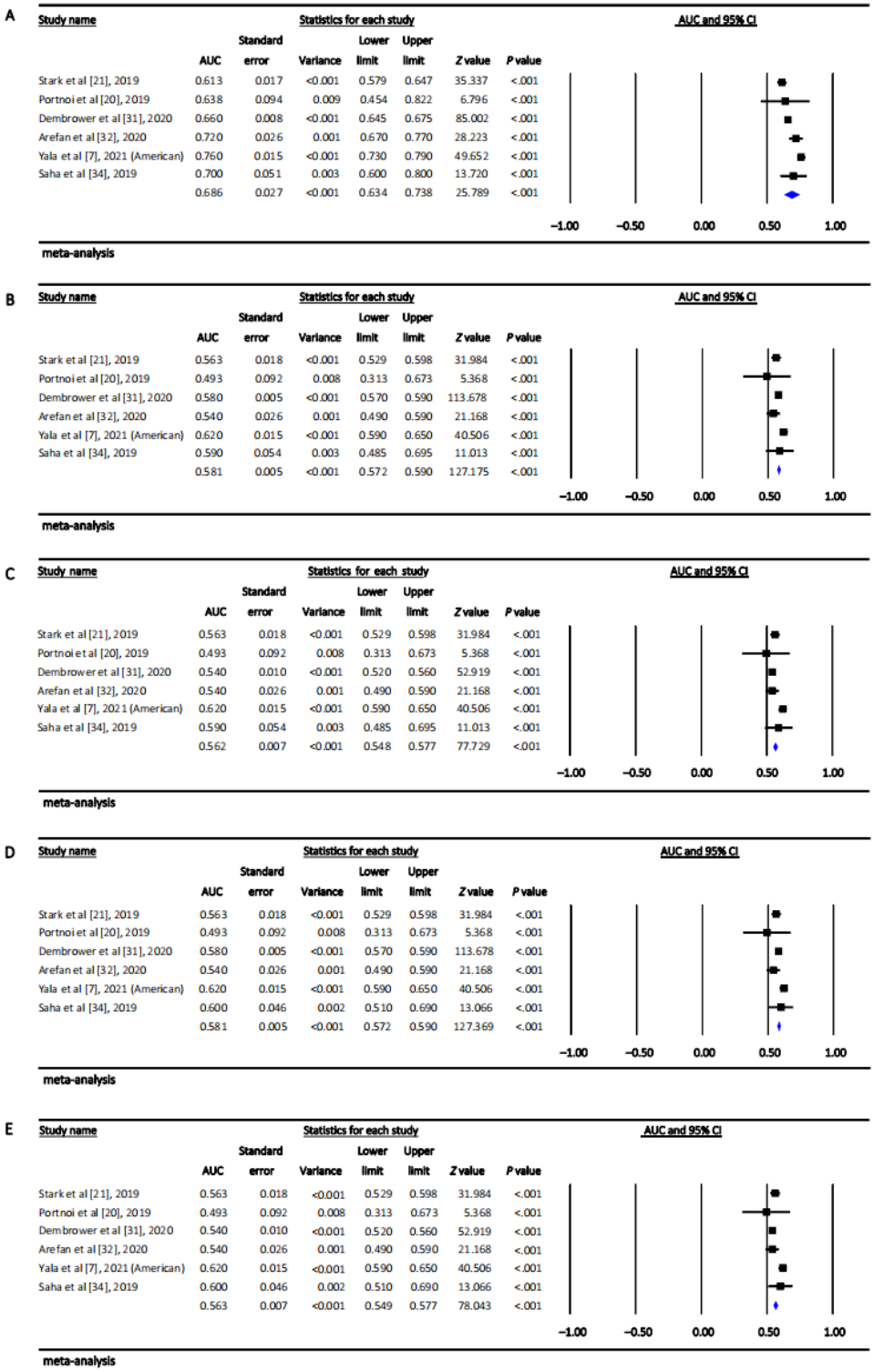


Figure 3. Forest plot of the pooled area under the curve in head-to-head comparisons of (A) machine learning models and (B,C,D,E) traditional risk factor–based models [7,20,21,31,32,34]. AUC: area under the curve.



The pooled AUC of neural network–based breast cancer risk prediction models was 0.71 (95% CI 0.65-0.77; approximate 95% PI 0.57-0.87; Q=131.42; I^2=95.43%; P<.001) (Figure 4A), which was higher than that of nonneural network–based optimal risk prediction models (0.68, 95% CI 0.56-0.81; approximate 95% PI 0.53-0.81; Q=1268.99; I^2=99.45%; P<.001) (Figure 4B). When stratified by the presence or absence of incorporation of imaging features, the pooled AUCs in models incorporated with imaging features and those in models not incorporated with imaging features were 0.73 (95% CI 0.67-0.79) and 0.61 (95% CI 0.57-0.64) ($P_{\text{heterogeneity}}$=.001), respectively (Table 4). Subgroup analysis also showed that the pooled AUC in models not incorporated with genetic risk factors was not significantly lower than that in models incorporated with genetic risk factors (0.71 vs 0.76, respectively; $P_{\text{heterogeneity}}$=.12) (Table 4). Our results also showed that models predicting short-term (≤5 year) breast cancer risk had a slightly higher pooled AUC than those

predicting long-term risk (0.72 vs 0.66, respectively), although the difference was not significant ($P_{heterogeneity}$=.10) (Table 4).

The funnel plot indicated that there was no publication bias, with an Egger regression coefficient of –3.85 (*P*=.46) (Multimedia Appendix 4). According to the trim-and-fill method, 2 studies had to be trimmed, and the adjusted pooled AUC was 0.75 (95% CI 0.69-0.82) after trimming (Multimedia Appendix 4).

Figure 4. Forest plot of the pooled area under the curve of the (A) neural network–based breast cancer risk prediction model and (B) nonneural network–based optimal risk prediction model [7,20,21,31,32]. AUC: area under the curve.

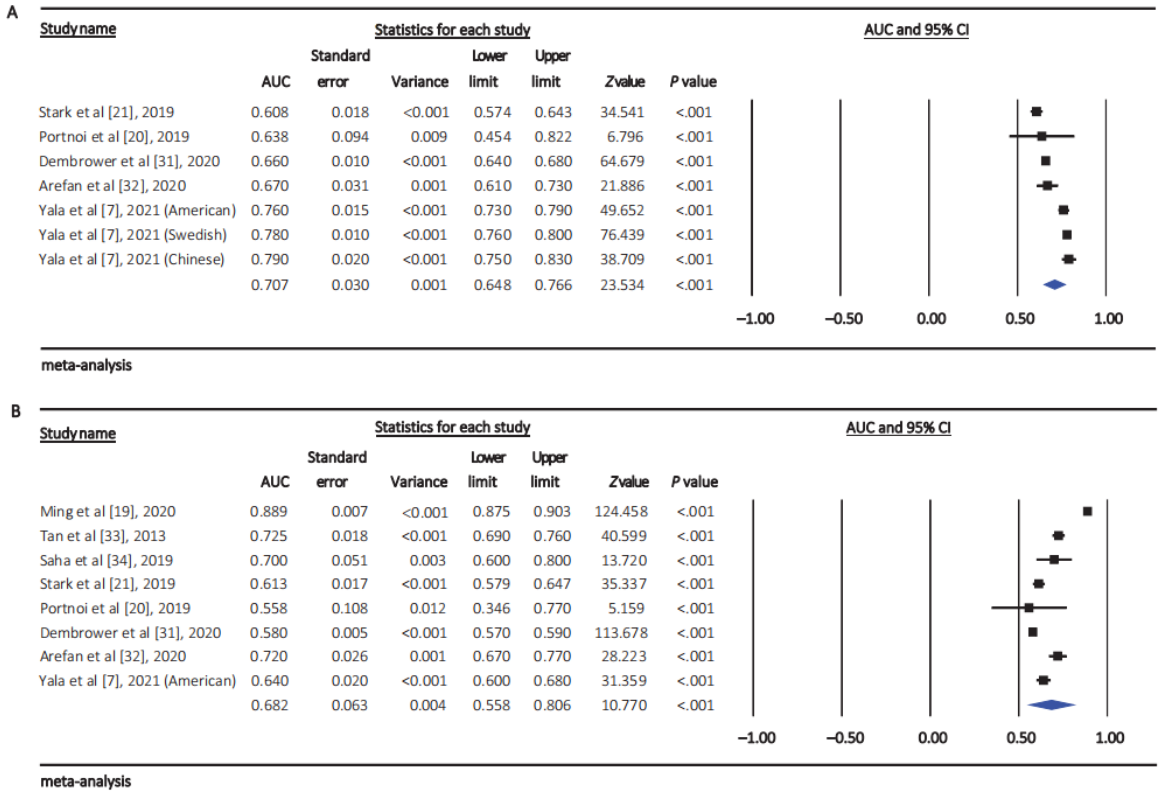


Table 4. Subgroup analysis.

Model, subgroup	Area under the curve (95% CI)	$P_{heterogeneity}$ value
Model with/without imaging features		.001
Model incorporated with imaging features	0.73 (0.67-0.79)	
Model not incorporated with imaging features	0.61 (0.57-0.64)	
Model with/without genetic risk factors		.12
Model incorporated with genetic risk factors	0.76 (0.73-0.80)	
Model not incorporated with genetic risk factors	0.71 (0.65-0.77)	
Model prediction of risk		.10
Model predicting short-term risk	0.72 (0.65-0.78)	
Model predicting long-term risk	0.66 (0.64-0.67)	

Discussion

Principal Findings

In this meta-analysis, 8 studies showed that the pooled AUC of machine learning–based breast cancer risk prediction models was 0.73 (95% CI 0.66-0.80). The results of head-to-head comparison of the performance difference in 2 types of models trained by the same data set showed that machine learning models had a slightly higher advantage than the traditional risk factor–based models in predicting future breast cancer risk. Machine learning approaches have the potential to achieve better accuracy and incorporate different types of information, including traditional risk factors, imaging features, genetic data, and clinical factors. However, of note, the predictive ability of the machine learning models showed substantial heterogeneity among the studies included in this review.

Machine learning represents a data-driven method; it has the ability to learn from past examples and detect hard-to-discern patterns from large and noisy data sets and model nonlinear and more complex relationships by employing a variety of statistical,

probabilistic, and optimization techniques [35]. This capability of machine learning algorithms offers a possibility for the investigation and development of risk prediction and diagnostic prediction models in cancer research [36]. It is evident that the use of machine learning methods can improve our understanding of cancer occurrence and progression [35,37]. Thus, developing machine learning–based breast cancer risk prediction models with improved discriminatory power can stratify women into different risk groups, which are useful for guiding the choice for personalized breast cancer screening in order to achieve a good balance in the risk benefit and cost benefit for breast cancer screening.

In our stratified analysis, neural network–based breast cancer risk prediction models incorporating imaging features showed superior performance. This result suggests that the incorporation of imaging inputs in machine learning models can deliver more accurate breast cancer risk prediction. Previous breast cancer risk assessments have already recognized the importance of imaging features in mammography [10,12], but the existing model was based on the underlying pattern that was assessed visually by radiologists, and the whole image was subjectively summarized as a density score on mammography as the model input [38]. It is unlikely that the single value of the density score would be able to take maximum advantage of the imaging features. The other human-specified features may not be able to capture all the risk-relevant information in the image. However, the flexibility of the neural networks might allow the extraction of more information from both finer patterns as well as the overall image characteristics, which can improve the accuracy of the prediction models.

The findings in this study showed that neural network–based models that predicted short-term (≤5 year) breast cancer risk had slightly better discriminatory accuracy than models predicting long-term risk, although confidence intervals overlapped. Improvement of public health literacy and the popularization of healthy lifestyles motivated more opportunities for women in their lifetime to participate in breast cancer prevention and screening and modify their identified modifiable risk factors associated with breast cancer. Unlike many currently known risk factors that do not change and maintain constant risk values, short-term risk factors may change over time. The cumulative effect of these changes may reduce the incidence of breast cancer. Therefore, it is unreasonable to predict the long-term risk of breast cancer by using these risk factors, which may lead to high probability of false-positive recall.

Model Reliability and Clinical Feasibility

Our study showed several issues regarding machine learning model reliability. The PROBAST analysis indicated that machine learning models have technical pitfalls. First, most machine learning models did not report sufficient statistical analysis information, and only few studies [7,31] provided the details for model reproduction. Second, many machine learning models showed a poor calibration analysis, indicating that the assessment of their utility was problematic, leading to inaccurate evaluation of the future breast cancer risk. Third, only 1 study [7] reported machine learning models that were externally validated in different ethnic populations. Six neural network–based models incorporated many complex imaging features, which may cause clinicians or public physicians to be unable to quickly and conveniently calculate the breast cancer risk by machine learning models manually. This may also be why few studies carry out external validation of the machine learning models. Due to the complexity of the machine learning model algorithms, many studies included many different types of predictors into the model construction, which may lead to an overfitting of the machine learning models [39]. However, only few development studies [7,21,34] reported the details for these predictor selection processes, which may lower the clinical feasibility of the machine learning models.

Limitations

This review had several limitations. First, most of the included studies [19,31-34] did not provide the expected/observed ratio or other indicators that could evaluate the calibration of the risk prediction model; therefore, this meta-analysis could not comprehensively review the calibration of the machine learning–based breast cancer risk prediction models. Second, substantial heterogeneity was presented in this systematic review, which impeded us from making further rigorous comparisons. The heterogeneity can be partially explained but could not be markedly diminished by different risk predicting times, with or without the incorporation of imaging features and genetic risk factors. The results of meta-analysis can only be interpreted carefully within the context. Third, the pooled results of the machine learning prediction model were analyzed based on most of the included studies that had high ROB [19-21,31-34]. The reason that these studies are rated as high ROB were that complexities in the data were not assessed or the calculation formulas of the predictors and their weights were not reported in the final model. These parameters, the so-called “black boxes,” are almost never presented in the original studies. Moreover, we performed a head-to-head fair comparison of the performance difference between 2 types of models trained by same data set, and the results showed that machine learning models had a slightly higher advantage in predicting future breast cancer risk. Lastly, we mainly focus on the statistical measures of model performance and did not discuss how to meta-analyze the clinical measures of performance such as net benefit. Hence, further research on how to meta-analyze net benefit estimates should be performed.

Conclusions

In summary, machine learning–based breast cancer risk prediction models had a slightly higher advantage in predicting future breast cancer risk than traditional risk factor–based models in head-to-head comparisons of the performance under the same experimental settings. However, machine learning–based breast cancer risk prediction models had some technical pitfalls, and their clinical feasibility and reliability were unsatisfactory. Future research may be worthwhile to obtain individual participant data to investigate in more detail how the machine learning models perform across different populations and subgroups. We also suggest that they could be considered to be implemented by pooling with breast cancer screening programs and to help developing optimal screening strategies, especially screening intervals.

Acknowledgments

The authors thank Professor Yuan Wang, who provided suggestions for the analysis and for editing this manuscript. This study was supported by the National Natural Science Foundation of China (grants 71804124, 71904142, 72104179).

Authors' Contributions

QZ and YW conceptualized the data. Shu Li and YJ curated the data. YG performed the formal analysis and wrote the original draft. YG, Shu Li, and LZ performed the methodology. SS and XX administered the project. Shuqian Li supervised this study. YG, Shu Li, and HY reviewed and edited the manuscript. All authors read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 17 KB - publichealth_v8i12e35750_app1.docx]

Multimedia Appendix 2
Details of risk of bias and the clinical applicability of the included studies.
[DOCX File , 16 KB - publichealth_v8i12e35750_app2.docx]

Multimedia Appendix 3
Sensitivity analysis of the pooled area under the curve of the machine learning–based breast cancer risk prediction models.
[DOCX File , 15 KB - publichealth_v8i12e35750_app3.docx]

Multimedia Appendix 4
Funnel plot of the discrimination of (A) machine learning–based breast cancer risk prediction model and (B) funnel plot adjusted by the trim-and-fill method. AUC: area under the curve.
[PNG File , 40 KB - publichealth_v8i12e35750_app4.png]

References

1. 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 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, International Agency for Research on Cancer Handbook Working Group. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med 2015 Jun 11;372(24):2353-2358. [doi: 10.1056/NEJMsr1504363] [Medline: 26039523]
3. Massat NJ, Dibden A, Parmar D, Cuzick J, Sasieni PD, Duffy SW. Impact of Screening on Breast Cancer Mortality: The UK Program 20 Years On. Cancer Epidemiol Biomarkers Prev 2016 Mar;25(3):455-462. [doi: 10.1158/1055-9965.EPI-15-0803] [Medline: 26646362]
4. Mühlberger N, Sroczynski G, Gogollari A, Jahn B, Pashayan N, Steyerberg E, et al. Cost effectiveness of breast cancer screening and prevention: a systematic review with a focus on risk-adapted strategies. Eur J Health Econ 2021 Nov;22(8):1311-1344. [doi: 10.1007/s10198-021-01338-5] [Medline: 34342797]
5. Arnold M, Pfeifer K, Quante AS. Is risk-stratified breast cancer screening economically efficient in Germany? PLoS One 2019;14(5):e0217213 [FREE Full text] [doi: 10.1371/journal.pone.0217213] [Medline: 31120970]
6. Brentnall AR, Cuzick J, Buist DSM, Bowles EJA. Long-term Accuracy of Breast Cancer Risk Assessment Combining Classic Risk Factors and Breast Density. JAMA Oncol 2018 Sep 01;4(9):e180174 [FREE Full text] [doi: 10.1001/jamaoncol.2018.0174] [Medline: 29621362]
7. Yala A, Mikhael PG, Strand F, Lin G, Smith K, Wan Y, et al. Toward robust mammography-based models for breast cancer risk. Sci Transl Med 2021 Jan 27;13(578):eaba4373. [doi: 10.1126/scitranslmed.aba4373] [Medline: 33504648]
8. Wang X, Huang Y, Li L, Dai H, Song F, Chen K. Assessment of performance of the Gail model for predicting breast cancer risk: a systematic review and meta-analysis with trial sequential analysis. Breast Cancer Res 2018 Mar 13;20(1):18 [FREE Full text] [doi: 10.1186/s13058-018-0947-5] [Medline: 29534738]
9. Amir E, Evans DG, Shenton A, Lalloo F, Moran A, Boggis C, et al. Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. J Med Genet 2003 Nov;40(11):807-814 [FREE Full text] [doi: 10.1136/jmg.40.11.807] [Medline: 14627668]

10. Brentnall AR, Harkness EF, Astley SM, Donnelly LS, Stavrinos P, Sampson S, et al. Mammographic density adds accuracy to both the Tyrer-Cuzick and Gail breast cancer risk models in a prospective UK screening cohort. Breast Cancer Res 2015 Dec 01;17(1):147 [FREE Full text] [doi: 10.1186/s13058-015-0653-5] [Medline: 26627479]
11. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res Treat 2012 Apr;132(2):365-377. [doi: 10.1007/s10549-011-1818-2] [Medline: 22037780]
12. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Ann Intern Med 2008 Mar 04;148(5):337-347 [FREE Full text] [doi: 10.7326/0003-4819-148-5-200803040-00004] [Medline: 18316752]
13. Gail MH, Costantino JP, Pee D, Bondy M, Newman L, Selvan M, et al. Projecting individualized absolute invasive breast cancer risk in African American women. J Natl Cancer Inst 2007 Dec 05;99(23):1782-1792. [doi: 10.1093/jnci/djm223] [Medline: 18042936]
14. Matsuno RK, Costantino JP, Ziegler RG, Anderson GL, Li H, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. J Natl Cancer Inst 2011 Jun 22;103(12):951-961 [FREE Full text] [doi: 10.1093/jnci/djr154] [Medline: 21562243]
15. Boggs DA, Rosenberg L, Adams-Campbell LL, Palmer JR. Prospective approach to breast cancer risk prediction in African American women: the black women's health study model. J Clin Oncol 2015 Mar 20;33(9):1038-1044 [FREE Full text] [doi: 10.1200/JCO.2014.57.2750] [Medline: 25624428]
16. Kim G, Bahl M. Assessing Risk of Breast Cancer: A Review of Risk Prediction Models. J Breast Imaging 2021;3(2):144-155 [FREE Full text] [doi: 10.1093/jbi/wbab001] [Medline: 33778488]
17. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219 [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]
18. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002;35(5-6):352-359 [FREE Full text] [doi: 10.1016/s1532-0464(03)00034-0] [Medline: 12968784]
19. Ming C, Viassolo V, Probst-Hensch N, Dinov ID, Chappuis PO, Katapodi MC. Machine learning-based lifetime breast cancer risk reclassification compared with the BOADICEA model: impact on screening recommendations. Br J Cancer 2020 Sep;123(5):860-867 [FREE Full text] [doi: 10.1038/s41416-020-0937-0] [Medline: 32565540]
20. Portnoi T, Yala A, Schuster T, Barzilay R, Dontchos B, Lamb L, et al. Deep Learning Model to Assess Cancer Risk on the Basis of a Breast MR Image Alone. American Journal of Roentgenology 2019 Jul;213(1):227-233. [doi: 10.2214/ajr.18.20813]
21. Stark GF, Hart GR, Nartowt BJ, Deng J. Predicting breast cancer risk using personal health data and machine learning models. PLoS One 2019;14(12):e0226765 [FREE Full text] [doi: 10.1371/journal.pone.0226765] [Medline: 31881042]
22. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
23. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014 Oct;11(10):e1001744 [FREE Full text] [doi: 10.1371/journal.pmed.1001744] [Medline: 25314315]
24. Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2017 Jan 05;356:i6460 [FREE Full text] [doi: 10.1136/bmj.i6460] [Medline: 28057641]
25. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, PROBAST Group†. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019 Jan 01;170(1):51-58 [FREE Full text] [doi: 10.7326/M18-1376] [Medline: 30596875]
26. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med 2019 Jan 01;170(1):W1-W33 [FREE Full text] [doi: 10.7326/M18-1377] [Medline: 30596876]
27. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
28. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011 Feb 10;342:d549. [doi: 10.1136/bmj.d549] [Medline: 21310794]
29. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
30. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
31. Dembrower K, Liu Y, Azizpour H, Eklund M, Smith K, Lindholm P, et al. Comparison of a Deep Learning Risk Score and Standard Mammographic Density Score for Breast Cancer Risk Prediction. Radiology 2020 Feb;294(2):265-272. [doi: 10.1148/radiol.2019190872] [Medline: 31845842]

32. Arefan D, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu S. Deep learning modeling using normal mammograms for predicting breast cancer risk. Med Phys 2020 Jan;47(1):110-118 [FREE Full text] [doi: 10.1002/mp.13886] [Medline: 31667873]
33. Tan M, Zheng B, Ramalingam P, Gur D. Prediction of near-term breast cancer risk based on bilateral mammographic feature asymmetry. Acad Radiol 2013 Dec;20(12):1542-1550 [FREE Full text] [doi: 10.1016/j.acra.2013.08.020] [Medline: 24200481]
34. Saha A, Grimm LJ, Ghate SV, Kim CE, Soo MS, Yoon SC, et al. Machine learning-based prediction of future breast cancer using algorithmically measured background parenchymal enhancement on high-risk screening MRI. J Magn Reson Imaging 2019 Aug;50(2):456-464 [FREE Full text] [doi: 10.1002/jmri.26636] [Medline: 30648316]
35. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inform 2007 Feb 11;2:59-77 [FREE Full text] [Medline: 19458758]
36. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8-17 [FREE Full text] [doi: 10.1016/j.csbj.2014.11.005] [Medline: 25750696]
37. Hutt S, Mihaies D, Karteris E, Michael A, Payne AM, Chatterjee J. Statistical Meta-Analysis of Risk Factors for Endometrial Cancer and Development of a Risk Prediction Model Using an Artificial Neural Network Algorithm. Cancers (Basel) 2021 Jul 22;13(15):3689 [FREE Full text] [doi: 10.3390/cancers13153689] [Medline: 34359595]
38. Tan M, Zheng B, Leader JK, Gur D. Association Between Changes in Mammographic Image Features and Risk for Near-Term Breast Cancer Development. IEEE Trans. Med. Imaging 2016 Jul;35(7):1719-1728. [doi: 10.1109/tmi.2016.2527619]
39. Sun Z, Dong W, Shi H, Ma H, Cheng L, Huang Z. Comparing Machine Learning Models and Statistical Models for Predicting Heart Failure Events: A Systematic Review and Meta-Analysis. Front Cardiovasc Med 2022;9:812276 [FREE Full text] [doi: 10.3389/fcvm.2022.812276] [Medline: 35463786]

Abbreviations

AUC: area under the curve
PI: prediction interval
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PROBAST: Prediction Model Risk of Bias Assessment Tool
ROB: risk of bias

Edited by A Mavragani, H Bradley; submitted 17.12.21; peer-reviewed by A Clift, A Spini; comments to author 10.05.22; revised version received 17.06.22; accepted 25.11.22; published 29.12.22.

Please cite as:
Gao Y, Li S, Jin Y, Zhou L, Sun S, Xu X, Li S, Yang H, Zhang Q, Wang Y
An Assessment of the Predictive Performance of Current Machine Learning–Based Breast Cancer Risk Prediction Models: Systematic Review
JMIR Public Health Surveill 2022;8(12):e35750
URL: https://publichealth.jmir.org/2022/12/e35750
doi:10.2196/35750
PMID:36426919

Original Paper

A Smartphone-Based Platform Assisted by Artificial Intelligence for Reading and Reporting Rapid Diagnostic Tests: Evaluation Study in SARS-CoV-2 Lateral Flow Immunoassays

David Bermejo-Peláez[1*], PhD; Daniel Marcos-Mencía[2*], MD; Elisa Álamo[1], MSc; Nuria Pérez-Panizo[3,4], NP; Adriana Mousa[1], MSc; Elena Dacal[1], PhD; Lin Lin[1,5], MSc; Alexander Vladimirov[1], MSc; Daniel Cuadrado[1], MSc; Jesús Mateos-Nozal[3,4], MD, PhD; Juan Carlos Galán[2,4,6], PhD; Beatriz Romero-Hernandez[2,4,6], PhD; Rafael Cantón[2,4,7], PhD; Miguel Luengo-Oroz[1], PhD; Mario Rodriguez-Dominguez[2,4,6], PhD

[1]Spotlab, Madrid, Spain

[2]Servicio de Microbiología, Hospital Universitario Ramon y Cajal, Madrid, Spain

[3]Servicio de Geriatría, Hospital Universitario Ramon y Cajal, Madrid, Spain

[4]Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Madrid, Spain

[5]Biomedical Image Technologies, ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

[6]CIBER en Epidemiología y Salud Pública (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain

[7]CIBER en Enfermedades Infecciosas (CIBERINFEC), Instituto de Salud Carlos III, Madrid, Spain

[*]these authors contributed equally

Corresponding Author:
Miguel Luengo-Oroz, PhD
Spotlab
P.º de Juan XXIII, 36B
Madrid, 28040
Spain
Phone: 34 916256927
Email: miguel@spotlab.ai

Abstract

Background: Rapid diagnostic tests (RDTs) are being widely used to manage COVID-19 pandemic. However, many results remain unreported or unconfirmed, altering a correct epidemiological surveillance.

Objective: Our aim was to evaluate an artificial intelligence–based smartphone app, connected to a cloud web platform, to automatically and objectively read RDT results and assess its impact on COVID-19 pandemic management.

Methods: Overall, 252 human sera were used to inoculate a total of 1165 RDTs for training and validation purposes. We then conducted two field studies to assess the performance on real-world scenarios by testing 172 antibody RDTs at two nursing homes and 96 antigen RDTs at one hospital emergency department.

Results: Field studies demonstrated high levels of sensitivity (100%) and specificity (94.4%, CI 92.8%-96.1%) for reading IgG band of COVID-19 antibody RDTs compared to visual readings from health workers. Sensitivity of detecting IgM test bands was 100%, and specificity was 95.8% (CI 94.3%-97.3%). All COVID-19 antigen RDTs were correctly read by the app.

Conclusions: The proposed reading system is automatic, reducing variability and uncertainty associated with RDTs interpretation and can be used to read different RDT brands. The web platform serves as a real-time epidemiological tracking tool and facilitates reporting of positive RDTs to relevant health authorities.

(JMIR Public Health Surveill 2022;8(12):e38533) doi:10.2196/38533

KEYWORDS

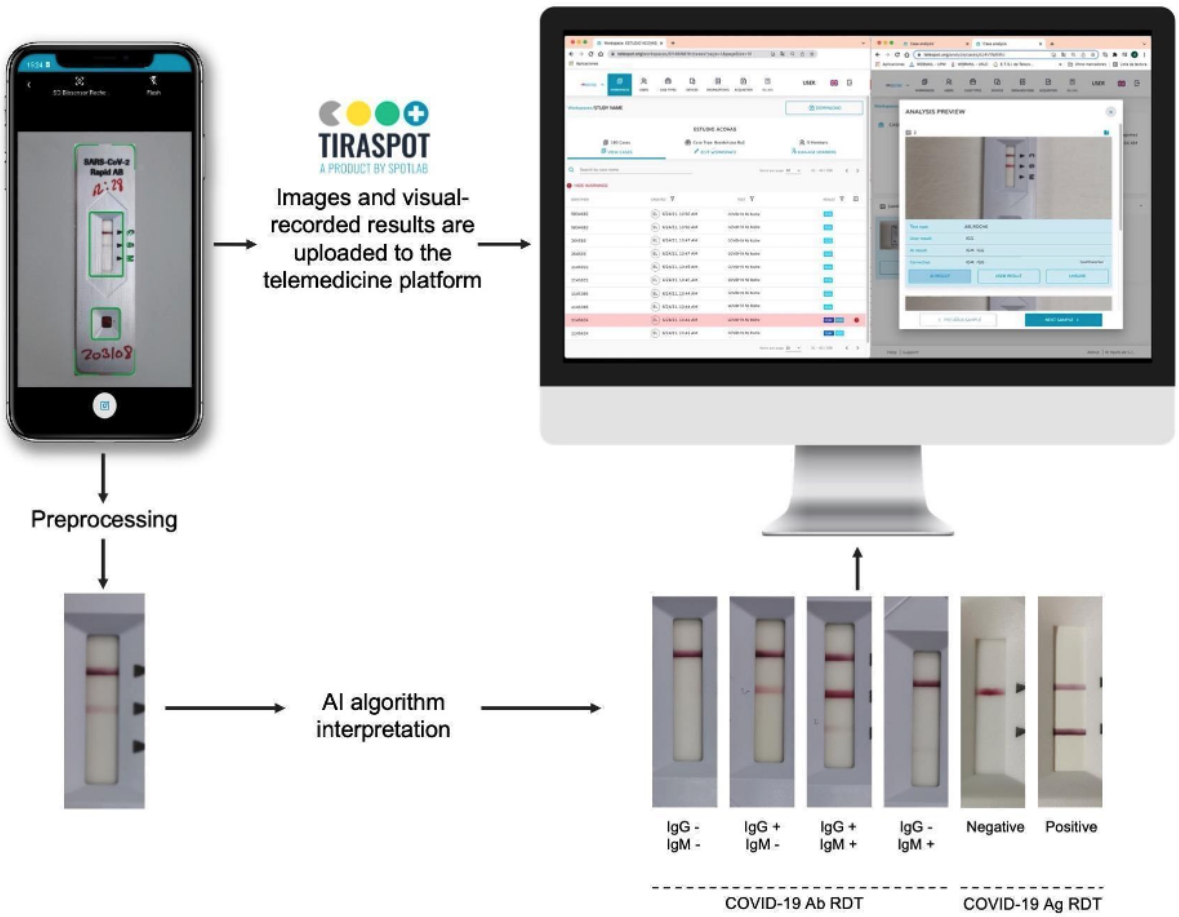
rapid diagnostic test; artificial intelligence; AI; telemedicine platform; COVID-19; rapid test; diagnostics; epidemiology; surveillance; automatic; automated; tracking

Introduction

To control COVID-19 pandemic, timely and accurate early-detection strategies of SARS-CoV-2 infections have been critical to slow down the spread of the virus. The use of rapid diagnostic tests (RDTs), both for detection of antibodies and antigens, has contributed to improve COVID-19 testing capacity, reducing costs of diagnosis, and allowing for fastest results [1]. First, COVID-19 RDTs were intended to be used just by professional health workers who have extensive experience in the use of this tool for different infectious diseases [2,3]. Later, multiple health ministries approved home testing kits, improving the accessibility to testing and taking pressure off health institutions. Nevertheless, self-testing strategies have some limitations; the general population is not familiar with the use of RDTs; and a minimum training is needed for sampling, testing, and result interpretation. Furthermore, as it has been seen during the latest waves [4], many results go unreported, impairing posttesting counseling and epidemiological surveillance.

Combining RDTs with digital tools, artificial intelligence (AI) and mobile health approaches can help standardize result interpretation and facilitate immediate reporting and monitoring of results [5]. Several works have been proposed to automatically interpret photographs of RDTs using different image processing approaches, from classical methods, such as morphology-based methods, to more sophisticated machine learning or deep learning methods [6-21]. Nevertheless, these approaches are not capable of handling 2-band and 3-band RDTs indistinctly, are not connected to a cloud platform that enables the collection of mass screening results, and many require additional hardware. In this paper, we describe the development and field validation of a mobile-based tool (exhaustively tested with a variety of phone models and different lighting conditions) that could be used with any smartphone for reading and reporting multiple types of SARS-CoV-2 RDTs and is connected to a real-time epidemiological monitoring web platform (Figure 1).

Figure 1. TiraSpot system is composed of (1) a mobile app for test digitization and result recording, (2) an artificial intelligence (AI) model for rapid diagnostic test (RDT) result interpretation, and (3) a web platform where all collected data can be visualized, allowing for result corrections in the cases in which a discrepancy exists between AI and user interpretation. Ab: antibody; Ag: antigen.



Methods

Procedure

This study was divided into 2 phases: first, the training and validation of an AI algorithm for the automatic interpretation of RDTs; second, 2 field studies to assess the performance of the AI-based system for reading both COVID-19 antibody and antigen RDTs in real-world scenarios.

Ethics Approval

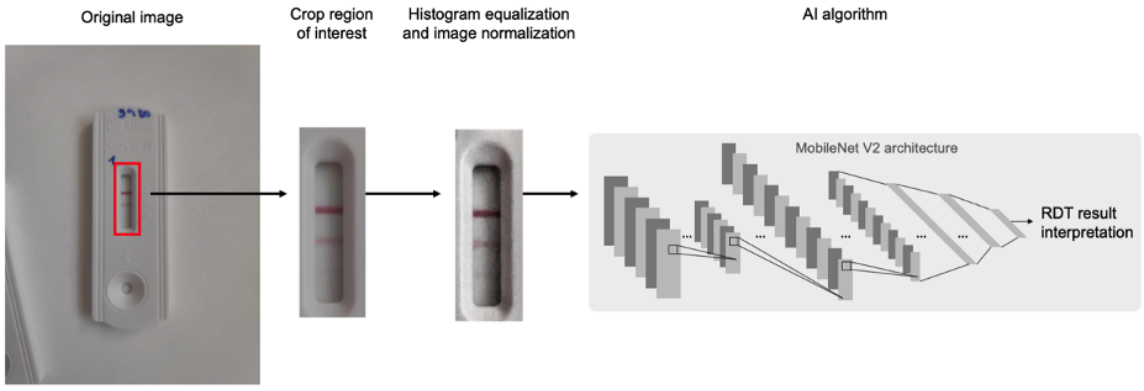
Ethics approval for the study was obtained from the Clinical Research Ethics Committee of the Ramón y Cajal University Hospital (127/21).

Algorithm Training and Validation Data Set

Ensuring standardized image acquisition is a key step in developing robust AI algorithms. With this purpose, all inoculated RDTs were digitized using the TiraSpot mobile app (Spotlab), which guarantees image quality and correct positioning of RDTs in the image by using a simple augmented reality system that displays a mask with the exact geometry of a given RDT in the screen of the smartphone, helping users to correctly align the RDT before making the picture (screenshot of the mobile app is presented in Figure 1). Each RDT brand has its own mask that guides the user to take a standardized picture. In addition, after the picture is taken, the user is presented with the picture and asked to confirm that it is aligned and on focus. If the user rejects the picture, the user is allowed to take another one. The mobile app also allows users to record sample metadata, which together with the images and their initial visual interpretation are uploaded to the cloud platform. To gain robustness and generalizability, a total of 11 different smartphone models, ranging from low- to high-range devices, were used in this study.

An AI algorithm was developed to predict test results based on a picture of the RDT. With this purpose, the image is first preprocessed by cropping the original image to extract a region of interest that contains the relevant part of the picture (strip of the RDT). Then, image normalization and contrast enhancement (Contrast Limited Adaptive Histogram Equalization method) were applied to highlight faint bands. Finally, the processed region of interest is introduced into a convolutional neural network (MobileNet V2 [22]), which then predicts the test result (Figure 2).

Figure 2. Image processing pipeline. Original image acquired with the TiraSpot app is first cropped to extract region of interest. The cropped image is then preprocessed and introduced to a convolutional neural network, which predicts rapid diagnostic test (RDT) result. AI: artificial intelligence.



For generating the training image data set, 12 human sera from patients with positive SARS-CoV-2 polymerase chain reaction tests (infected between March and May 2020) and with a positive enzyme-linked immunosorbent assay (ELISA) test were used. Each serum sample was serially diluted with reference human sera (H5667; Sigma-Aldrich) until it reached a negative result when inoculated in a COVID-19 antibody test. Each dilution was tested in 3 replicates for each of the 3 brands tested (ie, 2019-nCoV IgG/IgM Rapid Test Cassette, Hangzhou AllTest Biotech Co., Ltd.; Panbio COVID-19 IgG/IgM Rapid Test Device, Abbott; and UNscience COVID-19 IgG/IgM Rapid Test, Wuhan UNscience Biotechnology Co., Ltd.), resulting in 433 RDTs inoculated (61 positive for both IgG and IgM; 166 positive for IgG and negative for IgM; 43 negative for IgG and positive for IgM; and 164 negative for both IgG and IgM). Additionally, 12 COVID-19 antigen RDTs (Panbio COVID-19 Ag Rapid Test Device, Abbott; 6 positive and 6 negative) were also included to train the algorithm to read not only 3-band tests (such as the COVID-19 antibody tests used in this study) but also 2-band RDTs, such as COVID-19 antigen tests. The entire training data set consisted of 3614 images.

For collecting the independent validation data set, 240 human sera samples independent from the ones used for training were used to inoculate 720 COVID-19 antibody RDTs (each serum was tested in triplicate using the aforementioned brands). The samples were selected ensuring all possible results are well represented along the data set (108 positive for both IgG and IgM; 321 positive for IgG and negative for IgM; 27 negative for IgG and positive for IgM; and 264 negative for both IgG and IgM).

Each RDT was visually read by multiple observers (3 to 5), and the ground truth was established as the majority result from

total analyzers. All sera samples were collected between May and June 2020.

Field Validation Studies

The workflow for the field studies was as follows: a health professional digitized the RDTs by using the app and was asked for recording the visual interpretation of the test result; images were uploaded to the cloud platform and processed by the AI algorithm; and discrepancies between the interpretation made by the health professional and that obtained by the algorithm were subsequently reviewed by an external health professional through the platform.

The first field study used the system as part of a seroprevalence study conducted in two nursing homes in Madrid, Spain. A total of 172 vaccinated health care personnel were included in this study; a finger-prick blood sample was taken from them and inoculated into SARS-CoV-2 Rapid Antibody Test (Roche). A trained nurse digitized the RDTs and recorded their results using the app.

The second field validation study tested the system to read also COVID-19 antigen tests (ie, Panbio COVID-19 Ag Rapid Test Device, Abbot) composed of 2 bands (ie, control and test). This study was carried out at the emergency department of the Ramón y Cajal Hospital in Madrid, Spain, where 96 individuals' nasal swabs were inoculated in antigen tests and digitized by experienced health professionals using the app.

All images were acquired in very diverse real-world conditions involving different users, including different environmental illuminations (eg, different lighting color temperatures and a wide range of lighting), and using different smartphone models that ranged from low- to high-range devices. This was done with the purpose of developing and validating the robustness of the algorithm that may change in real life.

Results

AI Algorithm Training and App Validation

All images acquired with the app were uploaded to a cloud platform, where the AI algorithm processed the photographs to predict the result interpretations. As shown in Table 1 (part 1), when comparing the visual interpretations (used as ground truth) against the AI algorithm, the performance was high for all brands of RDT tested, obtaining a mean sensitivity and specificity of 98% and 100%, respectively, for the detection of the IgG band; and a mean sensitivity and specificity of 80% and 89%, respectively, for the detection of the IgM band. No significant differences were found in algorithm performance between different smartphone models or across different lighting conditions, pointing out the robustness of the readout algorithm.

Table 1. Performance of the artificial intelligence algorithm for predicting rapid diagnostic test (RDT) results with respect to human visual reading in (1) the validation set, (2) the field study for reading antibody (Ab) RDTs, and (3) in the field study when reading antigen (Ag) RDTs.

Evaluation data	AUC[a] (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Tests, n	
				Negative	Positive
1: RDT manufacturer (Ab) and band					
Abbott					
IgG	99.5 (98.7-100)	96.4 (94.1-98.8)	100 (100-100)	94	145
IgM	92.5 (85.4-99.6)	80.8 (75.8-85.8)	90.7 (87.0-94.3)	184	55
UNScience					
IgG	100 (100-100)	100 (100-100)	100 (100-100)	100	140
IgM	89.5 (83.7-95.2)	80.0 (74.9-85.1)	88.6 (84.6-92.6)	214	26
AllTest					
IgG	99.8 (99.4-100)	97.9 (96.1-99.7)	100 (100-100)	96	144
IgM	90.6 (85.0-96.1)	79.6 (74.5-84.7)	86.0 (81.6-90.4)	186	54
Global					
IgG	99.8 (99.5-100)	98.1 (97.1-99.1)	100 (100-100)	290	429
IgM	90.8 (87.4-94.3)	80.0 (77.1-82.9)	89.0 (86.2-90.9)	584	135
2: RDT manufacturer (Ab) and band					
Roche					
IgG	100 (100-100)	100 (100-100)	94.4 (92.8-96.1)	18	154
IgM	99.6 (96.0-100)	100 (100-100)	95.8 (94.3-97.3)	166	6
3: RDT manufacturer (Ag) and band					
Abbott					
Test	100 (100,100)	100 (100,100)	100 (100,100)	68	28

[a]AUC: area under the curve.

Validation in Real-world Scenarios

From the 172 RDTs used in this study (5 positive for both IgG and IgM; 149 positive for IgG and negative for IgM; 1 negative for IgG and positive for IgM; and 17 negative for both IgG and IgM), we only found 9 discrepancies between test result interpretations made by health professionals and those made by the AI algorithm. From these 9 cases, 2 were incorrectly classified by the algorithm due to an incorrect image acquisition with the app. The remaining discrepant cases were further reviewed by a second professional, and the AI-based system allowed for the detection and modification of the result with respect to the initial health professional interpretation in 4 cases by confirming the result predicted by the algorithm.

The overall performance of the algorithm with respect to the ground truth is shown in Table 1 (part 2). It should be noted that the performance of the system is high even when used with an RDT different from those used for training the algorithm, suggesting its potential use with any RDT on the market. The slight disparity in the performance of IgM band identification in antibody RDTs between the validation set and this field study may be explained by the presence of very faint signals that were almost invisible in the photographs.

Regarding the second field study for reading COVID-19 antigen RDTs, we found that all tests used and digitized using the TiraSpot app (ie, 58 negative and 30 positive) were correctly interpreted by the proposed system (Table 1, part 3), demonstrating that the system can also be applied for reading 2-band (ie, control and test) and 3-band (ie, IgG, IgM, and control) tests.

Discussion

We described the usefulness of an app for reading and result interpretation of lateral flow RDTs for SARS-CoV-2 testing. The results are sent to a cloud platform that allows for case identification and confirmation, quality control, and real-time monitoring.

Our AI algorithm demonstrates excellent performance, especially in prospective validation of real-life scenarios and for both antibody and antigen detection tests. The algorithm performed as well in RDT brands that were not used at all for training purposes, making the solution suitable for other RDTs, including other diseases. Compared with previous studies [6-21], our system is able to identify individual bands of the RDTs, allowing for complex result reading and sending them in real-time to a cloud platform. A requirement and limitation of

the proposed system is the correct acquisition of the image (acquisition error in the field studies was <0.8%).

In conclusion, the use of TiraSpot (Figure 1) is a useful tool for reporting, real-time monitoring, and quality control, as the results can be reviewed by specialists when needed. This is especially important in contexts where massive testing is to be done and the likelihood of subjectivity and errors in the interpretation of the result is higher. It is also important in the validation of self-diagnostic tests performed by untrained users, as it avoids the loss of information in case the user does not report it, and it provides an efficient system to confirm and report data, which has been a key challenge during the latest pandemic waves.

Acknowledgments

This study has been partially funded by JYLAG Foundation (Fribourg, Switzerland). DBP was supported by grant PTQ2020-011340/AEI/10.13039/501100011033 funded by the Spanish State Investigation Agency. LL was supported by a predoctoral grant (IND2019/TIC-17167) by Comunidad de Madrid, Spain.

Data Availability

The raw data supporting the results of this study will be made available upon reasonable request from the corresponding author.

Authors' Contributions

DBP, DMM, EA, JMN, JCG, BRH, RC, MLO, and MRD conceived and designed the study. DBP, DMM, EA, NPP, AM, ED, LL, and MRD acquired data. DBP and EA analyzed the data. EA, DC, and AV worked on the mobile app and web-based tool. DBP, DMM, and EA wrote the original draft. All authors interpreted data and revised the manuscript.

Conflicts of Interest

DBP, EA, AM, ED, LL, AV, DC, and MLO work for Spotlab. The rest of the authors declare no competing interests.

References

1. Borges LP, Martins AF, Silva BDM, Dias BDP, Gonçalves RL, Souza DRVD, et al. Rapid diagnosis of COVID-19 in the first year of the pandemic: a systematic review. Int Immunopharmacol 2021 Dec;101(Pt A):108144 [FREE Full text] [doi: 10.1016/j.intimp.2021.108144] [Medline: 34607235]
2. Weber NC, Klepser ME, Akers JM, Klepser DG, Adams AJ. Use of CLIA-waived point-of-care tests for infectious diseases in community pharmacies in the United States. Expert Rev Mol Diagn 2016 Nov 27;16(2):253-264. [doi: 10.1586/14737159.2015.1116388] [Medline: 26560318]
3. Bell SFE, Lemoire J, Debattista J, Redmond AM, Driver G, Durkin I, et al. Online HIV self-testing (HIVST) dissemination by an Australian community peer HIV organisation: a scalable way to increase access to testing, particularly for suboptimal testers. Int J Environ Res Public Health 2021 Oct 26;18(21):11252 [FREE Full text] [doi: 10.3390/ijerph182111252] [Medline: 34769771]
4. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature 2021 Apr 09;592(7854):438-443. [doi: 10.1038/s41586-021-03402-9] [Medline: 33690265]
5. Kadam R, White W, Banks N, Katz Z, Dittrich S, Kelly-Cirino C. Target Product Profile for a mobile app to read rapid diagnostic tests to strengthen infectious disease surveillance. PLoS One 2020 Jan 29;15(1):e0228311 [FREE Full text] [doi: 10.1371/journal.pone.0228311] [Medline: 31995628]
6. Turbé V, Herbst C, Mngomezulu T, Meshkinfamfard S, Dlamini N, Mhlongo T, et al. Deep learning of HIV field-based rapid tests. Nat Med 2021 Jul 17;27(7):1165-1170 [FREE Full text] [doi: 10.1038/s41591-021-01384-9] [Medline: 34140702]
7. Mendels D, Dortet L, Emeraud C, Oueslati S, Girlich D, Ronat J, et al. Using artificial intelligence to improve COVID-19 rapid diagnostic test result interpretation. Proc Natl Acad Sci U S A 2021 Mar 23;118(12):e2019893118 [FREE Full text] [doi: 10.1073/pnas.2019893118] [Medline: 33674422]
8. Wang W, Stafford JW, Keller MD. Customization and testing of a mobile reader app for an open-access SARS-CoV-2 antigen lateral flow assay. Proceedings Volume 11950, Optics and Biophotonics in Low-Resource Settings VIII 2022 Mar 02:5009-5010. [doi: 10.1117/12.2609795]
9. Park C. Supporting smartphone-based image capture of rapid diagnostic tests in low-resource settings. In: Proceedings of the 2020 International Conference on Information and Communication Technologies and Development. 2020 Presented at: ICTD2020; June 17-20; New York, NY p. 1-11. [doi: 10.1145/3392561.3394630]
10. Mudanyali O, Dimitrov S, Sikora U, Padmanabhan S, Navruz I, Ozcan A. Integrated rapid-diagnostic-test reader platform on a cellphone. Lab Chip 2012 Aug 07;12(15):2678-2686 [FREE Full text] [doi: 10.1039/c2lc40235a] [Medline: 22596243]
11. Xiao W, Huang C, Xu F, Yan J, Bian H, Fu Q, et al. A simple and compact smartphone-based device for the quantitative readout of colloidal gold lateral flow immunoassay strips. Sens Actuators B Chem 2018 Aug 01;266:63-70 [FREE Full text] [doi: 10.1016/j.snb.2018.03.110] [Medline: 32288251]

12. Paula Vaz Cardoso L, Dias RF, Freitas AA, Hungria EM, Oliveira RM, Collovati M, et al. Development of a quantitative rapid diagnostic test for multibacillary leprosy using smart phone technology. BMC Infect Dis 2013 Oct 23;13(1):497 [FREE Full text] [doi: 10.1186/1471-2334-13-497] [Medline: 24152601]
13. Guo W, Zhang Y, Hu X, Zhang T, Liang M, Yang X, et al. Region growing algorithm combined with fast peak detection for segmenting colloidal gold immunochromatographic strip images. IEEE Access 2019;7:169715-169723. [doi: 10.1109/access.2019.2955510]
14. Miikki K, Miikki L, Wiklund J, Karakoç A. Augmented sensitivity of at-home rapid SARS-CoV-2 antigen test (RAT) kits with computer vision: a framework and proof of concept. BioMed 2022 Apr 14;2(2):199-209. [doi: 10.3390/biomed2020018]
15. Mujtaba D, Mahapatra NR. Lateral flow test interpretation with residual networks. 2021 Presented at: International Conference on Computational Science and Computational Intelligence (CSCI); Dec 15-17; Las Vegas, NV. [doi: 10.1109/CSCI54926.2021.00261]
16. Wong NCK, Meshkinfamfard S, Turbé V, Whitaker M, Moshe M, Bardanzellu A, et al. Machine learning to support visual auditing of home-based lateral flow immunoassay self-test results for SARS-CoV-2 antibodies. Commun Med (Lond) 2022 Jul 06;2(1):78 [FREE Full text] [doi: 10.1038/s43856-022-00146-z] [Medline: 35814295]
17. Wang W, Hu L, Keller MD. Effects of image capture and correction approaches on quantifying results of lateral flow assays with mobile phones. Proceedings Volume 11632, Optics and Biophotonics in Low-Resource Settings VII 2021 Mar 05:320-321. [doi: 10.1117/12.2582396]
18. Richardson S, Kohn MA, Bollyky J, Parsonnet J. Validity of at-home rapid antigen lateral flow assay and artificial intelligence read to detect SARS-CoV-2. Diagn Microbiol Infect Dis 2022 Nov;104(3):115763 [FREE Full text] [doi: 10.1016/j.diagmicrobio.2022.115763] [Medline: 36070629]
19. Lee S, Yoo YK, Wee KW, Kim C, Lee NE, Kim KH, et al. Machine-learning-assisted lateral flow assay for COVID-19 and influenza detection. SSRN Journal 2022:1-26 [FREE Full text] [doi: 10.2139/ssrn.4073623]
20. Han G, Koo HJ, Ki H, Kim M. Paper/soluble polymer hybrid-based lateral flow biosensing platform for high-performance point-of-care testing. ACS Appl Mater Interfaces 2020 Aug 05;12(31):34564-34575. [doi: 10.1021/acsami.0c07893] [Medline: 32666783]
21. Hoque Tania M, Lwin KT, Shabut AM, Najlah M, Chin J, Hossain M. Intelligent image-based colourimetric tests using machine learning framework for lateral flow assays. Expert Syst Appl 2020 Jan;139:112843. [doi: 10.1016/j.eswa.2019.112843]
22. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: inverted residuals and linear bottlenecks. 2018 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23; Salt Lake City, UT. [doi: 10.1109/cvpr.2018.00474]

Abbreviations

AI: artificial intelligence
RDT: rapid diagnostic test

Edited by T Sanchez, A Mavragani; submitted 03.05.22; peer-reviewed by P Sarajlic, T Scherr; comments to author 29.08.22; revised version received 16.09.22; accepted 13.10.22; published 30.12.22.

Please cite as:
Bermejo-Peláez D, Marcos-Mencía D, Álamo E, Pérez-Panizo N, Mousa A, Dacal E, Lin L, Vladimirov A, Cuadrado D, Mateos-Nozal J, Galán JC, Romero-Hernandez B, Cantón R, Luengo-Oroz M, Rodriguez-Dominguez M
A Smartphone-Based Platform Assisted by Artificial Intelligence for Reading and Reporting Rapid Diagnostic Tests: Evaluation Study in SARS-CoV-2 Lateral Flow Immunoassays
JMIR Public Health Surveill 2022;8(12):e38533
URL: https://publichealth.jmir.org/2022/12/e38533
doi:10.2196/38533
PMID:36265136

Review

Risks, Epidemics, and Prevention Measures of Infectious Diseases in Major Sports Events: Scoping Review

Xiangyu Yan[1], PhD; Yian Fang[1], PhD; Yongjie Li[2], PhD; Zhongwei Jia[1,3,4], PhD; Bo Zhang[1], PhD

[1]School of Public Health, Peking University, Beijing, China

[2]School of Basic Medical Sciences, Peking University, Beijing, China

[3]Center for Intelligent Public Health, Institute for Artificial Intelligence, Peking University, Beijing, China

[4]Center for Drug Abuse Control and Prevention, National Institute of Health Data Science, Peking University, Beijing, China

Corresponding Author:
Bo Zhang, PhD
School of Public Health
Peking University
38 Xueyuan Rd
Haidian District
Beijing, 100191
China
Phone: 86 18811186101
Email: zhangbo0136@pku.edu.cn

Abstract

Background: Major sports events are the focus of the world. However, the gathering of crowds during these events creates huge risks of infectious diseases transmission, posing a significant public health threat.

Objective: The aim of this study was to systematically review the epidemiological characteristics and prevention measures of infectious diseases at major sports events.

Methods: The procedure of this scoping review followed Arksey and O'Malley's five-step methodological framework. Electronic databases, including PubMed, Web of Science, Scopus, and Embase, were searched systematically. The general information (ie, publication year, study type) of each study, sports events' features (ie, date and host location), infectious diseases' epidemiological characteristics (ie, epidemics, risk factors), prevention measures, and surveillance paradigm were extracted, categorized, and summarized.

Results: A total of 24,460 articles were retrieved from the databases and 358 studies were included in the final data synthesis based on selection criteria. A rapid growth of studies was found over recent years. The number of studies investigating epidemics and risk factors for sports events increased from 16/254 (6.3%) before 2000 to 201/254 (79.1%) after 2010. Studies focusing on prevention measures of infectious diseases accounted for 85.0% (238/280) of the articles published after 2010. A variety of infectious diseases have been reported, including respiratory tract infection, gastrointestinal infection, vector-borne infection, blood-borne infection, and water-contact infection. Among them, respiratory tract infections were the most concerning diseases (250/358, 69.8%). Besides some routine prevention measures targeted at risk factors of different diseases, strengthening surveillance was highlighted in the literature. The surveillance system appeared to have gone through three stages of development, including manual archiving, network-based systems, and automated intelligent platforms.

Conclusions: This critical summary and collation of previous empirical evidence is meaningful to provide references for holding major sports events. It is essential to improve the surveillance techniques for timely detection of the emergence of epidemics and to improve risk perception in future practice.

(JMIR Public Health Surveill 2022;8(12):e40042) doi:10.2196/40042

KEYWORDS

major sports event; epidemic; risk factor; prevention; surveillance; scoping review

Introduction

Major sports events, including national or international multisport events (eg, the Olympic Games) and single-sport events (eg, International Federation of Association Football [FIFA] World Cup), attract the focus of the world. However, the gathering of crowds from various parts of the world poses challenges to public health, especially for global infectious disease prevention and control [1]. Moreover, the threats of infectious disease exist over the entire course of the games, including the competition stage and the journey before and after the sports events [2,3]. Several factors play roles in public health security of these sports events, including close contact of the attendees in confined and crowded spaces, the demographics and disease exposure history of the participants, their mobility patterns, the event setting, and climate conditions [2-4]. The incoming and outgoing travel patterns of international participants attending these events can accelerate disease transmission among a large number of people, potentially leading to a pandemic within a short period [5,6].

Many recent studies have raised concerns about the potential risks of infectious disease in major sports events and their impacts on public health on a global scale [7]. During the 1991 International Special Olympics Games, 16 outbreak-associated secondary measles cases were reported in athletes, spectators, and volunteers [7]. An outbreak of 36 influenza cases among participants was recorded at the 2002 Winter Olympics held in Salt Lake City, Utah, United States [8]. Outbreaks of norovirus were reported at many sports events such as the 2006 Football World Cup and 2018 Winter Olympics [9,10]. In addition, many other types of infectious diseases might also threaten public health security in major sports events, such as leptospirosis infection in triathlons, spread of blood-borne diseases in boxing, and small-scale transmission of insect-borne diseases at World Cup events [11-13].

To better prevent and control infectious disease outbreaks, routine prevention and control measures, new technologies for symptom surveillance, environmental sampling, and virus testing are used in major sports events. Routine prevention and control measures include reminding participants to pay attention to personal hygiene, and strengthening publicity, education, and cooperation between health institutions [14,15]. For infectious disease surveillance and management, before the 1990s, major sports events mainly used traditional methods relying on routine manual filing of reportable diseases by clinicians [16-21]. The development of network communication technology makes it possible to quickly and automatically monitor infectious diseases based on a variety of data sources, including clinical data from health providers, drug sales from pharmacies, outbreak reports, emergency and urgent care data, websites and social media posts, environmental data, and travel flights, to name a few [22-24].

Overall, previous studies have provided abundant evidence of infectious disease epidemics, risk factors, and prevention measures among major sports events around the world. However, the evidence remains very fragmented, with no study systematically summarizing the prevention and control of infectious diseases in major sporting events. To fill this gap, the aim of this scoping review was to provide an overview and critical summary of existing evidence on the epidemics and risk factors, prevention measures, and surveillance paradigm of infectious diseases in major sports events, which could help to provide guidance for the prevention and control of infectious diseases at national and international major sports events, especially during the COVID-19 pandemic.

Methods

Design and Search Strategy

The procedure of the scoping review followed Arksey and O'Malley's [25] five-step methodological framework, including (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) extraction and charting the data; and (5) collating, summarizing, and reporting the results. The results are reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [26].

A systematic search of electronic databases (PubMed, Web of Science, Scopus, and Embase) was performed from the earliest record to October 15, 2022. All study designs were included. The search strategy contained two parts: (1) the sports event and its related terms, including "athletic" or "Olympic" or "Paralympic" or "World cup" or "championship" or "marathon" or "mass gatherings" or "stadium" or "sports venues" or "gym"; and (2) terms covering topics of the infectious diseases and their transmission, including "infection" or "infestation" or "infectious" or "transmission" or "communicable disease" or "community spread" (see Multimedia Appendix 1 for the detailed search strategy).

Study Selection

After eliminating duplicates, a two-step study selection procedure was performed. The first step was the preliminary selection through titles, abstracts, and keywords. Studies were excluded if (1) the content of the study was not relevant to infectious diseases or (2) not relevant to sports events. The second step was based on examination of the full text, and studies were excluded if (1) there was no available full text or (2) the article was not written in English. The type of article was not restricted, and comments, letters, or replies related to the research theme were all included in this review. Studies were screened independently by two reviewers (BZ, XY) against the above criteria. Disagreements were resolved through discussion or via a third reviewer (YL).

Data Extraction

Two authors (BZ, XY) reviewed the included studies and analyzed infectious diseases in major sports events. The general information (ie, publication year, study type) was extracted. Data related to the sports events' features (ie, type, date, and host location); key findings of each study, including infectious disease characteristics (ie, types, epidemics, risk factors), prevention measures, and surveillance paradigm of infectious diseases, were also extracted. According to the extracted data, literature quality assessment was performed based on the following two criteria: (1) specifies the type of sport and the

scale of the sport event, along with the host location and time of the event; (2) specifies the main types of infectious diseases in the sports event. If either of the two criteria was not met, the study was considered to be of low quality and was excluded.

Data Synthesis

Descriptive analysis was performed for statistics of related infectious diseases (eg, respiratory tract infection, vector-borne infection) by time period. According to the sports types, the included sports events were divided into four categories, including multisport (eg, the Olympic Games), ball (eg, football, basketball, tennis), race (eg, running, marathon, bike race), and other sports types (eg, swimming, fighting). This classification also referred to the sports categories of the world's greatest sports events from a previous report [27]. Maps were used to present the geographical distribution of studies on different infectious diseases around the world. Key findings of the included studies were summarized and categorized into two main topics, including (1) epidemics and risk factors and (2) prevention measures, which were determined by the studies' primary aims and outcomes, with subcategories identified where appropriate. Statistical analyses were performed using SPSS version 21.0 (IBM Corp), and Tableau 2021 (Tableau Software) was used for mapping.

Results

General Characteristics of Included Studies

We identified 24,460 articles through a database search. Following the removal of duplicates and screening for eligibility, 358 studies were ultimately included in the review (Figure 1). The basic information of the included articles is provided in Tables S1 and S2 of Multimedia Appendix 1. Figure 2 shows the numbers of publication and their themes over recent years, demonstrating the rapid growth of studies in this area. Only 16 of 254 studies (6.3%) investigated epidemics and risk factors of infectious disease in major sports events before 2000, which increased to 201 (79.1%) after 2011. Studies focusing on prevention measures accounted for 85.0% (238/280) of the articles published after 2010. The main types of sports studied over the years were multisport events and ball games (Figure 2). Among the 358 studies, the top five most studied major sports events were the 2016 Rio Olympics (n=46, 12.8%), 2020 Tokyo Olympics (n=42, 11.7%), 2012 London Olympics (n=19, 5.3%), 2014 FIFA World Cup (n=16, 4.5%), and 2010 FIFA World Cup (n=14, 3.9%) (Multimedia Appendix 1 Table S3).

Based on the main transmission routes, the infectious diseases reported were classified into five categories: respiratory tract infection, gastrointestinal infection, vector-borne infection, blood-borne infection, and water-contact infection. Among them, over half (250/358, 69.8%) focused on respiratory tract infections, with COVID-19, measles, and influenza as the main concerns (Figure 2). Studies on vector-borne diseases such as Zika and dengue showed the fastest rate of increase in recent years, from 23.8% (5/21) published before 2000 to 30.4% (85/280) published after 2010 (Figure 2). Moreover, a considerable number of studies have focused on gastrointestinal infections (eg, *Salmonella*, norovirus), blood-borne infections (eg, HIV/AIDS, hepatitis B), and water-contact infections (eg, leptospirosis) (Figure 2).

Several studies focused on the major international sports events' host countries (Figure 3A). The 1996 Olympics held in Atlanta, Georgia, United States; the 2000 Olympics in Sydney, Australia; the 2008 Olympics in Beijing, China; and the 2010 FIFA World Cup in South Africa were the key events of focus of the corresponding periods (Figure 3B, C). After 2010, based on the Olympics and the FIFA World Cup, most studies concentrated in Brazil, the United Kingdom, Japan, and China (Figure 3D). Among them, studies related to Brazil mainly focused on vector-borne infections, as the 2014 FIFA World Cup and the 2016 Olympic Games both took place in Brazil where Zika, dengue, and other vector-borne infections were more prevalent, whereas COVID-19 has been the focus of the Tokyo and Beijing Olympics and Paralympics (Figure 3D).

Figure 1. Flow of selection process for eligible studies for inclusion. WOS: Web of Science.

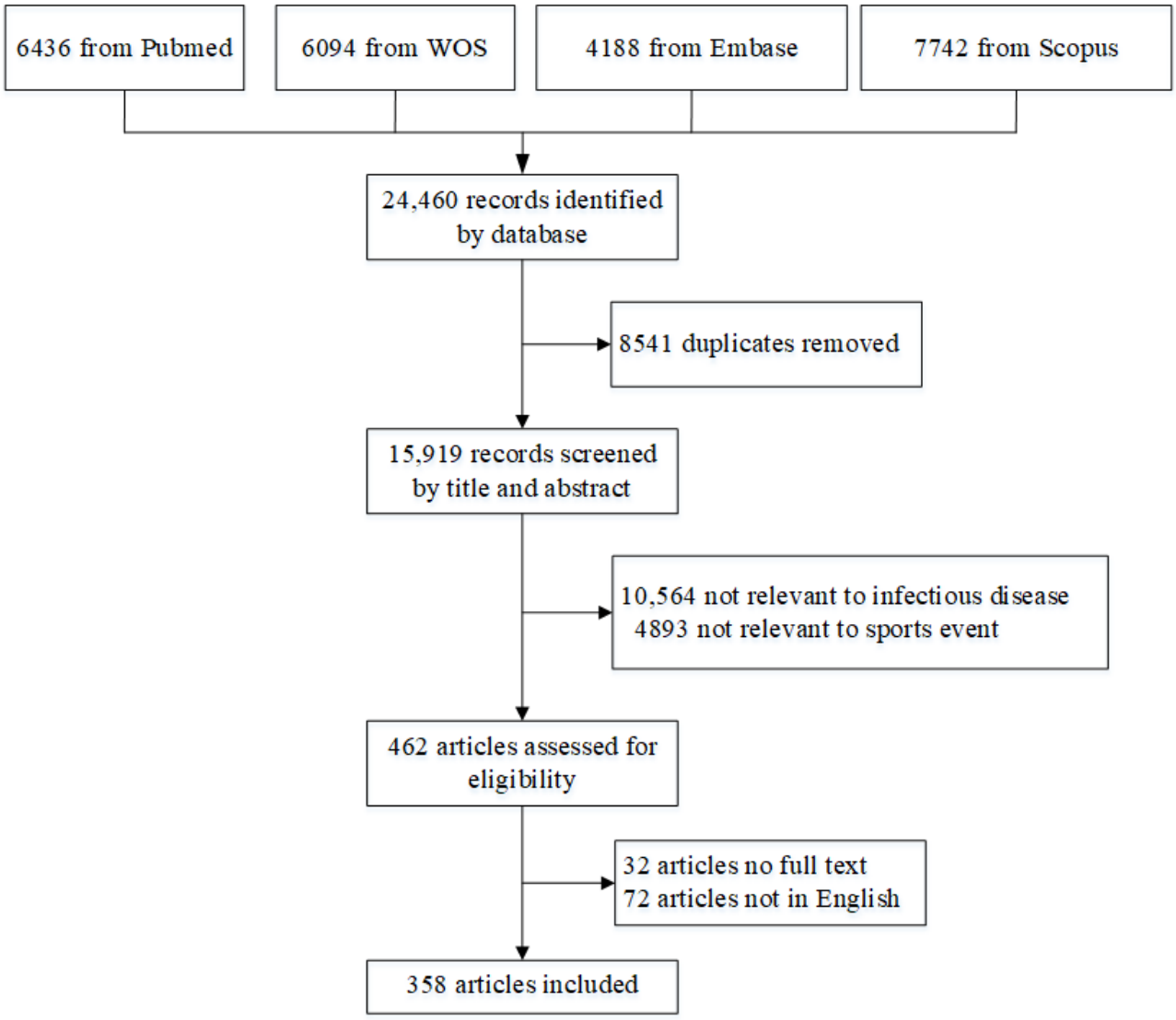


Figure 2. Number of studies on epidemics and prevention of infectious diseases in major sports events over time.

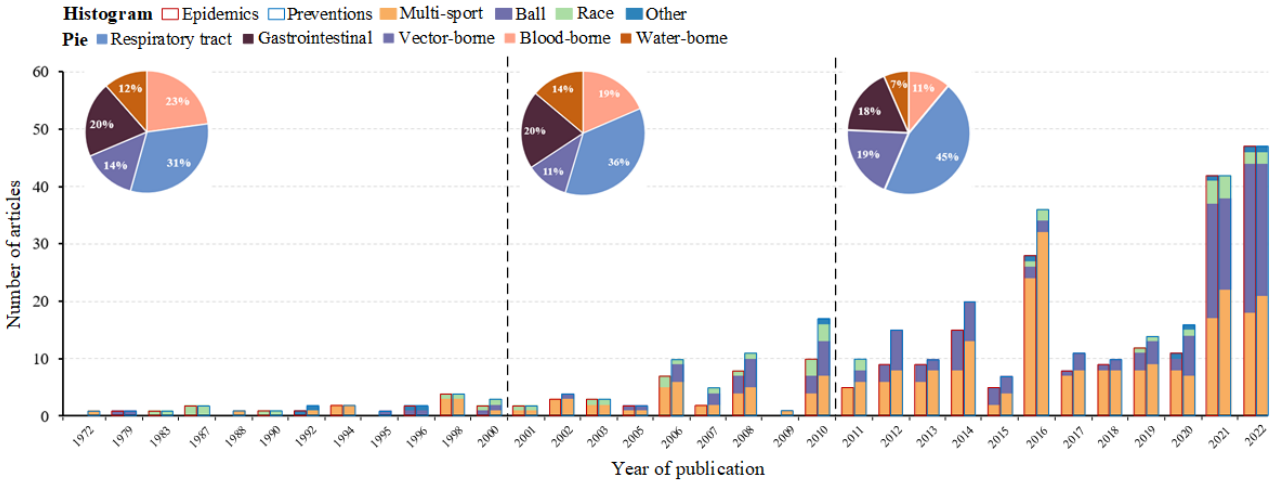
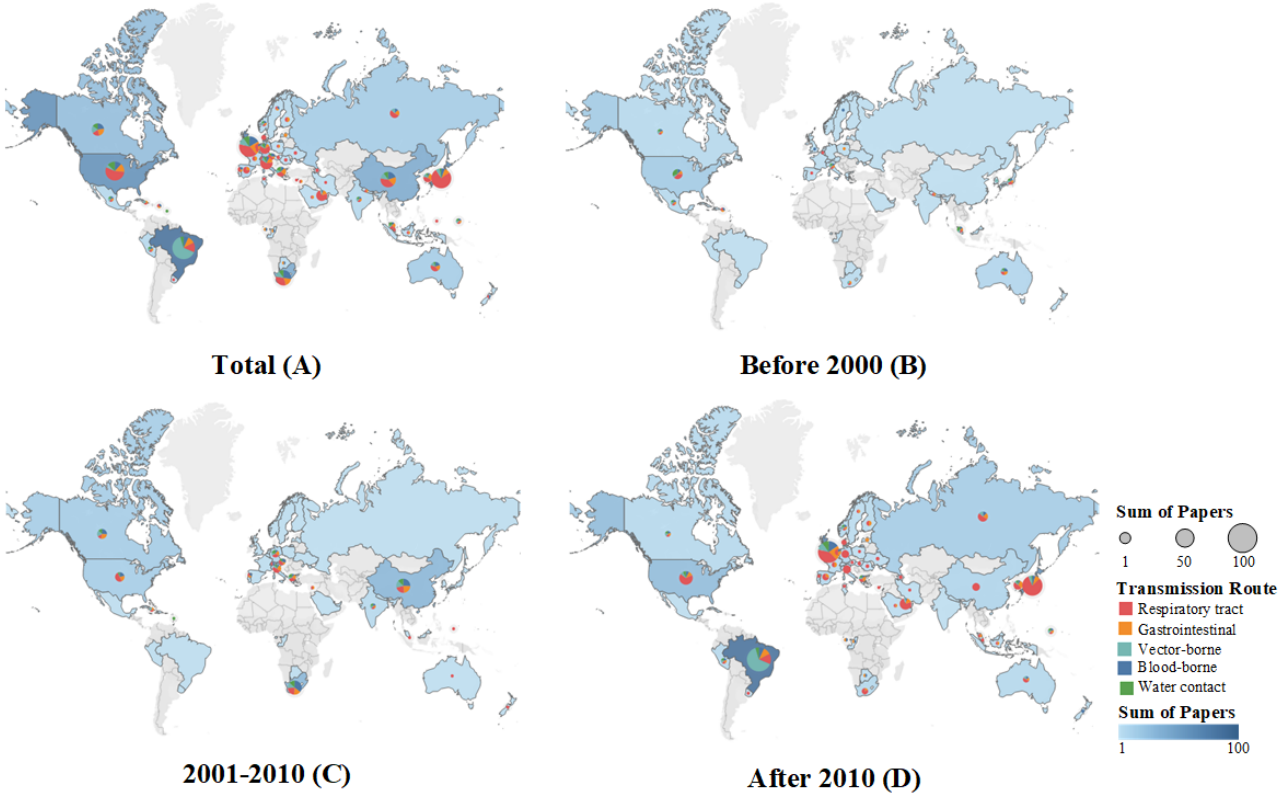


Figure 3. Geographical distribution of studies on infectious diseases in major sports events.



Epidemics and Risk Factors

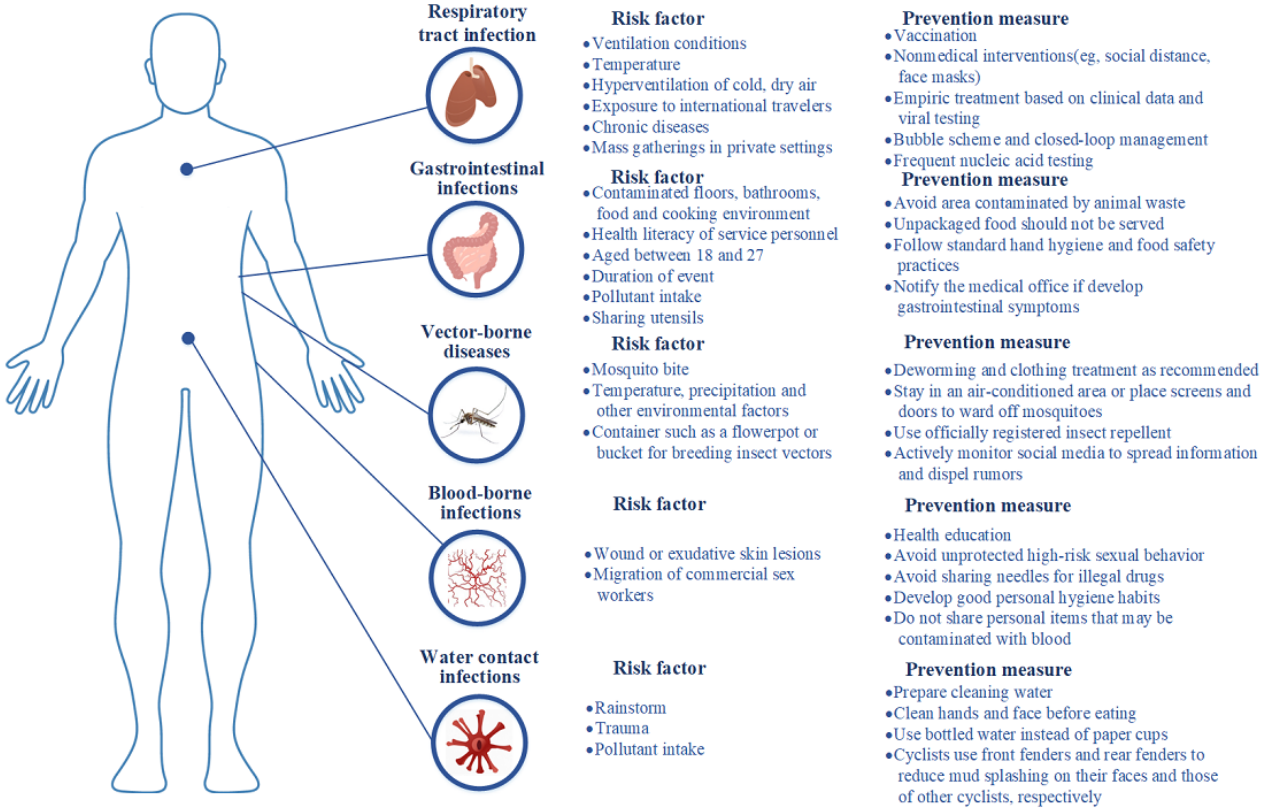
Respiratory Tract Infections

Respiratory tract infections such as influenza and measles are the most common infections in major sporting events. In the 2018 Winter Olympics, Team Finland reported that respiratory tract infections caused by coronaviruses, influenza B virus, respiratory syncytial virus A, rhinovirus, and human metapneumovirus affected 45% of the athletes and 32% of staff members of the service team [28]. Measles is also a highly infectious and acute viral respiratory tract infection that caused small outbreaks in sports events such as the 1991 International Special Olympics Games, 1991 International Gymnastics Competition, and 2007 International Youth Sporting Event [7]. At the end of 2019, the sudden outbreak of COVID-19 disrupted the pace of human life, and many sports events were forced to be cancelled or postponed. Thailand first reported SARS-CoV-2 transmission in a boxing sport stadium, with 27 individuals infected [29]. During the Tokyo Olympics and Paralympics, the positive rates of COVID-19 tests were 0.10% and 0.03% for games-related people arriving at airports and people living in the Tokyo 2020 bubble, respectively, including 24 of 11,476 Olympic athletes and 13 of 4303 Paralympic athletes who tested positive [30]. As for the Beijing 2022 Winter Olympics and Paralympics, 284 of the 16,092 games-related people tested positive at the airport and 179 people tested positive in the closed loop with no cluster epidemic [31] (see Tables S1 and S2 of Multimedia Appendix 1).

Risk factors for respiratory infectious diseases have mainly included three aspects. First, respiratory tract infections could easily be transmitted in confined places; therefore, ventilation conditions, temperature, cold, and dry air in the competition grounds were highlighted as risk factors for respiratory tract infections, because these conditions are suitable for viruses to survive and colonize the human respiratory tract [8]. Second, individuals with routine exposure to international travelers might be at greater risk of respiratory tract infections [32]. Third, people with chronic diseases or weak immunity are more susceptible to respiratory tract infection, as well as its aggravation or complications [32] (Figure 4).

For the ongoing COVID-19 epidemic, the emerging viral variants and the relatively low vaccination rate brought challenges to the holding of major sports events [33]. When holding the Tokyo Olympics, only 3.4% of the population in Japan had been immunized [33]. Another risk for the spread of COVID-19 is the uncontrolled gathering for activities occurring outside the venue during the sports events. Two studies indicated that the increase in COVID-19 cases and cluster epidemics during the 2020 European Football Championship (EURO) were more likely to occur as a result of aggregations and celebrations in private settings, pubs/other public buildings, or public squares rather than from the official football events [34,35]. In addition, previous studies found that athletes had a higher risk of COVID-19 than staff members during professional football competitions [36,37] (Figure 4).

Figure 4. Risk factors and prevention measures in major sports events.



Gastrointestinal Infection

Norovirus and *Salmonella* were the main pathogens that caused gastrointestinal infections in major sports events. In the 2015 Obstacle Adventure race, 866 of 1264 adults reported acute gastroenteritis infections [38]. In the 2018 PyeongChang Winter Olympics, symptoms associated with the gastrointestinal tract, including diarrhea, vomiting, and dyspepsia, were the second most common reported [10]; an outbreak of norovirus infections was recognized, starting with security staff, and noroviruses were detected in samples from food handlers [39] (Multimedia Appendix 1 Tables S1 and S2).

Dirty floors and bathrooms, contaminated food and unhygienic cooking, and serving staff with little awareness of hygiene were all highlighted as risk factors for gastrointestinal infections [38]. Younger athletes (aged 18-27 years) had a significantly higher risk of acute gastrointestinal infection [38]. Ingestion of mud during sports events was associated with gastrointestinal infection [38]. In addition, the pathogens could be transmitted by sharing water bottles or food boxes [40] (Figure 4).

Vector-Borne Infections

During the 2014 FIFA World Cup and the 2016 Rio Olympics Games, vector-borne diseases such as Zika and dengue aroused wide public concern. Due to the efforts of implementing highly effective preventive measures against vectors, there were no epidemic incidents during either event in Brazil [41,42] (Multimedia Appendix 1 Tables S1 and S2).

The primary vectors of Zika, *Aedes* mosquitoes, breed larvae in standing water. Potential mosquito breeding sites (eg, tires, flowerpots, buckets, and other similar containers) were highlighted as risk factors for disease spread [43]. In addition, temperature, precipitation, and other environmental factors had an impact on vector-borne disease infections [44,45] (Figure 4).

Blood-Borne Infections

There was no statistically significant increase in clinic attendance for blood-borne infections (eg, HIV, hepatitis B virus, or hepatitis C virus infections) during sports events in previous years. Sexual transmission was the main transmission route for such infections. The distribution of condoms and health-related messages about safer sex might have contributed to the successful control of blood-borne infections during major sports events [46-48] (Multimedia Appendix 1 Tables S1 and S2).

Injuries and close contacts during competition were highlighted as risk factors for blood-borne pathogen transmission. The pathogen could transmit through bleeding wounds or exudative skin lesions of an infected athlete to the injured skin or mucous membrane of another athlete [12]. Migration was also an important risk factor. Internal and international migration of commercial sex workers might become an important vector for the transmission of HIV [3] (Figure 4).

Water-Contact Infections

The largest outbreak of leptospirosis was reported in the 1998 Springfield triathlon, held in the United States. Among 834 athletes in the Springfield triathlon, 98 (11.8%) met the definition for a suspected case [11] (Multimedia Appendix 1 Tables S1 and S2). Heavy rains that preceded the triathlon were likely to cause leptospiral contamination of the lake. Urine from small mammals (eg, rodents such as mice and rats), wild boars,

or domestic pigs was proposed as a possible source of *Leptospira*. Swimming or wading in fresh unchlorinated water containing contaminated animal urine was another potential cause of leptospirosis [11,49,50] (Figure 4).

Prevention Measures

Routine Preventive Measures

Vaccination emerged as the most effective preventive measure. All delegation members, staff, volunteers, and accompanying visitors should be appropriately vaccinated according to the recommendations of their respective nations before arrival [51,52]. Under the influence of a global pandemic of infectious diseases, local and state health departments should work quickly with organizing committees and governing bodies to establish plans for evaluation, treatment, and prophylaxis, and to ensure that vaccination clinics are functioning appropriately [51] (Figure 4, Multimedia Appendix 1 Tables S1 and S2).

The strategy for respiratory tract infections prevention can integrate empiric treatment based on clinical data and viral testing, along with a public health surveillance approach, including daily review of all viral test results from the polyclinic and reports of symptoms in the community [8]. Nonpharmacological interventions such as social distancing, facemask wearing, and ventilation efficiency enhanced in the stadium are also needed to minimize the risk of community transmission [53,54] (Figure 4, Multimedia Appendix 1 Tables S1 and S2). To prevent COVID-19 transmission, no spectators were allowed and a bubble scheme for games-related people was implemented in some international sports events represented by the Tokyo Olympics and Paralympics, which consisted of a series of measures that separated participants from the general public [55]. Such a closed-loop management strategy was also implemented in the Beijing Winter Olympics and Paralympics [56]. Another strengthening preventive measure was frequent SARS-CoV-2 testing. During the Tokyo Olympics and Paralympics, all participants had to undergo laboratory-based saliva antigen screening every day, and relevant isolation and close-contacts management were implemented for people with positive nucleic acid testing results [30,55]. In addition, it is strongly recommended that games-related people should be vaccinated against COVID-19. More than 80% of athletes and staff were vaccinated in the Tokyo Olympics and Paralympics [55].

For gastroenteritis infections, standard hand hygiene and food safety precautions were recommended, such as eating appropriately cooked food and drinking bottled beverages. Participants were also advised to notify medical offices if gastrointestinal symptoms appeared during or after the competition [23]. In the meantime, the awareness and education of the public as well as of health care professionals are warranted [57] (Figure 4, Multimedia Appendix 1 Tables S1 and S2).

The mainstay for vector-borne infections prevention is to avoid mosquito bites [43,58]. Participants must take extra care to follow recommendations from the sports event organizing committee for insect repellant and clothing treatment [59]. Some equipment was recommended, including rooms with air conditioning, window and door screens, mosquito bed nets, and official registered insect repellents, which are proven to be safe and effective [43] (Figure 4, Multimedia Appendix 1 Tables S1 and S2).

Emphasis on the prevention of transmission of blood-borne pathogens in athletes should focus on reducing the traditional modes of transmission and modification of off-the-field behavior, including the avoidance of unprotected high-risk sexual activity and needle sharing in the use of illicit drugs such as anabolic steroids or blood doping [60,61]. Athletes should also practice good personal hygiene and not share personal items that might be contaminated with blood, such as razors, toothbrushes, and nail clippers [60,61]. In addition, education regarding universal precautions when dealing with blood or body fluids should also be enhanced [60,61] (Figure 4, Multimedia Appendix 1 Tables S1 and S2).

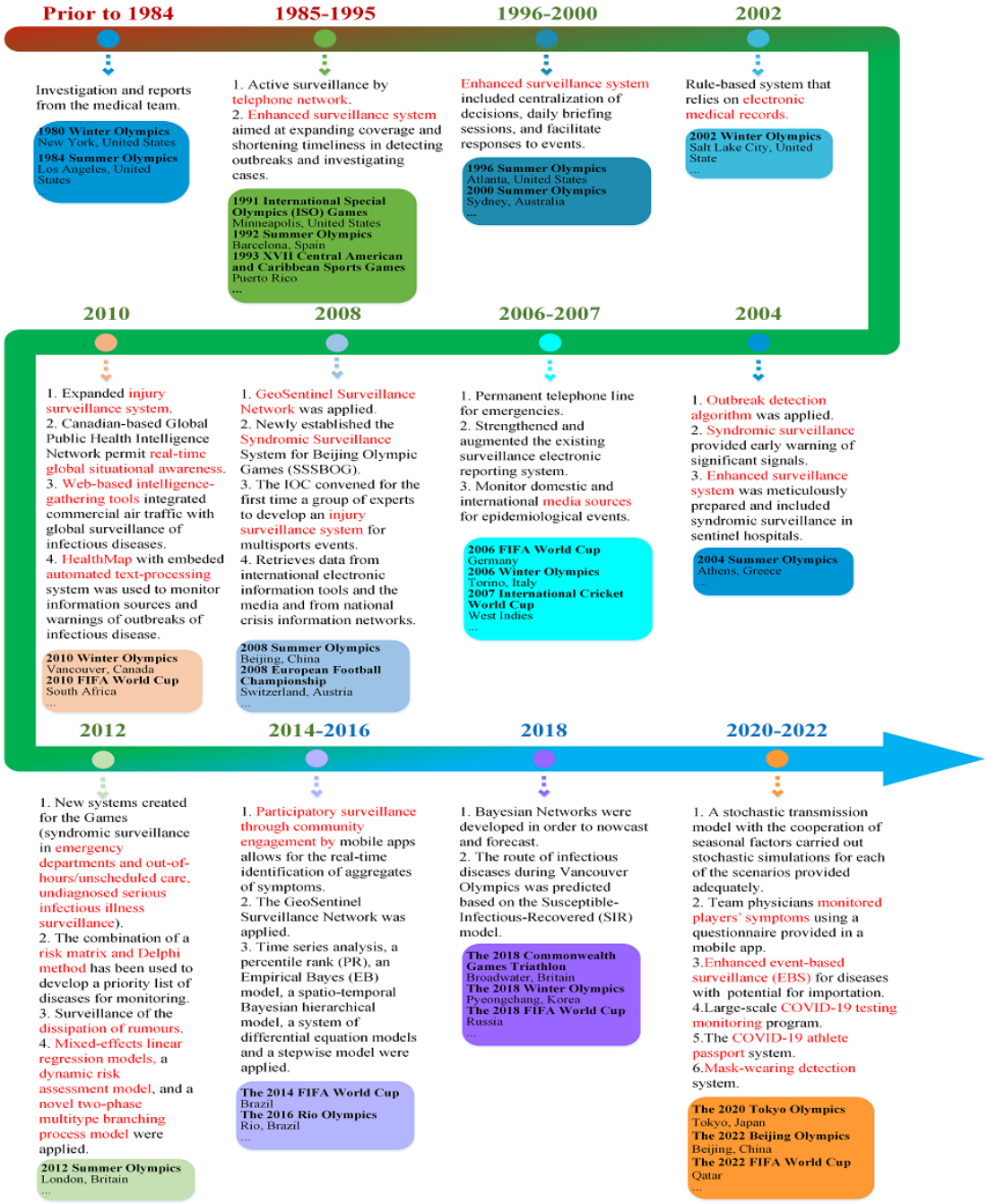
To prevent water-contact infections in eco-challenge and racing competitions, racers and organizers should be aware of the potential risk of inadvertent ingestion of muddy and possibly contaminated water during the race [62]. In general, planners of these competitions should consider building circuits to avoid areas heavily contaminated with animal feces [62]. In addition, clean running water should be available at stations to allow racers to clean mud off their hands and faces prior to eating and drinking [57]. An alternate form of water delivery should also be considered, such as bottles of water, which are less easily contaminated than paper cups [57]. It is also recommended that bike racers use front and rear fenders to reduce splashing of mud onto their and other riders' faces, respectively [57] (Figure 4, Multimedia Appendix 1 Tables S1 and S2).

Surveillance Paradigm

Before the 1990s, major sports events mainly used traditional methods for disease surveillance. Traditional disease surveillance relied on routine manual filing of reportable diseases by clinicians [16-21]. With the continuous development of information technology, infectious disease surveillance technology in major sports events has gone through three stages of development.

The first generation of a complete disease surveillance network for major sports events was applied in the 1984 Los Angeles Olympic Games. The whole network covered 46 hospitals, 90 physician offices, 4 university student health centers, 31 preschools, 3 Olympic Village polyclinics, and 24 Olympic first-aid stations [63]. During the Games, the International Olympic Committee made three phone calls per week with 198 participating sites. Each site used a disease report card to report disease information [63]. Early active surveillance systems reported structured and simple data, mainly including the name of surveillance sites, type of surveillance sites, date, and number of cases (grouped by disease type and age of patients) [64-66] (Figure 5).

Figure 5. Surveillance patterns in major sports events. FIFA: International Federation of Association Football; IOC: International Olympic Committee.



After 1995, the second generation of surveillance networks emerged with extensive application of internet technology in infectious disease surveillance that enhanced existing systems and integrated multisource data with richer data types that significantly improved real-time performance. During the 1996 Atlanta Olympic Games, the government designed and implemented "outside the fence" and "inside the fence" surveillance systems. The "outside the fence" system was an augmented surveillance system implemented for health conditions that occurred outside of Olympic venues, while the "inside the fence" surveillance system was established to enumerate clinical encounters in Olympic venues and at contract hospitals for Olympic athletes, official Olympic staff, and national delegations [67]. Continuous surveillance data collection allowed uninterrupted monitoring of disease epidemic trends.

In the following major sports events (eg, 2000 Sydney Olympics, 2006 Germany FIFA World Cup, 2007 International Cricket World Cup, 2008 EURO, 2008 Beijing Olympics), the infectious disease surveillance system continued using this mode, with two major improvements, including enhancing the existing domestic surveillance system and establishing syndromic surveillance targeting health-related symptoms [4-6,68,69]. Data structures and types were enriched in the form of electronic medical records. Automatic data acquisition technology improved the timeliness of information collection and reporting. Complex algorithms and models were used to predict outbreaks of infectious diseases. Several professional global surveillance

networks (eg, GeoSentinel, Healthmap) were set up, collecting clinician-based sentinel surveillance data (diagnosis, travel itinerary, demographics) to support the prevention and control of infectious diseases in major sports events [22,70].

After 2016, mobile internet apps and artificial intelligence algorithms were applied to monitor participants' health status in sports events. In the 2016 Rio Olympics, a surveillance platform based on a mobile app was installed on participants' smartphones called "OlymTRIP," which monitored health status through a daily interactive check of user health status, including geolocalization data [71]. The app also provided information and advice about Zika infection. In case of feeling unwell, participants could contact doctors through the app, and doctors could also track the health status of the participants in a real-time manner using a web-based platform.

In addition to mobile phones, there are many types of wearable devices for infectious disease monitoring. Using an agent-based model, the data collected by wearable devices were applied to explore the feasibility of using tracking data from a football match to assess interpersonal contact between individuals by calculating two measures of respiratory exposure [72]. More multisource data and more advanced technologies were integrated into the monitoring system, including online public opinion monitoring based on natural language processing technology, risk classification monitoring of international passengers based on global aviation data, and the risk early warning model of infectious diseases [22,24,73]. In addition, enhanced event-based surveillance was undertaken by the National Institute of Infectious Diseases, Japan, to identify infectious diseases occurring overseas (excluding COVID-19) that had potential for importation during the Tokyo Olympics [74].

In response to the challenges of COVID-19, some surveillance measures focusing on COVID-19 were developed and implemented. The COVID-19 athlete passport system was used during the Tokyo Olympic games to present a novel way of managing athletes' previous exposure, testing results, and vaccination status, which could be used to tailor appropriate precautions to each athlete [75]. A large-scale COVID-19 monitoring program involving daily testing was instituted during the US 2020 National Football League season [76]. For enhanced surveillance activities during EURO 2020, a specific identifier was marked in the system for each case within the observation period that might have had an association with the games to facilitate communication and timely case detection by local health authorities [77]. In addition, some image-based technologies of mask-wearing detection could contribute to monitoring the athletes' and spectators' mask-wearing behaviors during sports events [78] (Figure 5).

Discussion

Principal Findings

The most prominent characteristics of major sports events are mass gathering and cross-regional mass mobility, which may contribute to a pathogen's cross-regional, national, and even continental transmission. In recent years, more and more research on this topic has emerged, with a 13-fold increase in the number of research papers published since 2010 compared to that before 2000. There were a variety of infectious diseases of concern during major sporting events. Among them, the most predominant infectious diseases were respiratory tract infections, followed by gastrointestinal infections, vector-borne infections, blood-borne infections, and water-contact infections. It was noteworthy that massive outbreaks of infectious diseases in major sports events were generally scarce, mainly due to thorough prevention and control measures.

Major sports events are the hotbed of almost all types of infectious diseases. Therefore, in the process of prevention and control of infectious diseases in sports events, targeted prevention and control measures should be taken according to the main risk factors of infectious diseases. For vaccine-preventable diseases, routine vaccination coverage should be improved. The environmental conditions of the sports events venue should be monitored to protect the upper respiratory tract of participants. Food and drinking-water hygiene should be ensured to reduce the risk of gastrointestinal infection. For vector-borne diseases, the hidden danger of breeding insects should be eliminated. For blood-borne diseases, especially sexually transmitted infections, health education should be strengthened to reduce risky sexual behaviors. For water-contact infectious diseases prevention, bacteria and microorganisms in the water should be monitored to eliminate the risk of infection, and providing clean water for racing is essential. Previous studies have mainly focused on the disease epidemic and prevention measures for athletes and related personnel in host countries during sports events, whereas few studies have provided evidence on the follow-up monitoring of returning participants after major sports events. In future research, participating countries should strengthen health status monitoring of sports events–relevant individuals after they return, and, if necessary, form a closed loop for management to prevent transnational epidemics.

The control of infectious diseases in major sports events cannot be separated from the support of surveillance technology. The number of studies on surveillance of infectious diseases in major sports events in recent years also presented explosive growth. Additionally, improvements in technology (eg, mathematical modeling, big data, and artificial intelligence) have enhanced the ability to monitor the risk factors related to multiple key points during the epidemic course of infectious disease, and to realize early warning and prediction of infectious disease outbreaks. However, the application of virtual reality, big data, artificial intelligence, and other technologies in the prevention and control of COVID-19 in major sports events is still in its infancy, and the platform architecture and application mode of emerging technologies warrant further research. In the future, intelligent surveillance should be equipped with advanced tools to collect multisource heterogeneous global mass data in real time, and detect and warn about various risk factors through intelligent algorithms and models so as to realize the advance of the prevention point. The target of the development of surveillance technology in future major sports events is to realize the rapid identification of new outbreaks of emerging infectious

diseases and achieve the goal of gathering data faster than the epidemic situation.

Currently, we are nearly 3 years into the COVID-19 pandemic, resulting in many sports events having been suspended, postponed, or scaled back, which are now being gradually resumed. Based on the experiences of the Tokyo and Beijing Olympic Games, strict management measures, represented by a bubble scheme, closed-loop management, and frequent testing, could effectively prevent the spread of COVID-19 in the Olympic Village and the communities of host cities during the sports events [30,31]. However, international spectators were not allowed to attend the previous two Olympic Games. The gathering of spectators from all over the world, mutation of SARS-CoV-2, and the loosening of COVID-19 prevention and control policies in many countries are still putting pressure on the efficiency and safety of holding the upcoming World Cup in Qatar and other international sports events.

In the last year, the emergence of monkeypox outbreaks around the world has also posed challenges to the holding of international sports events. The transmission of monkeypox could occur through multiple routes, including close contact with respiratory droplets, infected lesions, body fluids, contaminated materials, and sexual behaviors, and the virus has a long incubation period [79]. Therefore, researchers are warning organizers and participants of the upcoming FIFA World Cup 2022 in Qatar to pay attention to monkeypox prevention and control, and to establish reliable communication channels among health authorities, visitors, and the local population [79]. Thus, the prevention and control of the risk of infectious diseases in major sports events is a continuously evolving topic. With the emergence of new infectious diseases and the change of pandemic risk of existing infectious diseases in host and participants countries, researchers, organizers, and participants of major sports events should always be vigilant.

Study Limitation

There is one key limitation of this study. Articles written in languages other than English were not included in this study, whereas relevant information about infectious diseases related to sports events might be published in the local media in the native language of the host countries. However, most major international sports events (such as the World Cup and Olympic Games) involve publications written in English. Our aim was to provide an overview and critical summary of existing evidence on infectious diseases in major sports events, and these English articles could cover this content. Although there may be some publications published in the local language in host countries of sports events, the key information of the sports events might not be notably missing if these publications were not included.

Conclusion

This scoping review provides an overview and critical summary of existing evidence on the risk factors, epidemics, prevention measures, and surveillance paradigm of infectious diseases in major sports events. We observed an increase in relevant research in recent years, with international sports events such as the Olympics and FIFA World Cup as the main focus. A variety of infectious diseases of concern during major sporting events were identified, including respiratory tract infections, gastrointestinal infections, vector-borne infections, blood-borne infections, and water-contact infections, among which respiratory tract infections were the most common. Surveillance of infectious diseases during major sports events has made great progress in recent years. In particular, progress has been made to update the surveillance paradigm from a manual archive to a network-based system, and finally to automated intelligent platforms gathering multisource data that are analyzed in a timely manner by multiple algorithms and mathematical models. In the future, it will be essential to strengthen the monitoring of various risk factors of infectious diseases, improve the accuracy of intelligent algorithms and models, and ensure the timely detection of the emergence of an epidemic to effectively mitigate the risk.

Acknowledgments

We appreciate the support of the Global Health and Infectious Diseases Group, Peking University. This work was supported by the National Natural Science Foundation of China (72104008), China Postdoctoral Science Foundation (2022M710250), and the Beijing Organizing Committee for the 2022 Olympic and Paralympic Winter Games (2021YFF0306001).

Authors' Contributions

XY, ZJ, and BZ were responsible for the study design; XY, YL, and BZ contributed to the study selection and data extraction; XY, YF, and BZ contributed to writing the manuscript. All authors have reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy and detailed study characteristics (Tables S1-S3).
[DOCX File , 425 KB - publichealth_v8i12e40042_app1.docx]

References

1. Major international sports events. Topend Sports. URL: https://www.topendsports.com/events/sport-events.htm [accessed 2021-10-20]
2. Daly P, Gustafson R. Public health recommendations for athletes attending sporting events. Clin J Sport Med 2011 Jan;21(1):67-70. [doi: 10.1097/JSM.0b013e31820758dc] [Medline: 21200174]
3. Smith DR, Wei N. Traveling to China for the Beijing Olympic Games: internal migration of the commercial sex workforce may be an important HIV risk factor for travelers. Travel Med Infect Dis 2008 May;6(3):160; author reply 161. [doi: 10.1016/j.tmaid.2008.01.010] [Medline: 18486075]
4. Payne L, Arias P, Kreidl P, Ciotti M, Coulombier D. Preparedness activities ahead of the Beijing 2008 Olympic Games--enhancing EU epidemic intelligence. Euro Surveill 2008 Aug 07;13(32):18947 [FREE Full text] [Medline: 18761898]
5. Leggat PA, Shaw MT. From Sydney to Athens: preparing for the Olympic and Paralympic Games. Travel Med Infect Dis 2003 Nov;1(4):201-203. [doi: 10.1016/j.tmaid.2003.11.004] [Medline: 17291918]
6. Williams CJ, Schenkel K, Eckmanns T, Altmann D, Krause G. FIFA World Cup 2006 in Germany: enhanced surveillance improved timeliness and detection. Epidemiol Infect 2009 Apr;137(4):597-605. [doi: 10.1017/S095026880800112X] [Medline: 18667109]
7. Ehresmann KR, Hedberg CW, Grimm MB, Norton CA, MacDonald KL, Osterholm MT. An outbreak of measles at an international sporting event with airborne transmission in a domed stadium. J Infect Dis 1995 Mar 01;171(3):679-683. [doi: 10.1093/infdis/171.3.679] [Medline: 7876616]
8. Gundlapalli AV, Rubin MA, Samore MH, Lopansri B, Lahey T, McGuire HL, et al. Influenza, Winter Olympiad, 2002. Emerg Infect Dis 2006 Jan;12(1):144-146 [FREE Full text] [doi: 10.3201/eid1201.050645] [Medline: 16494733]
9. Schenkel K, Williams C, Eckmanns T, Poggensee G, Benzler J, Josephsen J, et al. Enhanced surveillance of infectious diseases: the 2006 FIFA World Cup experience, Germany. Euro Surveill 2006 Dec;11(12):15-16. [doi: 10.2807/esm.11.12.00670-en] [Medline: 29208141]
10. Kim K, Jang JY, Moon G, Shim H, Jung PY, Kim S, et al. Experiences of the emergency department at the Pyeongchang Polyclinic during the 2018 PyeongChang Winter Olympic Games. Yonsei Med J 2019 May;60(5):474-480 [FREE Full text] [doi: 10.3349/ymj.2019.60.5.474] [Medline: 31016910]
11. Morgan J, Bornstein SL, Karpati AM, Bruce M, Bolin CA, Austin CC, Leptospirosis Working Group. Outbreak of leptospirosis among triathlon participants and community residents in Springfield, Illinois, 1998. Clin Infect Dis 2002 Jun 15;34(12):1593-1599. [doi: 10.1086/340615] [Medline: 12032894]
12. Gunby P. Boxing: AIDS? JAMA 1988 Mar 18;259(11):1613-1614. [doi: 10.1001/jama.259.11.1613] [Medline: 3343763]
13. Wilson M, Chen L, Han P, Keystone JS, Cramer JP, Segurado A, GeoSentinel Surveillance Network. Illness in travelers returned from Brazil: the GeoSentinel experience and implications for the 2014 FIFA World Cup and the 2016 Summer Olympics. Clin Infect Dis 2014 May;58(10):1347-1356 [FREE Full text] [doi: 10.1093/cid/ciu122] [Medline: 24585698]
14. Dove J, Gage A, Kriz P, Tabaddor RR, Owens BD. COVID-19 and review of current recommendations for return to athletic play. R I Med J 2020 Sep 01;103(7):15-20 [FREE Full text] [Medline: 32872685]
15. Herzog M, Fraser M, Register-Mihalik J, Kerr ZY. Epidemiology of skin infections in men's wrestling: analysis of 2009-2010 through 2013-2014 National Collegiate Athletic Association Surveillance data. J Athl Train 2017 May;52(5):457-463 [FREE Full text] [doi: 10.4085/1062-6050-52.2.16] [Medline: 28362160]
16. Shute P, Jeffries DJ, Maddocks AC. Scrum-pox caused by herpes simplex virus. Br Med J 1979 Dec 22;2(6205):1629 [FREE Full text] [doi: 10.1136/bmj.2.6205.1629] [Medline: 534919]
17. Stiles MH. Medical preparation for the Olympic Games. JAMA 1968 Sep 09;205(11):771-774. [Medline: 5695285]
18. Haid B. Medical care at winter olympic games. Logistics. JAMA 1972 Aug 28;221(9):990-992. [Medline: 5068304]
19. Laurin CA, Létourneau G. Medical report of the Montreal Olympic Games. Am J Sports Med 1978 Apr 23;6(2):54-61. [doi: 10.1177/036354657800600204] [Medline: 646008]
20. Gustafson TL, Booth AL, Fricker RS, Cureton E, Fowinkle EW, Hutcheson RH, et al. Disease surveillance and emergency services at the 1982 World's Fair. Am J Public Health 1987 Jul;77(7):861-863. [doi: 10.2105/ajph.77.7.861] [Medline: 3592044]
21. Weiss B, Mascola L, Fannin S. Public health at the 1984 Summer Olympics: the Los Angeles County experience. Am J Public Health 1988 Jun;78(6):686-688. [doi: 10.2105/ajph.78.6.686] [Medline: 3285705]
22. Alshammari, Mikler. Big data opportunities for disease outbreaks detection in global mass gatherings. 2018 Presented at: ICBDE '18: 2018 International Conference on Big Data and Education; March 9-11, 2018; Honolulu, HI p. 16-21. [doi: 10.1145/3206157.3206160]
23. Thackway SV, Delpech VC, Jorm LR, McAnulty JM, Visotina M. Monitoring acute diseases during the Sydney 2000 Olympic and Paralympic Games. Med J Aust 2000 Sep 18;173(6):318-321. [doi: 10.5694/j.1326-5377.2000.tb125666.x] [Medline: 11061404]
24. Khan K, Freifeld CC, Wang J, Mekaru SR, Kossowsky D, Sonricker AL, et al. Preparing for infectious disease threats at mass gatherings: the case of the Vancouver 2010 Olympic Winter Games. CMAJ 2010 Apr 06;182(6):579-583 [FREE Full text] [doi: 10.1503/cmaj.100093] [Medline: 20181726]

25. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Method 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
26. 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. [doi: 10.7326/m18-0850]
27. Khan E. 10 Greatest sport events – world's major sporting events. WondersList. URL: https://www.wonderslist.com/10-greatest-sport-events/ [accessed 2022-10-13]
28. Valtonen M, Waris M, Vuorinen T, Eerola E, Hakanen AJ, Mjosund K, et al. Common cold in Team Finland during 2018 Winter Olympic Games (PyeongChang): epidemiology, diagnosis including molecular point-of-care testing (POCT) and treatment. Br J Sports Med 2019 Sep 29;53(17):1093-1098 [FREE Full text] [doi: 10.1136/bjsports-2018-100487] [Medline: 31142472]
29. Jindahra P, Wongboonsin K, Wongboonsin P. Demographic and initial outbreak patterns of COVID-19 in Thailand. J Popul Res 2022 Nov 09;39(4):567-588 [FREE Full text] [doi: 10.1007/s12546-021-09276-y] [Medline: 34776786]
30. Akashi H, Shimada S, Tamura T, Chinda E, Kokudo N. SARS-CoV-2 infections in close contacts of positive cases in the Olympic and Paralympic Village at the 2021 Tokyo Olympic and Paralympic Games. JAMA 2022 Mar 08;327(10):978-980 [FREE Full text] [doi: 10.1001/jama.2022.0818] [Medline: 35113161]
31. Kong L, Duan M, Shi J, Hong J, Zhou X, Yang X, et al. Optimization of COVID-19 prevention and control measures during the Beijing 2022 Winter Olympics: a model-based study. Infect Dis Poverty 2022 Sep 06;11(1):95 [FREE Full text] [doi: 10.1186/s40249-022-01019-2] [Medline: 36068625]
32. Chen T, Kutty P, Lowe L, Hunt EA, Blostein J, Espinoza R, et al. Measles outbreak associated with an international youth sporting event in the United States, 2007. Pediatr Infect Dis J 2010 Sep;29(9):794-800. [doi: 10.1097/INF.0b013e3181dbaacf] [Medline: 20400927]
33. The Lancet. We need a global conversation on the 2020 Olympic Games. Lancet 2021 Jun 12;397(10291):2225. [doi: 10.1016/S0140-6736(21)01293-9] [Medline: 34119048]
34. Marsh K, Griffiths E, Young J, Gibb CA, McMenamin J. Contributions of the EURO 2020 football championship events to a third wave of SARS-CoV-2 in Scotland, 11 June to 7 July 2021. Euro Surveill 2021 Aug;26(31):2100707 [FREE Full text] [doi: 10.2807/1560-7917.ES.2021.26.31.2100707] [Medline: 34355691]
35. Riccardo F, Frisicale EM, Guzzetta G, Ferraro F, Merler S, Maringhini G, et al. Winning during a pandemic: epidemiology of SARS-CoV-2 during EURO2020 in Italy. Epidemiol Infect 2022 Apr 22;150:e166 [FREE Full text] [doi: 10.1017/S0950268822000723] [Medline: 35450542]
36. Gualano B, Brito GM, Pinto AJ, Lemes IR, Matos LDNJ, de Sá Pinto AL, Coalition SPORT-COVID-19. High SARS-CoV-2 infection rate after resuming professional football in São Paulo, Brazil. Br J Sports Med 2021 Jul 05;56(18):1004-1007 [FREE Full text] [doi: 10.1136/bjsports-2021-104431] [Medline: 34226184]
37. Papagiannis D, Laios T, Tryposkiadis K, Kouriotis K, Roussis X, Basdekis G, et al. COVID-19 infection among elite football players: a nationwide prospective cohort study. Vaccines 2022 Apr 19;10(5):634 [FREE Full text] [doi: 10.3390/vaccines10050634] [Medline: 35632390]
38. Six C, Aboukais S, Giron S, D'Oliveira JC, Peloux-Petiot F, Franke F, et al. Outbreak of diarrhoeal illness in participants in an obstacle adventure race, Alpes-Maritimes, France, June 2015. Euro Surveill 2016 Jun 09;21(23):30253 [FREE Full text] [doi: 10.2807/1560-7917.ES.2016.21.23.30253] [Medline: 27311488]
39. Jeong M, Song Y, Ju S, Kim SH, Kwak HS, An ES. Surveillance to prevent the spread of Norovirus outbreak from asymptomatic food handlers during the PyeongChang 2018 Olympics. J Food Prot 2021 Oct 01;84(10):1819-1823. [doi: 10.4315/JFP-21-136] [Medline: 34115864]
40. Edouard P, Junge A, Sorg M, Timpka T, Branco P. Illnesses during 11 international athletics championships between 2009 and 2017: incidence, characteristics and sex-specific and discipline-specific differences. Br J Sports Med 2019 Sep;53(18):1174-1182. [doi: 10.1136/bjsports-2018-100131] [Medline: 30862705]
41. Sheikh K. Zika-proofing the Olympics. Sci Am 2016 Aug 19;315(2):19. [doi: 10.1038/scientificamerican0816-19] [Medline: 27459560]
42. Wiwanitkit V. Zika virus infection and the Olympic Games 2016. Rev Clin Esp 2017 Apr;217(3):178-179. [doi: 10.1016/j.rce.2016.12.005] [Medline: 28117035]
43. Toresdahl BG, Asif IM. Update on Zika virus: considerations for the traveling athlete. Sports Health 2016 Sep 20;8(5):438-443 [FREE Full text] [doi: 10.1177/1941738116659060] [Medline: 27436751]
44. Shadgan B, Khodaee M. Smart clothing is the best protection against Zika during Rio Olympic and Paralympic Games. Clin J Sport Med 2016 Jul;26(4):345-346. [doi: 10.1097/JSM.0000000000000365] [Medline: 27359297]
45. Shadgan B, Pakravan A, Zaeimkohan H, Shahpar FM, Khodaee M. Zika and Rio Olympic Games. Curr Sports Med Rep 2016;15(4):298-300. [doi: 10.1249/JSR.0000000000000278] [Medline: 27399828]
46. Deering KN, Chettiar J, Chan K, Taylor M, Montaner JS, Shannon K. Sex work and the public health impacts of the 2010 Olympic Games. Sex Transm Infect 2012 Jun 21;88(4):301-303 [FREE Full text] [doi: 10.1136/sextrans-2011-050235] [Medline: 22436199]
47. Pavli A, Maltezou HC. Travelling to Greece for the summer 2011 Special Olympics. Travel Med Infect Dis 2011 May;9(3):135-141. [doi: 10.1016/j.tmaid.2011.04.001] [Medline: 21570357]

48. Gaines J, Sotir MJ, Cunningham TJ, Harvey KA, Lee CV, Stoney RJ, et al. Health and safety issues for travelers attending the World Cup and Summer Olympic and Paralympic Games in Brazil, 2014 to 2016. JAMA Intern Med 2014 Aug 01;174(8):1383-1390 [FREE Full text] [doi: 10.1001/jamainternmed.2014.2227] [Medline: 24887552]
49. Hochedez P, Rosine J, Théodose R, Abel S, Bourhy P, Picardeau M, et al. Outbreak of leptospirosis after a race in the tropical forest of Martinique. Am J Trop Med Hyg 2011 Apr;84(4):621-626 [FREE Full text] [doi: 10.4269/ajtmh.2011.10-0502] [Medline: 21460020]
50. Radl C, Müller M, Revilla-Fernandez S, Karner-Zuser S, de Martin A, Schauer U, et al. Outbreak of leptospirosis among triathlon participants in Langau, Austria, 2010. Wien Klin Wochenschr 2011 Dec 24;123(23-24):751-755. [doi: 10.1007/s00508-011-0100-2] [Medline: 22105111]
51. Centers for Disease Control (CDC). Measles at an international gymnastics competition--Indiana, 1991. MMWR Morb Mortal Wkly Rep 1992 Feb 21;41(7):109-111. [Medline: 1736086]
52. Young M, Fricker P, Maughan R, MacAuley D. The travelling athlete: issues relating to the Commonwealth Games, Malaysia, 1998. Br J Sports Med 1998 Mar;32(1):77-81 [FREE Full text] [Medline: 9562173]
53. Zumla A, McCloskey B, Bin Saeed AA, Dar O, Al Otabi B, Perlmann S, et al. What is the experience from previous mass gathering events? Lessons for Zika virus and the Olympics 2016. Int J Infect Dis 2016 Jun;47:1-4 [FREE Full text] [doi: 10.1016/j.ijid.2016.06.010] [Medline: 27321962]
54. Furusawa K, Tajima F, Okawa H, Takahashi M, Ogata H. The incidence of post-race symptoms of upper respiratory tract infection in wheelchair marathon racers. Spinal Cord 2007 Jul 6;45(7):513-517. [doi: 10.1038/sj.sc.3102028] [Medline: 17279097]
55. Yashio T, Murayama A, Kami M, Ozaki A, Tanimoto T, Rodriguez-Morales AJ. COVID-19 infection during the Olympic and Paralympic Games Tokyo 2020. Travel Med Infect Dis 2021 Nov;44:102205 [FREE Full text] [doi: 10.1016/j.tmaid.2021.102205] [Medline: 34785374]
56. Wang X, Cai Y, Zhang B, Zhang X, Wang L, Yan X, et al. Cost-effectiveness analysis on COVID-19 surveillance strategy of large-scale sports competition. Infect Dis Poverty 2022 Mar 18;11(1):32 [FREE Full text] [doi: 10.1186/s40249-022-00955-3] [Medline: 35303953]
57. Stuart TL, Sandhu J, Stirling R, Corder J, Ellis A, Misa P, et al. Campylobacteriosis outbreak associated with ingestion of mud during a mountain bike race. Epidemiol Infect 2010 Dec;138(12):1695-1703. [doi: 10.1017/S095026881000049X] [Medline: 20334726]
58. Milne CJ, Shaw MTM. Travelling to China for the Beijing 2008 Olympic Games. Br J Sports Med 2008 May 01;42(5):321-326. [doi: 10.1136/bjsm.2007.038794] [Medline: 18381825]
59. Petersen E, Wilson ME, Touch S, McCloskey B, Mwaba P, Bates M, et al. Rapid spread of Zika virus in the Americas--implications for public health preparedness for mass gatherings at the 2016 Brazil Olympic Games. Int J Infect Dis 2016 Mar;44:11-15 [FREE Full text] [doi: 10.1016/j.ijid.2016.02.001] [Medline: 26854199]
60. Shaw MT, Leggat PA, Borwein S. Travelling to china for the beijing 2008 olympic and paralympic games. Travel Med Infect Dis 2007 Nov;5(6):365-373. [doi: 10.1016/j.tmaid.2007.09.003] [Medline: 17983975]
61. Wilson ME, Chen LH. Health risks among travelers to Brazil: implications for the 2014 FIFA World Cup and 2016 Olympic Games. Travel Med Infect Dis 2014 May;12(3):205-207 [FREE Full text] [doi: 10.1016/j.tmaid.2014.04.001] [Medline: 24786668]
62. Stern E, Galloway R, Shadomy S, Wannemuehler K, Atrubin D, Blackmore C, et al. Outbreak of leptospirosis among Adventure Race participants in Florida, 2005. Clin Infect Dis 2010 Mar 15;50(6):843-849. [doi: 10.1086/650578] [Medline: 20146629]
63. Weiss B, Mascola L, Fannin S. Public health at the 1984 Summer Olympics: the Los Angeles County experience. Am J Public Health 1988 Jun;78(6):686-688. [doi: 10.2105/ajph.78.6.686] [Medline: 3285705]
64. Reintjes R, Kistemann T, MacLehose L, McKee M, Gill N, Weinberg J, et al. Detection and response to a meningococcal disease outbreak following a youth football tournament with teams from four European countries. Int J Hyg Environ Health 2002 May;205(4):291-296. [doi: 10.1078/1438-4639-00156] [Medline: 12068748]
65. Centers for Disease Control and Prevention (CDC). Update: leptospirosis and unexplained acute febrile illness among athletes participating in triathlons--Illinois and Wisconsin, 1998. MMWR Morb Mortal Wkly Rep 1998 Aug 21;47(32):673-676 [FREE Full text] [Medline: 9729034]
66. Centers for Disease Control and Prevention. Outbreak of acute febrile illness among athletes participating in triathlons—Wisconsin and Illinois, 1998. JAMA 1998 Nov 04;280(17):1473. [doi: 10.1001/jama.280.17.1473]
67. Meehan P, Toomey KE, Drinnon J, Cunningham S, Anderson N, Baker E. Public health response for the 1996 Olympic Games. JAMA 1998 May 13;279(18):1469-1473. [doi: 10.1001/jama.279.18.1469] [Medline: 9600482]
68. Editorial team. Health risks during the Cricket World Cup in the Caribbean: surveillance and assessment in the French departements. Euro Surveill 2007 Mar 08;12(3):E070308.2. [doi: 10.2807/esw.12.10.03151-en] [Medline: 17439782]
69. Kreidl P, Buxbaum P, Santos-O'Connor F, Payne L, Strauss R, Hrabcik H, et al. 2008 European Football Championship--ECDC epidemic intelligence support. Euro Surveill 2008 Aug 07;13(32):18946 [FREE Full text] [Medline: 18761899]

70. Davis X, MacDonald S, Borwein S, Freedman DO, Kozarsky PE, von Sonnenburg F, GeoSentinel Surveillance Network. Health risks in travelers to China: the GeoSentinel experience and implications for the 2008 Beijing Olympics. Am J Trop Med Hyg 2008 Jul;79(1):4-8. [Medline: 18606756]
71. Rodriguez-Valero N, Luengo Oroz M, Cuadrado Sanchez D, Vladimirov A, Espriu M, Vera I, et al. Mobile based surveillance platform for detecting Zika virus among Spanish Delegates attending the Rio de Janeiro Olympic Games. PLoS One 2018 Aug 22;13(8):e0201943 [FREE Full text] [doi: 10.1371/journal.pone.0201943] [Medline: 30133492]
72. Gonçalves B, Mendes R, Folgado H, Figueiredo P, Travassos B, Barros H, et al. Can tracking data help in assessing interpersonal contact exposure in team sports during the COVID-19 pandemic? Sensors 2020 Oct 29;20(21):6163 [FREE Full text] [doi: 10.3390/s20216163] [Medline: 33138115]
73. Gundlapalli AV, Olson J, Smith SP, Baza M, Hausam RR, Eutropius L, et al. Hospital electronic medical record-based public health surveillance system deployed during the 2002 Winter Olympic Games. Am J Infect Control 2007 Apr;35(3):163-171. [doi: 10.1016/j.ajic.2006.08.003] [Medline: 17433939]
74. Kasamatsu A, Ota M, Shimada T, Fukusumi M, Yamagishi T, Samuel A, et al. Enhanced event-based surveillance for imported diseases during the Tokyo 2020 Olympic and Paralympic Games. Western Pac Surveill Response J 2021 Oct 01;12(4):13 [FREE Full text] [doi: 10.5365/wpsar.2021.12.4.903] [Medline: 35251745]
75. McLarnon M, Heron N. The COVID-19 athlete passport: a tool for managing athlete COVID-19 status surrounding the Tokyo 2020 Olympic games. Phys Sportsmed 2021 Nov 14;49(4):367-370. [doi: 10.1080/00913847.2021.1885964] [Medline: 33535866]
76. Mack CD, Osterholm M, Wasserman EB, Petruski-Ivleva N, Anderson DJ, Myers E, et al. Optimizing SARS-CoV-2 surveillance in the United States: insights from the National Football League Occupational Health Program. Ann Intern Med 2021 Aug;174(8):1081-1089 [FREE Full text] [doi: 10.7326/M21-0319] [Medline: 34125571]
77. Heese H, Marquis A, Diercke M, Markus I, Böhm S, Metz J, et al. Results of the enhanced COVID-19 surveillance during UEFA EURO 2020 in Germany. Epidemiol Infect 2022 Mar 03;150:1-18 [FREE Full text] [doi: 10.1017/S0950268822000449] [Medline: 35236530]
78. Vest JR, Blackburn J, Cash-Goldwasser S, Peters Bergquist E, Embi PJ. Mask-wearing behavior at the 2021 NCAA men's basketball tournament. JAMA 2021 Oct 05;326(13):1324-1325 [FREE Full text] [doi: 10.1001/jama.2021.14057] [Medline: 34398222]
79. Farahat R, Setti M, Benmelouka A, Ali I, Umar TP, Alabdallat Y, et al. Monkeypox emergence and hosting a safe FIFA World Cup 2022 in Qatar: Challenges and recommendations. Int J Surg 2022 Oct;106:106935 [FREE Full text] [doi: 10.1016/j.ijsu.2022.106935] [Medline: 36130706]

Abbreviations

EURO: European Football Championship
FIFA: International Federation of Association Football
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by G Eysenbach; submitted 02.06.22; peer-reviewed by L Zhang; comments to author 10.10.22; revised version received 01.11.22; accepted 04.11.22; published 02.12.22.

Please cite as:
Yan X, Fang Y, Li Y, Jia Z, Zhang B
Risks, Epidemics, and Prevention Measures of Infectious Diseases in Major Sports Events: Scoping Review
JMIR Public Health Surveill 2022;8(12):e40042
URL: https://publichealth.jmir.org/2022/12/e40042
doi:10.2196/40042
PMID:36459401

Original Paper

Digitalizing and Upgrading Severe Acute Respiratory Infections Surveillance in Malta: System Development

John Paul Cauchi[1], MSc, MD, PhD; Maria-Louise Borg[1], MSc, MD; Aušra Džiugytė[1], BSc, MSc; Jessica Attard[1], BSc, MSc; Tanya Melillo[1], MSc, MD, PhD; Graziella Zahra[2], DMedSc; Christopher Barbara[2], DLSHTM, MSc, MD; Michael Spiteri[3], MD; Allan Drago[3], MD, MRCS; Luke Zammit[3], MD; Joseph Debono[4], MD; Jorgen Souness[4], BSc, MSc; Steve Agius[4], BSc, MBA; Sharon Young[4]; Alan Dimech[4]; Ian Chetcuti[4], BSc; Mark Camenzuli[4], BSc; Ivan Borg[4], BSc; Neville Calleja[5], MSc, MD, PhD; Lorraine Tabone[6], MD; Charmaine Gauci[7], MD, PhD; Pauline Vassallo[1], BChD, MSc, DDPH; Joaquin Baruch[1,8], DVSc, PhD

[1]Health Promotion and Disease Prevention Directorate, Msida, Malta

[2]Molecular Dianostics Pathology Department, Mater dei Hospital, Msida, Malta

[3]Emergency Department, Mater Dei Hospital, Msida, Malta

[4]Mater Dei Hospital, Msida, Malta

[5]Directorate for Health Information and Research, Msida, Malta

[6]Primary Health Care, Ministry of Health, Floriana, Malta

[7]Superintendent of Public Health, Ministry of Health, Msida, Malta

[8]European Programme for Intervention Epidemiology Training program, European Centre for Disease Prevention and Control, Solna, Sweden

Corresponding Author:
John Paul Cauchi, MSc, MD, PhD
Health Promotion and Disease Prevention Directorate
1, St. Luke's Square, Gwardamanġa
Msida, PTA 1010
Malta
Phone: 356 79899999
Email: john-paul.cauchi@gov.mt

Abstract

Background: In late 2020, the European Centre for Disease Prevention and Control and Epiconcept started implementing a surveillance system for severe acute respiratory infections (SARI) across Europe.

Objective: We sought to describe the process of digitizing and upgrading SARI surveillance in Malta, an island country with a centralized health system, during the COVID-19 pandemic from February to November 2021. We described the characteristics of people included in the surveillance system and compared different SARI case definitions, including their advantages and disadvantages. This study also discusses the process, output, and future for SARI and other public health surveillance opportunities.

Methods: Malta has one main public hospital where, on admission, patient data are entered into electronic records as free text. Symptoms and comorbidities are manually extracted from these records, whereas other data are collected from registers. Collected data are formatted to produce weekly and monthly reports to inform public health actions. From October 2020 to February 2021, we established an analogue incidence-based system for SARI surveillance. From February 2021 onward, we mapped key stakeholders and digitized most surveillance processes.

Results: By November 30, 2021, 903 SARI cases were reported, with 380 (42.1%) positive for SARS-CoV-2. Of all SARI hospitalizations, 69 (7.6%) were admitted to the intensive care unit, 769 (85.2%) were discharged, 27 (3%) are still being treated, and 107 (11.8%) died. Among the 107 patients who died, 96 (89.7%) had more than one underlying condition, the most common of which were hypertension (n=57, 53.3%) and chronic heart disease (n=49, 45.8%).

Conclusions: The implementation of enhanced SARI surveillance in Malta was completed by the end of May 2021, allowing the monitoring of SARI incidence and patient characteristics. A future shift to register-based surveillance should improve SARI detection through automated processes.

(JMIR Public Health Surveill 2022;8(12):e37669) doi:10.2196/37669

KEYWORDS

surveillance; public health; epidemiology; COVID-19; disease prevention; disease surveillance; digital health; health system; pandemic; public hospital; patient data; health data; electronic record; monitoring

Introduction

Severe Acute Respiratory Infections Surveillance

In 2020, the European Centre for Disease Prevention and Control (ECDC) invited countries across Europe to collaborate in setting up a surveillance system for patients with severe acute respiratory infections (SARI). In general terms, a patient with SARI is defined as a patient who is hospitalized presenting with acute respiratory symptoms (see more details below). Public health surveillance systems are vital tools to monitor the incidence and severity of infectious diseases in a population [1,2]. Most European countries had already set up a surveillance system monitoring influenza-like illness and regularly report to the European Influenza Surveillance Network [3].

With the onset of COVID-19, an international surveillance program (European SARI Network [E-SARI-Net]) was rolled out in the World Health Organization (WHO) European Region, focusing on SARI and encouraging European collaboration between public health authorities while also monitoring known and novel illnesses, such as influenza and COVID-19. The primary objectives of this international surveillance program are as follows: (1) to monitor incidence trends and describe SARI cases by etiology and demographics; (2) to describe the intensity and activity of SARI; (3) to identify, describe, and monitor at-risk groups for severe disease; and (4) to detect unusual and unexpected events (eg, new pathogens). Secondary objectives are as follows: (1) to assess the impact of SARI on health systems and the impact of interventions on SARI incidence and (2) to assess and analyze the SARI burden of disease.

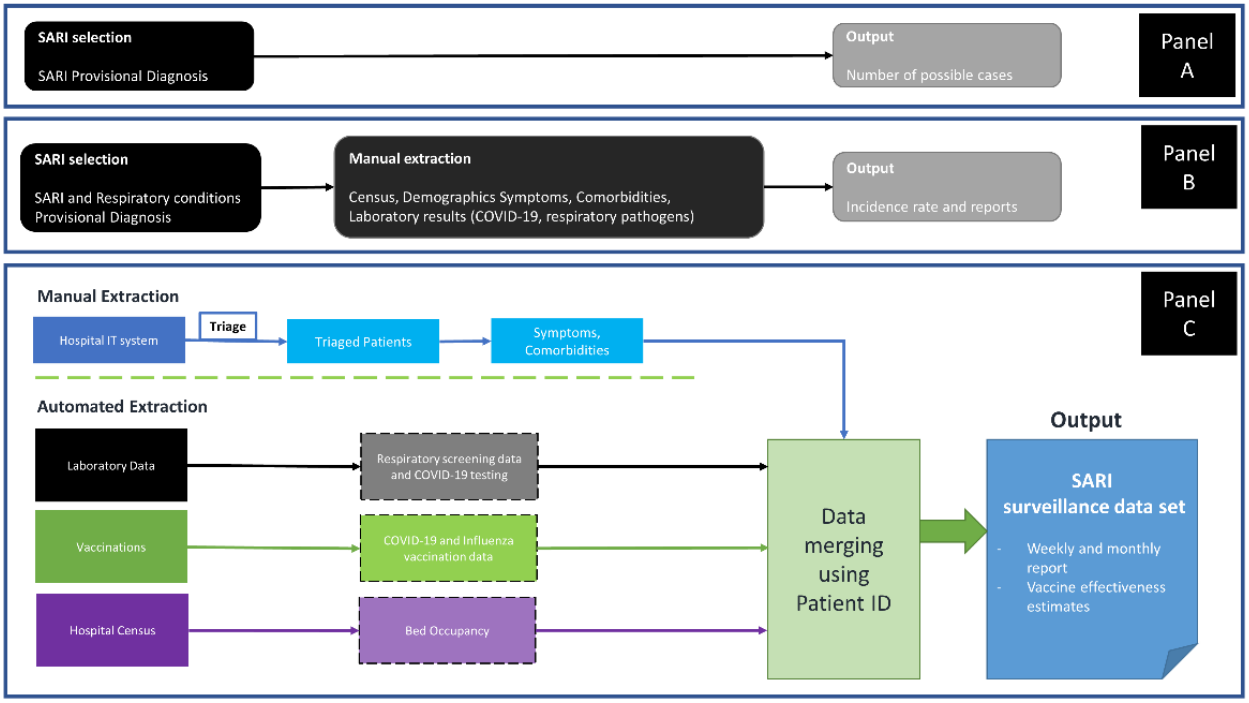
The rollout and establishment of the surveillance system were supervised by Epiconcept [4], who liaised with existing and newly established SARI surveillance teams in several countries. In the E-SARI-Net, countries were invited to collect questionnaire-based or register-based data. A questionnaire-based system obtains information directly from patients with SARI or from their medical notes, whereas a register-based system collects data directly entered into registers, which can then be extracted for monitoring and analysis. Questionnaire-based systems can be rolled out and updated easily; however, they require staff and time. In some countries such as Germany, where SARI surveillance has been established successfully for many years, register-based surveillance is the preferred data collection method as it is easily extractable, fast, and requires minimal investment in staff [5].

Background Information in Malta

Malta is a central Mediterranean island country and the smallest country in the European Union. It had a population of 516,100 at the end of 2020 [6], giving it one of the world's highest population densities at over 1376 people/km^2 [7]. All SARI cases in Malta are admitted to a central hospital: Mater Dei Hospital (MDH). Over the course of the last 2 and a half years, the following milestones occurred. (1) Before the pandemic, patient records were paper-based and stored in folders, which made data extraction nearly impossible, with access restricted to those with physical access to those files. (2) Since the start of the COVID-19 pandemic in Malta (March 2020), MDH has established and improved its capacity to carry out surveillance and data accessibility, with hospital staff entering medical admission notes electronically as free text in an accessible database. (3) Since October 19, 2020, the hospital started identifying SARI hospitalized cases based on a provisional diagnosis provided by clinicians. Through this process, an initial incidence-based system was put in place (Figure 1, Panel A).

In November 2020, Malta accepted to be part of the E-SARI-Net. Throughout this paper, we narrate the process of setting up a comprehensive SARI surveillance system in Malta and describe its output, along with challenges that we encountered and future steps in improving the system. We will also discuss the advantages and disadvantages of different SARI case definitions.

Figure 1. Flowchart describing SARI surveillance procedures in Malta. Panel A shows the system in place in October 2020. Panel B shows the system in place from February 1, 2021, which included data required by the ECDC protocol. Panel C shows the system currently in place, where manual extraction is required only for symptoms and comorbidities. ECDC: European Centre for Disease Prevention and Control; SARI: severe acute respiratory infections.



Methods

Study Details

This study describes the process of digitizing and upgrading SARI surveillance in Malta, an island country with a centralized health system, during the COVID-19 pandemic, focusing on SARI surveillance from February 1 to November 30, 2021. As aforementioned, Malta has gone from a paper-based system to one that is increasingly digitized, allowing the establishment of syndromic public health surveillance. This process is described in Figure 1. The results section includes descriptive data and a comparison of COVID-19 detection outcomes between different SARI definitions.

Definitions of Patients With SARI and COVID-19

The definition of a patient with SARI differs from one organization to another but holds key standard identifiers, with a patient with SARI being:

1. a patient with an acute respiratory condition, who has been
2. hospitalized for more than 24 hours, with
3. the onset of symptoms ≥10 days prior to admission, and presenting with
4. specific symptoms (see below).

The WHO SARI definition includes the following symptoms [8]: history of fever or measured fever of ≥38 °C and cough. The ECDC, although not having a specific definition for SARI, has a specific definition for patients with COVID-19—any person with at least one of the following symptoms: cough; fever; shortness of breath; and sudden onset of anosmia, ageusia, or dysgeusia [9].

In Malta, we have included these common criteria and the following symptoms to define a patient with SARI:

- Fever (>38.0 °C) OR feverishness (37.0 °C to 37.9 °C or reported history of fever) AND
- Acute onset of cough OR shortness of breath

This is a hybrid of WHO's [8] SARI definition and the ECDC definition of a patient who is COVID-19 positive [9]. The definition of a patient with COVID-19 follows the protocol established by the ECDC [9] and international guidelines: a positive case of COVID-19 is a patient who tested positive for the SARS-CoV-2 virus by polymerase chain reaction either during the first 2 days of hospitalization or up to 14 days prior to hospitalization.

To compare the WHO definition and our hybrid definition, we evaluated the percentage of patients with SARI who are COVID-19 positive that met our criteria by month and age category, among those eligible for triage (see below).

Establishing an SARI Surveillance System

Creating the Team

After signing the agreement with Epiconcept in late November 2020, data started being collected (see Multimedia Appendix 1 for the variables collected), and we established an SARI surveillance team. This team was comprised of a medical doctor, an EPIET fellow, and a local public health consultant; this team expanded over time. This team examined the ECDC protocol and commenced a mapping exercise of the key stakeholders and data sources.

Mapping Key Stakeholders

Until recently, Malta had limited efforts to consolidate data into a centralized system easily accessible for public health analysis. Consequently, a mapping exercise for critical stakeholders was undertaken: for example, contacting the vaccination team stakeholder, organizing a meeting with the team, and seeking permission to obtain their data. This exercise, although time-consuming, was beneficial in the long run as it brought together different expertise working together for a common goal—the establishment of the SARI surveillance system.

Manual Extraction of Data

On February 1, 2021, we started extracting data from the system that had been established in October 2020 (Figure 1, Panel A), this time including the data required by the ECDC protocol (Figure 1, Panel B). Here, clinical diagnoses provided by doctors specifically about either "SARI" or "Respiratory conditions" were flagged and then triaged according to the SARI definition above. Patients assigned a "SARI" provisional diagnosis were included automatically and classified as "confirmed" if they met the case definition or "probable" if they did not. Patients assigned a "Respiratory conditions" provisional diagnosis were only included if they met the case definition and were therefore all "confirmed." The distinction between "confirmed" and "possible" cases will be discussed further below. Past triage notes written by the doctor were read and their data were entered into a Microsoft Excel file to store the data. This required the manual extraction of data, which was time-consuming, as data had to be sourced from free-text entries in multiple different data sets. This system was still inefficient for surveillance, requiring the reading of patient records. However, it was an essential first step that permitted establishing a surveillance system.

A patient categorized as "hospitalized" is defined as a patient who has been admitted to hospital for more than 24 hours, as per WHO SARI protocol. Additionally, patient outcomes were also categorized in order of severity: hospitalized, admitted to intensive care unit (ICU), and death. Patients in the ICU and those who died were considered within the hospitalized cohort, and their counts in sections below are nested within the hospitalized cohort. Information on ICU admission was obtained from the daily hospital bed census, and death outcomes were obtained from the mortality register.

A Step Toward Automation: a Hybrid System

Currently, symptoms and comorbidities are collected from the IT system of the hospital's accident and emergency admissions department as described above. However, the rest of the information are collected through a merging of data sets (Figure 1, Panel C). Collected data are cleaned and matched using patient identification numbers to vaccination data (COVID-19 and influenza); laboratory data (COVID-19, influenza, and a respiratory panel of 32 pathogens); and the hospital census (to identify the patient's journey through hospital). The process of data cleaning and merging is carried out using R statistical software (version 4.1; R Foundation for Statistical Computing) [10], with scripts frequently updated to deal with the dynamic developments in the surveillance system.

Ethical Consideration

No ethical protocols are required in Malta for surveillance systems as long as data are anonymized and aggregated.

Results

Patient Demographics

From February 1 to November 30, 2021, 903 patients with SARI were detected. Figure 2 shows the epidemiological curve of weekly SARI admissions. The COVID-19 status of patients is shown in different colors; the proportion of patients who are COVID-19 positive matched the COVID-19 "wave" that occurred from March to April 2021 and the following, smaller COVID-19 "wave" in July 2021.

Table 1 describes the demographics and characteristics of people included in the surveillance system, including symptoms, underlying conditions, clinical conditions, and outcomes, grouped according to worsening severity of symptoms (from least to most severe: hospitalization, admission to ICU, death). It is important to note that those admitted to the ICU and those who died are nested within the hospitalized cohort and are not separate from it. Among the 903 cases, 380 (42.1%) were COVID-19 positive, and 107 (11.8%) died in hospital. In all, 615 (68.1%) patients were aged >60 years, and 531 (58.8%) were male. Among the 107 patients who died, 96 (89.7%) had more than one underlying condition, the most common of which were hypertension (n=57, 53.3%) and chronic heart disease (n=49, 45.8%).

Figure 3 shows the combination of SARI symptoms for confirmed SARI cases (n=735). The most common combination of symptoms is cough, shortness of breath, and fever, with 5.4% (49/903) of patients displaying all 3 symptoms concurrently. Other common co-occurring symptoms are tachypnea and chest pain.

Figure 2. Weekly SARI case counts according to COVID-19 status. SARI: severe acute respiratory infections.

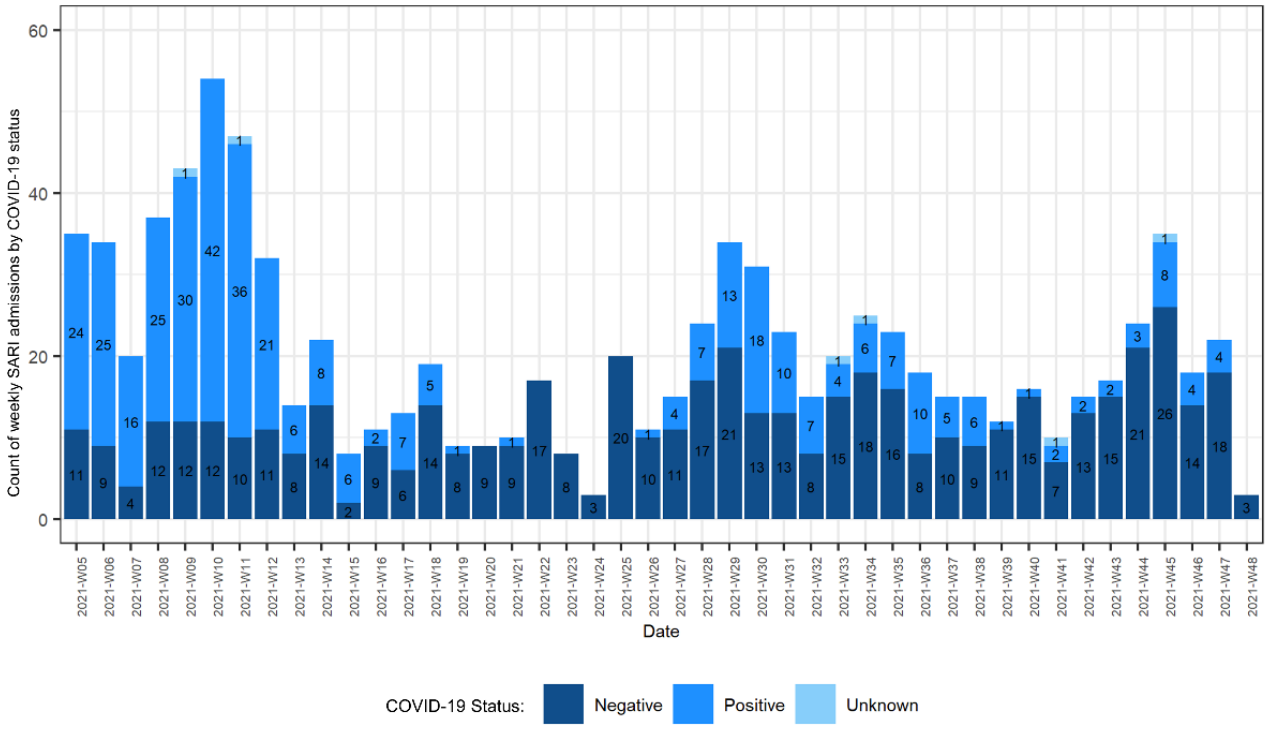


Table 1. Demographics and characteristics of patients with SARI[a] by severity categories, in Malta from February 1, 2021, to November 30, 2021.

Characteristic	Hospitalized (N=903)	ICU[b] admission (n=69)	Death (n=107)
Total in hospitalized category (N=903), n (%)[c]	903 (100)	69 (7.6)	107 (11.8)
Status, confirmed SARI cases[d], n (%)	735 (81.4)	52 (75.4)	80 (74.8)
Sex, n (%)			
Male	531 (58.8)	53 (76.8)	64 (59.8)
Female	369 (40.9)	16 (23.2)	43 (40.2)
Age (years), median (IQR)	69.5 (55-80)	63.0 (54-71)	78.0 (70-85)
Age group (years), n (%)			
<21	5 (0.6)	0 (0)	0 (0)
21-40	101 (11.2)	6 (8.7)	1 (0.9)
41-60	179 (19.8)	24 (34.8)	9 (8.4)
61-80	403 (44.6)	38 (55.1)	51 (47.7)
81+	212 (23.5)	1 (1.4)	46 (43)
Symptoms, n (%)			
Altered taste	14 (1.6)	0 (0)	0 (0)
Anosmia	28 (3.1)	3 (4.3)	2 (1.9)
Cough	651 (72.1)	56 (81.2)	65 (60.7)
Fever	488 (54)	44 (63.8)	49 (45.8)
Feverishness	470 (52)	23 (33.3)	58 (54.2)
Loss of taste	26 (2.9)	1 (1.4)	1 (0.9)
Malaise	26 (2.9)	3 (4.3)	1 (0.9)
Nausea	100 (11.1)	6 (8.7)	3 (2.8)
Shortness of breath	644 (71.3)	55 (79.7)	78 (72.9)
Vomiting	113 (12.5)	7 (10.1)	10 (9.3)
Underlying conditions, n (%)			
More than one underlying condition	854 (94.6)	65 (94.2)	96 (89.7)
Asthma	85 (9.4)	5 (7.2)	5 (4.7)
Chronic heart disease	269 (29.8)	13 (18.8)	49 (45.8)
Chronic lung disease	87 (9.6)	2 (2.9)	18 (16.8)
Chronic kidney disease	61 (6.8)	7 (10.1)	8 (7.5)
Cancer	104 (11.5)	4 (5.8)	23 (21.5)
Diabetes	230 (25.5)	25 (36.2)	28 (26.2)
Dyslipidemia	111 (12.3)	9 (13)	15 (14)
Hypertension	394 (43.6)	32 (46.4)	57 (53.3)
Obese	43 (4.8)	8 (11.6)	6 (5.6)
Clinical conditions, n (%)			
Acute respiratory distress syndrome	36 (4)	13 (18.8)	6 (5.6)
Pneumonia	662 (73.6)	63 (91.3)	58 (54.2)
Outcome, n (%)			
Discharged	769 (85.2)	46 (66.7)	N/A
Died	107 (11.8)	20 (30)	107 (100)
Still being treated	27 (3)	3 (4.3)	N/A

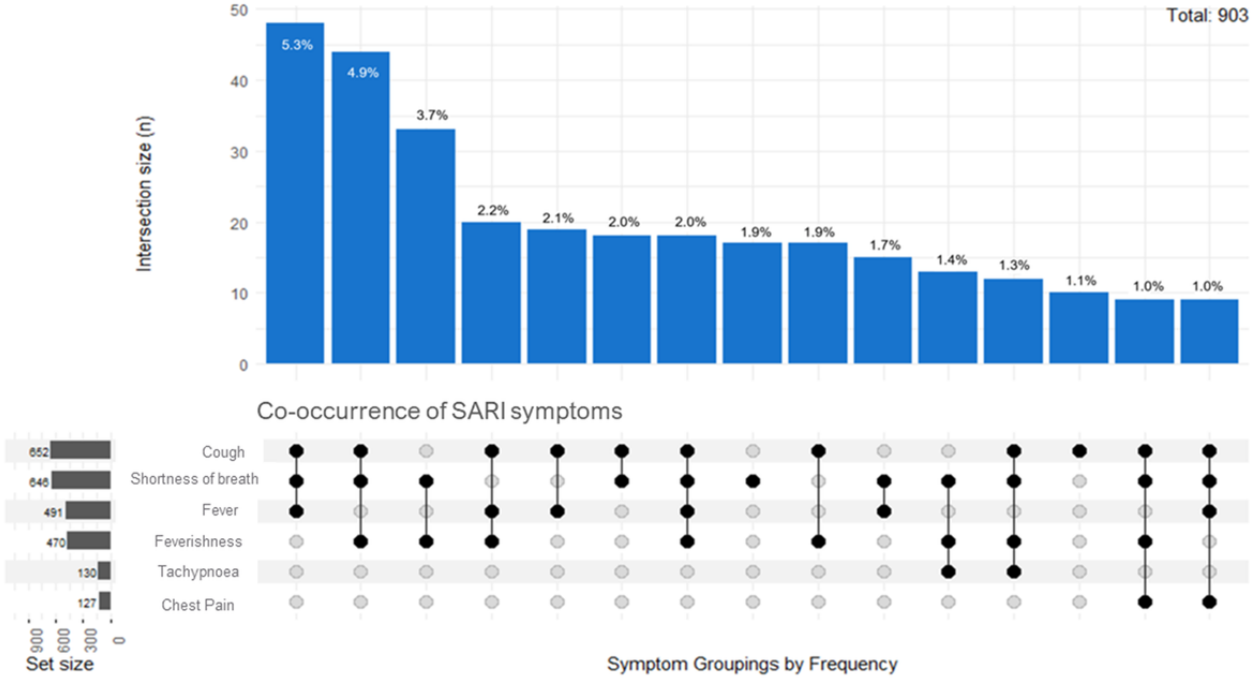
[a]SARI: severe acute respiratory infections.

[b]ICU: intensive care unit.

[c]This row shows row percentages, whereas other rows show column percentages.

[d]"Confirmed" SARI cases are patients who matched the SARI triage definition; these are in contrast to "possible" SARI cases, which are patients diagnosed as having SARI by the doctor who had missing symptomatic data provided.

Figure 3. Upset plot showing the most common symptom groupings for patients with severe acute respiratory infections (SARI).



Patients Who Are COVID-19 Positive

Although SARI surveillance is not exclusive to COVID-19, it was a substantial cause for SARI admissions during the period described here. Table 2 shows the characteristics of patients with SARI testing positive for SARS-CoV-2 (n=380). COVID-19 accounted for 42.2% (380/900) of all SARI hospitalizations, 86.9% (60/69) of all SARI ICU admissions, and 41.1% (44/107) of SARI deaths.

Table 2. Demographics and characteristics of patients with SARI[a] who are COVID-19 positive by severity categories, in Malta from February 1, 2021, to November 30, 2021, excluding those with unknown COVID-19 status (n=3). Total number of patients included in this table is 900.

Characteristic	Hospitalized (n=380)	ICU[b] admission (n=60)	Death (n=44)
COVID-19 positive, n/N (%)[c]	380/900 (42.2)	60/69 (86.96)	44/107 (41.1)
Sex, n (%)			
Male	244 (64.2)	45 (0.75)	30 (68.2)
Female	136 (35.8)	15 (0.25)	14 (31.8)
Age (years), median (IQR)	69.5 (55-80)	63.0 (54-71)	78 (70-85)
Age group (years), n (%)			
<21	4 (1.1)	0 (0)	0 (0)
21-40	57 (15)	4 (6.7)	0 (0)
41-60	104 (27.4)	20 (33.3)	3 (5)
61-80	179 (47.1)	35 (58.3)	26 (59.1)
81+	36 (9.5)	1 (1.7)	15 (34.1)
Symptoms, n (%)			
Altered taste	13 (3.4)	0 (0)	0 (0)
Anosmia	25 (6.6)	3 (5)	2 (4.5)
Cough	285 (75)	50 (83.3)	30 (68.2)
Fever	192 (50.5)	39 (65)	18 (40.9)
Feverishness	167 (43.9)	21 (35)	21 (47.7)
Loss of taste	24 (6.3)	1 (1.7)	1 (2.3)
Malaise	15 (3.9)	3 (5)	0 (0)
Nausea	60 (15.8)	6 (10)	1 (2.3)
Shortness of breath	263 (69.2)	47 (78.3)	30 (68.2)
Vomiting	49 (12.9)	5 (8.3)	1 (2.3)
Underlying conditions, n (%)			
Asthma	32 (8.4)	4 (6.7)	4 (9.1)
Cancer	23 (6.1)	4 (6.7)	7 (15.9)
Chronic heart disease	68 (17.9)	11 (18.3)	22 (50)
Chronic lung disease	12 (3.2)	1 (1.7)	4 (9.1)
Chronic kidney disease	21 (5.5)	6 (10)	5 (11.4)
Diabetes	98 (25.8)	22 (36.7)	14 (31.8)
Dyslipidemia	48 (12.6)	7 (11.7)	5 (11.4)
Hypertension	140 (36.8)	27 (45)	30 (68.2)
Obese	19 (5)	8 (13.3)	4 (9.1)
Clinical conditions, n (%)			
Acute respiratory distress syndrome	18 (4.7)	11 (18.3)	3 (6.8)
Pneumonia	348 (91.6)	55 (91.7)	31 (70.5)

[a]SARI: severe acute respiratory infections.

[b]ICU: intensive care unit.

[c]This row shows row percentages, whereas other rows show column percentages.

Comparing Malta's SARI Definition With the WHO SARI Definition

Figures 4 and 5 compare the current definition we use for SARI, which includes both "confirmed" and "possible" cases adopted by the Maltese team. In contrast, the "confirmed" SARI panel compares "confirmed" SARI cases to the WHO definition by month and age group. Our SARI case definition, including both "confirmed" and "possible" cases, identified substantially more cases, with twice as many patients with COVID-19 detected over the winter months compared to the WHO SARI case definition. However, the "confirmed" cases by themselves, although still an improvement from the WHO SARI case definition, improved at most only 5-15 percentage points (Figure 4). Figure 5 consistently shows a similar pattern for the definition including both "confirmed" and "possible" cases compared to the WHO definition. However, with "confirmed" cases alone, the improvement was only of 5-10 percentage points higher than the WHO definition. (Figure 5).

Figure 4. Comparing SARI definitions: patients with COVID-19 detected by month from February 1 to November 30, 2021. Percent of patients who are COVID-19 positive detected by triage using (A) Maltese SARI "confirmed" and "possible" definitions, (B) Maltese SARI "confirmed" definition, and (C) the World Health Organization's definition. SARI: severe acute respiratory infections.

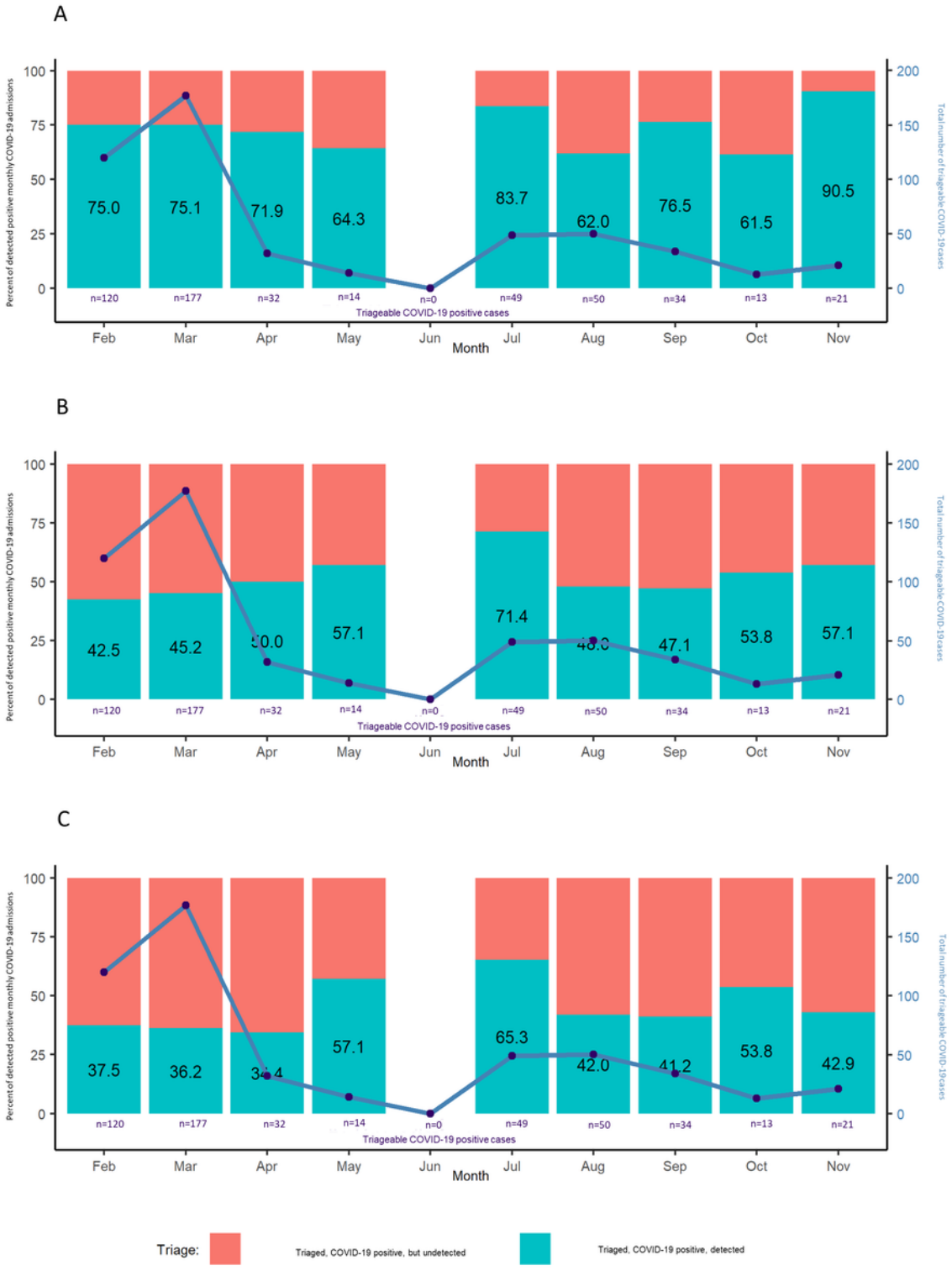
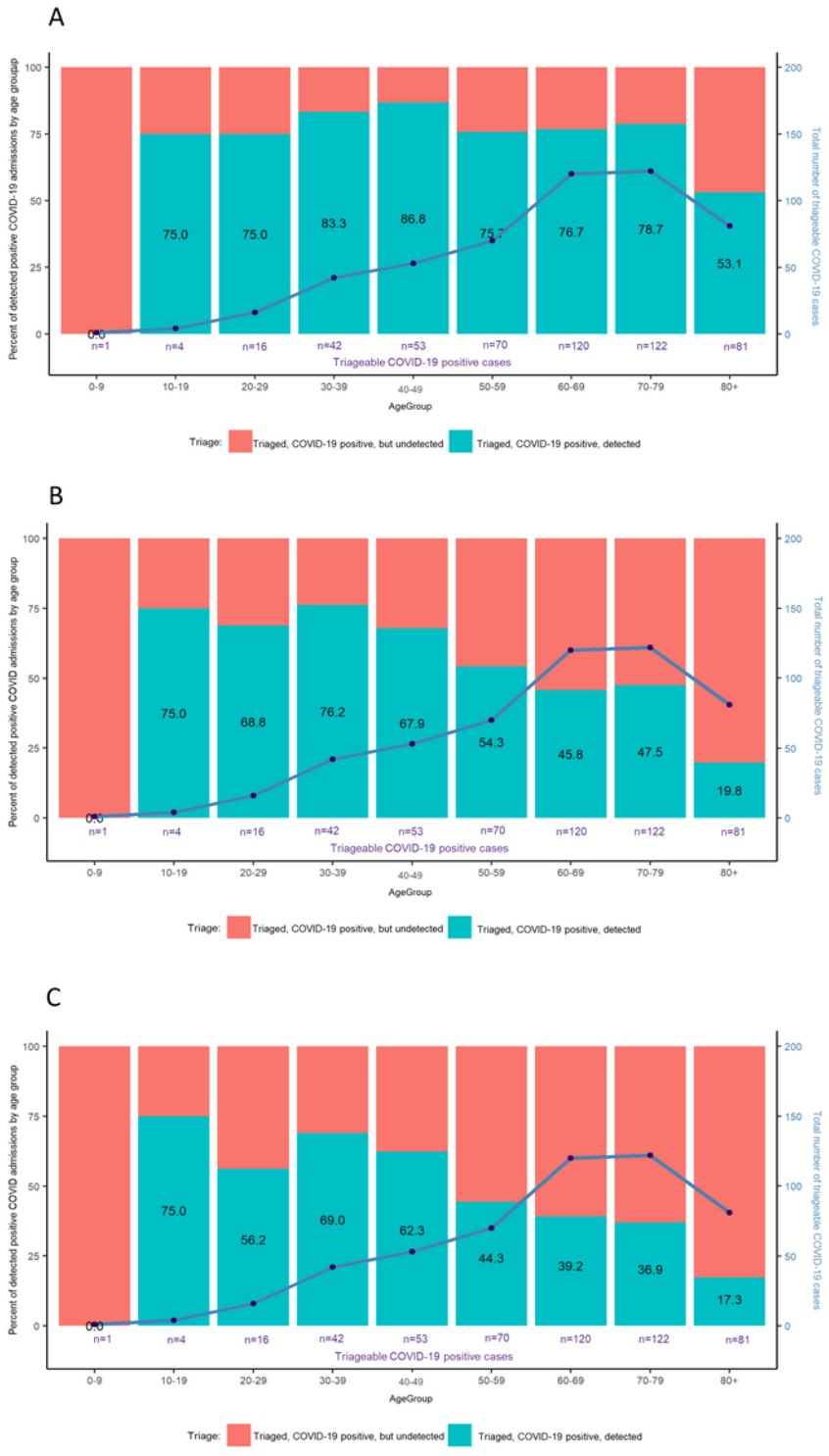


Figure 5. Comparing SARI definitions: patients with COVID-19 detected by age group from February 1 to November 30, 2021. Percent of patients who are COVID-19 positive detected by triage using (A) Maltese SARI "confirmed" and "possible" definitions, (B) Maltese SARI "confirmed" definition, and (C) the World Health Organization's definition. SARI: severe acute respiratory infections.



Surveillance Outputs

The SARI surveillance team sends case-based and aggregate data locally on a weekly and monthly basis. We also submit accumulated SARI surveillance data to the ECDC. Currently, weekly and monthly data are sent to the ECDC and pooled for specific studies for analysis, including the I-MOVE-COVID-19 study for vaccine effectiveness [11-13].

Data collected through SARI surveillance included extensive COVID-19 data, allowing Malta to conduct its vaccine effectiveness study. Besides the periodic data, our team has also generated:

- 7 monthly SARI reports
- 8 monthly vaccine effectiveness reports
- 10 weekly SARI reports

As a team, we regularly meet stakeholders who we invite to monthly SARI meetings, where ideas for collaboration, improvements to the system, and research outputs are discussed. Additionally, these reports and outputs serve to inform public health management and policy in a dynamic manner [14], informing the government on current respiratory illness outbreaks and local vaccine effectiveness. Currently, vaccine effectiveness against both infection and hospitalization are being analyzed to inform the Ministry of Health on the success of COVID-19 vaccination campaigns. This information will also advise government policy on COVID-19 booster doses in the coming months and years. In addition, E-SARI-Net countries meet monthly to discuss challenges they are facing in the process and adjust the international protocols; these meetings also offer spaces where ideas and experiences are shared for countries to learn from one another's challenges and adopt a different approach in their work.

Discussion

Building Bridges

Setting up a digital, standardized SARI surveillance system was challenging, considering that Malta has gone from paper-based data entry to digitizing its emergency department medical notes within 3 years. However, the biggest challenge to this setup was primarily mapping and identifying key stakeholders and bringing people together amid a busy schedule due to the COVID-19 pandemic under way.

The SARI surveillance team in Malta delivers weekly and monthly reports on SARI to interested stakeholders within Malta's health system, with monthly meetings to discuss emerging concerns in respiratory illness. These meetings have been key to the program's success: a space where ideas for data and surveillance processes are discussed and stakeholders are listened to, with the aim of constant improvements to the system. Additionally, reports and meetings also inform decision makers on the policies they can adapt based on the data, which is the key aim of SARI surveillance.

Principal Findings

Demographics

First, there is a gendered difference among patients with SARI, with more male patients (531/903, 58.8%) entering the hospital than female patients (369/903, 40.9%). This difference is more pronounced for patients who are COVID-19 positive, where men (244/380, 64.2%) are hospitalized in larger numbers than women (136/380, 35.8%). Although gender has an association with COVID-19 incidence, the severity of COVID-19 presentation differs according to gender, but the reasons for this difference have yet to be fully explained [15,16]. Second, median age differences for severity categories can be explained by policies at MDH. Those in the older age groups are less likely to be admitted for intensive care, given limited bed space and lower chances of positive outcomes from intensive care. Consequently, the median age of those admitted to intensive care is lower than all hospitalized for SARI.

Adoption of the Maltese SARI "Hybrid" Definition

The commencement of SARI surveillance coincided with the worst wave (at the time) of the COVID-19 pandemic in Malta. At first, Malta used the WHO SARI definition in its basis for the inclusion of patients [8]. Following numerous discussions among countries participating in the E-SARI-Net, we expanded its inclusion criteria to include feverishness and shortness of breath. Although an improvement over the WHO criteria, symptoms-based triage inclusion is only marginally better than the WHO SARI definition as compared to the comprehensive system our team has adopted, which heavily relies on clinical diagnosis. This can point to either (1) clinician bias, where COVID-19 cases are more likely to be registered as SARI; (2) poor symptom reporting; or (3) both. Clinician bias is highly likely, especially during the winter months (from February to March and in November 2021) when Malta experienced a COVID-19 wave. This form of bias, called "availability bias," is one that occurs where a particular diagnosis is given due to the likelihood of it being assigned, given the circumstances or presence in mind [17]. Poor symptom reporting is also increasingly likely during winter months when, under time pressure, clinicians might take a shortened account of symptom history. In the case of older age groups, the observed variability can be attributed to the broad spectrum of COVID-19 symptom presentations or poor patient communication [18]. These findings further motivate the shift to a register-based surveillance system, where symptom reporting will be mandatory, which will likely improve COVID-19 case detection.

Comparison With Prior SARI Surveillance

Compared to prior SARI surveillance carried out in Malta, the current system is increasingly comprehensive, including data pertaining to laboratory data, vaccination status, and severity of outcome, among other parameters. This improves the capacity for dynamic decision-making by public health policy makers while allowing Malta to carry out other analysis, such as vaccine effectiveness studies, which were not possible before. Additionally, the current form of surveillance allows Malta to share its data on a European level, for pooled analysis.

Although syndromic parameters are still extracted manually, Malta is aiming to digitize the system in the first half of 2022. The shift of the IT system at MDH to a register-based system is being piloted now and will allow automated extraction of symptom and comorbidity data. This last step will allow us to fully automate our last remaining manual extraction point (symptoms and comorbidities). Through precomposed and easily upgradeable R scripts, this would make SARI surveillance increasingly sustainable [19]. Consequently, we expect the volume of patients with SARI registered in the database to increase substantially as more accurate data are provided, given that currently symptoms are only mentioned at the clinician's discretion.

The SARI surveillance team has recently incorporated a new stakeholder: the Influenza vaccination team. These data will be included in pooled international influenza vaccine effectiveness studies, such as I-MOVE [20,21]. The collection of treatment data (and resulting outcomes) would also be a positive addition to the surveillance program. This could greatly expand Malta's

research potential, permitting cohort studies or case-control studies to examine the benefit and effectiveness of treatment provided for the patient.

Strengths and Limitations

Our current system has limitations. First, we only triage patients assigned a provisional diagnosis by clinicians. This will be addressed by the shift to register-based surveillance with mandatory symptomatic entry of SARI-defining symptoms. Second, patients provisionally diagnosed as "SARI" were automatically entered into the database. Among these patients, those who did not meet the triage definition were classified as "possible" SARI cases, whereas all patients meeting SARI definitions were classified as "confirmed." This was an important factor in recruitment—due to the novelty of the system, urgency of data input on busy days at the emergency department, and the "free-text" nature of entering data, specific symptoms useful for accurate SARI triage might have been missing (eg, fever or cough). These limitations are Malta's primary motivator to move toward register-based surveillance [22], where the entry of SARI diagnostic criteria would be mandatory at the patient-admission level and extraction is symptoms-based.

However, our system has many strengths. Over the past few months, it has fostered a spirit of cooperation and collaboration among various departments or teams that might have been too busy to collaborate otherwise. This essential outcome is often overlooked but is very much at the core of a surveillance system. Without stakeholders' trust, neither data provision nor its output would be possible. This is a very positive outcome, and its success might spur further collaboration among different stakeholders that could substantially improve Malta's public health capacity. We also managed to run a system with few resources, which will ensure long-term sustainability.

Conclusion

Since February 1, 2021, Malta has successfully established an SARI surveillance system. Although there is much room for improvement, its outcomes substantially improve Malta's ability to monitor patients with SARI, informing decision-making processes. This is even more important in the context of the COVID-19 pandemic, which was an essential driver in establishing SARI surveillance. Future changes, including the shift to register-based surveillance, should substantially increase both the catchment of SARI as well as our capacity to detect and respond to new disease threats. The establishment of the SARI surveillance system also encourages the development of dynamic policy-making and public health interventions, which can be more effective both on a population level and can make a more efficient use of public health resources. The findings of this study also encourage the examination of the standard WHO definitions of SARI for surveillance teams, who may choose to adapt or expand the definition of SARI based on the context of public health data available to them for better catchment.

Acknowledgments

We would like to acknowledge the input provided by Aftab Jasir (JB's EPIET frontline coordinator) on the draft manuscript. We would also like to acknowledge the support provided by Mater Dei Hospital staff and public health personnel.

The data of the study were originally collected as part of the project "Establishing Severe Acute Respiratory Infections (SARI) surveillance and performing hospital-based COVID-19 transmission studies," funded by the European Centre for Disease Prevention and Control through a service contract with Epiconcept (ECD.11236 and Amendment Nº 1 ECD.11810).

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of data collected for Malta's SARI surveillance system, including possible values and sources.
[DOCX File , 27 KB - publichealth_v8i12e37669_app1.docx]

References

1. Declich S, Carter AO. Public health surveillance: historical origins, methods and evaluation. Bull World Health Organ 1994;72(2):285-304 [FREE Full text] [Medline: 8205649]
2. Teutsch SM, Thacker SB. Planning a public health surveillance system. Epidemiol Bull 1995 Mar;16(1):1-6. [Medline: 7794696]
3. Sentinel surveillance. European Centre for Disease Prevention and Control. URL: https://www.ecdc.europa.eu/en/seasonal-influenza/surveillance-and-disease-data/facts-sentinel-surveillance [accessed 2021-12-30]
4. Epiconcept - Smart health: epidemiology - e-Health - data science. Epiconcept. URL: https://www.epiconcept.fr/en/ [accessed 2021-12-30]
5. Buda S, Tolksdorf K, Schuler E, Kuhlen R, Haas W. Establishing an ICD-10 code based SARI-surveillance in Germany - description of the system and first results from five recent influenza seasons. BMC Public Health 2017 Jun 30;17(1):612 [FREE Full text] [doi: 10.1186/s12889-017-4515-1] [Medline: 28666433]
6. World Population Day: 11 July 2020. National Statistics Office Malta. 2020 Jul 10. URL: https://nso.gov.mt/en/News_Releases/Documents/2020/07/News2020_114.pdf [accessed 2022-11-10]

7. Countries with the highest population density 2019. Statista. 2019. URL: https://www.statista.com/statistics/264683/top-fifty-countries-with-the-highest-population-density/ [accessed 2022-05-11]
8. Surveillance case definitions for ILI and SARI. World Health Organization. 2021. URL: https://www.who.int/teams/global-influenza-programme/surveillance-and-monitoring/case-definitions-for-ili-and-sari [accessed 2021-12-29]
9. Core protocol for ECDC studies of COVID-19 vaccine effectiveness against hospitalisation with Severe Acute Respiratory Infection laboratory-confirmed with SARS-CoV-2, version 1.0. European Centre for Disease Prevention and Control. 2021 Oct. URL: https://www.ecdc.europa.eu/sites/default/files/documents/Core-protocol-for-ECDC-studies-of-COVID-19-vaccine-effectiveness-against-hospitalisation-with-SARI.pdf [accessed 2022-11-10]
10. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2020. URL: https://www.R-project.org/ [accessed 2022-11-10]
11. Kissling E, Hooiveld M, Sandonis Martín V, Martínez-Baz I, William N, Vilcu A, I-MOVE-COVID-19 primary care study team, I-MOVE-COVID-19 primary care study team (in addition to authors above). Vaccine effectiveness against symptomatic SARS-CoV-2 infection in adults aged 65 years and older in primary care: I-MOVE-COVID-19 project, Europe, December 2020 to May 2021. Euro Surveill 2021 Jul;26(29):2100670 [FREE Full text] [doi: 10.2807/1560-7917.ES.2021.26.29.2100670] [Medline: 34296676]
12. Multidisciplinary European network for research, prevention and control of the COVID-19 Pandemic. European Commission. URL: https://cordis.europa.eu/project/id/101003673 [accessed 2021-12-29]
13. Rose AMC, Kissling E, Valenciano M. European study of COVID-19 vaccine effectiveness against hospitalised SARI patients laboratory-confirmed with SARS-CoV-2: I-MOVE-COVID-19 generic protocol. Zenodo 2021 Feb 22 [FREE Full text] [doi: 10.5281/zenodo.4555315]
14. Post LA, Issa TZ, Boctor MJ, Moss CB, Murphy RL, Ison MG, et al. Dynamic public health surveillance to track and mitigate the US COVID-19 epidemic: longitudinal trend analysis study. J Med Internet Res 2020 Dec 03;22(12):e24286 [FREE Full text] [doi: 10.2196/24286] [Medline: 33216726]
15. Jin J, Bai P, He W, Wu F, Liu X, Han D, et al. Gender differences in patients with COVID-19: focus on severity and mortality. Front Public Health 2020 Apr 29;8:152 [FREE Full text] [doi: 10.3389/fpubh.2020.00152] [Medline: 32411652]
16. Mukherjee S, Pahan K. Is COVID-19 gender-sensitive? J Neuroimmune Pharmacol 2021 Mar 06;16(1):38-47 [FREE Full text] [doi: 10.1007/s11481-020-09974-z] [Medline: 33405098]
17. Coleman JJ, Manavi K, Marson EJ, Botkai AH, Sapey E. COVID-19: to be or not to be; that is the diagnostic question. Postgrad Med J 2020 Jul;96(1137):392-398 [FREE Full text] [doi: 10.1136/postgradmedj-2020-137979] [Medline: 32522844]
18. ISARIC Clinical Characterisation Group. COVID-19 symptoms at hospital admission vary with age and sex: results from the ISARIC prospective multinational observational study. Infection 2021 Oct;49(5):889-905 [FREE Full text] [doi: 10.1007/s15010-021-01599-5] [Medline: 34170486]
19. Marbus SD, van der Hoek W, van Dissel JT, van Gageldonk-Lafeber AB. Experience of establishing severe acute respiratory surveillance in the Netherlands: evaluation and challenges. Public Health Pract (Oxf) 2020 Nov;1:100014 [FREE Full text] [doi: 10.1016/j.puhip.2020.100014] [Medline: 34171043]
20. Kissling E, Valenciano M, Buchholz U, Larrauri A, Cohen JM, Nunes B, et al. Influenza vaccine effectiveness estimates in Europe in a season with three influenza type/subtypes circulating: the I-MOVE multicentre case-control study, influenza season 2012/13. Euro Surveill 2014 Feb 13;19(6):20701 [FREE Full text] [doi: 10.2807/1560-7917.es2014.19.6.20701] [Medline: 24556348]
21. Valenciano M, Ciancio B, I-MOVE study team. I-MOVE: a European network to measure the effectiveness of influenza vaccines. Euro Surveill 2012 Sep 27;17(39):20281 [FREE Full text] [doi: 10.2807/ese.17.39.20281-en] [Medline: 23041022]
22. Gilbert R, Cliffe SJ. Public health surveillance. In: Regmi K, Gee I, editors. Public Health Intelligence. Cham, Switzerland: Springer; Apr 27, 2016:91-110.

Abbreviations

E-SARI-Net: European SARI Network
ECDC: European Centre for Disease Prevention and Control
ICU: intensive care unit
MDH: Mater Dei Hospital
SARI: severe acute respiratory infections
WHO: World Health Organization

Edited by T Sanchez, A Mavragani; submitted 03.03.22; peer-reviewed by J Xu, DW Waqar Ali; comments to author 10.05.22; revised version received 17.05.22; accepted 29.09.22; published 05.12.22.

Please cite as:
Cauchi JP, Borg ML, Džiugytė A, Attard J, Melillo T, Zahra G, Barbara C, Spiteri M, Drago A, Zammit L, Debono J, Souness J, Agius S, Young S, Dimech A, Chetcuti I, Camenzuli M, Borg I, Calleja N, Tabone L, Gauci C, Vassallo P, Baruch J
Digitalizing and Upgrading Severe Acute Respiratory Infections Surveillance in Malta: System Development
JMIR Public Health Surveill 2022;8(12):e37669
URL: https://publichealth.jmir.org/2022/12/e37669
doi:10.2196/37669
PMID:36227157

Original Paper

The Global, Regional, and National Burden and Trends of NAFLD in 204 Countries and Territories: An Analysis From Global Burden of Disease 2019

Huilong Chen[1*], PhD; Yuan Zhan[2*], MD; Jinxiang Zhang[3], MD; Sheng Cheng[4], MD; Yuhao Zhou[2], MD; Liyuan Chen[5], MD; Zhilin Zeng[1], PhD

[1]Department and Institute of Infectious Diseases, Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Department of Respiratory and Critical Care Medicine, National Clinical Research Center of Respiratory Disease, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[3]Department of Spine Surgery, The Second Hospital affiliated to Guangdong Medical University, Guangdong Medical University, Zhanjiang, China

[4]Department of Respiratory and Critical Care Medicine, University-Town Hospital, Chongqing Medical University, Chongqing, China

[5]Department of Obstetrics and Gynecology, Wuhan No1 Hospital, Wuhan, China

[*]these authors contributed equally

Corresponding Author:
Zhilin Zeng, PhD
Department and Institute of Infectious Diseases
Tongji Hospital of Tongji Medical College
Huazhong University of Science and Technology
1095 Jiefang Road
Wuhan, 430030
China
Phone: 86 13720336545
Email: ulfnssul@sina.com

Abstract

Background: Nonalcoholic fatty liver disease (NAFLD) poses a substantial socioeconomic burden and is becoming the fastest growing driver of chronic liver disease, potentially accompanied by a poor prognosis.

Objective: We aim to elucidate the global and regional epidemiologic changes in NAFLD during the past 30 years and explore the interconnected diseases.

Methods: Data on NAFLD incidence, prevalence, death, and disability-adjusted life-years (DALYs) were extracted from the Global Burden of Disease Study 2019. The age-standardized incident rate (ASIR), age-standardized prevalent rate (ASPR), age-standardized death rate (ASDR), and age-standardized DALYs were calculated to eliminate the confounding effects of age when comparing the epidemiologic changes between different geographical regions. In addition, we also investigated the correlation between the NAFLD burden and the sociodemographic index (SDI). Finally, the associations of the 3 common comorbidities with NAFLD were determined.

Results: Globally, the incidence and prevalence of NAFLD both increased drastically during the past 3 decades (incidence: from 88,180 in 1990 to 172,330 in 2019, prevalence: from 561,370,000 in 1990 to 1,235,700,000 in 2019), mainly affecting young adults who were aged from 15 to 49 years. The ASIR increased slightly from 1.94 per 100,000 population in 1990 to 2.08 per 100,000 population in 2019, while ASPR increased from 12,070 per 100,000 population in 1990 to 15,020 per 100,000 population in 2019. In addition, the number of deaths and DALYs attributable to NAFLD increased significantly as well from 93,760 in 1990 to 168,970 in 2019 and from 2,711,270 in 1990 to 4,417,280 in 2019, respectively. However, the ASDR and age-standardized DALYs presented decreasing trends with values of estimated annual percentage change equaling to –0.67 and –0.82, respectively (ASDR: from 2.39 per 100,000 population in 1990 to 2.09 per 100,000 population in 2019; age-standardized DALYs: from 63.28 per 100,000 population in 1990 to 53.33 per 100,000 population in 2019). Thereinto, the burden of death and DALYs dominated the patients with NAFLD who are older than 50 years. Moreover, SDI appeared to have obvious negative associations with ASPR, ASDR, and age-standardized DALYs among 21 regions and 204 countries, although there is no marked

association with ASIR. Finally, we found that the incidence and prevalence of NAFLD were positively related to those of diabetes mellitus type 2, stroke, and ischemic heart disease.

Conclusions: NAFLD is leading to increasingly serious health challenges worldwide. The morbidity presented a clear shift toward the young populations, while the heavier burden of death and DALYs in NAFLD was observed in the aged populations and in regions with relatively low SDI. Comprehensive acquisition of the epidemiologic pattern for NAFLD and the identification of high-risk comorbidities may help policy makers and clinical physicians develop cost-effective prevention and control strategies, especially in countries with a high NAFLD burden.

(JMIR Public Health Surveill 2022;8(12):e34809) doi:10.2196/34809

KEYWORDS

non-alcoholic fatty liver disease; Global Burden of Disease Study 2019; epidemiologic change; diabetes mellitus type 2; stroke; ischemic heart disease; incidence; prevalence; mortality; disability-adjusted life-years

Introduction

Nonalcoholic fatty liver disease (NAFLD), characterized by the accumulation of fat (hepatic steatosis) in more than 5% of hepatocytes, is currently recognized as an important driver leading to an increasing burden of chronic liver disease (CLD) worldwide and thus far lacks effective pharmacological therapies [1-4]. Among individuals with NAFLD, some will develop nonalcoholic steatohepatitis (NASH) and potentially progress to end-stage liver cirrhosis and carcinoma, with possible requirements for liver transplants and a poor prognosis [5-8]. In this context, many articles have investigated the epidemic pattern and attributable risk factors for NAFLD [2,3,9] to provide beneficial references for the prevention and control of this disease and thereby alleviate the global and regional socioeconomic burden of NAFLD.

In fact, tremendous heterogeneity of the NAFLD burden is observed around the world, and NAFLD has become the most rapidly growing contributor to liver mortality and morbidity [10]. Substantial NAFLD burdens have been reported successively in Asia, the Middle East, North Africa, Canada, and the United Kingdom [2,11-13], all of which indicate an urgent need for the systematic management and control of this disease. In addition to the liver insult itself, concomitant complications and comorbidities in other organs and systems related to NAFLD have likewise been investigated intensely [14]. The development of NAFLD necessitates the retention of intrahepatic triacylglycerol (IHTAG), whereas IHTAG has been reported to be strongly associated with obesity, insulin resistance, and diabetes mellitus type 2 (DM2) [15]. Self-evident associations between NAFLD and these disorders have been demonstrated in the clinic, and NAFLD patients tend to have DM2, dyslipidemia, and hypertension, which increase the susceptibility to cardiovascular complications [14,16,17]. The common factors contributing to the death of patients with NAFLD are often manifested in various complications, such as stroke and cardiovascular emergencies [18]. Accordingly, a greater understanding of the risk factors or possible comorbidities of NAFLD may help to decrease morbidity and mortality and thereby alleviate the disease burden.

To systematically and comprehensively grasp the global and regional socioeconomic burden of NAFLD during the past 30 years and to explore the interconnected diseases, we summarized the incidence, prevalence, mortality, and disability-adjusted life years (DALY) from the Global Burden of Disease 2019 (GBD 2019) study in this study. This updated epidemiologic pattern and potential risk factors for NAFLD are expected to benefit the development of efficient prevention and control policies, especially for those dedicated to the clinical treatment of NAFLD and public health care.

Methods

Data Acquisition

GBD 2019 provides the information about the burdens of 369 diseases and injuries along with 87 risk factors in the globe, different geographic areas, and 204 countries and territories [19]. Data on the NAFLD burden from 1990 to 2019, including its incidence, prevalence, death, DALY, and their corresponding age-standardized rates (ASRs), were acquired from the Global Health Data Exchange GBD results tool [20]. Meanwhile, information about the distributions of sex and age and related comorbidities, including DM2, stroke, and ischemic heart disease (IHD), was also obtained. The rates were standardized according to the GBD world population and were reported per 100,000 person-years. For the current report, we used the GBD results tool to extract the estimates and their 95% certainty intervals (CIs) for the prevalence of cases, deaths, and DALYs as measures of the NAFLD burden from 1990 to 2019 by region and country. To better exhibit the age distribution of the NAFLD burden, the patients were classified into 3 groups, namely, those aged 15 to 49 years, 50 to 69 years, and above 70 years (no data for those under 15 years). The sociodemographic index (SDI) is a composite indicator of social background and economic conditions that influence health outcomes in each location. In short, it is the geometric mean of 0 to 1 indices of total fertility rate for those younger than 25 years old, mean education for those 15 years old and older, and lag - distributed income per capita [19], which has been reported to be correlated with the incidence and mortality of diseases. Based on that, 204 countries and their territories were classified into 5 groups according to the SDI values calculated in each year from 1990 to 2019 (low-SDI, low-middle-SDI, middle-SDI, high-middle-SDI, and high-SDI; Multimedia Appendix 1), to explore the association between NAFLD burden and social development degrees in different regions.

Statistical Analysis

The incidence, prevalence, death, and DALYs and their corresponding ASRs were the main metrics characterizing the NAFLD burden and were compared at the global, regional, and country levels. CIs were calculated from 1000 estimates for each parameter, and 95% CIs were defined by the 25th and 975th values of the ordered 1000 estimates; 95% CI excluding 0 was considered statistically significant. To investigate the dynamic changes in the NAFLD burden, we further calculated the estimated annual percentage change (EAPC) to delineate the temporal trend in different ASRs for the NAFLD burden. Moreover, we constructed a regression model fitting the natural logarithm of the ASR with the calendar year, namely, ln (ASRs) = $\alpha + \beta \times$ calendar year + ε, to estimate the EAPC with its 95% CI based on the formula of $100 \times (\exp[\beta] - 1)$. If the EAPC value and its 95% CI were both above zero, the changed trend of ASR was considered upwards and vice versa. Otherwise, the ASR was considered stable over time [21]. Finally, we examined the association between the ASRs of NAFLD and the corresponding SDI value of each year using Pearson correlation analysis, as well as associations between the incidence or prevalence and DM2/stroke/IHD. All statistical analyses were performed using GraphPad Prism (version 8; GraphPad Software).

Ethics Approval

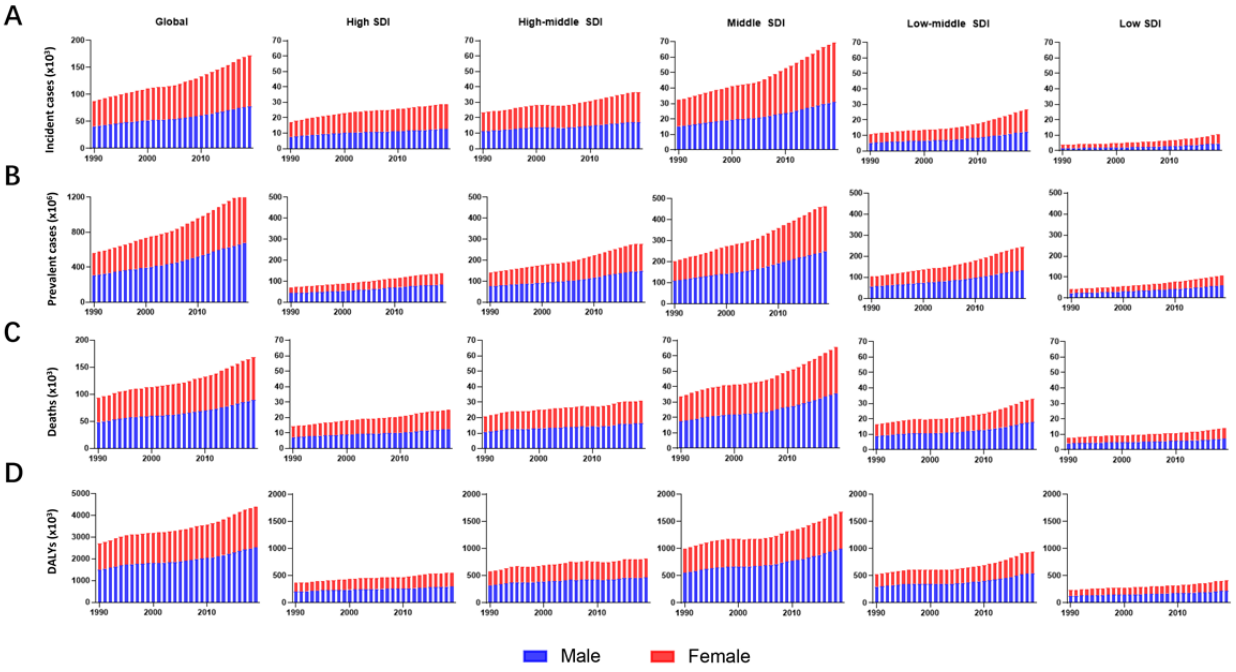
Ethics approval was waived, as all the data in this study was obtained from GBD 2019 study.

Results

Incidence of NAFLD

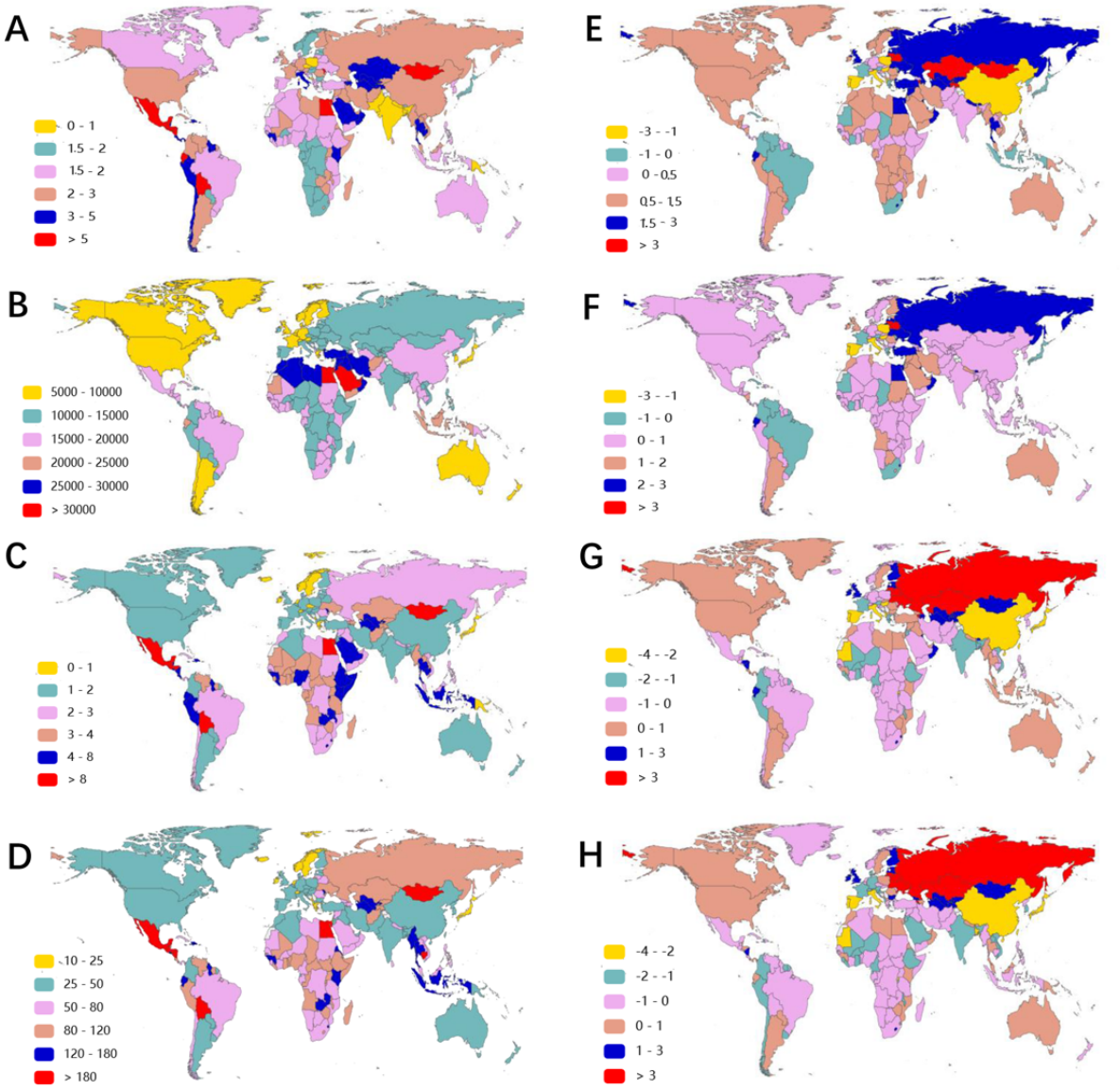
Globally, the incidence of NAFLD increased sharply in the past 30 years from 88,180 (95% CI 62,300-128,320) in 1990 to 172,330 (95% CI 125,780-243,640) in 2019, while there were no obvious changes after standardization by age (EAPC 0.1, 95% CI 0.04 to 0.23; Table S1 in Multimedia Appendix 2). As shown in Figure 1A and Table S1 in Multimedia Appendix 2, there were elevations in the incidence in both sexes, with a slightly higher incidence in women. Meanwhile, various SDI regions presented with gradually increasing NAFLD incidences, which mainly affected the low-middle (from 10,950 in 1990 to 26,640 in 2019) and middle SDI regions (from 32,460 in 1990 to 69,670 in 2019). However, there were no clear changes in age-standardized incidence rate (ASIR) among different SDI regions or by sex (Image A in Multimedia Appendix 3).

Figure 1. The trends of changes in NAFLD incidence, prevalence, deaths, and DALY from 1990 to 2019. The trends of changes in (A) incidences, (B) prevalence, (C) deaths, and (D) DALY (D) are shown. Blue bars represent males, and red bars represent females. DALYs: disability-adjusted life years; NAFLD: nonalcoholic fatty liver disease; SDI: sociodemographic index.



Although the overall incidences displayed upward conditions, ASIR of NAFLD and its changed trend presented immense heterogeneity among different countries and territories (Figure 2A,E; Table S1 in Multimedia Appendix 2). Specifically, the top 3 ASIRs of NAFLD were Central Latin America (6.88 per 100,000 population), Andean Latin America (5.62 per 100,000 population), and Central Asia (4.19 per 100,000 population). Central Asia, Eastern Europe, and the Middle East presented with great increases in ASIR changes during the past 30 years, with relatively higher EAPCs of ASIRs. East Asia had a negative change trend in ASIR (Figure 2E). Furthermore, from the country's perspective, Mongolia had the highest ASIR in 2019 (12.65 per 100,000 population), and Papua New Guinea had the lowest (0.45 per 100,000 population). Finally, we analyzed the associations between SDI and ASIR among 21 regions (r=0.11, P<.01) and 204 countries (r=0.11, P<.001), which presented no obvious correlations (Figure 3A,E).

Figure 2. The age-standardized rates of NAFLD in 2019 and EAPC of NAFLD ASRs from 1990 to 2019 in 204 countries and territories. The (A) ASIR, (B) ASPR, (C) ASDR, and (D) age-standardized DALY of NAFLD around the world in 2019 are shown. The EAPC of (E) ASIR, (F) ASPR, (G) ASDR, and (H) age-standardized DALY in the past 30 years are shown. ASDR: age-standardized death rate; ASIR: age-standardized incident rate; ASPR: age-standardized prevalent rate; ASRs: age-standardized rates; DALY: disability-adjusted life years; EAPC: estimated annual percentage change; NAFLD: nonalcoholic fatty liver disease.

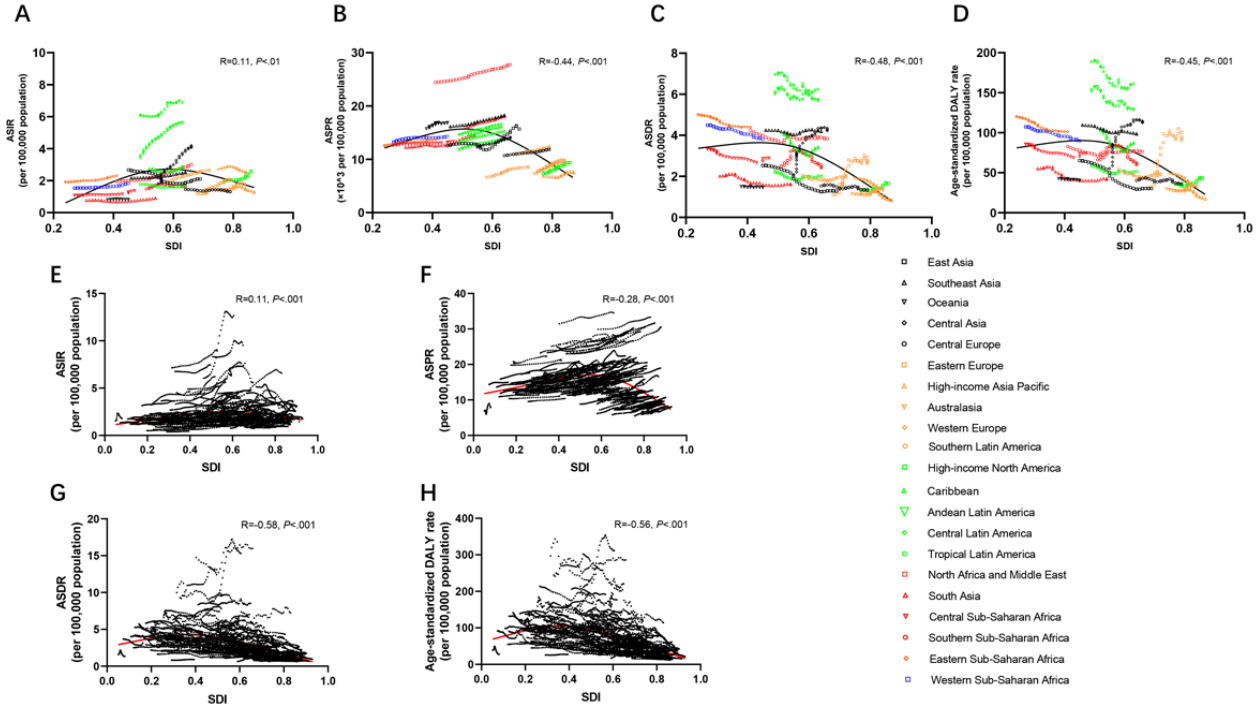


NAFLD Prevalence

In the past 3 decades, the prevalence of NAFLD increased by more than 2-fold at the global level from 561,370,000 (95% CI 498,430,000-633,300,000) in 1990 to 1,235,700,000 (95% CI 1,109,540,000-1,378,530,000) in 2019 (Table S2 in Multimedia Appendix 2). EAPC, indicating the temporal trend of NAFLD age-standardized prevalence rate (ASPR), also presented significant upregulation (0.77, 95% CI 0.69-0.85). There were no marked differences between sexes or among various SDI regions regarding the prevalence and its change trend, all showing obvious upward changes (Figure 1B; Image B in Multimedia Appendix 3, and Table S2 in Multimedia Appendix 2).

In Table S2 in Multimedia Appendix 2, east Asia, south Asia, north Africa, and the Middle East maintained the highest prevalence globally in 1990 and 2019. However, the top 3 prevalences after age standardization were in north Africa and the Middle East (27,750 per 100,000 population), southeast Asia (18,300 per 100,000 population), and southern Sub-Saharan Africa (18,080 per 100,000 population; Figure 2B; Table S2 in Multimedia Appendix 2). Meanwhile, the regions with a higher EAPC of ASPR were around the Mediterranean (Figure 2F). Egypt had the highest ASPR in 2019 among all countries (34.69 per 100,000 population). Moreover, there were slightly negative correlations between ASPR and SDI among 21 regions (r=0.44, P<.001) and 204 countries (r=0.28, P<.001; Figures 3B and 3F), which demonstrated that the more developed the region was, the lower the ASPR.

Figure 3. Correlation analyses between ASRs of NAFLD and SDI in 21 regions and 204 territories from 1990 to 2019. The SDI presented no obvious correlation with the (A, E) ASIR and negative correlations with (B, F) ASPR, (C, G) ASDR, and (D, H) age-standardized DALY in 21 regions and 204 territories. ASDR: age-standardized death rate; ASIR: age-standardized incident rate; ASPR: age-standardized prevalent rate; ASR: age-standardized rate; DALY: disability-adjusted life years; NAFLD: nonalcoholic fatty liver disease; SDI: sociodemographic index.



NAFLD-Related Mortalities

Although the global deaths due to NAFLD increased by almost 2-fold (from 93,760 in 1990 to 168,970 in 2019), the age-standardized death rate appeared to descend from 2.39 per 100,000 population (95% CI 1.84-3.05) in 1990 to 2.09 per 100,000 population (95% CI 1.61-2.60) in 2019, as demonstrated by the EAPC of ASDR with a negative value (0.67, 95% CI 0.76 to 0.57; Table S3 in Multimedia Appendix 2). After the classification by sex and SDI value, the cases of death sharply increased and the ASDR visibly decreased among all groups, which is similar to that observed at the global level (Figure 1C, Image C in Multimedia Appendix 3, and Table S3 in Multimedia Appendix 2).

Throughout various regions, NAFLD mortality and ASDR revealed extreme variations. East and south Asia had the highest mortality in both 1990 and 2019, while the top 3 ASDRs were Central Latin America, Andean Latin America, and eastern Sub-Saharan Africa (Figure 2C; Table S3 in Multimedia Appendix 2). However, owing to the relatively high ASDR in 1990, the temporal changes displayed negative trends in most regions. The most significant reductions in ASDR were observed in east Asia (EAPC 3.04, 95% CI 3.4 to 2.69), high-income Asia Pacific (EAPC 1.66, 95% CI 1.97 to 1.34), and western Europe (EAPC 1.32, 95% CI 1.41 to 1.24; Figure 2G and Table S3 in Multimedia Appendix 2). Similar to ASPR, Egypt had the highest ASDR in 2019 (15.97 per 100,000 population). In addition, correlation analyses demonstrated that SDI had a negative association with ASDR among 21 regions (r=0.48, P<.001) and 204 countries (r=0.58, P<.001) (Figures 3C and 3G), which might indicate higher ASDR in developing territories.

NAFLD DALY

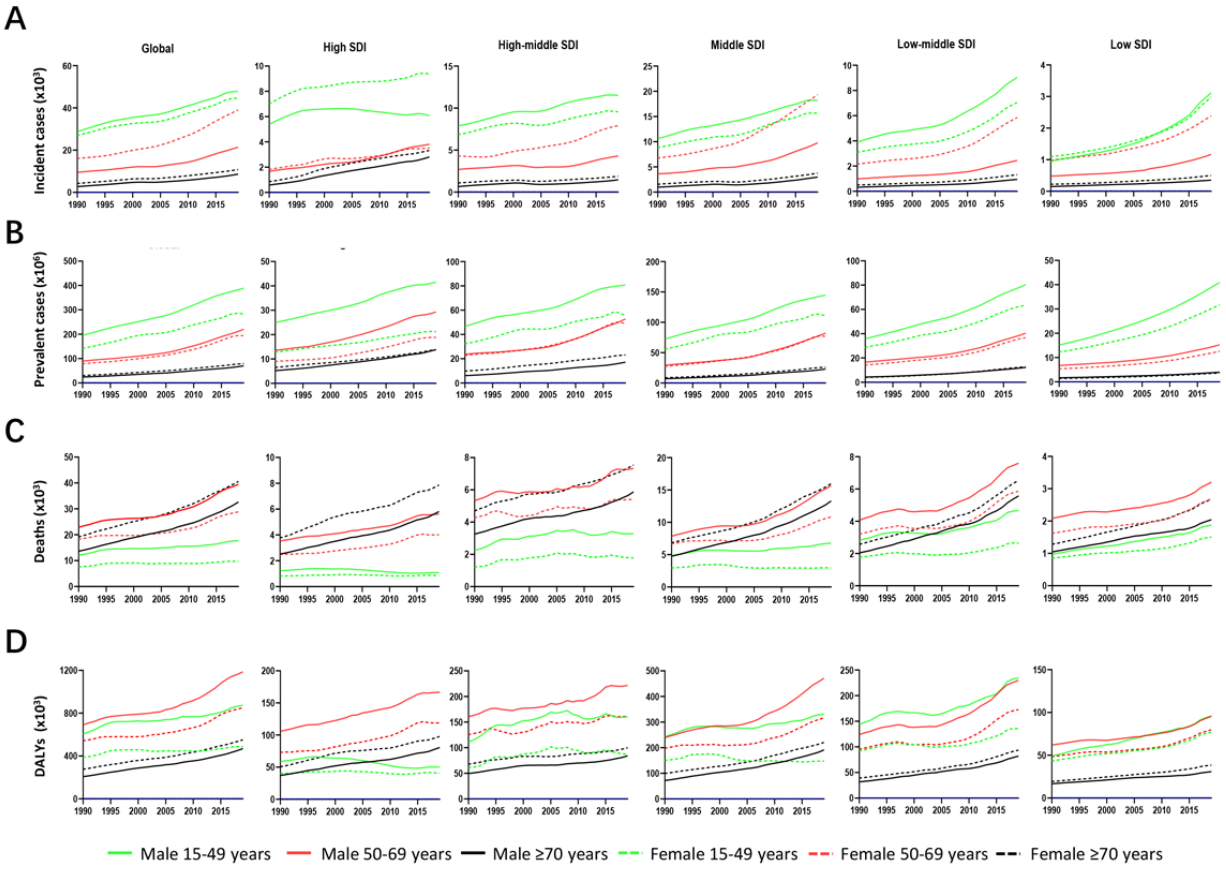
DALY is a critical parameter assessing disease burden, including years of life lost (YLL) owing to premature death and years lived with disability (YLD). As shown in Table S4 in Multimedia Appendix 2, global DALYs regarding NAFLD were elevated from 2,711,270 (95% CI 2,078,580-3,478,940) to 4,417,280 (95% CI 3,348,220-5,671,200). However, age-standardized DALYs presented decreased changes, with an EAPC of 0.82 (95% CI 0.93 to 0.70). In Figure 1D, the DALYs in both sexes and among different SDI regions exhibited obvious increases. However, the age-standardized DALYs all presented decreasing trends, with a more significant decline in women relative to men and in the middle SDI region relative to other SDI regions (Image D in Multimedia Appendix 3 and Table S4 in Multimedia Appendix 2).

Similar to the epidemiologic pattern of NAFLD-related death, the highest DALYs affected east and south Asia. The top 3 age-standardized DALYs are in Central Latin America (161.39 per 100,000 population), Andean Latin America (128.67 per 100,000 population), and Central Asia (112.39 per 100,000 population) (Figure 2D). The decline in age-standardized DALY was most pronounced in east Asia (EAPC 3.42, 95% CI 3.82 to 3.03), the high-income Asia Pacific (EAPC 2.15, 95% CI 2.48 to 1.81), and western Europe (EAPC 1.62, 95% CI 1.71 to 1.53), whereas eastern Europe and central Asia presented significant elevations (Figure 2H and Table S4 in Multimedia Appendix 2). Finally, we found that age-standardized DALY had a marked negative correlation with SDI among 21 territories (r=0.45, P<.001) (Figure 3D) or among 204 countries (r=0.56, P<.001) (Figure 3H).

Age Distribution

As depicted in Figure 4, all age groups presented gradually increasing trends in incidence, prevalence, death, and DALY, especially in low-middle SDI regions, regardless of sex and SDI values. Nevertheless, there was a certain heterogeneity in the specific age distribution. Young adults (aged from 15 to 49 years) dominated the NAFLD incidence and prevalence over the past 30 years, with global male incidence changing from 28,785 in 1990 to 47,862 in 2019 and with global male prevalence changing from 195,268,990 in 1990 to 388,787,021 in 2019 (Figures 4A and 4B and Multimedia Appendix 4), while the incidence and prevalence among females were slightly less than those among males. Interestingly, in high-SDI regions, NAFLD incidence was always higher in females than in males (females: 6,847 in 1990 to 9555 in 2019; males: 5399 in 1990 to 6082 in 2019; Figure 4A and Multimedia Appendix 4). In terms of NAFLD-related deaths, those aged 50 to 69 years were mostly men, and the elderly (aged above 70 years) held the dominant place among women, whereas young adults had relatively low mortality, which might correlate with the chronic and slow progression of NAFLD (Figure 4C). The NAFLD DALYs mainly influenced the quinquagenarian males with highest DALYs in 2019 of 1,186,057, but in relatively underdeveloped regions, the DALYs of young adults remained close to the quinquagenarians (Figure 4D and Multimedia Appendix 4). Consequently, age may be a vital factor affecting the NAFLD burden in various regions.

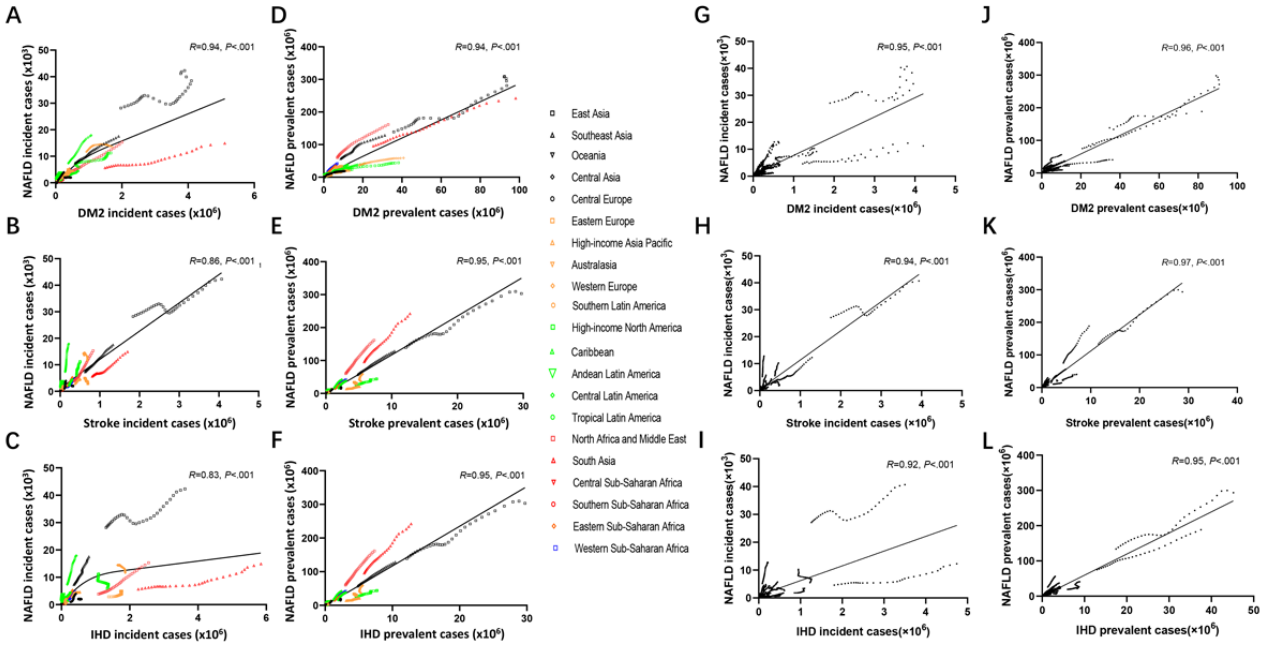
Figure 4. The trends of changes in NAFLD incidence, prevalence, deaths, and DALYs from 1990 to 2019 in different age groups. The trends of changes in (A) incidence, (B) prevalence, (C) deaths, and (D) DALYs. DALY: disability-adjusted life years; NAFLD: nonalcoholic fatty liver disease; SDI: sociodemographic index.



Associations Between NAFLD and Other Interconnected Diseases

To further explore this hazardous disease, we analyzed the associations between NAFLD and 3 common interconnected disorders [22]. Interestingly, we found that the incidence of NAFLD presented strongly positive correlations with that of DM2, stroke, and IHD, both in 21 territories (r=0.94, P<.001; r=0.86, P<.001; r=0.83, P<.001; Figures 5A-5C) and 204 countries (r=0.95, P<.001; r=0.94, P<.001; r=0.92, P<.001; Figures 5G-5I). Meanwhile, the same positive associations were observed in the prevalence of NAFLD and DM2, stroke, and IHD in 21 territories (r=0.94, P<.001; r=0.95, P<.001; r=0.95, P<.001; Figures 5D-5F) and in 204 countries (r=0.96, P<.001; r=0.97, P<.001; r=0.95, P<.001; Figures 5J-5L).

Figure 5. The associations of incidence and prevalence between NAFLD and other comorbidities in 21 regions and 204 territories from 1990 to 2019. NAFLD presented strongly positive correlations of incidence/prevalence with (A, D, G, J) DM2, (B, E, H, K) stroke, and (C, F, I, L) IHD in 21 regions and 204 territories from 1990 to 2019. DM2: diabetes mellitus type 2; IHD: ischemic heart disease; NAFLD: nonalcoholic fatty liver disease.



Discussion

Principal Findings

In this study, we comprehensively assessed the global burden of NAFLD and compared its associations with common correlative diseases. In general, the increasing disease burden caused by NAFLD has placed heavy pressure on contemporary society at a gradual pace over recent decades, with corresponding epidemiological parameters showing upward changes, including increases in incidence, prevalence, deaths, and DALYs. A previous study [23] suggested a significant shift in NAFLD burden toward a younger population, which was echoed by our findings. However, after age standardization, there were no obvious upregulations in NAFLD incidence, even though declining alterations in ASDR and age-standardized DALYs were observed. Meanwhile, we uncovered distinctly negative correlations between the SDI and ASPR, ASDR, and age-standardized DALYs. Finally, we confirmed 3 strongly relevant diseases accompanied by NAFLD incidence and prevalence, namely, DM2, stroke, and IHD. Therefore, the systematic understanding of global epidemiologic patterns for NAFLD and its interrelated disorders may be valuable for the development of corresponding prevention and control strategies, especially for public health policy makers and clinical physicians.

Compared to 1990, the overall incidence and prevalence of NAFLD in 2019 increased by approximately 2-fold, mainly impacting low- and middle-SDI regions, such as some countries in the Middle East and north Africa [2,24]. Furthermore, the same rising socioeconomic burden of NAFLD has influenced relatively developed regions. Williams reported that the prevalence of NAFLD was up to 46% in the United States [25], while other European countries, including Italy, Greece, and the United Kingdom, presented with a markedly increased incidence and prevalence, resulting in an increasing socioeconomic burden [26-28]. In addition, relatively young patients dominated NAFLD morbidity and their prevalence rapidly increased, especially in low-middle-SDI regions. Given the pathogenesis of NAFLD associated with fat accumulation in hepatocytes and the growing obesity among youngsters [1,29], this seems reasonable to explain the increasing number of young patients. Interestingly, unlike the increased number of NAFLD cases, ASIR presented no marked upregulation, which might partly be associated with changes in sociodemographic structure in different areas and countries [30,31]. The reason behind it remains under investigation and requires more research.

NAFLD is a complex and multifactor disorder that is affected by metabolic and environmental factors, along with genetic and epigenetic predispositions involving multiple organs and diverse mechanisms [32]. The exact contribution of each factor to the development of NAFLD is unknown, requiring further investigation, and it may vary by geographic location, which is associated with the great heterogeneity of the NAFLD prevalence in different districts. Recently, metabolic imbalances have been gradually considered the predominant risk factor, and an international expert group has agreed to change the name of NAFLD to metabolic (dysfunction)–associated fatty liver disease (MAFLD) [33,34]. DM2, as the most prevalent metabolic disease worldwide, was found, in our study, to correlate significantly with the NAFLD incidence and prevalence, which is concurrent with previous studies [35,36] and further highlights the vital role of metabolic dysfunction in NAFLD. Meanwhile, owing to alterations in the diet structure in modern life, populations of individuals with obesity are increasing at a rapid pace, which is regarded as the main risk factor for diabetes and fuels NAFLD-related morbidity. In fact, NAFLD may in turn be a pathogenic component of the development of DM2, and the bidirectional relationship between

NAFLD and type 2 diabetes remains controversial and needs more exploration [37]. However, active prevention and control of obesity and diabetes can help alleviate the development of NAFLD to some extent [38,39].

NAFLD poses a substantial threat to individual health, with the number of deaths and DALYs increasing dramatically from 1990 to 2019, which was primarily driven by population growth and aging worldwide, specifically in low-middle–income countries [30]. In the meantime, the patients older than 50 years had the most deaths and DALYs, regardless of sex or the different SDI regions, which could be expected because of the aging population and the worse response to therapy among the elderly population. However, the age-standardized death and DALY rates were decreased globally and were particularly lower in higher-SDI regions. With rapid progression of society and elevation of health care, we primarily regard that the early diagnosis and prompt treatment could improve the survival and prognosis of patients with NAFLD. Additionally, Allen et al [23] previously reported the shift in NAFLD incidence toward a younger population who accounted for the majority of patients with NAFLD , which was consistent with our results. Meanwhile, considering the chronic course of NAFLD [40], it could be easily concluded that there showed relatively low deaths and DALY in young adults. Therefore, the global ASDR and age-standardized DALY presented downward changes after age standardization. Besides, the association analyses showed significantly negative relations between SDI and ASDR or age-standardized DALY, indicating a more serious NAFLD burden in lower-income regions. In specific, we observed relatively high ASDR and age-standardized DALY in Latin America, north Africa, and the Middle East. High-income Asia Pacific, Central Europe, and high-income North America had lower ASDRs and age-standardized DALY. Differences in access to health care and medical level have remained the fundamental factor for immense heterogeneity in the disease's mortality across countries [41]. Therefore, most NAFLD-related deaths could partly be reduced in high-income countries through easy access to better health care and a stronger health infrastructure, such as early-stage identification of NAFLD and education of patients. Of course, other factors contributing to the decreases in the ASRs of NAFLD death and DALYs remained to be investigated.

In the general population, more than 10% of all patients with NAFLD may develop NASH [42], which is characterized by steatosis, hepatocellular ballooning, lobular inflammation, and often fibrosis [43]. During the response to tissue damage, hepatocytes are replaced by type I collagen produced by stellate cells, leading to the progression of NASH toward fibrosis and cirrhosis with overt clinical consequences [44-46]. Furthermore, patients with NASH have been reported to be highly susceptible to liver cancer [47]. Failure to recognize high-risk individuals and provide prompt treatment for NAFLD might lead to progression to NASH and even cirrhosis or carcinoma with a poor prognosis and high mortality, especially in low-income countries. Meanwhile, it also partly explained why the ASDR and age-standardized DALY of NAFLD were lower in regions with higher SDI values than in those with lower SDI values. Accordingly, to decrease the mortality and DALYs of NAFLD, preventing its development into NASH and then cirrhosis or liver cancer is essential.

In addition to hepatic causes, cardiovascular disease is the leading cause of death in patients with NAFLD, with mortality up to approximately 20% [48]. In this study, we showed that common cardiovascular diseases, including stroke and ischemic heart disease, had strong positive associations with the incidence and prevalence of NAFLD, which can be attributable to shared risk factors between these 2 diseases, such as dyslipidemia, insulin resistance, hypertension, and obesity [49,50]. Consequently, co-occurring stroke and IHD in patients with NAFLD merits considerable attention for better prevention of cardiovascular events and lower mortality.

Limitations

There remain some obvious limitations of this study. First, we relied heavily on GBD estimates for this study. The accuracy of the GBD estimates was limited by the quality and availability of each country's vital registration system and a mass of undefined NAFLD cases in their registry data. Moreover, other possible intermediate factors or confounders during the correlation analyses were not included for adjustment owing to lack of data on other relevant parameters. Other potential interconnected diseases with NAFLD, such as dyslipidemia that are unavailable in GBD 2019 database, await further investigation. In addition, we preliminarily conducted the association analyses and calculated the potential correlation coefficients via the GBD data set but cannot perform the validation on correlation coefficients due to lacking other analogous study populations covering almost all regions and countries in the world. Finally, subgroup analyses were not performed among patients with NAFLD grouped on the basis of whether they had comorbid metabolic syndrome, were on medication therapy, or had other factors.

Conclusions

The global burden of NAFLD is gradually increased and is predicted to continue to increase in the future. The morbidity presented a clear shift toward young populations. Meanwhile, higher age-standardized death and DALY rates can be observed in aged individuals and low-SDI regions. Furthermore, NAFLD presented strong correlations with three high-risk comorbidities, namely, DM2, stroke, and ischemic heart disease. Therefore, the development of cost-effective global and regional strategies to mitigate NAFLD morbidity and mortality, alleviate the socioeconomic burden, and prevent risky interconnected diseases are urgently required by policy makers and clinical physicians.

Acknowledgments

We are grateful to all the GBD 2019 staff for their contribution. This work was funded by the National Natural Science Foundation of China (grant numbers 81800041 and 82000078).

Data Availability

The data sets used in the present study are available in the GBD 2019 study.

Authors' Contributions

ZLZ had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design was prepared by ZLZ and LYC. Acquisition, analysis, and interpretation of data were performed by HLC, YZ, JXZ, SC, and YHZ revised the manuscript. Drafting of the manuscript was done by HLC and YZ. ZLZ and LYC revised the manuscript, statistical analysis was performed by HLC and YZ, and ZLZ and LYC supervised the data. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original data for Fig. 3.
[XLSX File (Microsoft Excel File), 541 KB - publichealth_v8i12e34809_app1.xlsx]

Multimedia Appendix 2
Supplementary Tables.
[DOCX File , 56 KB - publichealth_v8i12e34809_app2.docx]

Multimedia Appendix 3
The change trends of NAFLD ASIR, ASPR, ASDR and age-standardized DALYs from 1990 to 2019. The change trends of ASIR (A), ASPR (B), ASDR (C) and age-standardized DALYs (D) were displayed. Blue bars represent males and red bars represent females. NAFLD: nonalcoholic fatty liver disease; ASIR: age-standardized incident rate; ASPR: age-standardized prevalent rate; ASDR: age-standardized death rate; DALYs: disability-adjusted life years; SDI: social-demographic index.
[PNG File , 178 KB - publichealth_v8i12e34809_app3.png]

Multimedia Appendix 4
Original data for Fig. 4.
[XLSX File (Microsoft Excel File), 349 KB - publichealth_v8i12e34809_app4.xlsx]

References

1. Sanyal AJ, Brunt EM, Kleiner DE, Kowdley KV, Chalasani N, Lavine JE, et al. Endpoints and clinical trial design for nonalcoholic steatohepatitis. Hepatology 2011 Jul;54(1):344-353 [FREE Full text] [doi: 10.1002/hep.24376] [Medline: 21520200]
2. Golabi P, Paik JM, AlQahtani S, Younossi Y, Tuncer G, Younossi ZM. Burden of non-alcoholic fatty liver disease in Asia, the Middle East and North Africa: data from Global Burden of Disease 2009-2019. J Hepatol 2021 Oct;75(4):795-809. [doi: 10.1016/j.jhep.2021.05.022] [Medline: 34081959]
3. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 2018 Jan;15(1):11-20. [doi: 10.1038/nrgastro.2017.109] [Medline: 28930295]
4. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018 Jan;67(1):328-357. [doi: 10.1002/hep.29367] [Medline: 28714183]
5. Golabi P, Rhea L, Henry L, Younossi ZM. Hepatocellular carcinoma and non-alcoholic fatty liver disease. Hepatol Int 2019 Nov;13(6):688-694. [doi: 10.1007/s12072-019-09995-8] [Medline: 31701393]
6. Younossi ZM, Stepanova M, Rafiq N, Makhlouf H, Younoszai Z, Agrawal R, et al. Pathologic criteria for nonalcoholic steatohepatitis: interprotocol agreement and ability to predict liver-related mortality. Hepatology 2011 Jun;53(6):1874-1882. [doi: 10.1002/hep.24268] [Medline: 21360720]
7. Angulo P, Kleiner DE, Dam-Larsen S, Adams LA, Bjornsson ES, Charatcharoenwitthaya P, et al. Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. Gastroenterology 2015 Aug;149(2):389-397.e10 [FREE Full text] [doi: 10.1053/j.gastro.2015.04.043] [Medline: 25935633]
8. Wong RJ, Aguilar M, Cheung R, Perumpail RB, Harrison SA, Younossi ZM, et al. Nonalcoholic steatohepatitis is the second leading etiology of liver disease among adults awaiting liver transplantation in the United States. Gastroenterology 2015 Mar;148(3):547-555. [doi: 10.1053/j.gastro.2014.11.039] [Medline: 25461851]

9. Li J, Zou B, Yeo YH, Feng Y, Xie X, Lee DH, et al. Prevalence, incidence, and outcome of non-alcoholic fatty liver disease in Asia, 1999-2019: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2019 May;4(5):389-398. [doi: 10.1016/S2468-1253(19)30039-1] [Medline: 30902670]
10. Paik JM, Golabi P, Younossi Y, Mishra A, Younossi ZM. Changes in the Global Burden of Chronic Liver Diseases From 2012 to 2017: the growing impact of NAFLD. Hepatology 2020 Nov;72(5):1605-1616. [doi: 10.1002/hep.31173] [Medline: 32043613]
11. Sanai FM, Al Khathlan A, Al Fadhli A, Jazzar AS, Hashim AM, Mansour E, et al. Clinical and economic burden of nonalcoholic steatohepatitis in Saudi Arabia, United Arab Emirates and Kuwait. Hepatol Int 2021 Aug;15(4):912-921 [FREE Full text] [doi: 10.1007/s12072-021-10182-x] [Medline: 33822317]
12. Flemming JA, Djerboua M, Groome PA, Booth CM, Terrault NA. NAFLD and alcohol-associated liver disease will be responsible for almost all new diagnoses of cirrhosis in Canada by 2040. Hepatology 2021 Dec;74(6):3330-3344. [doi: 10.1002/hep.32032] [Medline: 34174003]
13. Morgan A, Hartmanis S, Tsochatzis E, Newsome PN, Ryder SD, Elliott R, et al. Disease burden and economic impact of diagnosed non-alcoholic steatohepatitis (NASH) in the United Kingdom (UK) in 2018. Eur J Health Econ 2021 Jun;22(4):505-518 [FREE Full text] [doi: 10.1007/s10198-020-01256-y] [Medline: 33751289]
14. Chan W, Tan AT, Vethakkan SR, Tah P, Vijayananthan A, Goh K. Non-alcoholic fatty liver disease in diabetics—prevalence and predictive factors in a multiracial hospital clinic population in Malaysia. J Gastroenterol Hepatol 2013 Aug;28(8):1375-1383. [doi: 10.1111/jgh.12204] [Medline: 23517307]
15. Hodson L, Gunn PJ. The regulation of hepatic fatty acid synthesis and partitioning: the effect of nutritional state. Nat Rev Endocrinol 2019 Dec;15(12):689-700. [doi: 10.1038/s41574-019-0256-9] [Medline: 31554932]
16. Vernon G, Baranova A, Younossi ZM. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. Aliment Pharmacol Ther 2011 Aug;34(3):274-285 [FREE Full text] [doi: 10.1111/j.1365-2036.2011.04724.x] [Medline: 21623852]
17. Portillo Sanchez P, Bril F, Maximos M, Lomonaco R, Biernacki D, Orsak B, et al. High prevalence of nonalcoholic fatty liver disease in patients with type 2 diabetes mellitus and normal plasma aminotransferase levels. J Clin Endocrinol Metab 2014 Oct 10;100(5):2231-2238. [doi: 10.1210/jc.2014-2739] [Medline: 25303488]
18. Alexander M, Loomis AK, van der Lei J, Duarte-Salles T, Prieto-Alhambra D, Ansell D, et al. Non-alcoholic fatty liver disease and risk of incident acute myocardial infarction and stroke: findings from matched cohort study of 18 million European adults. BMJ 2019 Oct 08;367:l5367 [FREE Full text] [doi: 10.1136/bmj.l5367] [Medline: 31594780]
19. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
20. GBR results. Institute for Health Metrics and Evaluation. URL: http://ghdx.healthdata.org/gbd-results-tool [accessed 2022-12-02]
21. Fan J, Liu Z, Mao X, Tong X, Zhang T, Suo C, et al. Global trends in the incidence and mortality of esophageal cancer from 1990 to 2017. Cancer Med 2020 Sep;9(18):6875-6887 [FREE Full text] [doi: 10.1002/cam4.3338] [Medline: 32750217]
22. Ferguson D, Finck BN. Emerging therapeutic approaches for the treatment of NAFLD and type 2 diabetes mellitus. Nat Rev Endocrinol 2021 Aug;17(8):484-495. [doi: 10.1038/s41574-021-00507-z] [Medline: 34131333]
23. Allen AM, Therneau TM, Larson JJ, Coward A, Somers VK, Kamath PS. Nonalcoholic fatty liver disease incidence and impact on metabolic burden and death: a 20 year-community study. Hepatology 2018 May;67(5):1726-1736 [FREE Full text] [doi: 10.1002/hep.29546] [Medline: 28941364]
24. Sanai FM, Abaalkhail F, Hasan F, Farooqi MH, Nahdi NA, Younossi ZM. Management of nonalcoholic fatty liver disease in the Middle East. World J Gastroenterol 2020 Jul 07;26(25):3528-3541 [FREE Full text] [doi: 10.3748/wjg.v26.i25.3528] [Medline: 32742124]
25. Williams CD, Stengel J, Asike MI, Torres DM, Shaw J, Contreras M, et al. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. Gastroenterology 2011 Jan;140(1):124-131. [doi: 10.1053/j.gastro.2010.09.038]
26. Soresi M, Noto D, Cefalù AB, Martini S, Vigna GB, Fonda M, Metabolic Syndrome Study Group. Nonalcoholic fatty liver and metabolic syndrome in Italy: results from a multicentric study of the Italian Arteriosclerosis society. Acta Diabetol 2013 Apr;50(2):241-249. [doi: 10.1007/s00592-012-0406-1] [Medline: 22684314]
27. Zois CD, Baltayiannis GH, Bekiari A, Goussia A, Karayiannis P, Doukas M, et al. Steatosis and steatohepatitis in postmortem material from Northwestern Greece. World J Gastroenterol 2010 Aug 21;16(31):3944-3949 [FREE Full text] [doi: 10.3748/wjg.v16.i31.3944] [Medline: 20712056]
28. Alazawi W, Mathur R, Abeysekera K, Hull S, Boomla K, Robson J, et al. Ethnicity and the diagnosis gap in liver disease: a population-based study. Br J Gen Pract 2014 Nov;64(628):e694-e702 [FREE Full text] [doi: 10.3399/bjgp14X682273] [Medline: 25348993]
29. Fan J, Kim S, Wong VW. New trends on obesity and NAFLD in Asia. J Hepatol 2017 Oct;67(4):862-873. [doi: 10.1016/j.jhep.2017.06.003] [Medline: 28642059]

30. GBD 2017 Population and Fertility Collaborators. Population and fertility by age and sex for 195 countries and territories, 1950-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1995-2051 [FREE Full text] [doi: 10.1016/S0140-6736(18)32278-5] [Medline: 30496106]
31. Vollset SE, Goren E, Yuan C, Cao J, Smith AE, Hsiao T, et al. Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study. Lancet 2020 Oct 17;396(10258):1285-1306 [FREE Full text] [doi: 10.1016/S0140-6736(20)30677-2] [Medline: 32679112]
32. Juanola O, Martínez-López S, Francés R, Gómez-Hurtado I. Non-alcoholic fatty Liver disease: metabolic, genetic, epigenetic and environmental risk factors. Int J Environ Res Public Health 2021 May 14;18(10):5227. [doi: 10.3390/ijerph18105227] [Medline: 34069012]
33. Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: an international expert consensus statement. J Hepatol 2020 Jul;73(1):202-209. [doi: 10.1016/j.jhep.2020.03.039] [Medline: 32278004]
34. Eslam M, Sanyal AJ, George J, International Consensus Panel. MAFLD: a consensus-driven proposed nomenclature for metabolic associated fatty liver disease. Gastroenterology 2020 May;158(7):1999-2014.e1. [doi: 10.1053/j.gastro.2019.11.312] [Medline: 32044314]
35. Mishra A, Younossi ZM. Epidemiology and natural history of non-alcoholic fatty liver disease. J Clin Exp Hepatol 2012 Jun;2(2):135-144 [FREE Full text] [doi: 10.1016/S0973-6883(12)60102-9] [Medline: 25755422]
36. Leite NC, Villela-Nogueira CA, Pannain VLN, Bottino AC, Rezende GFM, Cardoso CRL, et al. Histopathological stages of nonalcoholic fatty liver disease in type 2 diabetes: prevalences and correlated factors. Liver Int 2011 May;31(5):700-706. [doi: 10.1111/j.1478-3231.2011.02482.x] [Medline: 21457442]
37. Tilg H, Moschen AR, Roden M. NAFLD and diabetes mellitus. Nat Rev Gastroenterol Hepatol 2017 Jan;14(1):32-42. [doi: 10.1038/nrgastro.2016.147] [Medline: 27729660]
38. Schmid A, Arians M, Karrasch T, Pons-Kühnemann J, Schäffler A, Roderfeld M, et al. Improvement of type 2 diabetes mellitus and attenuation of NAFLD are associated with the success of obesity therapy. J Clin Med 2022 Mar 22;11(7):1756 [FREE Full text] [doi: 10.3390/jcm11071756] [Medline: 35407364]
39. Kumar V, Xin X, Ma J, Tan C, Osna N, Mahato RI. Therapeutic targets, novel drugs, and delivery systems for diabetes associated NAFLD and liver fibrosis. Adv Drug Deliv Rev 2021 Sep;176:113888 [FREE Full text] [doi: 10.1016/j.addr.2021.113888] [Medline: 34314787]
40. Ekstedt M, Franzén LE, Mathiesen UL, Thorelius L, Holmqvist M, Bodemar G, et al. Long-term follow-up of patients with NAFLD and elevated liver enzymes. Hepatology 2006 Oct;44(4):865-873. [doi: 10.1002/hep.21327] [Medline: 17006923]
41. GBD 2017 Cirrhosis Collaborators. The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol 2020 Mar;5(3):245-266 [FREE Full text] [doi: 10.1016/S2468-1253(19)30349-8] [Medline: 31981519]
42. Delli Bovi AP, Marciano F, Mandato C, Siano MA, Savoia M, Vajro P. Oxidative stress in non-alcoholic fatty liver disease. An updated mini review. Front Med (Lausanne) 2021;8:595371 [FREE Full text] [doi: 10.3389/fmed.2021.595371] [Medline: 33718398]
43. Kleiner DE, Makhlouf HR. Histology of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis in adults and children. Clin Liver Dis 2016 May;20(2):293-312 [FREE Full text] [doi: 10.1016/j.cld.2015.10.011] [Medline: 27063270]
44. Argo CK, Caldwell SH. Epidemiology and natural history of non-alcoholic steatohepatitis. Clin Liver Dis 2009 Nov;13(4):511-531. [doi: 10.1016/j.cld.2009.07.005] [Medline: 19818302]
45. Starley BQ, Calcagno CJ, Harrison SA. Nonalcoholic fatty liver disease and hepatocellular carcinoma: a weighty connection. Hepatology 2010 May;51(5):1820-1832. [doi: 10.1002/hep.23594] [Medline: 20432259]
46. Ratziu V, Bellentani S, Cortez-Pinto H, Day C, Marchesini G. A position statement on NAFLD/NASH based on the EASL 2009 special conference. J Hepatol 2010 Aug;53(2):372-384 [FREE Full text] [doi: 10.1016/j.jhep.2010.04.008] [Medline: 20494470]
47. Zhang T, Xu J, Ye L, Lin X, Xu Y, Pan X, et al. Age, gender and geographic differences in Global Health Burden of cirrhosis and liver cancer due to nonalcoholic steatohepatitis. J Cancer 2021;12(10):2855-2865 [FREE Full text] [doi: 10.7150/jca.52282] [Medline: 33854586]
48. Kim D, Adejumo AC, Yoo ER, Iqbal U, Li AA, Pham EA, et al. Trends in mortality From extrahepatic complications in patients with chronic liver disease, from 2007 through 2017. Gastroenterology 2019 Oct;157(4):1055-1066.e11. [doi: 10.1053/j.gastro.2019.06.026] [Medline: 31251928]
49. Stepanova M, Younossi ZM. Independent association between nonalcoholic fatty liver disease and cardiovascular disease in the US population. Clin Gastroenterol Hepatol 2012 Jun;10(6):646-650. [doi: 10.1016/j.cgh.2011.12.039] [Medline: 22245962]
50. Marchesini G, Bugianesi E, Forlani G, Cerrelli F, Lenzi M, Manini R, et al. Nonalcoholic fatty liver, steatohepatitis, and the metabolic syndrome. Hepatology 2003 Apr;37(4):917-923. [doi: 10.1053/jhep.2003.50161] [Medline: 12668987]

Abbreviations

ASIR: age-standardized incidence rate
ASR: age-standardized rate
ASPR: age-standardized prevalence rate
CI: certainty interval
CLD: chronic liver disease
DALY: disability-adjusted life years
DM2: diabetes mellitus type 2
EAPC: estimated annual percentage change
GBD 2019: Global Burden of Disease 2019
IHD: ischemic heart disease
IHTAG: intrahepatic triacylglycerol
MAFLD: metabolic (dysfunction)–associated fatty liver disease
NAFLD: nonalcoholic fatty liver disease
NASH: nonalcoholic steatohepatitis
SDI: sociodemographic index
YLD: years lived with disability
YLL: years of life lost

Edited by A Mokdad; submitted 09.11.21; peer-reviewed by C Shah, K Fitzner, P Nguyen; comments to author 18.05.22; revised version received 22.06.22; accepted 25.10.22; published 12.12.22.

Please cite as:
Chen H, Zhan Y, Zhang J, Cheng S, Zhou Y, Chen L, Zeng Z
The Global, Regional, and National Burden and Trends of NAFLD in 204 Countries and Territories: An Analysis From Global Burden of Disease 2019
JMIR Public Health Surveill 2022;8(12):e34809
URL: https://publichealth.jmir.org/2022/12/e34809
doi:10.2196/34809
PMID:36508249

Original Paper

Representativeness, Vaccination Uptake, and COVID-19 Clinical Outcomes 2020-2021 in the UK Oxford-Royal College of General Practitioners Research and Surveillance Network: Cohort Profile Summary

Meredith Leston[1*], MPH; William H Elson[1*], BSc, MSc, MBBS; Conall Watson[2*], PhD; Anissa Lakhani[2], BSc, MPH; Carole Aspden[1], BA; Clare R Bankhead[1], DPhil; Ray Borrow[3], PhD; Elizabeth Button[1], MSc; Rachel Byford[1], BA; Alex J Elliot[4], PhD; Xuejuan Fan[1], PhD; Uy Hoang[1], MD (Res); Ezra Linley[3], PhD; Jack Macartney[1]; Brian D Nicholson[1], MBChB, DPhil; Cecilia Okusi[1], MRES; Mary Ramsay[2], MBBS; Gillian Smith[4], MBBS; Sue Smith[4], MPH; Mark Thomas[5], MA; Dan Todkill[4], MBChB, PhD; Ruby SM Tsang[1], PhD; William Victor[5], BSc; Alice J Williams[1], BA, MST, PhD; John Williams[1], MBBChir, MSc; Maria Zambon[6], PhD; Gary Howsam[5], MBBS; Gayatri Amirthalingam[2*], BSc, MBBS, MPH; Jamie Lopez-Bernal[2*], MBBS; F D Richard Hobbs[1*], FMedSci; Simon de Lusignan[1*], MD (Res)

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2]Immunisation and Vaccine Preventable Diseases Division, UK Health Security Agency, Colindale, London, United Kingdom

[3]Vaccine Evaluation Unit, UK Health Security Agency, Manchester Royal Infirmary, Manchester, United Kingdom

[4]Real-time Syndromic Surveillance Team, Field Service, UK Health Security Agency, Birmingham, United Kingdom

[5]Royal College of General Practitioners, London, United Kingdom

[6]Reference Microbiology, UK Health Security Agency, Colindale, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Simon de Lusignan, MD (Res)
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 01865 617855
Email: simon.delusignan@phc.ox.ac.uk

Abstract

Background: The Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) is one of Europe's oldest sentinel systems, working with the UK Health Security Agency (UKHSA) and its predecessor bodies for 55 years. Its surveillance report now runs twice weekly, supplemented by online observatories. In addition to conducting sentinel surveillance from a nationally representative group of practices, the RSC is now also providing data for syndromic surveillance.

Objective: The aim of this study was to describe the cohort profile at the start of the 2021-2022 surveillance season and recent changes to our surveillance practice.

Methods: The RSC's pseudonymized primary care data, linked to hospital and other data, are held in the Oxford-RCGP Clinical Informatics Digital Hub, a Trusted Research Environment. We describe the RSC's cohort profile as of September 2021, divided into a Primary Care Sentinel Cohort (PCSC)—collecting virological and serological specimens—and a larger group of syndromic surveillance general practices (SSGPs). We report changes to our sampling strategy that brings the RSC into alignment with European Centre for Disease Control guidance and then compare our cohort's sociodemographic characteristics with Office for National Statistics data. We further describe influenza and COVID-19 vaccine coverage for the 2020-2021 season (week 40 of

2020 to week 39 of 2021), with the latter differentiated by vaccine brand. Finally, we report COVID-19–related outcomes in terms of hospitalization, intensive care unit (ICU) admission, and death.

Results: As a response to COVID-19, the RSC grew from just over 500 PCSC practices in 2019 to 1879 practices in 2021 (PCSC, n=938; SSGP, n=1203). This represents 28.6% of English general practices and 30.59% (17,299,780/56,550,136) of the population. In the reporting period, the PCSC collected >8000 virology and >23,000 serology samples. The RSC population was broadly representative of the national population in terms of age, gender, ethnicity, National Health Service Region, socioeconomic status, obesity, and smoking habit. The RSC captured vaccine coverage data for influenza (n=5.4 million) and COVID-19, reporting dose one (n=11.9 million), two (n=11 million), and three (n=0.4 million) for the latter as well as brand-specific uptake data (AstraZeneca vaccine, n=11.6 million; Pfizer, n=10.8 million; and Moderna, n=0.7 million). The median (IQR) number of COVID-19 hospitalizations and ICU admissions was 1181 (559-1559) and 115 (50-174) per week, respectively.

Conclusions: The RSC is broadly representative of the national population; its PCSC is geographically representative and its SSGPs are newly supporting UKHSA syndromic surveillance efforts. The network captures vaccine coverage and has expanded from reporting primary care attendances to providing data on onward hospital outcomes and deaths. The challenge remains to increase virological and serological sampling to monitor the effectiveness and waning of all vaccines available in a timely manner.

(JMIR Public Health Surveill 2022;8(12):e39141) doi:10.2196/39141

KEYWORDS

cohort profile; computerized medical record systems; general practice; influenza; COVID-19; sentinel surveillance; syndromic surveillance; serology; virology; public health; digital surveillance; vaccination; primary care data; health data; cohort; virus; immunology; surveillance; representation; uptake; outcome; hospital; sampling; monitoring

Introduction

The emergence of SARS-CoV-2 and the resultant COVID-19 pandemic has reinforced the importance of continuous respiratory disease surveillance. However, processing routine health data comes with considerable challenges, requiring sophisticated digital infrastructure and data linkage to secondary data sources. The benefits afforded by such surveillance are contingent on data quality and timeliness.

The Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) is now completing its 55th year of surveillance [1-3]. The University of Oxford's Nuffield Department of Primary Care Health Sciences (NDPCHS) became the academic home for the RSC in 2019. As part of Oxford's response to the emerging COVID-19 pandemic, the NDPCHS rapidly scaled up the RSC from its base of 500 practices in 2019 to over 1800 in 2021. What differentiates the RSC from other comparable disease surveillance networks is its integration of serology and virology sampling, and its close links between the RSC and network member practices.

The RSC works in partnership with the UK Health Security Agency (UKHSA), formerly Public Health England. RSC-led outputs include a surveillance report (Weekly Returns), published every week over the last 55 years, which increased to twice-weekly since the start of the pandemic, and an annual report [4]. Both are freely available online. In addition to these key outputs, there are a range of observatories providing contemporary national data; of note are our COVID-19 and mortality observatories. The RSC also provides weekly data for the European Centre for Disease Control (ECDC) on behalf of UKHSA.

RSC contributions to UKHSA intelligence include a large pseudonymized data set that enables vaccine effectiveness (eg, influenza vaccine) to be monitored. Of particular value is the differentiable data that enable subgroups of interest to be studied [5]. RSC data are stored securely in the Oxford-Royal College of General Practitioners Clinical Information Digital Hub (ORCHID) Trusted Research Environment (TRE). The ORCHID TRE offers multiple platforms for approved researchers to make use of. The surveillance platform (ORCHID-S) is the most developed, providing extended primary care surveillance. A trials platform (ORCHID-T) is in development, which will enable trial case identification and follow-up of consented patients, and an emergent epidemiology platform (ORCHID-E) that provides access to contemporary fully anonymized linked health data [6]. The ORCHID TRE facilitates the curation of clinical code sets and digital phenotypes used in computerized medical records (CMR) surveillance and research.

The objective of this study was to report the scale and representativeness of the RSC following a period of substantial growth, demonstrating the network's value for monitoring vaccine uptake and a range of health outcomes, including hospitalization, intensive care unit (ICU) admission, and mortality. Specifically, we describe the representativeness of the RSC population at the end of the 2020-2021 surveillance season (September 30, 2021), its sentinel sampling, and data-linking procedures. This report also includes a description of the clinical informatics that underpins the network and our updated sentinel sampling criteria for the 2021-2022 season.

Methods

Surveillance Overview

The surveillance processes for using ORCHID TRE data were set out in 2020 [6]. Since that time, several additional features have been added: the growth in the size of the network; more frequent linkage to hospital and other data; the harnessing of a new daily data flow to support UKHSA's real-time syndromic surveillance [7]; extending serosurveillance to evaluate COVID-19 immunity, with a particular focus on indications of

waning immunity in vaccine risk groups; and adapting our practice liaison work to support member practices remotely.

Unless stated otherwise, our surveillance year starts on International Standards Organization (ISO) week 40 of 2020 and runs to the end of week 39 of the subsequent year. This approximates to the start of October 2020 to the end of September in the following year.

RSC practices are divided into two subcategories: the Primary Care Sentinel Cohort (PCSC) and syndromic surveillance general practices (SSGPs). There is overlap between these two groups of practices with 264 appearing in both.

PCSC Characteristics

The PCSC is the longest established part of the RSC, with practices providing twice-weekly extracts of pseudonymized routine primary care data to UKHSA. Practice recruitment is nationally representative. A subset of the PCSC collects virology swabs and/or blood samples for virological and serological analysis. Only virological results are reported back to PCSC practices and the patient, although it must be stressed that this is not a diagnostic service. Practices are reimbursed for each virology and serology sample they submit for testing and code accurately. Both virology and serology samples are transported to UKHSA reference laboratories where they are analyzed and stored in dedicated biobanks for onward research.

From March 2022, the virology sampling criteria broadened to align with those of the ECDC. All patients presenting with symptoms consistent with influenza-like-illness (ILI), acute respiratory illness, or COVID-19 are now eligible for virological sampling (see Multimedia Appendix 1 for a visualization of overlapping symptoms that would confer eligibility for swabbing). Prior to this date, only those presenting with ILI, bronchitis or bronchiolitis (in those under 5 years of age), or COVID-19 were eligible. This new schema is included in our formal commissioning letter to our PCSC practices (see Multimedia Appendix 2).

Virology samples are obtained either in practice or within the patient's own home via kits ordered online (supplied by Take A Test UK, an initiative of the nonprofit Saving Lives). This virological sampling is carried out in volunteer practices within the PCSC. Patients are eligible for swabbing if they present within 10 days of symptom onset (except within 14 days of a patient having received their live attenuated influenza vaccine). Samples are processed within the UKHSA's Respiratory Virus Unit in Colindale, London. The UKHSA reference lab conducts an extended panel of tests for the presence of influenza A and B, respiratory syncytial virus A and B, COVID-19, metapneumoviruses, and (additionally this season) seasonal coronaviruses.

Serology samples are obtained opportunistically from volunteer patients attending for routine blood tests at sampling practices in the PCSC. These patients may be having blood tests as part of an acute illness or chronic disease management or prevention, leading to an overrepresentation of risk groups. These samples are primarily used to estimate population exposure to COVID-19; however, serological surveillance has also been successfully applied to monitoring levels of population exposure to influenza and diphtheria. This work also collects data on incidences of rare clotting events postvaccination via platelet factor 4 levels. Current serological tests for COVID-19–related antibodies include spike (S) antibodies (indicative of previous infection or vaccination) and nucleocapsid (N) antibodies (indicative of previous infection). Serology results are linked to RSC primary care records to compare SARS-CoV-2 S and N antibody results to existing patient health data and national seroprevalence studies [5]. Serology samples are managed by UKHSA's Vaccine Evaluation Unit in Manchester.

SSGP Characteristics

UKHSA's national real-time syndromic surveillance service uses an existing combination of National Health Service (NHS) 111 calls and online assessments, ambulance dispatch calls, emergency department attendances, and general practitioner (GP) in- and out-of-hours consultations to monitor and identify trends that may indicate impending public health issues and events that might need intervention. SSGPs (all utilizing Egton Medical Information Systems [EMIS] Health clinical services) supply daily data to UKHSA to supplement and enhance the existing GP in-hours component of syndromic surveillance with a focus on respiratory disease. However, like the RSC's Weekly Return, this will soon supply data on a wider range of diseases of interest [8]. This work is in its development and pilot stages.

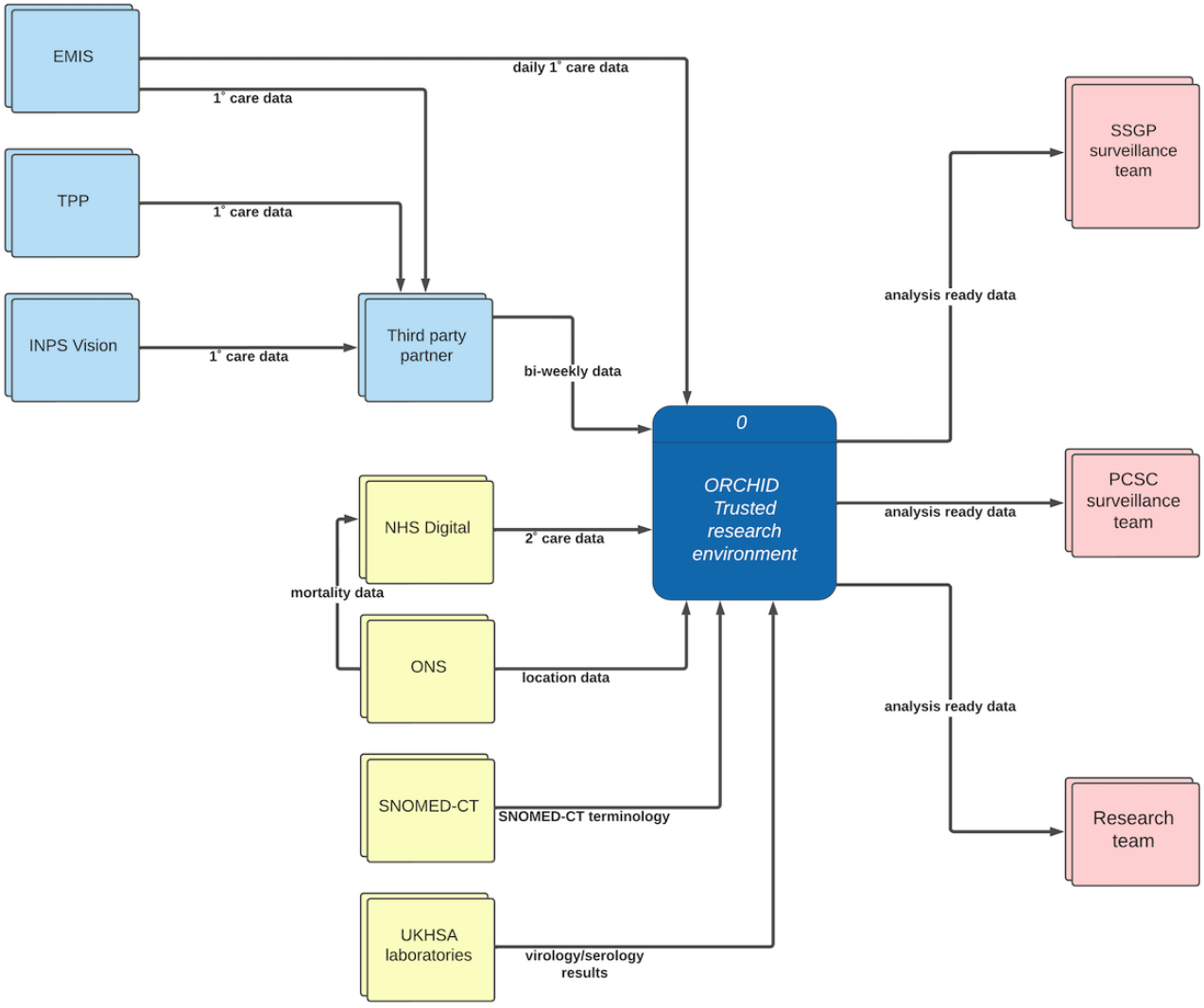
Data Sources

Figure 1 summarizes the health-related data sources used by the RSC and the constituent data flows of the ORCHID TRE.

The ORCHID TRE receives pseudonymized data from general practices (providing access to routine primary care data), NHS Digital (providing access to hospital, Office of National Statistics [ONS] death certificate data, and other national data sets), and UKHSA laboratories (who supply virology and serology results). Limitations on data usage are determined on a user-specific or project-specific basis. However, as set out in our data sharing agreement with practices, linking to other data sets is only permitted when done with the intention of enabling Health Surveillance, Quality Improvement, Research and/or Education (SQUIRE principles). The ORCHID TRE accepts data from any brand of CMR system. We currently receive data from EMIS, The Phoenix Partnership (TPP) System One, and In Practice Systems (INPS)-Vision. These data linkages are listed in full within Multimedia Appendix 3.

PCSC CMRs required for surveillance purposes are extracted twice weekly using a third-party company on behalf of the RSC. This process is managed by formal data sharing and service level agreements. Patient CMRs are pseudonymized using a nonreversible "hash" algorithm as close to the source as possible. Meanwhile, as stated, SSGP CMRs support UKHSA's syndromic surveillance service by providing a direct feed of in-hours GP data to augment out-of-hours GP data, NHS 111, ambulance, and emergency department attendance data. Patients who decline to share their data are excluded from either extraction process.

Figure 1. Data flow diagram for Oxford-Royal College of General Practitioners Clinical Information Digital Hub (ORCHID) Trusted Research Environment (TRE). Pale blue boxes represent the principal primary care data sources, pale yellow represents secondary data sources, and pale red represents data outputs. The dark blue box represents the data processing within the ORCHID TRE. Egton Medical Information Systems (EMIS) Health and The Phoenix Partnership (TPP) are primary care software services and act as data processors on behalf of National Health Service (NHS) primary care providers. INPS: In Practice Systems; ONS: Office for National Statistics; PCSC: Primary Care Sentinel Cohort; SNOMED-CT: Systematized Nomenclature of Medicine Clinical Terms; SSGP: syndromic surveillance general practices; UKHSA: UK Health Security Agency.



Scope of Data Collection

Data Capture

Data are captured at an individual pseudonymized level and patients' general practice, which links a record to that practice's relevant Lower Super Output Area (LSOA) level (minimum population of 1000, mean population of 1500), NHS administrative area, and NHS Region.

Sociodemographic Data

Sociodemographic data include age, gender, ethnicity, socioeconomic status (SES), NHS Region, rurality, smoking status, and obesity. Ethnicity is grouped into five categories: Asian, Black, white, mixed, and other. An ontology is used to maximize identification. SES is measured using the Index of Multiple Deprivation, a metric that is revealed by the LSOA level. NHS Region is defined using NHS Region mapping and is divided into seven areas: East of England, London, Midlands, North East and Yorkshire, North West, South East, and South West. Rurality is measured using ONS measures of population density, and divided into rural, town and city, and conurbation categories. Smoking status is categorized into nonsmoker, current smoker, and exsmoker. Obesity is categorized by BMI intervals (kg/m^2), whereby <18.5 denotes underweight, 18.5-24.9 denotes normal weight, 25-29.9 denotes overweight, 30-34.9 denotes obese (obesity class I), 35-39.9 denotes obese class II, and 40 or above denotes morbid obesity (obesity class III).

Vaccine Uptake Data

The RSC provides data on vaccination uptake to its partners. Vaccine uptake data for COVID-19 are derived from RSC's linkage to the National Immunisation Management Service (NIMS) [9]. Remaining vaccination data come directly from the primary care record.

Health Outcomes

For respiratory infectious disease surveillance and research, key health outcomes of interest are medically attended ILI/COVID-19 (qualified by admission to hospital, admission to ICU, and death). Relevant data are provided within CMRs or through linkages made with hospital data sets (eg, Health Episode Statistics) and ONS death registries via NHS Digital. These outcomes can be observed at a disease-specific level, and can be linked to reveal the entire patient journey and the variables that can predict excess risk.

Comorbidities and Other Variables

We provide a proxy variable to indicate consultation frequency and attendance. We also utilize the electronic frailty index for patients aged over 65 years and the Cambridge multimorbidity score within our data [10-12] (R Tsang, unpublished data, October 2022). Where possible, we provide a measure of household size to control for household bias when monitoring disease spread. This is done by applying a "household key" to RSC pseudonymized records. Here, groups of individuals are identified as living in a common address by flagging where records' first line of an address and postcodes match. This matching is done at the point of data extraction from the GP system so that personal data are never revealed. This unique tool has been used in household transmission studies of acute gastroenteritis, influenza, and acute respiratory illnesses [13,14].

Data Quality

The data quality within ORCHID TRE is underpinned by practice engagement, data capture, cleaning, aggregation, and analysis. The Practice Liaison Officers provide support and training, including on-site visits, for member practices. Other activities include personalized training, supply of sampling materials, webinars, and patient information, including information for patient participation groups. Practices are invited to access dedicated dashboards to monitor their sampling performance. We also publish overarching observatories with the goal of improving data quality among RSC members and provide specific support in data codification under our new *"Coding is caring"* initiative [15-17]. We have also started engaging with practice patient participation groups.

Our ontological mapping process recognizes that clinical concepts can be represented differently within a clinical terminology. To enable consistent and machine-readable identification of key outcomes and other variables [18], we use Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) as the principal terminology to develop code sets [19]. A curated code set is a list of relevant clinical codes that best represent a specific clinical idea. All code sets are stored in our variable library. Hospital and death data are primarily coded using the International Classification of Disease version 10, with procedures recorded using the Office of Population Census and Surveys classification version 5.1 [20,21]. Treatments are codified in line with the descriptions and codes utilized by the Dictionary of Medicines and Devices [22]. All health conditions included in PCSC and SSGP surveillance, and their respective SNOMED-CT codes and our variable library numbers, are provided in Multimedia Appendix 4.

Data Analysis

We accessed the secure ORCHID TRE using R version 4.2.0 to undertake all analyses [23]. We aggregated patient- and practice-level data to summarize characteristics of the network on October 8, 2021, including the number of practices that had agreed to share data, the number of participants actively supplying data, and key demographic variables. To establish network growth over the pandemic, we compared these data with historic records. Vaccine uptake figures were generated by aggregating primary care records linked to NIMS data. Hospitalization and ICU figures were created by aggregating primary care records linked to Hospital Episode Statistics data and mortality figures were generated using primary care data linked to ONS mortality figures.

Ethical Considerations

We work within relevant legislation and research and information governance frameworks and are fully compliant with the University of Oxford's ethical standards. The University is registered on the Information Commissioner's Office Data Protection Register and is compliant with the Data Protection Act, General Data Protection Regulation, and other key data privacy and protection legislations. As required by NHS Data Security Standard 3 in the Caldicott 3 Review, all research members of NDPCHS are required to complete Data Security Awareness modules on an annual basis.

The legal basis for RSC surveillance is Regulation 3 (health protection) of the Health Service (Control of Patient Information) Regulations 2002, with some of our work with UKHSA falling under Regulation 5 (Health Promotion) [24,25]. Other nonsurveillance studies that use ORCHID TRE data require appropriate ethical approval. For low-risk epidemiological studies, this is through Oxford University Medical Sciences Interdivisional Research and Ethics Committee, whereas for trials or other prospective studies involving contact with patients, it is through the Integrated Research Approval Service [26]. All nonsurveillance studies also must be approved by the independent Primary Care Hosted Research Datasets Independent Scientific Committee. The RSC also meets NHS Digital's stringent Data Security and Protection toolkit requirements. Finally, RSC activities are restricted to conform with the data sharing agreements with member practices, who, as stated, share data for SQUIRE purposes [6].

Results

Network Growth

The number of practices within the entire network prior to the emergence of SARS-CoV-2 (October 2019) was approximately 500 (all PCSC). The RSC network has grown substantially since that time. At the start of the reporting period (October 2020), however, the network included 1764 practices (879 PCSC practices, 1100 SSGPs); at its end (October 2021), it included 1879 practices (930 PCSC practices, 1203 SSGPs). By October 2021, there was an overlap of 262 practices supplying data to both groups. Table 1 shows the number of practices listed and

samples collected as of ISO week 39 of 2021 and those recorded nationally (ONS data).

Figure 2 presents a population pyramid of the age-sex profile of RSC data compared with the ONS report of the English national population's age-sex profile. There was a higher proportion of younger working-age adults, aged 25 to 40 years, in the RSC population than the ONS standard. We describe the sociodemographic characteristics of the RSC compared to national population data in Table 2. This revealed higher levels of nonwhite ethnicity and active smoking among the RSC membership, and lower levels of females and population in the Eastern region in the RSC.

Table 1. Summary of Research and Surveillance Centre (RSC) practice and population sizes compared to national Office for National Statistics (ONS) data.

Data type	All RSC practices	PCSC[a]	SSGP[b]	National data (ONS)
General practices, n (%)[c]	1879 (28.63)	938 (14.29)	1203 (18.33)	6563 (100)
Registered list size, n (%)[c]	17,299,780 (30.59)	8,414,204 (14.88)	12,356,618 (21.9%)	56,550,136 (100)
Virology sampling practices, n	245	245	—[d]	—
Virology specimens, n	8049	8049	—	—
Serology sampling practices, n	220	220	—	—
Serology specimens, n	23,879	23,879	—	—

[a]PCSC: Primary Care Sentinel Cohort.

[b]SSGP: syndromic surveillance general practice.

[c]Percentage of national data (final column).

[d]Not applicable.

Figure 2. Age-sex pyramid of the Research and Surveillance Centre population on October 2021 compared to Office for National Statistics (ONS) estimates for 2019. F: female; M: male; ORCHID: Oxford-Royal College of General Practitioners Clinical Informatics Digital Hub.

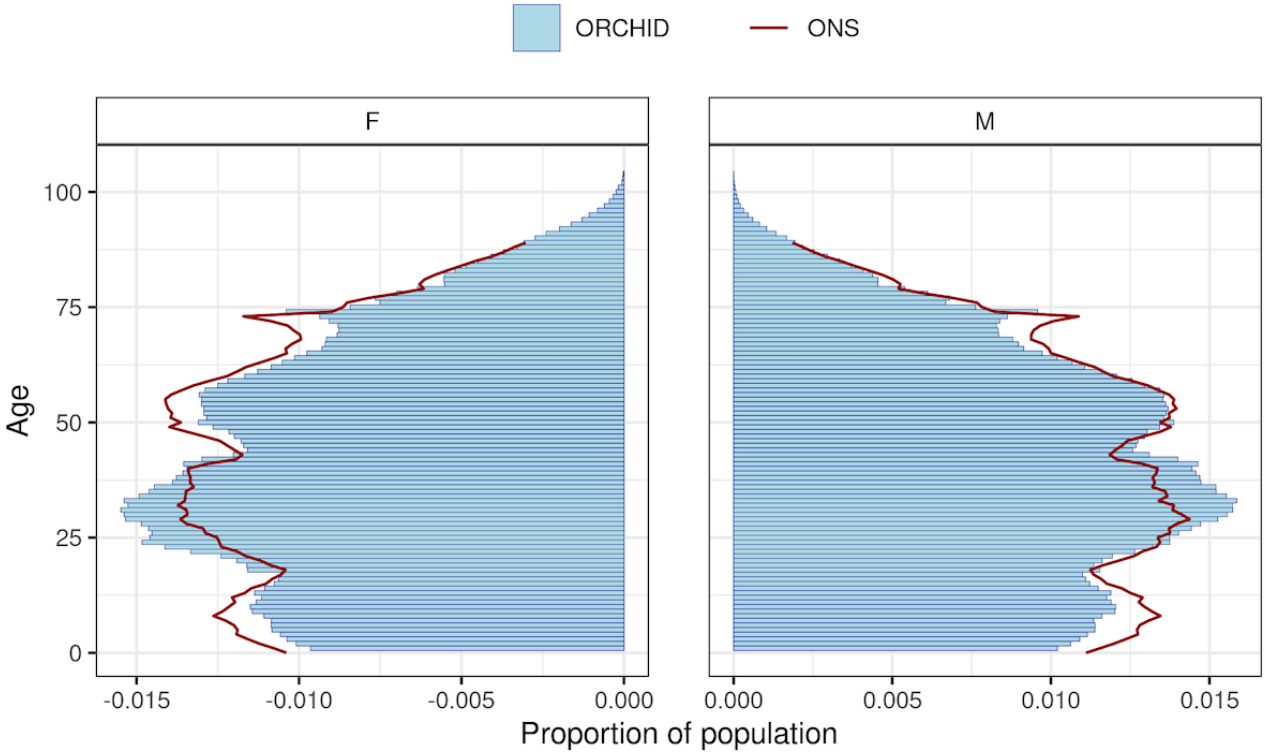


Table 2. Demographic characteristics of Research and Surveillance Centre (RSC) compared to Office of National Statistics (ONS) benchmark data.

Characteristics	All RSC surveillance practices[a,b]	Primary Care Sentinel Cohort[b]	Syndromic surveillance general practices[b]	English national data (ONS)
Age (years), median (IQR)[c]	39 (21-58)	39 (22-58)	38 (21-57)	40 (21-59)
Female, n (%)[c]	8,640,025 (49.9)	4,206,592 (50)	6,164,188 (49.9)	28,567,320 (50.5)
Ethnicity, n (%)[c]				
White	11,347,155 (81.5)	5,585,337 (83.3)	8,108,670 (81)	47,417,500 (84.2)
Asian	1,390,875 (10)	622,009 (9.3)	1,022,726 (10.2)	4,661,000 (8.3)
Black	611,387 (4.4)	243,921 (3.6)	460,248 (4.6)	2,066,100 (3.7)
Mixed	310,100 (2.2)	142,891 (2.1)	224,780 (2.2)	1,086,600 (1.9)
Other	266,960 (1.9)	113,805 (1.7)	198,491 (2)	1,055,800 (1.9)
NHS[d] Region, n (%)[c]				
South East	3,553,593 (20.5)	1,567,376 (18.6)	2,829,726 (22.9)	8,933,822 (15.8)
London	3,289,770 (19)	1,191,081 (14.2)	2,552,695 (20.7)	9,002,488 (15.9)
Midlands	2,791,986 (16.1)	1,171,549 (13.9)	2,076,374 (16.8)	10,658,558 (18.8)
North West	2,741,232 (15.8)	1,330,266 (15.8)	2,126,805 (17.2)	7,087,447 (12.5)
South West	2,124,626 (12.3)	1,523,323 (18.1)	1,164,724 (9.4)	5,665,799 (10)
North East Yorkshire	1,720,689 (9.9)	1,022,070 (12.1)	1,074,612 (8.7)	8,639,006 (15.3)
East of England	1,077,884 (6.2)	608,539 (7.2)	531,682 (4.3)	6,563,018 (11.6)
Index of multiple deprivation (IMD), n (%)[c]				
IMD1 (most deprived)	3,366,018 (19.5)	1,508,104 (17.9)	2,487,902 (20.1)	11,267,059 (20)
IMD2	3,483,363 (20.1)	1,648,030 (19.6)	2,480,510 (20.1)	11,576,973 (20.6)
IMD3	3,386,739 (19.6)	1,664,973 (19.8)	2,363,689 (19.1)	11,424,153 (20.3)
IMD4	3,408,200 (19.7)	1,764,757 (21)	2,398,673 (19.4)	11,117,694 (19.8)
IMD5 (least deprived)	3,652,825 (21.1)	1,826,114 (21.7)	2,625,435 (21.2)	10,901,082 (19.4)
Morbid obesity, n (%)[e]	434,514 (2.5)	208,995 (2.5)	312,654 (2.5)	(3.2)[f]
Smoking, n (%)[g]				
Active	2,304,974 (17.1)	1,106,166 (16.8)	1,659,899 (17.2)	4,897,952 (12.1)
Ex	3,236,336 (23.9)	1,608,336 (24.4)	2,297,591 (23.8)	10,645,963 (26.3)
Never	7,974,524 (59)	3,869,419 (58.8)	5,707,428 (59.1)	24,935,033 (61.6)

[a]Data represent a cross-sectional view of the state of the RSC network on October 8, 2021.

[b]Represent the percentage of nonmissing data.

[c]ONS data based on 2019 estimates.

[d]NHS: National Health Service.

[e]ONS data based on 2020 estimates of morbid obesity in those aged over 16 years.

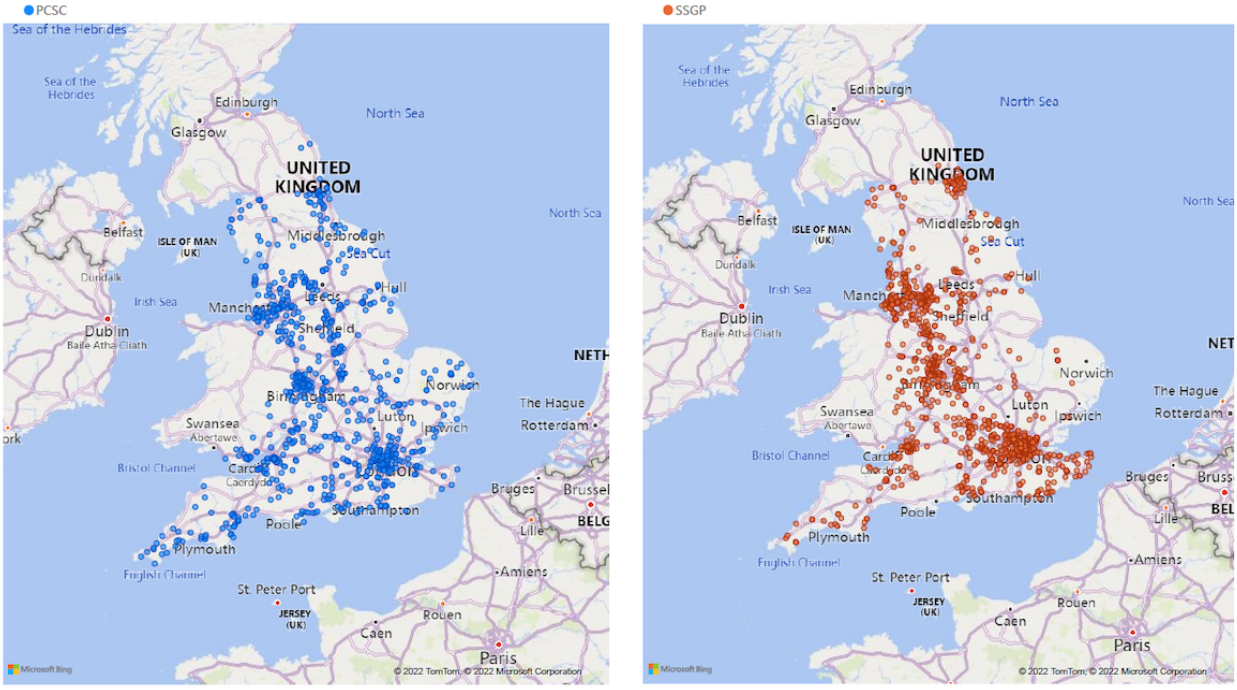
[f]Data based on a sample; therefore, only the percentage is provided.

[g]ONS data based on 2020 estimates of smoking status in those aged over 18 years.

Geographical Profile

The maps in Figure 3 demonstrate the national distribution of PCSC practices compared with SSGPs; note that these graphics omit the 264-practice overlap that exists between these two subcategories. The PCSC is recruited to be nationally represented, although has lower representation in the Eastern region, and the SSGPs are recruited from and follow the national distribution of the EMIS brand of the CMR system.

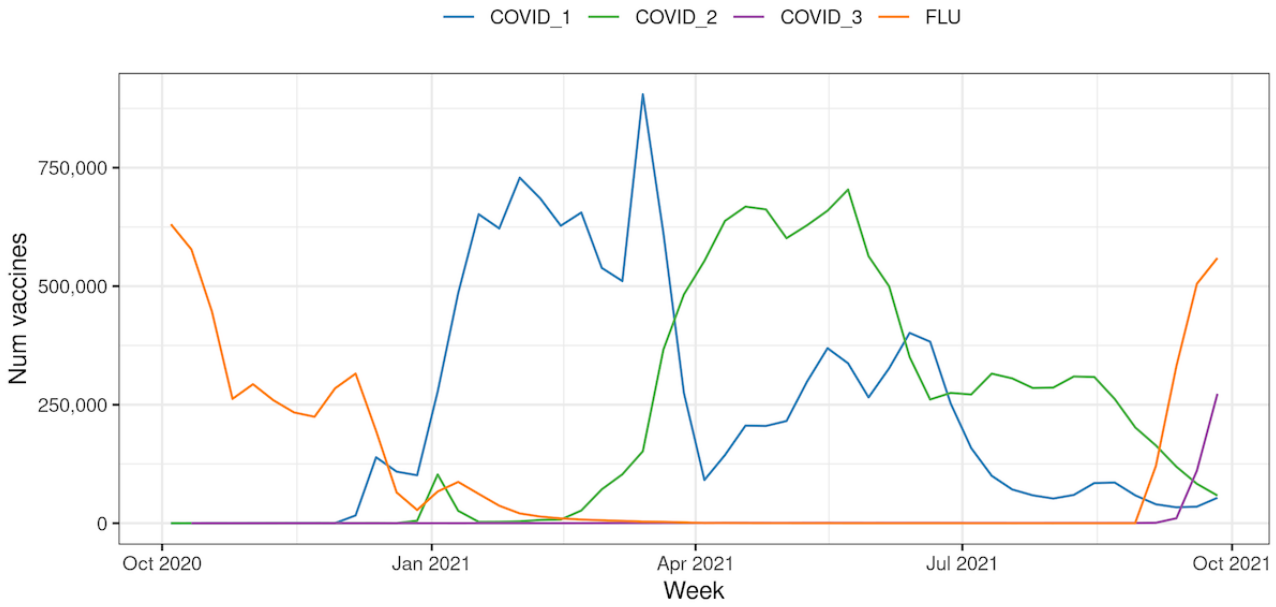
Figure 3. Geographical distribution of Primary Care Sentinel Cohort (PCSC) practices (left panel) and syndromic surveillance general practice (SSGP) practices (right panel).



Vaccine Uptake

RSC data provide a profile of national vaccine uptake; RSC uptake data can be differentiated by vaccine type, brand, and batch in addition to the demographics of those who have been vaccinated (Figure 4 and Multimedia Appendix 5). Across the RSC, there were 11,897,180 first, 10,992,049 second, and 397,986 third/booster doses of COVID-19 vaccine (a total of 23,287,215), and 5,387,169 doses of influenza vaccine administrations recorded between October 2020 and the end of September 2021. The brands of vaccine administered over this period were AstraZeneca (n=11,632,841), Pfizer (n=10,849,114), and Moderna (n=694,696).

Figure 4. Absolute weekly vaccine uptake for seasonal influenza and COVID-19 doses over time. Data represent the entire Research and Surveillance Centre network population.



Health Outcomes

Figure 5 reports the COVID-19–specific hospitalization and hospital occupancy; Figure 6 reports the death data, separating hospital and community mortality; and Multimedia Appendix 6 summarizes the ICU admission and ICU occupancy rates. Over this period, COVID-19 admissions varied from 104 to 6835 per week (median 1181, IQR 559-1559). Bed occupancy

over this period varied from 557 to 19,110 per week (median 3284, IQR 1512-7443). For the ICU, the equivalent data were admissions varying from 6 to 842 per week (median 115, IQR 50-174), with ICU occupancy varying from 103 to 2451 per week (median 504.5, IQR 213.5-813.0).

Figure 5. COVID-19 hospitalization data in the RSC network population, calculated using RSC data linked to Hospital Episode Statistics (HES) data. Left panel: Number of hospitalisations per week due to COVID-19. Right panel: Weekly hospital bed occupancy. An admission was defined as spending at least 1 night in hospital.

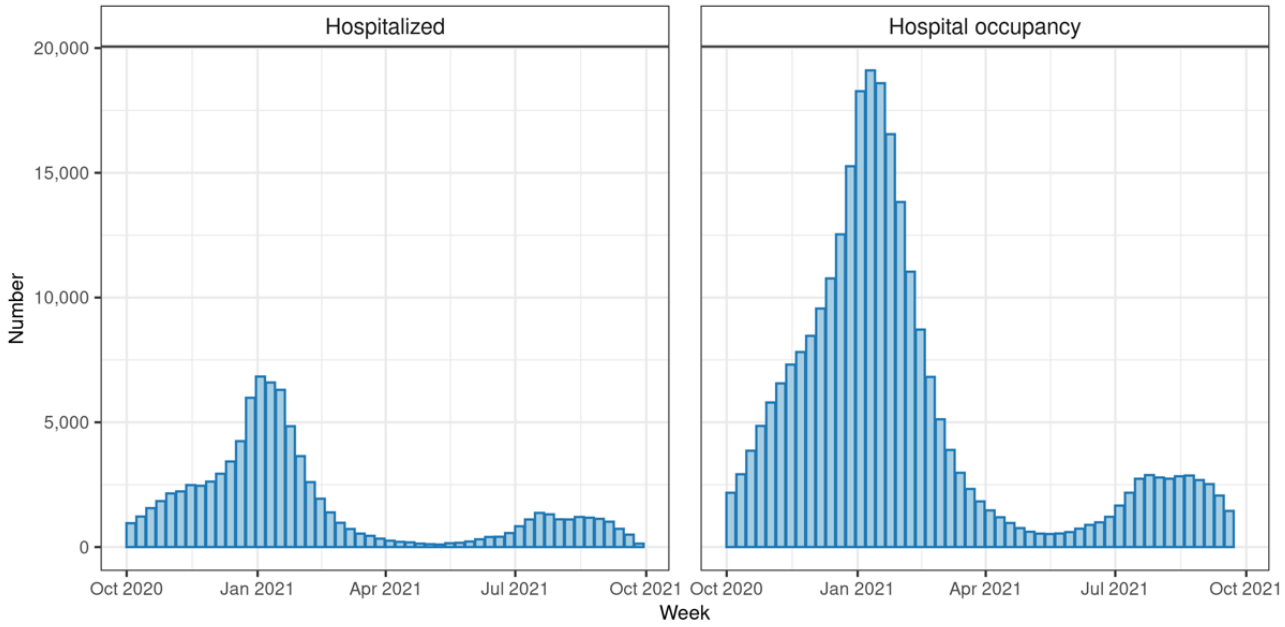
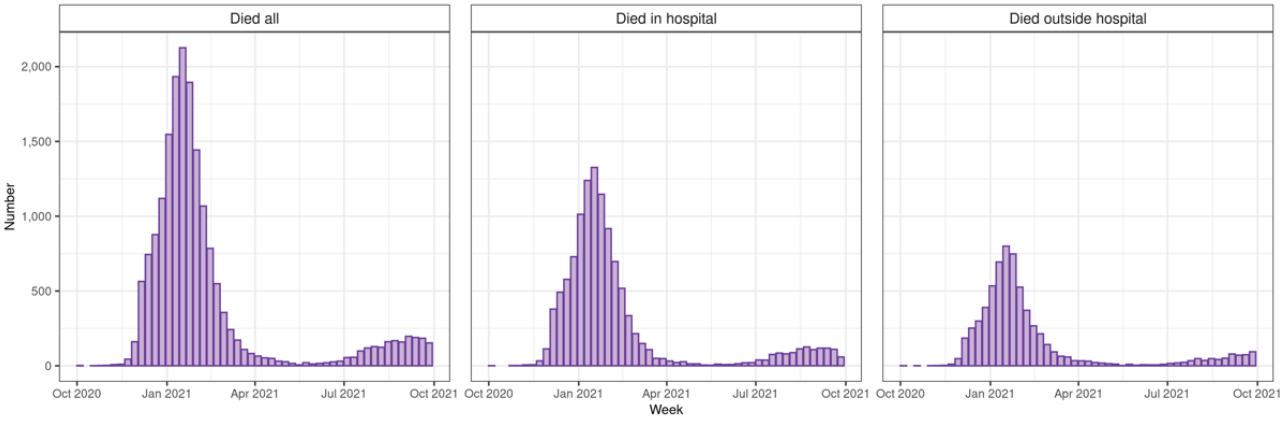


Figure 6. COVID-19 death data in the Research and Surveillance Centre (RSC) network population, calculated using RSC data linked to Office of National Statistics death data and COVID-19 Hospitalizations in England Surveillance System data. Most deaths occurring between October and December 2020 are not shown due to missing data. Left panel: All deaths. Middle panel: In-hospital deaths. Right panel: Out of hospital deaths.



Practice Visits and Patient Participation

Restrictions due to COVID-19 have limited our ability to conduct visits to practices. However, over this period, there were 25 virtual and 5 in-person visits. There were monthly newsletters, 5 training webinars, and the completion of 483 practice material requests to sampling practices. We have updated and provided practices with bespoke dashboards about influenza vaccination, their data quality, and—for sampling practices—numbers of virology swabs completed and serology samples taken. We provided practice members with new COVID-19 case and mortality observatories. We have commenced a pilot of direct engagement with five patient participation groups of sampling practices to test whether giving more direct health and surveillance information to patients and the public will enhance sampling. Eight practices left the network during the 2020-2021 season.

Discussion

Main Findings

This summary cohort profile demonstrates how the RSC provides contemporaneous data about disease patterns, emerging illnesses, circulating viral infections and their variants, population immunity, and vaccine coverage. The RSC has changed and broadened its virology sampling criteria during the specified reporting period.

The RSC has also grown significantly in recent years by adding SSGPs to its PCSC; historically, the RSC processed data from

approximately 1 million patients and we now have data from 17.3 million registered people, including nearly 12 million people who have had at least one COVID-19 vaccine dose. We are now starting a pilot of direct engagement with patients and the public. The innovation to informatics over the reporting period included consistently curated code sets of variables, improved data management (with more data linkages, including to hospital and death data), and improved data availability for subsequent analyses.

Implications of the Findings

The capability of the RSC has grown, embedding serology sampling alongside virology. However, there is more to be done in this area, particularly as national virological testing for COVID-19 is decommissioned and UKHSA may increasingly look toward sentinel sampling to compensate for this loss. We hope our broadened sampling criteria will result in the collection of a larger number of virology samples, as will building out our direct contacts with patients and the public.

Comparison With the Literature

RSC surveillance is comparable to that conducted internationally, and similar approaches to monitoring disease have been used in developed health systems [27]. These include the Canadian Primary Care Sentinel Surveillance Network, the US Influenza-like Illness Surveillance Program, the National Respiratory and Enteric Virus Surveillance System (also US-based), and the *Sentinelles* network in France [28-31].

Strengths and Limitations

The functions and outputs of the RSC are aligned with the priorities and ambitions specified by UKHSA in their Information and Intelligence Working Group "Public Health Surveillance: Towards a Public Health Surveillance Strategy for England" [32]. Effective surveillance is categorized via its ongoing nature, its timeliness, and the measures of population or group health status it provides against historical or geographical baselines. All of these are observable within the RSC results reported here. The wider objectives of a robust surveillance system include:

- Monitoring changes in infectious agents
- Providing early warning of seasonal disease activity and future emerging threats through an enhanced national GP syndromic surveillance system
- Identifying high-risk populations or areas to target interventions
- Evaluating the effectiveness of preventative health and health control measures
- Supporting health planning and the allocation of appropriate resources within the health care system
- Providing an archive of disease activity (or biological samples) for future reference and research.

The quality of surveillance is contingent on the extent to which its findings are generalizable to its underlying population. ORCHID TRE data now approximate to national data for age, gender, ethnicity, SES, rurality, smoking status, and obesity. This enables nuanced analyses for how certain demographics and individuals with protected characteristics are at higher risk for experiencing health inequalities and adverse outcomes. There remains scope to improve the representativeness of the RSC. The practices that volunteer are typically larger than average, are not equally distributed between regions, and are, overall, from slightly less deprived areas. There may also be other undetected forms of bias in our membership. We need to recruit more practices in the eastern region and target recruitment into the PCSC to geographically balance membership. However, pressures on primary care make recruitment and retention of practices challenging [33,34].

The UKHSA syndromic surveillance service also utilizes anonymized CMR data from an external feed of TPP primary care data; we will need to work closely with these providers to ensure overall national representativeness of this offering and the standardization of coding underpinning surveillance indicators. Signals that emerge through syndromic surveillance data can be validated against wider ORCHID TRE data and its opportunities for broader data linkage. Integrating SSGP data into the existing UKHSA syndromic surveillance program will also increase the application of ORCHID TRE data in a multihazard public health response, including surveillance of nonrespiratory infectious diseases (eg, gastrointestinal pathogens), environmental impacts (eg, heat waves), chemical incidents (eg, health impacts of industrial fires), and mass gatherings (eg, 2022 Birmingham Commonwealth Games).

We are aware of attendance bias in those who provide the RSC with samples. In virology sampling, it is well-established that some families and population groups attend more frequently with respiratory illnesses than is nationally representative [35,36]. Serology sampling is also seen to be biased toward those attending more regularly for blood tests, especially older people with chronic conditions [37].

Conclusions

This cohort profile describes the RSC's capabilities for conducting disease surveillance and vaccine effectiveness studies, and provides a guide to network components and capabilities. While there remains scope for improvement, the RSC is now stronger and larger than at any time in its 55-year history, particularly in terms of sampling performance. Its 2020-2021 end of surveillance year registered population was 17.3 million, accounting for 31% of the English national population. This population was seen to collectively receive over 23 million COVID-19 vaccination doses. We move through the 2021 to 2022 season with revised sampling criteria and, for the first time, daily data contributions to the UKHSA real-time syndromic surveillance service.

Acknowledgments

We are grateful to the general practices and patients who agree to share data with the RSC, and to EMIS, TPP, INPS, and Wellbeing for facilitating pseudonymized data extracts. FDRH acknowledges partial support as Director of the National Institute for Health Research (NIHR) Applied Research Collaboration Oxford Thames Valley, and Theme Lead of the NIHR Oxford Biomedical Research Centre. ML is a doctoral student at the Nuffield Department of Primary Care Health Sciences. Her research is funded via the Industrial CASE Studentship award, a Doctoral Training Partnership supported by both the Medical Research Council and EMIS Health. WE is a doctoral fellow at Nuffield Department of Primary Care. UKHSA is the principal funder of the RSC, and thus no specific funding was received for this work. SdL is also funded by Wellcome grant 212763/Z/18/Z. AJE and GS are affiliated with the NIHR Health Protection Research Unit (HPRU) in Gastrointestinal Infections at the University of Liverpool and the NIHR HPRU in Emergency Preparedness and Response at King's College London. The views expressed are those of the author(s) and not necessarily those of the NIHR, UK Health Security Agency, or the Department of Health and Social Care.

Conflicts of Interest

As specified above, ML's doctorate is partly funded by EMIS Health. MZ is a member of SAGE/NERVTAG/JCVI working/expert groups (unpaid positions) and chair of the charitable organization ISIRV (unpaid position). The Vaccine Evaluation Unit carries out contract research on behalf of UKHSA for GSK, Pfizer, and Sanofi. This is not directly related to the work in this publication. The Immunisation Department provides vaccine manufacturers (including Pfizer) with postmarketing surveillance reports about pneumococcal and meningococcal disease, which the companies are required to submit to the UK Licensing Authority in compliance with their Risk Management Strategy. A cost recovery charge is made for these reports. FDRH has also received occasional fees or expenses for speaking or consultancy on cardiovascular disease from AstraZeneca, BI, Bayer, BMS/Pfizer, and Novartis. SdL is the director of the RCGP-RSC; he has received vaccine-related research funding from AstraZeneca, GSK, Sanofi, Seqirus, and Takeda, and has been a member of advisory boards for AstraZeneca, Sanofi, and Seqirus. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Research and Surveillance Centre (RSC) Primary Care Sentinel Cohort (PCSC) Virology sampling criteria. ARI: acute respiratory illness; ILI: influenza-like illness; LRTI: lower respiratory tract infection; URTI: upper respiratory tract infection.
[DOCX File , 73 KB - publichealth_v8i12e39141_app1.docx]

Multimedia Appendix 2
Commissioning letter to Primary Care Sentinel Cohort (PCSC) practices.
[PDF File (Adobe PDF File), 2015 KB - publichealth_v8i12e39141_app2.pdf]

Multimedia Appendix 3
Oxford-Royal College of General Practitioners Clinical Information Digital Hub (ORCHID) linkage to other data sets.
[DOCX File , 34 KB - publichealth_v8i12e39141_app3.docx]

Multimedia Appendix 4
Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) code sets used for surveillance in the Oxford-Royal College of General Practitioners Clinical Information Digital Hub (ORCHID) Trusted Research Environment (TRE) variable library.
[DOCX File , 18 KB - publichealth_v8i12e39141_app4.docx]

Multimedia Appendix 5
Absolute weekly COVID-19 vaccine uptake over time differentiated by vaccine type.
[DOCX File , 124 KB - publichealth_v8i12e39141_app5.docx]

Multimedia Appendix 6
COVID-19 intensive care unit (ICU) data in the Research and Surveillance Centre (RSC) network population.
[DOCX File , 130 KB - publichealth_v8i12e39141_app6.docx]

References

1. de Lusignan S, Correa A, Smith GE, Yonova I, Pebody R, Ferreira F, et al. RCGP Research and Surveillance Centre: 50 years' surveillance of influenza, infections, and respiratory conditions. Br J Gen Pract 2017 Oct;67(663):440-441 [FREE Full text] [doi: 10.3399/bjgp17X692645] [Medline: 28963401]

2. Elliot AJ, Fleming DM. Surveillance of influenza-like illness in England and Wales during 1966-2006. Euro Surveill 2006;11(10):249-250. [Medline: 17130657]
3. Fleming DM, Elliot AJ. Lessons from 40 years' surveillance of influenza in England and Wales. Epidemiol Infect 2008 Jul;136(7):866-875. [doi: 10.1017/S0950268807009910] [Medline: 18047750]
4. de Lusignan S, Tsang RSM, Amirthalingam G, Akinyemi O, Sherlock J, Tripathy M, et al. Adverse events of interest following influenza vaccination, a comparison of cell culture-based with egg-based alternatives: English sentinel network annual report paper 2019/20. Lancet Reg Health Eur 2021 Mar;2:100029 [FREE Full text] [doi: 10.1016/j.lanepe.2021.100029] [Medline: 34557791]
5. Whitaker HJ, Tsang RSM, Byford R, Andrews NJ, Sherlock J, Sebastian Pillai P, et al. Pfizer-BioNTech and Oxford AstraZeneca COVID-19 vaccine effectiveness and immune response amongst individuals in clinical risk groups. J Infect 2022 May;84(5):675-683 [FREE Full text] [doi: 10.1016/j.jinf.2021.12.044] [Medline: 34990709]
6. de Lusignan S, Jones N, Dorward J, Byford R, Liyanage H, Briggs J, et al. The Oxford Royal College of General Practitioners Clinical Informatics Digital Hub: protocol to develop extended COVID-19 surveillance and trial platforms. JMIR Public Health Surveill 2020 Jul 02;6(3):e19773 [FREE Full text] [doi: 10.2196/19773] [Medline: 32484782]
7. Smith GE, Elliot AJ, Lake I, Edeghere O, Morbey R, Catchpole M, Public Health England Real-time Syndromic Surveillance Team. Syndromic surveillance: two decades experience of sustainable systems - its people not just data!. Epidemiol Infect 2019 Jan;147:e101 [FREE Full text] [doi: 10.1017/S0950268819000074] [Medline: 30869042]
8. Smith S, Morbey R, de Lusignan S, Pebody RG, Smith GE, Elliot AJ. Investigating regional variation of respiratory infections in a general practice syndromic surveillance system. J Public Health 2021 Jun 07;43(2):e153-e160. [doi: 10.1093/pubmed/fdaa014] [Medline: 32009178]
9. NHS Immunisation Management Service. National Health Service. URL: https://www.scwcsu.nhs.uk/services/nhs-immunisation-management-service/ [accessed 2022-03-14]
10. Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age Ageing 2016 May;45(3):353-360 [FREE Full text] [doi: 10.1093/ageing/afw039] [Medline: 26944937]
11. Fogg C, Fraser SDS, Roderick P, de Lusignan S, Clegg A, Brailsford S, Frailty Dynamics study team. The dynamics of frailty development and progression in older adults in primary care in England (2006-2017): a retrospective cohort profile. BMC Geriatr 2022 Jan 06;22(1):30 [FREE Full text] [doi: 10.1186/s12877-021-02684-y] [Medline: 34991479]
12. Payne RA, Mendonca SC, Elliott MN, Saunders CL, Edwards DA, Marshall M, et al. Development and validation of the Cambridge Multimorbidity Score. CMAJ 2020 Feb 03;192(5):E107-E114 [FREE Full text] [doi: 10.1503/cmaj.190757] [Medline: 32015079]
13. de Lusignan S, Konstantara E, Joy M, Sherlock J, Hoang U, Coyle R, et al. Incidence of household transmission of acute gastroenteritis (AGE) in a primary care sentinel network (1992-2017): cross-sectional and retrospective cohort study protocol. BMJ Open 2018 Aug 23;8(8):e022524 [FREE Full text] [doi: 10.1136/bmjopen-2018-022524] [Medline: 30139907]
14. Joy M, Hobbs FR, Bernal JL, Sherlock J, Amirthalingam G, McGagh D, et al. Excess mortality in the first COVID pandemic peak: cross-sectional analyses of the impact of age, sex, ethnicity, household size, and long-term conditions in people of known SARS-CoV-2 status in England. Br J Gen Pract 2020 Dec;70(701):e890-e898 [FREE Full text] [doi: 10.3399/bjgp20X713393] [Medline: 33077508]
15. Mayor N, Tsang R, Joy M, Hobbs FR, de Lusignan S. Long covid: coding is caring. BMJ 2021 May 19;373:n1262. [doi: 10.1136/bmj.n1262] [Medline: 34011495]
16. Liyanage H, Akinyemi O, Pathirannahelage S, Joy M, de Lusignan S. Near real time feedback of seasonal influenza vaccination and virological sampling: dashboard utilisation in a primary care sentinel network. Stud Health Technol Inform 2020 Jun 16;270:1339-1340. [doi: 10.3233/SHTI200431] [Medline: 32570648]
17. Pathirannehelage S, Kumarapeli P, Byford R, Yonova I, Ferreira F, de Lusignan S. Uptake of a dashboard designed to give realtime feedback to a sentinel network about key data required for influenza vaccine effectiveness studies. Stud Health Technol Inform 2018;247:161-165. [Medline: 29677943]
18. de Lusignan S. In this issue: Ontologies a key concept in informatics and key for open definitions of cases, exposures, and outcome measures. J Innov Health Inform 2015 Jul 10;22(2):170 [FREE Full text] [doi: 10.14236/jhi.v22i2.170] [Medline: 26245238]
19. Chang E, Mostafa J. The use of SNOMED CT, 2013-2020: a literature review. J Am Med Inform Assoc 2021 Aug 13;28(9):2017-2026 [FREE Full text] [doi: 10.1093/jamia/ocab084] [Medline: 34151978]
20. OPCS Classification of Interventions and Procedures. National Health Service. URL: https://www.datadictionary.nhs.uk/supporting_information/opcs_classification_of_interventions_and_procedures.html [accessed 2022-03-14]
21. International Statistical Classification of Diseases and Related Health Problems (ICD). World Health Organization. URL: https://www.who.int/standards/classifications/classification-of-diseases [accessed 2022-03-14]
22. Dictionary of medicines and devices (dm+d). NHS Business Services Authority. URL: https://www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/dictionary-medicines-and-devices-dmd [accessed 2022-03-14]
23. R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.R-project.org/ [accessed 2022-11-24]

24. Control of patient information (COPI) notice. National Health Service. URL: https://digital.nhs.uk/coronavirus/coronavirus-covid-19-response-information-governance-hub/control-of-patient-information-copi-notice [accessed 2022-03-14]
25. Taylor MJ. Legal bases for disclosing confidential patient information for public health: distinguishing between health protection and health improvement. Med Law Rev 2015 May 20;23(3):348-374 [FREE Full text] [doi: 10.1093/medlaw/fwv018] [Medline: 25995294]
26. Integrated Research Application System. NHS Health Research Authority. 2021. URL: https://www.hra.nhs.uk/about-us/committees-and-services/integrated-research-application-system/ [accessed 2022-03-14]
27. Liaw S, Kuziemsky C, Schreiber R, Jonnagaddala J, Liyanage H, Chittalia A, et al. Primary care informatics response to Covid-19 pandemic: adaptation, progress, and lessons from four countries with high ICT development. Yearb Med Inform 2021 Aug 21;30(1):44-55 [FREE Full text] [doi: 10.1055/s-0041-1726489] [Medline: 33882603]
28. Garies S, Birtwhistle R, Drummond N, Queenan J, Williamson T. Data Resource Profile: National electronic medical record data from the Canadian Primary Care Sentinel Surveillance Network (CPCSSN). Int J Epidemiol 2017 Aug 01;46(4):1091-1092f. [doi: 10.1093/ije/dyw248] [Medline: 28338877]
29. The National Respiratory and Enteric Virus Surveillance System (NREVSS). US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/surveillance/nrevss/index.html [accessed 2022-03-14]
30. Turbelin C, Boëlle PY. Open data in public health surveillance systems: a case study using the French Sentinelles network. Int J Med Inform 2013 Oct;82(10):1012-1021. [doi: 10.1016/j.ijmedinf.2013.06.009] [Medline: 23850384]
31. Flu View: U.S. Influenza Surveillance: Purpose and Methods. US Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/flu/weekly/overview.htm#anchor_1539281266932 [accessed 2022-03-14]
32. Public Health England: approach to surveillance. UK Department of Health. URL: https://www.gov.uk/government/publications/public-health-england-approach-to-surveillance/public-health-england-approach-to-surveillance [accessed 2022-03-14]
33. Razai MS, Majeed A. General practice in England: the current crisis, opportunities, and challenges. J Ambul Care Manage 2022;45(2):135-139. [doi: 10.1097/JAC.0000000000000410] [Medline: 35202030]
34. Borek AJ, Pilbeam C, Mableson H, Wanat M, Atkinson P, Sheard S, et al. Experiences and concerns of health workers throughout the first year of the COVID-19 pandemic in the UK: A longitudinal qualitative interview study. PLoS One 2022;17(3):e0264906 [FREE Full text] [doi: 10.1371/journal.pone.0264906] [Medline: 35294450]
35. de Lusignan S, Correa A, Pathirannehelage S, Byford R, Yonova I, Elliot AJ, et al. RCGP Research and Surveillance Centre Annual Report 2014-2015: disparities in presentations to primary care. Br J Gen Pract 2017 Jan;67(654):e29-e40 [FREE Full text] [doi: 10.3399/bjgp16X688573] [Medline: 27993900]
36. de Lusignan S, Correa A, Pebody R, Yonova I, Smith G, Byford R, et al. Incidence of lower respiratory tract infections and atopic conditions in boys and young male adults: Royal College of General Practitioners Research and Surveillance Centre Annual Report 2015-2016. JMIR Public Health Surveill 2018 Apr 30;4(2):e49 [FREE Full text] [doi: 10.2196/publichealth.9307] [Medline: 29712621]
37. de Lusignan S, Borrow R, Tripathy M, Linley E, Zambon M, Hoschler K, et al. Serological surveillance of influenza in an English sentinel network: pilot study protocol. BMJ Open 2019 Mar 08;9(3):e024285 [FREE Full text] [doi: 10.1136/bmjopen-2018-024285] [Medline: 30852535]

Abbreviations

CMR: computerized medical record
ECDC: European Centre for Disease Control
EMIS: Egton Medical Information System
GP: general practitioner
HPRU: Health Protection Research Unit
ICU: intensive care unit
ILI: influenza-like illness
INPS: In Practice Systems
ISO: International Standards Organization
LSOA: Lower Level Super Output Area
NDPCHS: Nuffield Department of Primary Care Health Sciences
NHS: National Health Service
NIHR: National Institute for Health Research
NIMS: National Immunisation Management Service
ONS: Office of National Statistics
ORCHID: Oxford-Royal College of General Practitioners Clinical Informatics Digital Hub
PCSC: Primary Care Sentinel Cohort
RGCP: Royal College of General Practitioners
RSC: Research and Surveillance Centre

SES: socioeconomic status
SNOMED-CT: Systematized Nomenclature of Medicine Clinical Terms
SQUIRE: Surveillance, Quality Improvement, Research, Education
SSGP: Syndromic Surveillance General Practices
TPP: The Phoenix Partnership
TRE: Trusted Research Environment
UKHSA: United Kingdom Health Security Agency

Edited by G Eysenbach; submitted 04.05.22; peer-reviewed by S Pesälä; comments to author 17.10.22; revised version received 27.10.22; accepted 02.11.22; published 19.12.22.

Please cite as:
Leston M, Elson WH, Watson C, Lakhani A, Aspden C, Bankhead CR, Borrow R, Button E, Byford R, Elliot AJ, Fan X, Hoang U, Linley E, Macartney J, Nicholson BD, Okusi C, Ramsay M, Smith G, Smith S, Thomas M, Todkill D, Tsang RSM, Victor W, Williams AJ, Williams J, Zambon M, Howsam G, Amirthalingam G, Lopez-Bernal J, Hobbs FDR, de Lusignan S
Representativeness, Vaccination Uptake, and COVID-19 Clinical Outcomes 2020-2021 in the UK Oxford-Royal College of General Practitioners Research and Surveillance Network: Cohort Profile Summary
JMIR Public Health Surveill 2022;8(12):e39141
URL: https://publichealth.jmir.org/2022/12/e39141
doi:10.2196/39141
PMID:36534462

Original Paper

The Role of Heterogenous Real-world Data for Dengue Surveillance in Martinique: Observational Retrospective Study

Emmanuelle Sylvestre[1,2], MD, PhD; Elsa Cécilia-Joseph[2], PhD; Guillaume Bouzillé[1], MD, PhD; Fatiha Najioullah[3], PharmD, PhD; Manuel Etienne[4], PhD; Fabrice Malouines[4]; Jacques Rosine[5], MPH; Sandrine Julié[6], MD; André Cabié[7,8,9], MD, PhD; Marc Cuggia[1], MD, PhD

[1]Laboratoire de Traitement du Signal et de l'Image (LTSI) - Unité Mixte de Recherche (UMR) 1099, Université de Rennes, Centre Hospitalier Universitaire Rennes, Institut national de la santé et de la recherche médicale (INSERM), Rennes, France

[2]Centre de Données Cliniques, Centre Hospitalier Universitaire Martinique, Fort-de-France, Martinique

[3]Laboratoire de Virologie, Centre Hospitalier Universitaire Martinique, Fort-de-France, Martinique

[4]Centre de Démoustication et de Recherche Entomologique, Collectivité Territoriale de la Martinique – Agence Régionale de Santé, Fort-de-France, Martinique

[5]Cellule Martinique, Santé Publique France, Saint-Maurice, France

[6]Département d'Information Médicale, Service de Santé Publique, Centre Hospitalier Universitaire Martinique, Fort-de-France, Martinique

[7]Infectious and Tropical Diseases Unit, Centre Hospitalier Universitaire Martinique, Fort-de-France, Martinique

[8]Centre d'Investigation Clinique (CIC)-1424, Centre Hospitalier Universitaire Martinique, Institut national de la santé et de la recherche médicale (INSERM), Fort-de-France, Martinique

[9]Pathogenesis and Control of Chronic and Emerging Infections (PCCEI), Université de Montpellier - Université des Antilles, Institut national de la santé et de la recherche médicale (INSERM) - Etablissement Français du Sang (EFS), Montpellier, France

Corresponding Author:
Emmanuelle Sylvestre, MD, PhD
Centre de Données Cliniques
Centre Hospitalier Universitaire Martinique
CS 90632
Fort-de-France, 97261
Martinique
Phone: 596 596553507
Email: emmanuelle.sylvestre@chu-martinique.fr

Abstract

Background: Traditionally, dengue prevention and control rely on vector control programs and reporting of symptomatic cases to a central health agency. However, case reporting is often delayed, and the true burden of dengue disease is often underestimated. Moreover, some countries do not have routine control measures for vector control. Therefore, researchers are constantly assessing novel data sources to improve traditional surveillance systems. These studies are mostly carried out in big territories and rarely in smaller endemic regions, such as Martinique and the Lesser Antilles.

Objective: The aim of this study was to determine whether heterogeneous real-world data sources could help reduce reporting delays and improve dengue monitoring in Martinique island, a small endemic region.

Methods: Heterogenous data sources (hospitalization data, entomological data, and Google Trends) and dengue surveillance reports for the last 14 years (January 2007 to February 2021) were analyzed to identify associations with dengue outbreaks and their time lags.

Results: The dengue hospitalization rate was the variable most strongly correlated with the increase in dengue positivity rate by real-time reverse transcription polymerase chain reaction (Pearson correlation coefficient=0.70) with a time lag of −3 weeks. Weekly entomological interventions were also correlated with the increase in dengue positivity rate by real-time reverse transcription polymerase chain reaction (Pearson correlation coefficient=0.59) with a time lag of −2 weeks. The most correlated query from Google Trends was the “Dengue” topic restricted to the Martinique region (Pearson correlation coefficient=0.637) with a time lag of −3 weeks.

Conclusions: Real-word data are valuable data sources for dengue surveillance in smaller territories. Many of these sources precede the increase in dengue cases by several weeks, and therefore can help to improve the ability of traditional surveillance

systems to provide an early response in dengue outbreaks. All these sources should be better integrated to improve the early response to dengue outbreaks and vector-borne diseases in smaller endemic territories.

(JMIR Public Health Surveill 2022;8(12):e37122) doi:10.2196/37122

KEYWORDS

dengue; surveillance; real-word data; Big Data; Caribbean; dengue-endemic region

Introduction

Dengue is one of the most important vector-borne diseases worldwide, with 390 million infections, 96 million symptomatic cases, and 20,000 estimated deaths per year in >125 countries [1,2]. The disease is mostly endemic in tropical and subtropical regions (ie, Southeast Asia, the Americas, and the Pacific), with 4 billion people at risk [3]. In Latin America and the Caribbean, morbidity and mortality increased from 400,519 cases and 92 deaths in 2000 to >3.1 million cases and 1534 deaths in 2019 [4,5]. Dengue prevention and control in these regions rely on 2 main approaches: vector control programs and traditional surveillance, which is based on passive detection of symptomatic cases (inpatients and outpatients) [4,6]. Although both approaches are effective, they are expensive and are hampered by the delay between case occurrence and case reporting. Furthermore, some countries do not have routine vector control measures [7] and national epidemiological surveillance systems tend to underestimate the true disease burden of dengue [8].

In Martinique, a French overseas territory in the Lesser Antilles with approximately 360,000 inhabitants, health authorities have launched the "Monitoring, warning and management of dengue outbreaks program" (Programme de surveillance, d'alerte et de gestion des épidémies de dengue [PSAGE]), in which vector control and traditional surveillance are combined. PSAGE identifies five main stages in dengue outbreaks: (1) sporadic transmission, (2) dengue clusters with or without an epidemiological link, (3) epidemic risk when the number of symptomatic cases is above the expected threshold, (4) dengue outbreak, and (5) return to normal. Vector surveillance still plays a role in this system; however, the change in PSAGE stage is mainly based on the number of symptomatic cases identified by general practitioners who are part of the French Sentinel Network surveillance system [9,10].

Surveillance systems are a key public health tool to detect early cases of emerging infectious diseases, prevent outbreaks [11] among populations, and implement measures to reduce transmission [12]. Traditional surveillance systems are often expensive because of the time and resources required to process data collected from public health networks [13]. To improve these systems and reduce the delay between diagnosis and reporting, researchers have evaluated novel data sources, especially real-world data (ie, data not collected in experimental conditions [14]), such as emergency department visits, mobile data, and internet-based systems [15-18]. Other studies on surveillance and forecasting, especially those using climate data [19-21], have also shown promising results. Scientists mostly rely on correlation methods to test these data sources [22,23], but other approaches have also been tested, for instance Naive Bayes methods [24,25]. Most of these studies were conducted in Asia (70% of the global dengue burden) [2]. Studies in the Americas concerned large territories or countries, such as Brazil and Mexico [24,25], and in the Caribbean, they focused on the bigger islands of the Greater Antilles [21,26].

The aim of this study, carried out in Martinique, was to investigate whether heterogeneous real-world data sources could help to reduce reporting delays and improve dengue monitoring in a smaller endemic region.

Methods

Data Sources

Overview

We used several types of data that had been routinely collected during the study period (from January 1, 2007, to February 28, 2021): epidemiological surveillance reports from the French National Public Health Agency (Santé Publique France), reimbursement claims and laboratory data from Martinique University Hospital, entomological data from the Martinique Mosquito Control and Entomological Research Center (Centre de Démoustication et de Recherche Entomologique [CEDRE]), and relative search volumes (RSV) from Google Trends. Entomological, clinical, and laboratory data are available within 24 to 48 hours. Google RSV and epidemiological surveillance data are available in real time and at the end of each week, respectively. All used data were anonymized.

Epidemiological Surveillance Data

We obtained weekly dengue surveillance reports from the French Public Health Agency. These reports are based on data collected by general practitioners from the French Sentinel Network. They also provide the official start and end dates of each dengue outbreak and the weekly PSAGE stage during the outbreak. These reports are not continuously published but only if the dengue risk level is above stage 1 (ie, the baseline stage). We used the PSAGE stage described in each report to create the PSAGE ordinal variable with 4 levels. Indeed, although the PSAGE program has 5 levels, stage 5 ("back to normal") was used only 5 times in the last 15 years, and experts prefer to use stage 1 ("sporadic transmission") after stage 4 ("dengue outbreak"). Moreover, when stage 5 was used, it was for 1 week, except once in 2021, when it lasted 2 weeks. Thus, we combined stages 1 and 5 into a single stage (stage 1 or 5, sporadic transmission).

Clinical and Laboratory Data

We obtained weekly aggregated data from Martinique University Hospital: (1) inpatient data (age and diagnoses associated with dengue disease or dengue symptoms), (2) administrative data (outpatient medical consultations, hospitalizations, and

emergency department visits), and (3) laboratory data—dengue virus (DENV) detection by real-time reverse transcriptase polymerase chain reaction (RT-PCR).

All included diagnoses were coded using the French version of the International Classification of Diseases, 10th edition (ICD-10):

- Dengue or severe dengue
- Possible coding errors associated with dengue: fever and unspecified viral hemorrhagic fever
- Severity symptoms: hemorrhage, shock, and dehydration
- Thrombocytopenia
- Hepatic symptoms: hepatitis, hepatomegaly, hepatic failure, and elevation of transaminase
- Neurological symptoms: encephalitis and encephalopathy

We selected these diagnoses with the help of infectious disease physicians. All reimbursement claims data were obtained from the Martinique University Hospital, where the only infectious disease department for the whole island is located. The relevant ICD-10 codes are listed in Multimedia Appendix 1.

We normalized all administrative data as follows:

where *x* is the weekly number of hospitalizations, consultations, or emergency department visits and min and are the minimum and maximum values observed in the data set, respectively.

For laboratory data, we used the DENV positivity rate determined by RT-PCR. Laboratory results were concerned about both inpatients and outpatients because the Martinique University Hospital is the reference center for DENV screening using RT-PCR in Martinique.

Entomological Data

We used data from the CEDRE surveillance database, such as the weekly number of entomological interventions and where they were carried out. Entomological interventions were defined as all vector control interventions and measures taken by CEDRE: information and education of the households, physical vector control (ie, eliminating mosquito sites, such as old containers filled with water), and chemical vector control with insecticides [26]. This agency manages entomological surveillance and vector control in Martinique and collects data on each intervention.

Google RSV

We used data from Google Trends [27], which provides real-time and archived information on Google queries from 2004 onward. These queries are normalized by Google as RSV by dividing the total search volume for a query in a geographic location by the total number of queries in that region at a given point in time [28]. We used this tool to retrieve information on the search interest for keywords associated with dengue during our selected time frame (January 2007 to February 2021). However, we could not retrieve weekly data for the Martinique region, especially for the first years of the study period, because there were not always enough RSV (as indicated by the Google error message “Sorry, not enough search volume to show graphs”). Therefore, we based our methodology to retrieve Google Trends data on previously published methodology frameworks, indicating that Google Trends data should be retrieved for exactly the same period as the other data under study and as a single data set rather than as individual queries for each year [29]. As data for our study period were only available at monthly intervals, we considered that interest was constant over each week of the month for each query.

For data retrieval, relevant keywords were selected with experts in the field. Normally, all spelling variations should be included in the research to limit the risk of missing data. However, in our case, combining all possible spelling variants of some keywords into a single query was impossible, and an error message from Google indicated that the available data were insufficient. Nevertheless, we retrieved results using the “topic” option from Google that includes various keywords associated with a category.

As Martinique (and the other islands in the Lesser Antilles) are small regions, we tried 2 strategies to explore the geographic region of our keywords: we selected “Martinique” as the region in the tool and we added “Martinique” as a keyword in our query, with the region selected as “worldwide.” Moreover, we selected our keywords in 3 different languages (French, English, and Spanish) because the Lesser Antilles is a multilingual region.

Data Processing

Clinical and laboratory data were already aggregated into a structured database and did not require data processing. Similarly, data from Google queries are normalized by Google as RSV. Conversely, most of the information in the CEDRE database was unstructured and required processing. Indeed, the CEDRE database is a comprehensive database with some structured data (eg, the date of an entomological intervention), but the details associated with entomological interventions (ie, the type of insecticide used or the number of old containers removed) were in free text; we needed this information to count the number of weekly interventions. Therefore, we used rule-based natural language processing methods (ie, part-of-speech tags) to process the data and extract relevant information for our study. All statistical analyses were performed using R (version 4.1.0; R Foundation for Statistical Computing) [30] (*tidytext* [31],*stopwords*, and *SnowballC* packages).

Statistical Analysis

A total of 4 dengue outbreaks were recorded in Martinique between 2007 and 2021: from August 20, 2007, to January 14, 2008; from February 22, 2010, to October 25, 2010; from July 22, 2013, to April 14, 2014; and from November 18, 2019, to February 8, 2021. The fourth dengue outbreak was the largest in Martinique over the last 20 years.

During the same period, there was a chikungunya outbreak in 2014, a Zika virus outbreak in 2016, the first COVID-19 wave in March 2020, and the second COVID-19 wave from September to December 2020. The last dengue outbreak was concomitant with the second COVID-19 wave. Consequently, the PSAGE stages did not vary much over the years, making it

difficult to study correlations with this categorical variable. Therefore, it is necessary to find a good continuous estimator for time series analyses. To this end, we assessed the DENV RT-PCR positive rate performance for PSAGE stage prediction using a repeated stratified k-fold cross-validation approach. First, we divided the data into 10 stratified folds, then built a logistic regression model to predict the PSAGE stages. Finally, we repeated the process 10 times and evaluated its performance. The original PSAGE variable was an ordinal variable with 4 levels, but the data were not evenly distributed among the levels. Therefore, we ran 4 binary logistic regression analyses, rather than a single multinomial regression model, to assess how RT-PCR positive rates can predict each level. We calculated the predicted probability of a PSAGE stage by using the following equation:

where X is the vector of the predictor values, β_1 is the vector of the regression coefficients, and β_0 is the intercept of the model. As the data set was imbalanced, we also used stratified sampling in the PSAGE stage for k-fold cross-validation.

The metrics used to assess the logistic model performance were accuracy, specificity, precision, recall, F_1-score, and area under the curve (AUC).

Accuracy assesses the overall effectiveness of the logistic regression model and can be defined as the ratio of the correct number of predictions to the total number of predictions:

where TP are true positive, TN are true negative, FP are false positive, and FN are false negative results.

Specificity is the model's capacity to predict that a week is not in the PSAGE stage and is defined as the ratio between correctly predicted negative classes and all items that are actually negative:

where TN is true negative, and FP is false positive.

Precision (or positive predictive value) is the agreement between the true stages and the stages predicted by the RT-PCR positive rate and is defined as the ratio between the correctly predicted positive classes and all items predicted to be positive:

where TP is true positive, and FP is false positive.

Recall (or sensitivity) is the model's capacity to identify the true stages and is defined as the ratio between correctly predicted positive classes and all items that are actually positive:

where TP is true positive, and FN is false negative.

The F_1-score is the harmonic mean of precision and recall. The AUC represents the capacity of the model to avoid false classification into a stage.

To investigate the association between the RT-PCR positive rate and each data source, we plotted their time series. Finally, for each source, we estimated the Pearson correlation coefficient (r) and the cross-correlations between the weekly data and the DENV RT-PCR positive rate. The aim of the cross-correlation function is to investigate the relationship between time series and their lag values [32]. In our case, we wanted to determine whether the increase in the studied variables was correlated with the DENV RT-PCR positive rate and whether it preceded it. All statistical analyses were performed using R (version 4.1.0) [30]. For cross-correlations, significance is determined graphically when the lines are above (or below) the dotted blue line.

Ethics Approval

This study was approved by the local Ethics Committee of Martinique University Hospital (approval number 2022/177).

Results

RT-PCR Positive Rate Performance

The accuracy and AUC values ranged between 0.83 and 0.95 and between 0.55 and 0.89, respectively. Overall, the model performed better at predicting sporadic transmission (stage 1 or 5: accuracy=0.83; AUC=0.84) and outbreak (stage 4: accuracy=0.89; AUC=0.89; Table 1).

Table 1. DENV[a] RT-PCR[b] positive rate and PSAGE[c] stage prediction.

Metrics	PSAGE			
	Stage 1 or 5	Stage 2	Stage 3	Stage 4
Accuracy	0.828	0.879	0.953	0.888
Specificity	0.616	1	1	0.96
Precision	0.827	—[d]	—	0.742
Recall	0.936	0	0	0.535
F_1-score	0.878	—	—	0.612
AUC[e]	0.838	0.546	0.828	0.888

[a]DENV: dengue virus.

[b]RT-PCR: real-time reverse transcriptase polymerase chain reaction.

[c]PSAGE: Programme de surveillance, d'alerte et de gestion des épidémies de dengue.

[d]Not enough data available to build a prediction model for these stages.

[e]AUC: area under the curve.

Hospital Data

We normalized all hospital data to plot the time series to consider the different scales. As children and adults can be affected differently depending on the dengue infection type (primary vs secondary), we stratified our data sets based on the ward type (adult or pediatric).

Administrative Data

Adult hospitalizations (P=.01) and emergency department visits (P<.001) were significantly correlated with the DENV RT-PCR positive rate. We also observed a significant cross-correlation at −3 and −5 weeks, suggesting that the increase in emergency department visits preceded the increase in the DENV RT-PCR positive rate by 3 to 5 weeks. Table 2 shows the correlations and cross-correlations between the administrative data and DENV detection rate by RT-PCR. All cross-correlations between administrative data and DENV RT-PCR positive rate are listed in Multimedia Appendix 2.

Table 2. Correlations and cross-correlations between administrative data and DENV[a] RT-PCR[b] positive rate.

Data	Correlation (95% CI)	*P* value	Max cross-correlation[c]	Time lag[d]
Hospitalizations				
Total (n=506,992)	−0.066 (−0.137 to 0.006)	.07	−0.091	−5 weeks
Adults (n=444,045)	−0.095 (−0.165 to −0.023)	*.01* [e]	−0.097	−4 weeks
Children (n=62,947)	0.067 (−0.004 to 0.139)	.06	*0.118* [f]	*−8 weeks*
Emergency department visits				
Total (n=1,082,343)	0.111 (0.039 to 0.181)	*.002*	*0.169*	*−5 weeks*
Adults (n=740,282)	0.181 (0.11 to 0.25)	*<.001*	0.216	−3 weeks
Children (n=342,061)	0.046 (−0.025 to 0.118)	.21	*0.107*	*−5 weeks*
Consultations				
Total (n=2,715,906)	−0.065 (−0.137 to 0.007)	.08	−0.067	−2 weeks
Adults (n=2,467,565)	−0.061 (−0.133 to 0.0105)	.09	−0.097	−5 weeks
Children (n=248,341)	−0.046 (−0.118 to 0.026)	.21	−0.087	−5 weeks

[a]DENV: dengue virus.

[b]RT-PCR: real-time reverse transcriptase polymerase chain reaction.

[c]Maximum cross-correlation.

[d]Time lag that results in the maximum cross-correlation.

[e]Italicized *P* values are significant.

[f]Italicized cross-correlations are statistically significant (details in Multimedia Appendices 2-5).

Inpatient Data

We normalized inpatient data as the percentage of each diagnosis among all diagnoses for that year. The percentage of dengue diagnoses among inpatients was significantly associated with an increase in the DENV RT-PCR positive rate. We also detected a significant cross-correlation at −3 weeks, indicating that the increase in dengue diagnoses among hospitalized people preceded the increase in DENV RT-PCR positive rates by 3 weeks (Table 3). All cross-correlations between dengue diagnoses in inpatients and DENV RT-PCR positive rates are listed in Multimedia Appendix 3.

Concerning dengue-related symptoms, thrombocytopenia and liver involvement in adults and children were associated with the DENV RT-PCR positive rate.

The significant cross-correlation, at time lags ranging between −2 and −5 weeks, indicated that the increase in thrombocytopenia and liver dysfunction preceded the increase in DENV RT-PCR positive rates by 3 to 5 weeks (Table 4). All cross-correlations between dengue symptoms among inpatients and DENV RT-PCR positive rate are listed in Multimedia Appendices 4 and 5.

The weekly hospitalization rates for dengue and thrombocytopenia during the study period are shown in Figure 1, with DENV RT-PCR positive rate, as a reference.

Table 3. Correlations between dengue diagnoses inpatients and DENV[a] RT-PCR[b] positive rate.

Data	Correlation (95% CI)	*P* value	Max cross-correlation[c]	Time lag[d]
Total	0.704 (0.665-0.738)	*<.001* [e]	*0.710* [f]	*−3 weeks*
Adults	0.698 (0.659-0.733)	*<.001*	*0.703*	*−3 weeks*
Children	0.672 (0.631-0.701)	*<.001*	*0.675*	*−3 weeks*

[a]DENV: dengue virus.

[b]RT-PCR: real-time reverse transcriptase polymerase chain reaction.

[c]Maximum cross-correlation.

[d]Time lag that results in the maximum cross-correlation.

[e]Italicized *P* values are significant.

[f]Italicized cross-correlations are statistically significant (details in Multimedia Appendices 2-5).

Table 4. Correlations between dengue symptoms among inpatients and dengue RT-PCR[a] positive rate.

Data	Correlation (95% CI)	*P* value	Max cross-correlation[b]	Time lag[c]
Symptoms				
Total	0.077 (0.005 to 0.148)	*.04* [d]	*0.081* [e]	*−4 weeks*
Adults	0.071 (−0.001 to 0.142)	.05	0.071	0 weeks
Children	0.093 (0.021 to 0.16)	*.01*	*0.127*	*−4 weeks*
Coding errors				
Total	−0.096 (−0.167 to –0.024)	*.009*	−0.098	−1 week
Adults	−0.072 (−0.143 to $–1.12 \times 10^{-4}$)	*.05*	−0.086	−2 weeks
Children	−0.043 (−0.115 to 0.0285)	.24	−0.043	0 weeks
Symptom severity				
Total	0.105 (0.033 to 0.175)	*.004*	0.105	0 weeks
Adults	0.068 (−0.004 to 0.139)	.07	0.068	0 weeks
Children	0.263 (0.195 to 0.329)	<.001	0.279	−4 weeks
Thrombocytopenia				
Total	0.281 (0.213 to 0.346)	*<.001*	*0.289*	*−2 weeks*
Adults	0.235 (0.166 to 0.302)	*<.001*	*0.242*	*−2 weeks*
Children	0.269 (0.201 to 0.335)	*<.001*	*0.288*	*−4 weeks*
Liver dysfunction symptoms				
Total	0.152 (0.081 to 0.222)	*<.001*	*0.179*	*−5 weeks*
Adults	0.123 (0.0517 to 0.193)	*<.001*	*0.153*	*−5 weeks*
Children	0.152 (0.081 to 0.222)	*<.001*	*0.147*	*−5 weeks*
Neurological symptoms				
Total	−0.028 (−0.100 to 0.0435)	.44	−0.061	−7 weeks
Adults	−0.045 (−0.117 to 0.0265)	.22	−0.068	−7 weeks
Children	0.029 (−0.0427 to 0.101)	.43	0.034	−6 weeks

[a]RT-PCR: real-time reverse transcriptase polymerase chain reaction.

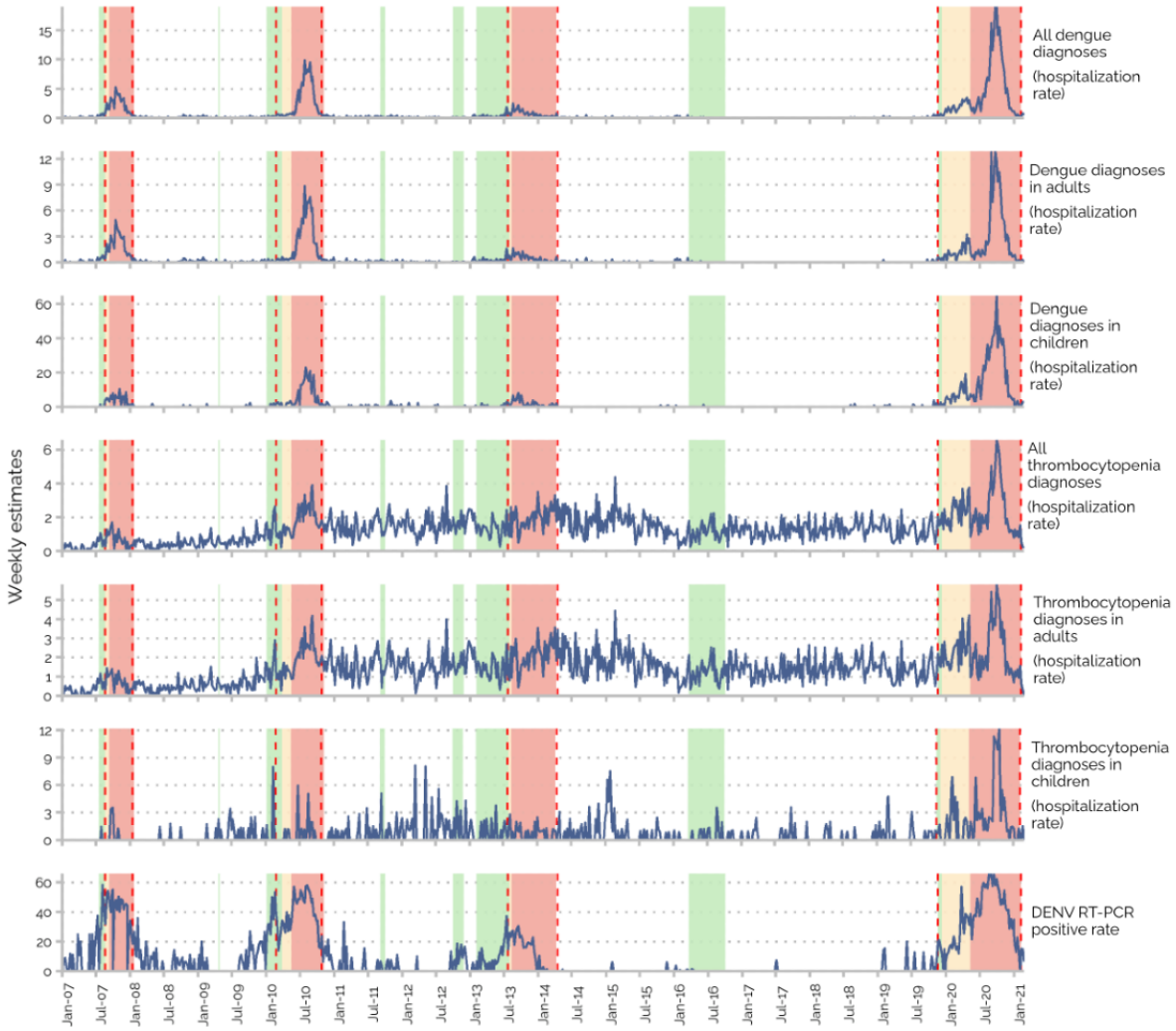
[b]Maximum cross-correlation.

[c]Time lag that results in the maximum cross-correlation.

[d]Italicized *P* values are significant.

[e]Italicized cross-correlations are statistically significant (details in Multimedia Appendices 2-5).

Figure 1. Weekly hospitalization rates for dengue and thrombocytopenia during the different Programme de surveillance, d'alerte et de gestion des épidémies de dengue (PSAGE) stages from January 2007 to February 2021. The DENV RT-PCR positive rate was used as a reference. Blue curves: weekly hospitalization rates for the indicated ICD-10 diagnoses. Green areas: PSAGE stage 2 (dengue clusters). Yellow areas: PSAGE stage 3 (epidemic risk). Red areas: PSAGE stage 4 (dengue outbreak). Red dashed lines: official dates of dengue outbreaks that were decided retrospectively by the French Public Health Agency at the end of each outbreak. DENV: dengue virus; RT-PCR: real-time reverse transcription polymerase chain reaction.



Entomological Data

The weekly number of entomological interventions was significantly (P<.001) associated with DENV RT-PCR positive rate (r=0.591; 95% CI 0.542-0.636). They were also significantly cross-correlated (0.627 at −2 weeks), indicating that their increase preceded an increase in the DENV RT-PCR positive rate by 2 weeks.

We did not find any significant correlation or cross-correlation between the intervention zones and the RT-PCR positive rate.

Google RSV

We considered that interest was constant over each week of the month for each query to compute our weekly data, but RSV could have high variability across weeks. Therefore, we also compared monthly RSV to monthly DENV RT-PCR positive rates to assess whether our approach had a high impact on the results.

Several Google keywords were significantly associated with the DENV RT-PCR positive rate. Overall, this association was stronger for the simplest combination of keywords, without spelling variations, especially for the keywords "dengue symptoms." We could not assess some keyword combinations because of the lack of data. Furthermore, when Google Trends provided "Topics," the results outperformed those obtained using manual combinations of keywords that included spelling, language, or accent variations. Keywords not restricted to the geographic region of "Martinique" (by using the Geographical region feature or by adding the keyword "Martinique" to the query) were not significantly associated with the DENV RT-PCR positive rate. We obtained the strongest significant cross-correlation using the topic "dengue" in the Martinique region (0.643 at the time lag of −3 weeks). This indicated that an increase in queries for this term in the Martinique region preceded the increase in the DENV RT-PCR positive rate by 3 weeks (Table 5). Conversely, we did not find any significant cross-correlation within meaningful time lag values for the term "mosquito" and its different spellings and language variations.

For monthly correlations, the results were similar to weekly results (Table 6). All weekly correlations between Google Trends keywords and DENV RT-PCR positive rates are listed in Multimedia Appendix 6. All monthly correlations between

Google Trends keywords and DENV RT-PCR positive rates are listed in Multimedia Appendix 7. All weekly cross-correlations between nonhospital data and DENV RT-PCR positive rate are listed in Multimedia Appendix 8. All monthly cross-correlations between Google Trends keywords and DENV RT-PCR positive rate are listed in Multimedia Appendix 9.

The weekly estimates for nonhospital data during the study period are displayed in Figure 2, with the DENV RT-PCR positive rate as a reference.

Table 5. Strongest correlations between Google Trends keywords and DENV[a] RT-PCR[b] positive rate.

Keywords	Correlation (95% CI)	*P* value	Max cross-correlation[c]	Time lag[d]
Dengue				
Keywords “dengue” + “dingue” and region “Martinique”	0.597 (0.548-0.641)	*<.001* [e]	*0.598* [f]	*−1 week*
Keywords “dengue” + “Martinique”	0.534 (0.480-0.583)	*<.001*	*0.611*	*−6 weeks*
Topic “dengue” and region “Martinique”	0.637 (0.591-0.677)	*<.001*	*0.643*	*−3 weeks*
Dengue symptoms				
Keyword “symptome dengue” and region “Martinique”	0.412 (0.351-0.47)	*<.001*	*0.435*	*−3 weeks*
Mosquito				
Keyword “mosquito” with various French spellings and region “Martinique”	0.200 (0.130-0.268)	*<.001*	0.200	0 weeks
Aedes				
Keywords “aedes” and region “Martinique”	0.339 (0.273-0.401)	*<.001*	*0.369*	*−3 weeks*
Topic “aedes” and region “Martinique”	0.214 (0.591-0.677)	*<.001*	*0.304*	*−7 weeks*

[a]DENV: dengue virus.

[b]RT-PCR: real-time reverse transcriptase polymerase chain reaction.

[c]Maximum cross-correlation.

[d]Time lag that results in the maximum cross-correlation.

[e]Italicized *P* values are significant.

[f]Italicized cross-correlations are statistically significant (details in Multimedia Appendices 2-5).

Table 6. Strongest monthly correlations between Google Trends keywords and DENV[a] RT-PCR[b] positive rate.

Keywords	Correlation (95% CI)	*P* value	Max cross-correlation[c]	Time lag[d]
Dengue				
Keywords “dengue” + “dingue” and region “Martinique”	0.632 (0.531-0.714)	*<.001* [e]	0.632	0 months
Keywords “dengue” + “Martinique”	0.592 (0.484-0.681)	*<.001*	*0.643* [f]	*−1 month*
Topic “dengue” and region “Martinique”	0.675 (0.583-0.749)	*<.001*	0.675	0 months
Dengue symptoms				
Keyword “symptome dengue” and region “Martinique”	0.436 (0.306-0.55)	*<.001*	*0.453*	*−1 month*
Mosquito				
Keyword “mosquito” with various French spellings and region “Martinique”	0.217 (0.004-0.068)	*<.001*	0.217	0 weeks
Aedes				
Keywords “aedes” and region “Martinique”	0.379 (0.243-0.501)	*<.001*	*0.394*	*−1 month*
Topic “aedes” and region “Martinique”	0.242 (0.095-0.379)	*<.001*	*0.313*	*−2 months*

[a]DENV: dengue virus.

[b]RT-PCR: real-time reverse transcriptase polymerase chain reaction.

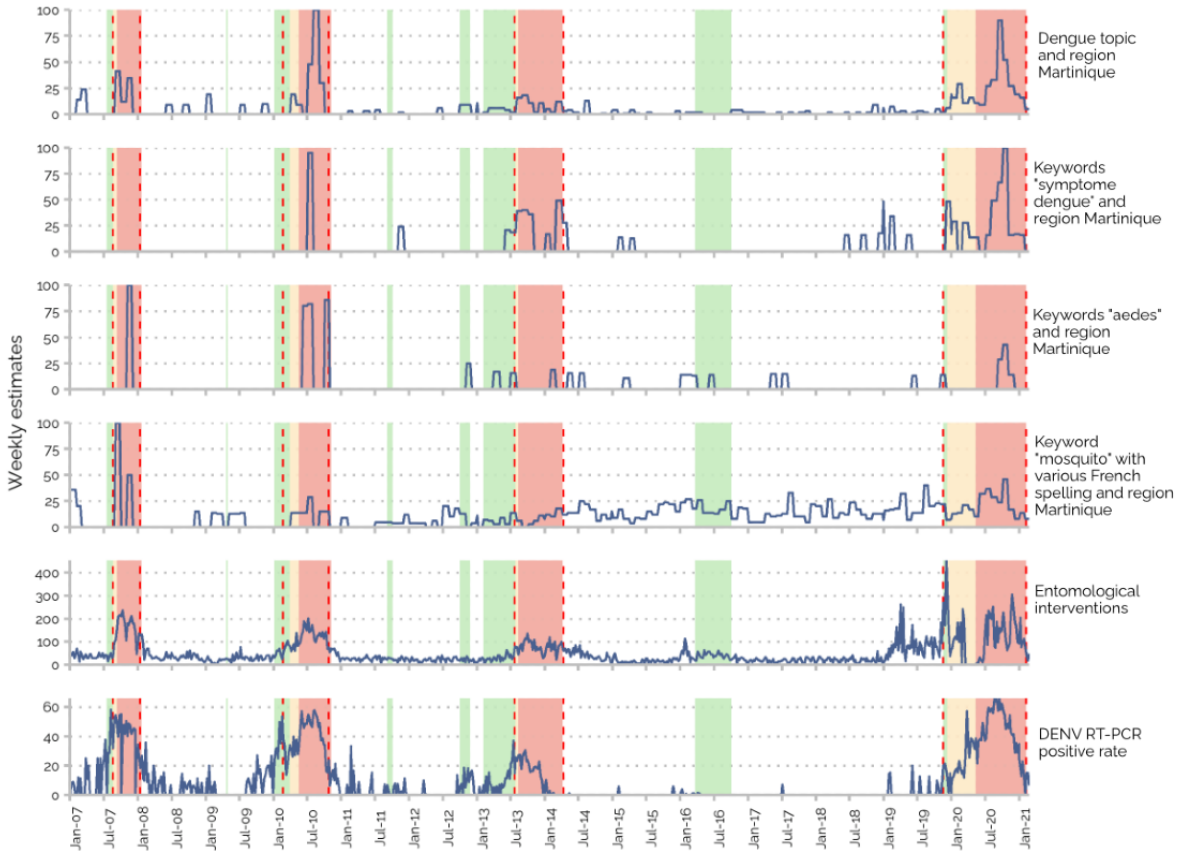
[c]Maximum cross-correlation.

[b]Time lag that results in the maximum cross-correlation.

[e]Italicized *P* values are significant.

[f]Italicized cross-correlations are statistically significant (details in Multimedia Appendices 2-5).

Figure 2. Weekly estimates for the indicated nonhospital data during the different Programme de surveillance, d'alerte et de gestion des épidémies de dengue (PSAGE) stages from January 2007 to February 2021. The DENV RT-PCR positive rate was used as a reference. Blue curves: weekly estimates for the strongest correlated Google keywords and entomological interventions. Green areas: PSAGE stage 2 (dengue clusters). Yellow areas: PSAGE stage 3 (epidemic risk). Red areas: PSAGE stage 4 (dengue outbreak). Red dashed lines: official dates of the outbreaks decided retrospectively by the French Public Health Agency at the end of each outbreak. DENV: dengue virus; RT-PCR: real-time reverse transcription polymerase chain reaction.



Discussion

Principal Findings

This study demonstrates the potential of real-world data for dengue outbreak monitoring. It indicates that multiple heterogeneous data sources, such as clinical data, vector data, and novel Big Data streams, should be leveraged simultaneously because they can all play a role in improving traditional dengue surveillance systems. Moreover, some data, such as the weekly hospitalization rates for thrombocytopenia, the weekly number of entomological interventions, and Google keywords, were not only significantly correlated with the weekly DENV RT-PCR positive rates, but their increase preceded the increase in RT-PCR positive results by 2 to 4 weeks.

An early response is crucial in dengue management because it can reduce mortality [18] and help stakeholders better anticipate needs and resources. In Martinique, the early signs identified in this study could be used to set up more hospital beds (including in the intensive care unit), increase staffing, particularly in emergency services and infectious diseases department, and increase the blood bank stock levels for patients with severe dengue who may need blood transfusions. Moreover, stakeholders could use them to justify requests for reinforcements from other territories (for Martinique, mostly from mainland France), medical equipment, and hospital staff. In addition, they could be used to notify earlier the Pan-American Health Organization, which is the Regional Office for the Americas of the World Health Organization [33], and help other islands to better prepare for an incoming outbreak.

Previous studies have already investigated the role of entomological data [34], inpatient data [35], and internet data streams [36] in dengue management, but few have assessed all these data sources simultaneously. In this study, we found that they should all be considered together rather than individually. Vector-based data tend to be underused [37], despite their central place in dengue surveillance, although we observed a rather strong correlation between the number of weekly entomological interventions and the increase in DENV RT-PCR positive rates. Therefore, they should be better integrated into the dengue surveillance system to improve its efficiency because both clinical surveillance and vector-based surveillance are essential for optimal dengue management [38]. The role of internet search engines in dengue surveillance has been frequently addressed in recent years [23,39]. Most studies were carried out in Asia and in larger American countries, such as Mexico and Brazil [24,25], and used a different approach based on weekly extracted data, which was not possible in our case. However, we found that even in Martinique, a smaller territory with a smaller population and, thus, with a lower data volume from internet data streams, Google queries were still correlated with an increase in the DENV RT-PCR positive rate. This means that they can also be used as part of surveillance systems across the

islands of the Lesser Antilles. However, the methodological framework [29] still needs to be adapted to the size of these territories, and the simplest keywords and Google topics, when available, should be preferred over multiple spelling variations. With these small adaptations, we propose a way to offset the limitations related to smaller territories to use internet data streams in this context because their interest in emerging disease surveillance has been demonstrated in previous studies [40,41]. Overall, for smaller territories, the challenge lies in the small population size that leads to a lower weekly signal variability, thus complicating covariance estimation (and consequently the use of correlation methods). As most studies evaluating real data sources for dengue surveillance were based on correlation methods [22,37,42], we needed to confirm that these approaches were still applicable using a smaller sample. Despite these limitations, we managed to identify relevant indicators from all data sources to improve monitoring.

Moreover, most studies on real-world data sources used symptomatic cases as gold standard [37]. However, in practice, public health authorities do not rely solely on symptomatic cases for decision-making during an outbreak. Here, we compared our data sources to the actual gold standard used by stakeholders for decision-making, which is based on objective and subjective parameters and found a reliable objective proxy (ie, the weekly DENV RT-PCR positive rates) to assess our variables. Finally, dengue hospitalizations and the symptoms associated with severe dengue cases (thrombocytopenia and liver dysfunction symptoms) should be closely monitored in inpatients, especially in children, because they tend to precede the DENV RT-PCR positive rate increase by several weeks.

Our study also highlighted homogenous time lags across different data sources, despite their heterogeneity. This further demonstrates the importance of considering them globally rather than individually, although some of these correlations were low or moderate. For instance, an increase in hospitalized patients with liver dysfunction symptoms could prompt physicians to pay closer attention to the dengue hospitalization rate because both precede the increase in DENV RT-PCR by 5 weeks and 3 weeks, respectively. The capacity to identify variables that precede the DENV RT-PCR positive rate increase is very relevant for dengue management because a rapid and early response can influence outbreak severity [18].

Limitations

Our method is promising but has some limitations. First, some correlations were very low, although they were statistically significant. Second, we did not include climate data because insufficient data were available for our time frame. Several studies have demonstrated the role of climate data (especially temperature, humidity, and rainfall) in dengue surveillance, but they were mostly carried out in Asian countries [43,44] and South America [45,46]. Few studies in the Caribbean region showed the role of rainfall and temperature in increasing the risk of dengue outbreaks. However, their time lags (between 7 weeks and 5 months) [47,48] were longer than the time lags we found for the other data sources. Nevertheless, this data source could have been relevant.

Third, our laboratory data did not include private sector biology laboratories, because they did not use RT-PCR techniques before the COVID-19 pandemic in 2020. Before this date, dengue diagnosis in private sector laboratories was based on NS1 antigen detection and needed sometimes to be confirmed by the more sensitive RT-PCR test at the hospital laboratory. It should be noted that the weekly number of DENV RT-PCR tests increased over time. Therefore, we used the weekly positive rate and not the weekly number of RT-PCR tests. Similarly, the World Health Organization dengue case classification and guidelines for hospitalization changed during the study period [49], and this may have influenced the results. Nevertheless, the rate of hospitalized patients with a dengue diagnosis was more strongly correlated with the DENV RT-PCR positive rate in our study.

Finally, concerning the entomological data, we only studied the correlation between the weekly number of interventions and the increase in the DENV RT-PCR positive rate, but we did not consider the number of mosquito clusters (ie, several clusters can be detected during 1 intervention). We focused on the simplest variable because vector control programs vary among the countries in this region [4,6], and we wanted to develop a common approach for all Caribbean territories. Furthermore, because entomological interventions tend to increase during outbreaks, we cannot rule out the influence of these practices on our results. Nevertheless, we could show that entomological interventions precede the increase in the DENV RT-PCR positive rate by 2 weeks.

Our approach does not intend to replace traditional monitoring systems based on syndromic surveillance, but to reduce the delays in these systems by leveraging data that are already routinely collected. These new data sources are readily available and can be easily implemented in the existing surveillance systems with minimal cost and training. However, their ability to predict future dengue outbreaks need to be thoroughly assessed, especially in smaller territories in the Lesser Antilles.

Conclusions

Our study shows that real-world data are valuable data sources for dengue surveillance in Martinique. Several heterogeneous data sources are relevant, from clinical data to vector control data and Google Trends data. Their increase precedes the increase in dengue cases by several weeks, and therefore, they can help to improve traditional surveillance systems to provide an early response to dengue outbreaks. By improving the integration of many different sources, we might better respond to dengue outbreaks in endemic regions, as well as to other types of vector-borne diseases such as Zika and chikungunya.

Conflicts of Interest

None declared.

Multimedia Appendix 1
International Classification of Diseases, 10th edition, codes of the selected diagnoses for inpatient data.
[DOCX File , 16 KB - publichealth_v8i12e37122_app1.docx]

Multimedia Appendix 2
Cross-correlations between administrative data and dengue virus real-time reverse transcription polymerase chain reaction positive rate.
[PNG File , 205 KB - publichealth_v8i12e37122_app2.png]

Multimedia Appendix 3
Cross-correlations between dengue diagnoses inpatients and dengue virus real-time reverse transcription polymerase chain reaction positive rate.
[PNG File , 70 KB - publichealth_v8i12e37122_app3.png]

Multimedia Appendix 4
Cross-correlations between general dengue symptoms among inpatients and dengue virus real-time reverse transcription polymerase chain reaction positive rate.
[PNG File , 204 KB - publichealth_v8i12e37122_app4.png]

Multimedia Appendix 5
Cross-correlations between specific dengue symptoms among inpatients and dengue virus real-time reverse transcription polymerase chain reaction positive rate.
[PNG File , 198 KB - publichealth_v8i12e37122_app5.png]

Multimedia Appendix 6
Weekly correlations between Google Trends keywords and dengue virus real-time reverse transcription polymerase chain reaction positive rate.
[DOCX File , 20 KB - publichealth_v8i12e37122_app6.docx]

Multimedia Appendix 7
Monthly correlations between Google Trends keywords and dengue virus real-time reverse transcription polymerase chain reaction positive rate.
[DOCX File , 20 KB - publichealth_v8i12e37122_app7.docx]

Multimedia Appendix 8
Weekly cross-correlations between nonhospital data and dengue virus real-time reverse transcription polymerase chain reaction positive rate.
[PNG File , 139 KB - publichealth_v8i12e37122_app8.png]

Multimedia Appendix 9
Monthly cross-correlations between Google Trends keywords and dengue virus real-time reverse transcription polymerase chain reaction positive rate.
[PNG File , 94 KB - publichealth_v8i12e37122_app9.png]

References

1. 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]
2. Stanaway JD, Shepard DS, Undurraga EA, Halasa YA, Coffeng LE, Brady OJ, et al. The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. Lancet Infect Dis 2016 Jun;16(6):712-723 [FREE Full text] [doi: 10.1016/S1473-3099(16)00026-8] [Medline: 26874619]
3. 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]
4. Torres JR, Orduna TA, Piña-Pozas M, Vázquez-Vega D, Sarti E. Epidemiological characteristics of dengue disease in Latin America and in the Caribbean: a systematic review of the literature. J Trop Med 2017;2017:8045435 [FREE Full text] [doi: 10.1155/2017/8045435] [Medline: 28392806]
5. Dengue. Pan American Health Organization. URL: https://www.paho.org/en/topics/dengue [accessed 2022-02-07]

6. Cafferata ML, Bardach A, Rey-Ares L, Alcaraz A, Cormick G, Gibbons L, et al. Dengue epidemiology and burden of disease in Latin America and the Caribbean: a systematic review of the literature and meta-analysis. Value Health Reg Issues 2013 Dec;2(3):347-356. [doi: 10.1016/j.vhri.2013.10.002] [Medline: 29702769]
7. Gómez-Dantés H, Willoquet JR. Dengue in the Americas: challenges for prevention and control. Cad Saude Publica 2009;25 Suppl 1:S19-S31 [FREE Full text] [doi: 10.1590/s0102-311x2009001300003] [Medline: 19287863]
8. Shepard DS, Undurraga EA, Betancourt-Cravioto M, Guzmán MG, Halstead SB, Harris E, et al. Approaches to refining estimates of global burden and economics of dengue. PLoS Negl Trop Dis 2014 Nov;8(11):e3306 [FREE Full text] [doi: 10.1371/journal.pntd.0003306] [Medline: 25412506]
9. Dussart D, Gustave J, Lagathu G, Koulman L, Lordinot M, Martial J, et al. Programme de surveillance, d'alerte et de gestion des épidémies de dengue (Psage dengue) en Martinique. Santé Publique France. 2006. URL: https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-a-transmission-vectorielle/dengue/documents/rapport-synthese/programme-de-surveillance-d-alerte-et-de-gestion-des-epidemies-de-dengue-psage-dengue-en-martinique [accessed 2022-02-05]
10. Flahault A, Blanchon T, Dorléans Y, Toubiana L, Vibert JF, Valleron AJ. Virtual surveillance of communicable diseases: a 20-year experience in France. Stat Methods Med Res 2006 Oct;15(5):413-421. [doi: 10.1177/0962280206071639] [Medline: 17089946]
11. Milinovich GJ, Williams GM, Clements AC, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis 2014 Feb;14(2):160-168 [FREE Full text] [doi: 10.1016/S1473-3099(13)70244-5] [Medline: 24290841]
12. Jajosky RA, Groseclose SL. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. BMC Public Health 2004 Jul 26;4:29 [FREE Full text] [doi: 10.1186/1471-2458-4-29] [Medline: 15274746]
13. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. CMAJ 2009 Apr 14;180(8):829-831 [FREE Full text] [doi: 10.1503/cmaj.090215] [Medline: 19364791]
14. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - what is it and what can it tell us? N Engl J Med 2016 Dec 08;375(23):2293-2297. [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]
15. Elliot AJ, Hughes HE, Hughes TC, Locker TE, Shannon T, Heyworth J, et al. Establishing an emergency department syndromic surveillance system to support the London 2012 Olympic and Paralympic Games. Emerg Med J 2012 Dec;29(12):954-960. [doi: 10.1136/emermed-2011-200684] [Medline: 22366039]
16. Heffernan R, Mostashari F, Das D, Besculides M, Rodriguez C, Greenko J, et al. New York City syndromic surveillance systems. MMWR Suppl 2004 Sep 24;53:23-27 [FREE Full text] [Medline: 15714622]
17. 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]
18. Katzelnick LC, Coloma J, Harris E. Dengue: knowledge gaps, unmet needs, and research priorities. Lancet Infect Dis 2017 Mar;17(3):e88-100 [FREE Full text] [doi: 10.1016/S1473-3099(16)30473-X] [Medline: 28185868]
19. Polwiang S. The time series seasonal patterns of dengue fever and associated weather variables in Bangkok (2003-2017). BMC Infect Dis 2020 Mar 12;20(1):208 [FREE Full text] [doi: 10.1186/s12879-020-4902-6] [Medline: 32164548]
20. Koplewitz G, Lu F, Clemente L, Buckee C, Santillana M. Predicting dengue incidence leveraging internet-based data sources. A case study in 20 cities in Brazil. PLoS Negl Trop Dis 2022 Jan;16(1):e0010071 [FREE Full text] [doi: 10.1371/journal.pntd.0010071] [Medline: 35073316]
21. Souza J, Leung CK, Cuzzocrea A. An innovative big data predictive analytics framework over hybrid big data sources with an application for disease analytics. In: Proceedings of the 34th International Conference on Advanced Information Networking and Applications. 2020 Presented at: AINA '20; April 15-17, 2020; Caserta, Italy p. 669-680. [doi: 10.1007/978-3-030-44041-1_59]
22. Anno S, Hara T, Kai H, Lee MA, Chang Y, Oyoshi K, et al. Spatiotemporal dengue fever hotspots associated with climatic factors in Taiwan including outbreak predictions based on machine-learning. Geospat Health 2019 Nov 06;14(2) [FREE Full text] [doi: 10.4081/gh.2019.771] [Medline: 31724367]
23. Husnayain A, Fuad A, Lazuardi L. Correlation between Google Trends on dengue fever and national surveillance report in Indonesia. Glob Health Action 2019;12(1):1552652 [FREE Full text] [doi: 10.1080/16549716.2018.1552652] [Medline: 31154985]
24. Gluskin RT, Johansson MA, Santillana M, Brownstein JS. Evaluation of Internet-based dengue query data: Google Dengue Trends. PLoS Negl Trop Dis 2014 Feb;8(2):e2713 [FREE Full text] [doi: 10.1371/journal.pntd.0002713] [Medline: 24587465]
25. Romero-Alvarez D, Parikh N, Osthus D, Martinez K, Generous N, Del Valle S, et al. Google Health Trends performance reflecting dengue incidence for the Brazilian states. BMC Infect Dis 2020 Mar 26;20(1):252 [FREE Full text] [doi: 10.1186/s12879-020-04957-0] [Medline: 32228508]
26. Saide PM, Prokopec GV, Clennon J, Che-Mendoza A, Pan American Health Organization. Technical document for the implementation of interventions based on generic operational scenarios for Aedes aegypti control. International Development Research Centre. 2019. URL: http://hdl.handle.net/10625/60176 [accessed 2022-06-25]
27. Google Trends. URL: https://trends.google.com [accessed 2022-01-25]

28. Hswen Y, Zhang A, Ventelou B. Estimation of asthma symptom onset using internet search queries: lag-time series analysis. JMIR Public Health Surveill 2021 May 10;7(5):e18593 [FREE Full text] [doi: 10.2196/18593] [Medline: 33970108]
29. Mavragani A, Ochoa G. Google trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
30. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: http://www.r-project.org/index.html [accessed 2021-12-20]
31. 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]
32. Shumway RH, Stoffer DS. Time Series Analysis and Its Applications: With R Examples. New York, NY, USA: Springer; 2011.
33. Who We Are. Pan American Health Organization. URL: https://www.paho.org/en/who-we-are [accessed 2022-06-29]
34. Marques-Toledo CA, Bendati MM, Codeço CT, Teixeira MM. Probability of dengue transmission and propagation in a non-endemic temperate area: conceptual model and decision risk levels for early alert, prevention and control. Parasit Vectors 2019 Jan 16;12(1):38 [FREE Full text] [doi: 10.1186/s13071-018-3280-z] [Medline: 30651125]
35. Mello-Román JD, Mello-Román JC, Gómez-Guerrero S, García-Torres M. Predictive models for the medical diagnosis of dengue: a case study in Paraguay. Comput Math Methods Med 2019 Jul 29;2019:7307803 [FREE Full text] [doi: 10.1155/2019/7307803] [Medline: 31485259]
36. Klein GH, Guidi Neto P, Tezza R. Big Data and social media: surveillance of networks as management tool. Saude Soc 2017 Mar;26(1):208-217. [doi: 10.1590/S0104-12902017164943]
37. Sylvestre E, Joachim C, Cécilia-Joseph E, Bouzillé G, Campillo-Gimenez B, Cuggia M, et al. Data-driven methods for dengue prediction and surveillance using real-world and Big Data: a systematic review. PLoS Negl Trop Dis 2022 Jan;16(1):e0010056 [FREE Full text] [doi: 10.1371/journal.pntd.0010056] [Medline: 34995281]
38. Comprehensive Guideline for Prevention and Control of Dengue and Dengue Haemorrhagic Fever. Revised and expanded edition. World Health Organization, Regional Office for South-East Asia. 2011. URL: https://apps.who.int/iris/handle/10665/204894 [accessed 2022-01-26]
39. Verma M, Kishore K, Kumar M, Sondh AR, Aggarwal G, Kathirvel S. Google search trends predicting disease outbreaks: an analysis from India. Healthc Inform Res 2018 Oct;24(4):300-308 [FREE Full text] [doi: 10.4258/hir.2018.24.4.300] [Medline: 30443418]
40. 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]
41. Gianfredi V, Bragazzi NL, Nucci D, Martini M, Rosselli R, Minelli L, et al. Harnessing big data for communicable tropical and sub-tropical disorders: implications from a systematic review of the literature. Front Public Health 2018 Mar 21;6:90 [FREE Full text] [doi: 10.3389/fpubh.2018.00090] [Medline: 29619364]
42. Syamsuddin M, Fakhruddin M, Sahetapy-Engel JT, Soewono E. Causality analysis of Google trends and dengue incidence in Bandung, Indonesia with linkage of digital data modeling: longitudinal observational study. J Med Internet Res 2020 Jul 24;22(7):e17633 [FREE Full text] [doi: 10.2196/17633] [Medline: 32706682]
43. Liu D, Guo S, Zou M, Chen C, Deng F, Xie Z, et al. A dengue fever predicting model based on Baidu search index data and climate data in South China. PLoS One 2019 Dec 30;14(12):e0226841 [FREE Full text] [doi: 10.1371/journal.pone.0226841] [Medline: 31887118]
44. Puengpreeda A, Yhusumrarn S, Sirikulvadhana S. Weekly forecasting model for dengue hemorrhagic fever outbreak in Thailand. Eng J 2020 May 31;24(3):71-87. [doi: 10.4186/ej.2020.24.3.71]
45. Flamand C, Fabregue M, Bringay S, Ardillon V, Quénel P, Desenclos JC, et al. Mining local climate data to assess spatiotemporal dengue fever epidemic patterns in French Guiana. J Am Med Inform Assoc 2014 Oct;21(e2):e232-e240 [FREE Full text] [doi: 10.1136/amiajnl-2013-002348] [Medline: 24549761]
46. Baquero OS, Santana LM, Chiaravalloti-Neto F. Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. PLoS One 2018 Apr 2;13(4):e0195065 [FREE Full text] [doi: 10.1371/journal.pone.0195065] [Medline: 29608586]
47. Gharbi M, Quenel P, Gustave J, Cassadou S, La Ruche G, Girdary L, et al. Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. BMC Infect Dis 2011 Jun 09;11:166 [FREE Full text] [doi: 10.1186/1471-2334-11-166] [Medline: 21658238]
48. Lowe R, Gasparrini A, Van Meerbeeck CJ, Lippi CA, Mahon R, Trotman AR, et al. Nonlinear and delayed impacts of climate on dengue risk in Barbados: a modelling study. PLoS Med 2018 Jul;15(7):e1002613 [FREE Full text] [doi: 10.1371/journal.pmed.1002613] [Medline: 30016319]
49. Dengue guidelines for diagnosis, treatment, prevention and control : new edition. World Health Organization. Geneva, Switzerland: World Health Organization; 2009. URL: https://apps.who.int/iris/handle/10665/44188 [accessed 2021-12-20]

Abbreviations

AUC: area under the curve

CEDRE: Centre de Démoustication et de Recherche en Entomologie
DENV: dengue virus
ICD-10: International Classification of Diseases, 10th edition
PSAGE: Programme de surveillance, d'alerte et de gestion des épidémies de dengue
RSV: relative search volumes
RT-PCR: real-time reverse transcriptase polymerase chain reaction

Edited by Y Khader; submitted 08.02.22; peer-reviewed by A Lamer, C Poirier, E Luellen, N Parikh; comments to author 19.03.22; revised version received 30.06.22; accepted 22.10.22; published 22.12.22.

Please cite as:
Sylvestre E, Cécilia-Joseph E, Bouzillé G, Najioullah F, Etienne M, Malouines F, Rosine J, Julié S, Cabié A, Cuggia M
The Role of Heterogenous Real-world Data for Dengue Surveillance in Martinique: Observational Retrospective Study
JMIR Public Health Surveill 2022;8(12):e37122
URL: https://publichealth.jmir.org/2022/12/e37122
doi:10.2196/37122
PMID:36548023

Original Paper

Machine Learning Techniques to Explore Clinical Presentations of COVID-19 Severity and to Test the Association With Unhealthy Opioid Use: Retrospective Cross-sectional Cohort Study

Hale M Thompson[1,2], MA, MSc, PhD; Brihat Sharma[1], MSc; Dale L Smith[1], PhD; Sameer Bhalla[3], MD; Ihuoma Erondu[1], MD; Aniruddha Hazra[4], MD; Yousaf Ilyas[1]; Paul Pachwicewicz[1], MSc; Neeral K Sheth[5], MD; Neeraj Chhabra[6], MSc, MD; Niranjan S Karnik[1], MD, PhD; Majid Afshar[7], MSc, MD

[1]Section of Community Behavioral Health, Department of Psychiatry and Behavioral Sciences, Rush University Medical Center, Chicago, IL, United States

[2]Center for Education, Research, and Advocacy, Department of Social and Behavioral Research, Howard Brown Health, Chicago, IL, United States

[3]Department of Internal Medicine, Rush University Medical Center, Chicago, IL, United States

[4]Section of Infectious Diseases and Global Health, Department of Medicine, University of Chicago, Chicago, IL, United States

[5]Department of Psychiatry and Behavioral Sciences, Rush University Medical Center, Chicago, IL, United States

[6]Department of Emergency Medicine, Rush University Medical College, Rush University Medical Center, Chicago, IL, United States

[7]Division of Pulmonary and Critical Care, Department of Medicine, School of Medicine and Public Health, University of Wisconsin, Madison, WI, United States

Corresponding Author:
Hale M Thompson, MA, MSc, PhD
Section of Community Behavioral Health
Department of Psychiatry and Behavioral Sciences
Rush University Medical Center
Suite 302
1645 W. Jackson Boulevard
Chicago, IL, 60612
United States
Phone: 1 4153108569
Email: hale_thompson@rush.edu

Abstract

Background: The COVID-19 pandemic has exacerbated health inequities in the United States. People with unhealthy opioid use (UOU) may face disproportionate challenges with COVID-19 precautions, and the pandemic has disrupted access to opioids and UOU treatments. UOU impairs the immunological, cardiovascular, pulmonary, renal, and neurological systems and may increase severity of outcomes for COVID-19.

Objective: We applied machine learning techniques to explore clinical presentations of hospitalized patients with UOU and COVID-19 and to test the association between UOU and COVID-19 disease severity.

Methods: This retrospective, cross-sectional cohort study was conducted based on data from 4110 electronic health record patient encounters at an academic health center in Chicago between January 1, 2020, and December 31, 2020. The inclusion criterion was an unplanned admission of a patient aged ≥18 years; encounters were counted as COVID-19-positive if there was a positive test for COVID-19 or 2 COVID-19 International Classification of Disease, Tenth Revision codes. Using a predefined cutoff with optimal sensitivity and specificity to identify UOU, we ran a machine learning UOU classifier on the data for patients with COVID-19 to estimate the subcohort of patients with UOU. Topic modeling was used to explore and compare the clinical presentations documented for 2 subgroups: encounters with UOU and COVID-19 and those with no UOU and COVID-19. Mixed effects logistic regression accounted for multiple encounters for some patients and tested the association between UOU and COVID-19 outcome severity. Severity was measured with 3 utilization metrics: low-severity unplanned admission, medium-severity unplanned admission and receiving mechanical ventilation, and high-severity unplanned admission with in-hospital death. All models controlled for age, sex, race/ethnicity, insurance status, and BMI.

Results: Topic modeling yielded 10 topics per subgroup and highlighted unique comorbidities associated with UOU and COVID-19 (eg, HIV) and no UOU and COVID-19 (eg, diabetes). In the regression analysis, each incremental increase in the

classifier's predicted probability of UOU was associated with 1.16 higher odds of COVID-19 outcome severity (odds ratio 1.16, 95% CI 1.04-1.29; *P*=.009).

Conclusions: Among patients hospitalized with COVID-19, UOU is an independent risk factor associated with greater outcome severity, including in-hospital death. Social determinants of health and opioid-related overdose are unique comorbidities in the clinical presentation of the UOU patient subgroup. Additional research is needed on the role of COVID-19 therapeutics and inpatient management of acute COVID-19 pneumonia for patients with UOU. Further research is needed to test associations between expanded evidence-based harm reduction strategies for UOU and vaccination rates, hospitalizations, and risks for overdose and death among people with UOU and COVID-19. Machine learning techniques may offer more exhaustive means for cohort discovery and a novel mixed methods approach to population health.

(JMIR Public Health Surveill 2022;8(12):e38158) doi:10.2196/38158

KEYWORDS

unhealthy opioid use; substance misuse; COVID-19; severity of illness; overdose; topic modeling; machine learning; opioid use; pandemic; health outcome; public health; disease severity; electronic health record; COVID-19 outcome; risk factor; patient data

Introduction

Background

The COVID-19 pandemic has illuminated health disparities and inequities in the United States [1-3]. Chronic illness and conditions like diabetes, hypertension, cancer, autoimmune disease, and obesity, often disproportionate in aging and in uninsured populations, are associated with more severe COVID-19 outcomes [4-9]. Derived from electronic health record (EHR) data that were deidentified and aggregated on the TriNetX Research Network platform, national cohort studies have established substantial evidence of increased risks for acquiring COVID-19 and having more severe outcomes for patients with diagnosed mental health disorders or substance use disorder (SUD) [10-12]. Patients with SUD and COVID-19 have a higher odds risk for hospitalization, receiving mechanical ventilation, and mortality [11,13]. Fully vaccinated patients with SUD also have a higher odds risk for COVID-19 breakthrough infections compared to patients with no SUD [12].

Patients with opioid use disorder (OUD) often have comorbidities, such as kidney, pulmonary, liver, cardiovascular, metabolic, and immune-related disorders, that lead to disproportionate susceptibility to COVID-19 [10]. Excessive opioid use has been shown to suppress the immune system and damage the lungs, leading to an impaired respiratory system. These comorbidities could explain the observed severity of clinical outcomes in patients with OUD [11]. In one national study, patients with OUD had the greatest odds risk for breakthrough COVID-19 among those with SUD, and this disparity widened when evaluating outcomes across strata of race/ethnicity and gender. African American patients with OUD displayed an increased risk for acquisition and adverse outcomes [10,12]. Prior to the pandemic, people who misuse opioids were already experiencing the highest number of overdose deaths ever reported [1]; the pandemic has since created new and exacerbated existing disruptions in access to treatment of OUD, further accelerating the rise in overdose deaths [14-17]. COVID-19 has stressed the capacities of emergency departments (EDs) and acute care settings to conduct, for example, manual screenings for SUD, widening treatment gaps for OUD [18]. The higher risk for infection and adverse outcomes, in combination with missed treatment opportunities and increasing overdose deaths, further compounds the negative effects of the pandemic in this already vulnerable population.

Patients who misuse opioids and experience other mental health conditions may struggle with social distancing and quarantine requirements. These patients frequently experience socioeconomic and societal disadvantages that result in crowded living spaces, such as encampments, homeless shelters, and incarceration [12,19]. Stigma around opioid misuse and implicit and structural biases of the health care system could also contribute to the severity of COVID-19 clinical outcomes seen in patients with OUD [20]. Mistrust of health care providers can delay treatment-seeking at the onset of symptoms, further exacerbating illness severity [16,21]. In addition, the pandemic has disrupted access to treatments like buprenorphine, as well as access to methadone, a highly regulated medication for OUD (MOUD) that is disproportionately prescribed to Medicaid patients and may be a driver of the increase in overdose deaths [22,23].

Objective

Our recent study of unhealthy alcohol use (UAU) among our COVID-19 patients guided our current aims and our use of the term "unhealthy opioid use" (UOU) [13]. Similar to opioid misuse, people with UOU may not have an OUD diagnosis; the US Preventive Services Task Force defines UOU as the consumption of illegally obtained opioids or the nonmedical consumption of prescription opioids [24]. To discern any unique clinical presentations of UOU and COVID-19, we conducted topic modeling from the clinical notes of the EHRs of 2 subcohorts of hospitalized patient encounters: (1) UOU and COVID-19 and (2) no UOU and COVID-19. Next, we tested the association between increasing probability of UOU with increased severity of COVID-19–related health outcomes. Our findings from this novel mixed methods approach may offer more effective COVID-19 prevention and treatment pathways, as well as more effective harm reduction resources and treatment planning for UOU.

Methods

Setting and Sample

This cross-sectional study took place at Rush University Medical Center (RUMC), a large academic health center on Chicago's West Side, and was conducted with data from 4110 inpatient EHR encounters between January 1, 2020, and December 31, 2020. The inclusion criteria were an unplanned admission of a patient aged ≥18 years and a COVID-19 diagnosis. Encounters were counted as COVID-19–positive according to the National COVID Cohort Collaborative phenotype; specifically, encounters were positive if there was a documented positive test for COVID-19 or if 2 or more COVID-19–related International Classification of Disease, Tenth Revision (ICD-10) codes were recorded in a single encounter or day [25]. Using a predefined cutoff with optimal sensitivity and specificity to identify UOU, we ran our Substance Misuse and Referral to Treatment Artificial Intelligence (SMART-AI) classifier on all EHR clinical notes for patients with COVID-19 to estimate a subcohort of patients with UOU and a subcohort with no UOU.

SMART-AI for Cohort Discovery and Natural Language Processing of Clinical Notes

The SMART-AI classifier is a multi-label convolutional neural network model that was developed and tested within RUMC and externally validated at the trauma center of another local academic health system [26]. SMART-AI demonstrated good face validity, with model features containing explicit mentions of opioid misuse, and demonstrated excellent test characteristics in identifying cases of UOU when validated against the Drug Abuse Screening Test [18,26]. During temporal validation, the sensitivity and specificity for opioid misuse were 0.87 (95% CI 0.84-0.90) and 0.99 (95% CI 0.99-0.99), respectively. The positive predictive value and negative predictive value were 0.76 (95% CI 0.72-0.88) and 0.99 (95% CI 0.99-0.99), respectively. The classifier was trained as a single model with binary outputs for alcohol, opioid-drug, and nonopioid-drug misuse and allows for deactivation of any label; in this study, only the opioid label operated for the purpose of subcohort discovery among the cohort of 2020 COVID-19 hospitalized patients, and the nonopioid drug and alcohol labels were deactivated.

Natural language processing of the sample's clinical notes used the Clinical Text and Knowledge Extraction System (cTAKES) version 4.0 [27]. The cTAKES is a natural language processing system designed for knowledge extraction from the EHR clinical narrative that is scalable, comprehensive, robust, and interoperable. The cTAKES recognizes words and phrases from the clinical narrative that represent domain concepts, or named entities, in the National Library of Medicine Unified Medical Language System metathesaurus of medical ontologies. These domain concepts have been mapped from clinical notes and standardized as concept unique identifiers (CUIs).

Ethical Considerations

This study was approved by the RUMC Institutional Review Board (18061108-IRB01). Our sample was drawn from retrospective encounters documented in the EHRs; these data were deidentified for both sets of analyses and did not require informed consent.

Topic Modeling to Identify Subcohort Clinical Presentations

A domain of unsupervised machine learning, topic modeling synthesizes unwieldy textual data into more concise and deliverable concepts and organizes them into domains, or topics, based on the patterned clustering of the concepts across a data set [28,29]. In our experiment, topic modeling mined the corpus of clinical notes in the EHRs for common groupings of terms, represented as standardized medical concepts, or CUIs. When conducted for each of the 2 subcohorts, this process clustered similar and correlated concepts into topic groupings derived from clinical notes during the 2020 pandemic year, delineating key clinical differences and similarities.

We used latent Dirichlet allocation (LDA) to model the corpus of clinical data from each subcohort. Although more recent models and techniques have achieved higher accuracy, LDA is one of the most effective unsupervised probabilistic topic models for text mining based on CUIs. LDA requires a predefined number of topics to model [29], and coherence value (CV) scores for each subcohort were derived in order to identify the number of topics with the best model fit. Ten topics were determined to be optimal and parsimonious (Figure S1 and Table S1 in Multimedia Appendix 1). Similar to a scree plot in factor analysis, the point at which the CV curve initially bends or plateaus for each subcohort is an indicator of the optimal topic number.

A panel of 6 clinical experts, from 3 academic health centers, including RUMC, in psychiatry, infectious disease, addiction medicine, nursing, pulmonology/critical care, and emergency medicine convened to review and summarize the 10 topics that contained clusters of medical concepts generated for each subcohort. Each topic was presented in word cloud format in order to visually highlight the high-frequency concepts that, in aggregate, formed the core idea or topic (for the complete set of 20 word clouds, see Figure S2 in Multimedia Appendix 2) [30]. Together, the group discussed and agreed upon the emergent topic of the 10 clusters of concepts for each of the patient subcohorts. These topics were written up for the panel's review, feedback, and to confirm consensus.

Measurement and Statistical Analysis

Measurements

To assess descriptive statistics and test associations with COVID-19 outcome severity, demographic and clinical data were extracted from the EHRs. The variables included age, sex, race/ethnicity, insurance status, length of stay in days, minimum oxygen saturation level, and BMI. COVID-19 severity was measured according to the maximum level of care that a patient received: (1) low severity was an unplanned admission without receiving mechanical ventilation; (2) medium severity was an unplanned admission with receiving invasive mechanical ventilation; and (3) high severity was an unplanned admission ending in death.

Primary Outcome Analysis

In order to accommodate some repeated observations and the ordered categorical nature of how severity was measured, mixed effects ordinal logistic regression analyses with random intercepts were conducted to predict COVID-19 severity status of the 2 COVID-19 subgroups. In the first analysis, the classifier's predictive probability of UOU for each encounter with COVID-19 was regressed onto the severity outcome (ie, low, medium, or high). A higher predictive probability from the classifier indicated a greater likelihood of UOU. In the second analysis, the severity outcome was dichotomized into low (unplanned admission only) and high (unplanned admission with ventilator or in-hospital death). The classifier estimation of UOU probability was log transformed due to strong positive skew in the distribution of probabilities. All models controlled for BMI, age, sex, race/ethnicity, and insurance status. Due to sparse data, the model did not control for smoking status. We also examined interactions between classifier status and these demographic characteristics to test for potential effect modification, though we did not identify any significant interactions, and they are not reported here. Among variables used in the analysis, BMI was missing for 601/4110 (14.6%) of the COVID-19 encounters. Because BMI was not missing at random and missingness was associated with higher outcome severity, complete case analysis was used. Analyses were conducted in Stata (version 17, StataCorp LLC).

Results

Descriptive characteristics of unplanned admissions in 2020 are presented in Table 1, stratified by UOU and COVID-19 (n=102) and no UOU and COVID-19 (n=4008), with *P* values provided for the chi-square and Kruskal-Wallis tests. Compared to the no UOU subgroup, the UOU subgroup was disproportionately younger (mean age 55.6, SD 14.6 years; *P*=.001), male (68/102, 66.7%; *P*=.002), Black (71/102, 69.6%; *P*<.001), and Medicaid-insured (67/102, 65.7%; *P*<.001). This group was also disproportionately discharged against medical advice (14/102, 13.7%, *P*<.001) and had a significantly shorter average length of stay (mean 6.8, SD 7.9 days; *P*<.001). This subgroup's BMI (mean 26.3, SD 7.0 kg/m^2; *P*<.001) and minimum level of oxygen saturation (mean 81.6%, SD 11.6%; *P*=.008) were also lower.

Table 1. Sample characteristics for a cohort with unplanned admissions at a Chicago academic health center between January 1 and December 31, 2020 (N=4110), comparing those with unhealthy opioid use and COVID-19 and those with no unhealthy opioid use and COVID-19. Test statistic values represent Kruskal-Wallis tests for continuous variables and the proportion of male patients and chi-square tests for other categorical variables.

Characteristics	UOU[a] (n=102)	No UOU (n=4008)	Test statistic value (*df*)	*P* value
Age (years), mean (SD)	55.6 (14.6)	59.4 (17.4)	10.7 (1)	.001
Sex (male), n (%)	68 (66.7)	2036 (50.8)	9.4 (1)	.002
Race/ethnicity, n (%)			39.5 (3)	<.001
Black	71 (69.6)	1674 (41.7)		
White	19 (18.6)	816 (20.4)		
Hispanic or Latinx	5 (4.9)	1187 (29.6)		
Other	7 (6.8)	331 (8.2)		
Insurance status, n (%)			41.7 (3)	<.001
Medicaid	67 (65.7)	1402 (34.9)		
Medicare	20 (19.6)	1366 (34.1)		
Private	13 (12.7)	906 (22.6)		
Other	2(1.9)	334 (8.3)		
Discharge status, n (%)			147.5 (4)	<.001
Home	42 (41.2)	1810 (45.2)		
Other	24 (23.5)	1236 (30.8)		
Long- or short-term care	13 (12.7)	617 (15.4)		
In-hospital death	9 (8.8)	312 (7.7)		
Against medical advice	14 (13.7)	33 (<1)		
Length of stay (days), mean (SD)	6.8 (7.9)	8.5 (10.1)	9.9 (1)	.002
BMI (kg/m^2), mean (SD)	26.3 (7.0)	32.0 (10.3)	36.1 (1)	<.001
Minimum oxygen saturation (%), mean (SD)	81.6 (11.6)	83.4 (12.4)	6.9 (1)	.008

[a]UOU: unhealthy opioid use.

Topic Modeling

Our panel characterized the 10 topics modeled from each of the 2 EHR patient encounter subcohorts with COVID-19 in 2020 (Table 2).

For the no UOU subcohort, concepts within each topic spanned a range of symptoms, comorbidities, and procedures indicative of moderate to high severity. The first topic was deemed a "classic hospitalized COVID patient" by the expert panel of physicians and advanced practice nurses and displayed several comorbidities and procedures, such as diabetes and intubation, respectively, associated with higher severity. The second topic was related to sepsis, followed by a topic for ordering procedures associated with COVID-19. Topics 4 through 6 were long-term intensive care unit (ICU) patients, chronic obstructive airway disease, and procedures and interventions to address acute respiratory failure and hypoxia, respectively. Topics 7 through 10 were neurology-related, followed by chronic conditions associated with severe COVID-19 (eg, diabetes, coronary artery disease, and heart failure), then COVID-19–related terms indicating less severity (eg, normal limits, c-reactive protein, and myalgia), and finally conditions highly susceptible to COVID-19, like cancer and organ transplantation.

Table 2. Topic modeling for 2020 hospital admissions comparing 10 topics for 2 COVID-19 patient encounter subcohorts: those with unhealthy opioid use and those with no unhealthy opioid use (N=4110). Subcohorts were identified using the Substance Misuse and Referral to Treatment Artificial Intelligence (SMART-AI) digital classifier for opioid misuse [26]. The topic numbers are labels and do not reflect a ranking of topics.

Topic	Concepts
Unhealthy opioid use (n=102)	
1	Cardiopulmonary illnesses and social determinants of health
2	Opioid misuse comorbidities
3	Renal and cardiac pathologies, HIV-related terms
4	Neurological comorbidities with altered mental status
5	Neurological workups with cardiac disturbances
6	Critically ill/intensive care unit patients
7	Risk for overdose with cardiopulmonary and respiratory distress
8	Chronic opioid misuse with respiratory distress
9	Opioid overdose patients
10	Chronic illness and traumatic injury
No unhealthy opioid use (n=4008)	
1	Classic COVID-19 hospitalization with severity
2	Sepsis-related, less clearly COVID-19 related
3	COVID-19 orders/procedures, moderate to severe neurological orders
4	Long-term intensive care unit patients
5	Chronic obstructive pulmonary disease
6	Interventions for acute respiratory failure
7	Very neurologically focused, less COVID-19 related
8	Chronic conditions associated with severe COVID-19
9	Less severe COVID-19 symptoms and measures
10	Chronic disease highly susceptible to COVID-19

In the UOU subcohort, topics indicated illness associated with both UOU and COVID-19, as well as social determinants of health. The first topic indicated a number of cardiac and pulmonary chronic illnesses that could increase risk for COVID-19 severity, plus methadone. The second topic was characterized as UOU comorbidities and included concepts like cocaine, methadone, suboxone, and anxiety. Topic 3 was renal and cardiac pathologies with some HIV-related concepts, followed by a topic related to neurological workups and altered mental status. Concepts related to fentanyl, cocaine, Narcan, magnetic resonance imaging, and computed tomography scans of the brain had small-to-medium sized weights relative to heavily weighted concepts for cerebrovascular accidents, angiograms, hemorrhage, stenosis, and seizures. Topic 5 was also deemed to be neurological-related but with blood and cardiac disturbances present, plus methadone. Topic 6 was deemed critical illnesses or ICU patients, with concepts like malnutrition, nutrition function, cardiac arrest, and severe or moderate adverse events prominent in the word cloud. The panel characterized topic 7 as overdose risk with cardiopulmonary disorders, and respiratory and reactive airway terms, like asthma and nebulizer, appeared alongside UOU terms, such as opioids and methadone. Topic 8 was characterized as chronic UOU with respiratory distress, while topic 9 indicated opioid overdose

with 3 heavily weighted concepts: Narcan, falls, and respiratory failure. The final topic for the UOU patients was much less distinct, with a mix of chronic illness– and traumatic injury–related concepts along with unhealthy substance use–related concepts like naloxone and liver cirrhosis.

Mixed Effects Ordinal Logistic Regression

In our test for an association of UOU with COVID-19 outcome severity, each incremental increase in SMART-AI's predicted probability of UOU was associated with higher severity of outcomes (odds ratio [OR] 1.16, 95% CI 1.04-1.29; *P*=.009; Figure 1 and Table 3). Age, sex, and BMI, but not race/ethnicity or insurance status, were also associated with severity status, with male, older, and higher-BMI participants having greater risk of being in more severe categories (Table 3). Results indicating greater severity for COVID-19 patients with UOU were also robust for the dichotomization of severity level into inpatients with no ventilator use or those with either ventilator use or in-hospital death. UOU status remained a predictor of severity in the adjusted analysis (OR 1.19, 95% CI 1.12-1.26; *P*<.001) for the composite dichotomous outcome. The distribution of type of unplanned admission via ED stratified by UOU or no UOU is shown in Figure 2. For admissions with UOU, 77/102 (75%) were ED to hospital admissions, 16/102 (16%) were ED to hospital admissions requiring invasive mechanical ventilation, and 9/102 (9%) were in-hospital deaths. For admissions with no UOU, 3260/4008 (81%) were ED to hospital admissions, 436/4008 (11%) were ED to hospital admissions requiring mechanical ventilation, and 312/4008 (8%) were in-hospital deaths (see Figure 2).

Figure 1. The increased probability of unhealthy opioid use (as a continuous scale) across patient encounters that included a diagnosis of COVID-19 in 2020 (N=4110) was associated with increased outcome severity, measured by unplanned admission via the emergency department at a large Chicago hospital. ED: emergency department. ln: natural log.

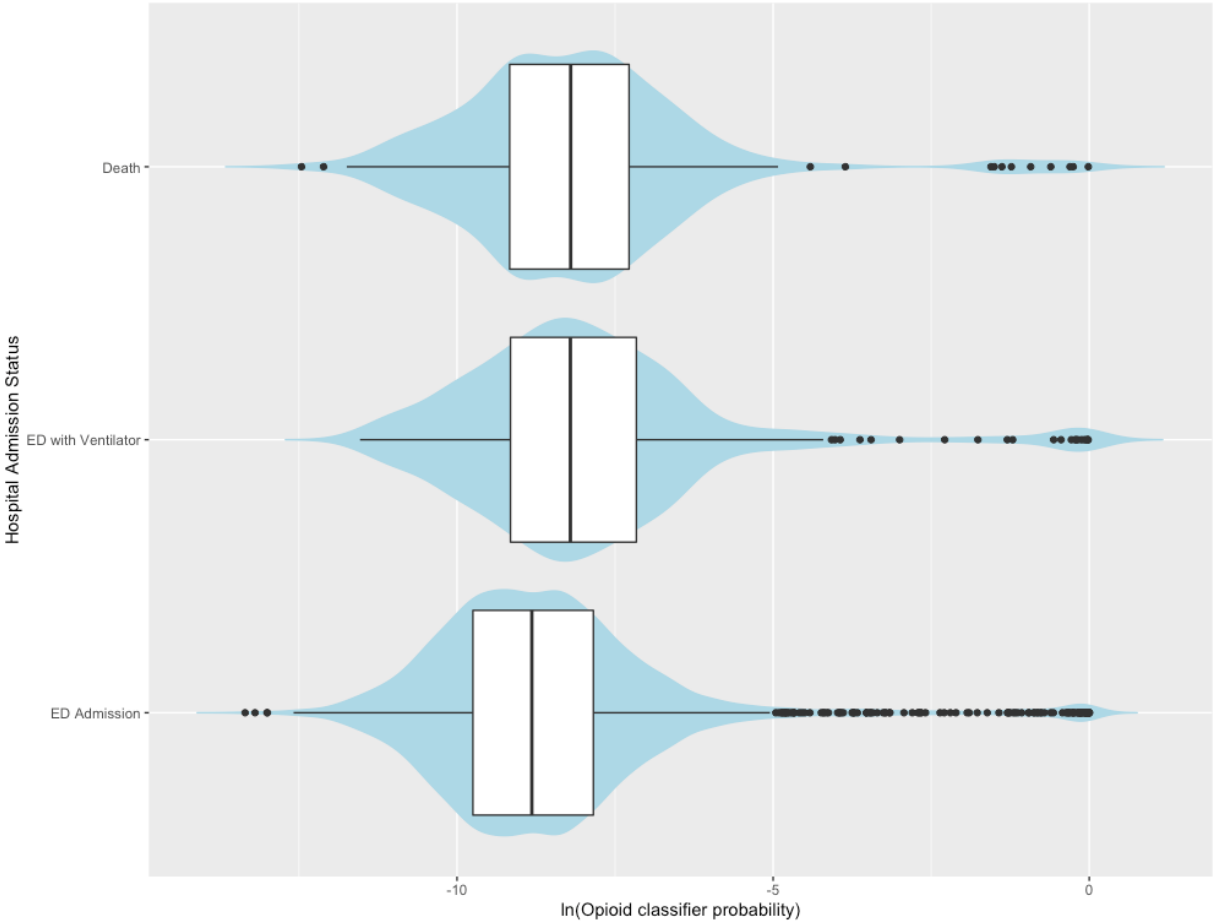


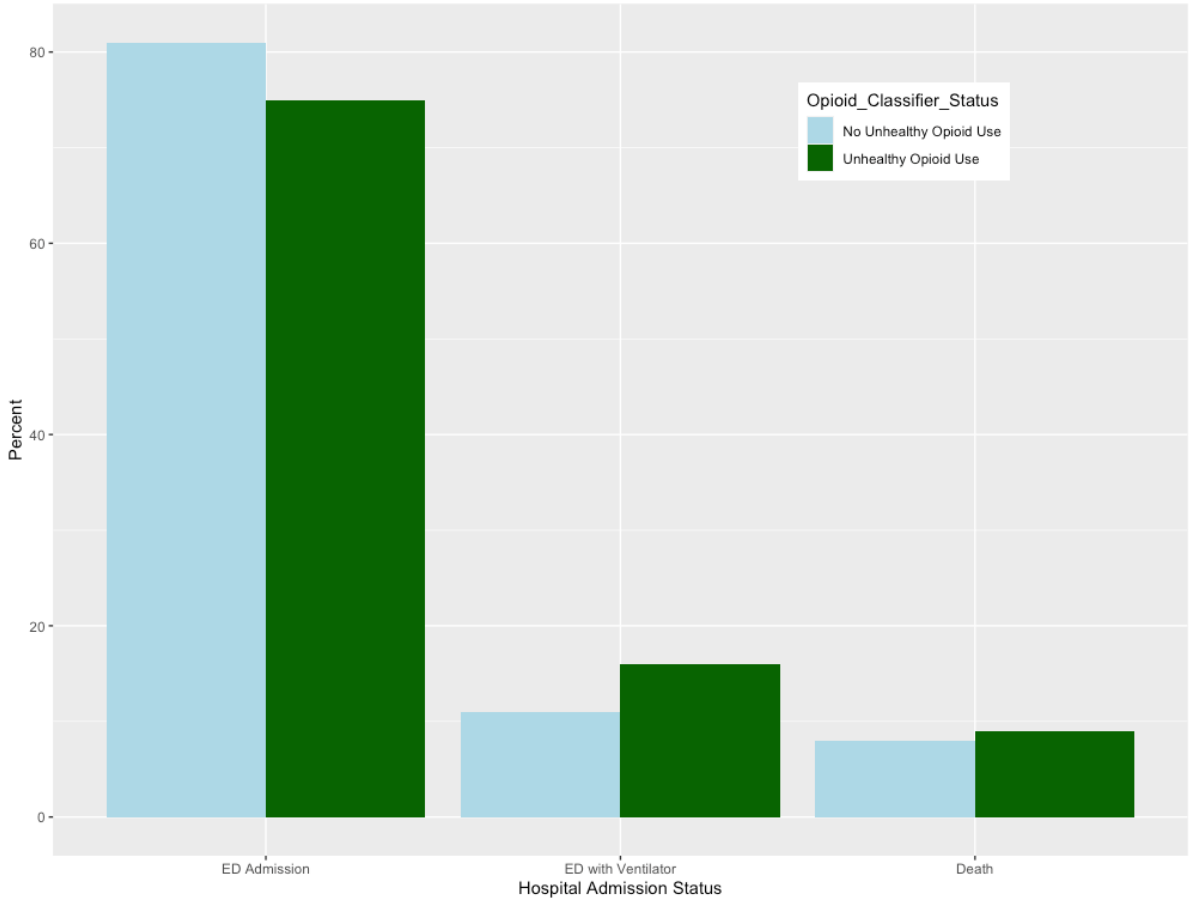
Table 3. Adjusted associations between unhealthy opioid use and outcome severity for hospitalized patient encounters carrying a diagnosis of COVID-19 in Chicago between January 1, 2020, and December 31, 2020 (N=4110).

Predictor	Odds ratio (95% CI)	*P* value
Unhealthy opioid use[a]	1.16 (1.04-1.29)	.009
Age	1.01 (1.01-1.02)	<.001
BMI	1.02 (1.01-1.03)	<.001
Sex[b]	0.75 (0.63-0.90)	.002
Race/ethnicity[b]		.21
Black	0.92 (0.73-1.18)	.52
Hispanic or Latinx	1.12 (0.84-1.45)	.39
Other	1.22 (0.86-1.72)	.27
Insurance[b]		.18
Medicare	0.87 (0.69-1.07)	.22
Private	0.79 (0.62-1.00)	.053
Other	0.78 (0.55-1.11)	.17

[a]Opioid misuse classifications were log transformed in this analysis.

[b]These rows report the *P* value for the omnibus effect for categorical predictors with more than 2 levels, and rows nested with them represent comparisons with the reference categories of male, non-Hispanic White, and Medicaid.

Figure 2. Unplanned hospital admission status via emergency department for patient encounters carrying a diagnosis for COVID-19 (N=4110) at a Chicago hospital in 2020, stratified by unhealthy opioid use and no unhealthy opioid use. ED: emergency department.



Discussion

Key Findings

Our study used SMART-AI, a validated substance misuse classifier, for UOU cohort discovery and to determine whether UOU was an independent predictor of COVID-19 outcome severity. Controlling for age, sex, race/ethnicity, insurance status, and BMI, the regression analysis demonstrated that UOU was an independent risk factor associated with increased severity of COVID-19 outcomes, measured in terms of hospital utilization. This "unhealthy opioid use" category expands the bounds for meeting the threshold for opioid misuse, traditionally a formal OUD diagnosis, and represents a unique contribution to recent studies documenting the association between OUD and COVID-19 outcome severity. As an open-source tool that has high accuracy and no major inequities across demographic subgroups for type I and II errors [26], SMART-AI is a useful and effective tool for both clinical screening and research into substance misuse. This analytic strategy integrating deep learning and unsupervised topic modeling is a novel mixed methods approach.

Our unique application of topic modeling enabled our expert panel to conduct a timely analysis of the 2020 COVID-19 patient data and to distinguish the clinical profile of COVID-19 patients hospitalized with UOU from those with COVID-19 who did not misuse opioids. Across both subgroups of COVID-19 admissions, topics reflected severity but with some distinctly different comorbidities that may have contributed to severity. The UOU subgroup had chronic and acute illnesses related to perivascular, pulmonary, HIV, and psychiatric comorbidities, as well as social determinants of health. The prominence of the Medicaid, methadone, and overdose concepts, for example, indicated a UOU subgroup with high poverty and limited access to health care and other resources who may have experienced medical emergencies due to disruptions in access to opioids or opioid treatments or increased exposure to the community spread of COVID-19 [22,23,31].

The no UOU and COVID-19 subgroup was distinguished by the presence of a sepsis topic and a topic related to less severe COVID-19 symptoms and measures. Consistent with that subgroup's higher mean BMI and older mean age were the prominence of age-related illnesses, like dementia and sepsis, and weight-related concepts, like diabetes and sleep apnea [32].

Comparisons With Other Work

Our analysis confirms the presence of a range of chronic illnesses associated with COVID-19 [2,3]. Although race/ethnicity and insurance status were not associated with severity in our analysis, this may be because COVID-19 disproportionately impacts populations on Medicaid or Medicare and Black and Latinx populations at every level of severity in our sample. Nonetheless, the UOU subgroup was disproportionately Medicaid-insured and Black. Further, the prevalence of the topic methadone, versus suboxone, across the UOU subgroup, for example, signals underresourced and underinsured patients who may experience challenges with social distancing and heightened difficulties with access to MOUDs [16]. The distinct presence of both an overdose topic and an overdose risk topic indicates that social determinants of health continue to play a role during the pandemic. The presence of these topics may also indicate disruptions in access to both MOUDs and illicit opioids; both types of disruptions may limit capacity to social distance and increase possible exposure to COVID-19 [14,31].

As with UAU, identified in a companion study conducted by members of our lab, UOU interferes with immune and respiratory functioning and may increase susceptibility to, as well as the severity of, COVID-19 [13]. Taken together, our studies' methods and findings inform a data-driven approach for timely and effective planning and deployment of resources to improve treatment pathways and outcomes for both unhealthy substance use and COVID-19 [19].

Limitations

These analyses have limitations. The use of SMART-AI for UOU subcohort discovery could have resulted in the possible misclassification of the cohorts with UOU and no UOU; although SMART-AI has high accuracy, classification also depends on the substance of the documentation in clinical notes.

The 2020 EHR encounter data predate vaccines and new variants of the virus; it is important for future research to index the evolving pandemic, vaccination rates among those with UOU, and changes in UOU and COVID-19 severity. The encounter data were cross-sectional and prevented causal inference of outcome severity. For example, the topic modeling experiment highlighted a distinct topic for opioid overdose and COVID-19. These patients may have been incidentally diagnosed with COVID-19 during hospitalization, complicating the interpretation of outcome severity as associated with COVID-19 rather than with an overdose. UOU also tends to drive higher discharges against medical advice, limiting interpretation of shorter average lengths of stay or low severity outcomes [33,34]. The regression analysis did not adjust for patient comorbidities, and the complete case analysis to address nonignorable missingness of BMI data may have inflated standard errors for the BMI covariate estimate.

Our topic modeling experiment was limited by the number of topics we chose (ie, 10). In addition to CV scores, parsimony and cognitive load for the panelists guided the determination of the optimal topic number. Although bidirectional encoder representations from transformers (BERT) has outperformed cTAKES in terms of distinguishing social and nonsocial sentences and concepts, BERT has a higher computational cost, and cTAKES protects patient privacy with the use of standardized concepts (ie, CUIs) [27,35].

Conclusion

The role of COVID-19 therapeutics and inpatient management of acute COVID-19 pneumonia remains unclear for patients with UOU [36-38]. The increased risks for severe outcomes, such as invasive ventilation and death, for patients with both COVID-19 and UOU warrants additional considerations for clinical practice and research priorities. Further research is needed to test associations between expanded evidence-based harm reduction strategies for treatment of UOU, vaccination rates, hospitalizations, and risks for overdose and death among

people with UOU and COVID-19 [22,23,39,40]. Machine learning techniques may offer more exhaustive means of cohort discovery and a novel mixed methods approach to population health.

Acknowledgments

This research was supported by the National Institutes of Health (award UG1DA049467). Additional support was provided by a COVID-19 grant supplement and research grants from the Agency for Healthcare Research and Quality (3K12HS026385-03S1), the National Institute on Drug Abuse (R01DA051464, 5R25DA035692-08), and Rush University Medical Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality or the National Institutes of Health.

Authors' Contributions

HMT contributed to funding acquisition, study design, analysis, and writing of the original draft and to reviewing and editing the manuscript. BS contributed to data curation, formal analysis, software creation, manuscript review, and editing. DLS contributed to formal analysis, contributing to the original draft results, and reviewing and editing the manuscript. SB contributed to manuscript review and writing. YI and PP contributed to data curation and manuscript review and editing. IE contributed to research and writing of the draft background and reviewing and editing the manuscript. AH, NKS, and NC contributed to analysis and review and editing of the manuscript. MA and NSK contributed to conceptualization, funding acquisition, methodology, analysis, and manuscript review and writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Topic modelling visualization (Figure S1) illustrating the change in coherence value scores per increase in five topics (Table S1), based on EHR data of two subgroups of unplanned admissions at Chicago academic health center in 2020: 1) COVID-19 patients with unhealthy opioid use, and 2) COVID-19 patients with no unhealthy opioid use.

[DOCX File , 41 KB - publichealth_v8i12e38158_app1.docx]

Multimedia Appendix 2

Topic modelling experiment to identify ten topics regarding the clinical presentations for each of two subgroups (n=4,110) of a cohort of unplanned admissions at a Chicago academic health center between January 1 and December 31, 2020 (N = 32,635): 1) admissions with COVID-19 and unhealthy opioid use (COV-UOU, n=102) and 2) admissions with COVID-19 and no unhealthy opioid use (COV-NO-UOU, n=4,008).

[PPTX File , 7441 KB - publichealth_v8i12e38158_app2.pptx]

References

1. Volkow ND, Blanco C. Research on substance use disorders during the COVID-19 pandemic. J Subst Abuse Treat 2021 Oct;129:108385 [FREE Full text] [doi: 10.1016/j.jsat.2021.108385] [Medline: 34080553]
2. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, the Northwell COVID-19 Research Consortium, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA 2020 May 26;323(20):2052-2059 [FREE Full text] [doi: 10.1001/jama.2020.6775] [Medline: 32320003]
3. Schwartz K. Trends in overall and non-COVID-19 hospital admissions. Oncology Practice Management 2020 Dec;10 [FREE Full text]
4. Yang F, Shi S, Zhu J, Shi J, Dai K, Chen X. Clinical characteristics and outcomes of cancer patients with COVID-19. J Med Virol 2020 Oct;92(10):2067-2073. [doi: 10.1002/jmv.25972] [Medline: 32369209]
5. Giannakoulis VG, Papoutsi E, Siempos II. Effect of cancer on clinical outcomes of patients with COVID-19: a meta-analysis of patient data. JCO Glob Oncol 2020 Jun;6:799-808 [FREE Full text] [doi: 10.1200/GO.20.00225] [Medline: 32511066]
6. Zhang H, Han H, He T, Labbe K, Hernandez A, Chen H, et al. Clinical characteristics and outcomes of COVID-19-infected cancer patients: a systematic review and meta-analysis. J Natl Cancer Inst 2021 Apr 06;113(4):371-380 [FREE Full text] [doi: 10.1093/jnci/djaa168] [Medline: 33136163]
7. Tan EH, Sena AG, Prats-Uribe A, You SC, Ahmed WUR, Kostka K, et al. COVID-19 in patients with autoimmune diseases: characteristics and outcomes in a multinational network of cohorts across three countries. Rheumatology (Oxford) 2021 Oct 09;60(SI):SI37-SI50 [FREE Full text] [doi: 10.1093/rheumatology/keab250] [Medline: 33725121]

8. Imam Z, Odish F, Gill I, O'Connor D, Armstrong J, Vanood A, et al. Older age and comorbidity are independent mortality predictors in a large cohort of 1305 COVID-19 patients in Michigan, United States. J Intern Med 2020 Oct 22;288(4):469-476 [FREE Full text] [doi: 10.1111/joim.13119] [Medline: 32498135]
9. Du Y, Lv Y, Zha W, Zhou N, Hong X. Association of body mass index (BMI) with critical COVID-19 and in-hospital mortality: A dose-response meta-analysis. Metabolism 2021 Apr;117:154373 [FREE Full text] [doi: 10.1016/j.metabol.2020.154373] [Medline: 32949592]
10. Wang Q, Xu R, Volkow ND. Increased risk of COVID-19 infection and mortality in people with mental disorders: analysis from electronic health records in the United States. World Psychiatry 2021 Feb;20(1):124-130 [FREE Full text] [doi: 10.1002/wps.20806] [Medline: 33026219]
11. Baillargeon J, Polychronopoulou E, Kuo Y, Raji MA. The impact of substance use disorder on COVID-19 outcomes. Psychiatr Serv 2021 May 01;72(5):578-581 [FREE Full text] [doi: 10.1176/appi.ps.202000534] [Medline: 33138712]
12. Wang L, Wang Q, Davis PB, Volkow ND, Xu R. Increased risk for COVID-19 breakthrough infection in fully vaccinated patients with substance use disorders in the United States between December 2020 and August 2021. World Psychiatry 2022 Feb;21(1):124-132 [FREE Full text] [doi: 10.1002/wps.20921] [Medline: 34612005]
13. Bhalla S, Sharma B, Smith D, Boley R, McCluskey C, Ilyas Y, et al. Investigating unhealthy alcohol use as an independent risk factor for increased COVID-19 disease severity: observational cross-sectional study. JMIR Public Health Surveill 2021 Nov 05;7(11):e33022 [FREE Full text] [doi: 10.2196/33022] [Medline: 34665758]
14. Sun Y, Bao Y, Kosten T, Strang J, Shi J, Lu L. Editorial: challenges to opioid use disorders during COVID-19. Am J Addict 2020 May;29(3):174-175 [FREE Full text] [doi: 10.1111/ajad.13031] [Medline: 32281130]
15. Goedel WC, Shapiro A, Cerdá M, Tsai JW, Hadland SE, Marshall BDL. Association of racial/ethnic segregation with treatment capacity for opioid use disorder in counties in the United States. JAMA Netw Open 2020 Apr 01;3(4):e203711 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3711] [Medline: 32320038]
16. Santoro T, Santoro J. Racial bias in the US opioid epidemic: a review of the history of systemic bias and implications for care. Cureus 2018 Dec 14;10(12):e3733 [FREE Full text] [doi: 10.7759/cureus.3733] [Medline: 30800543]
17. Ahmad FB, Cisewski JA, Rossen LM, Sutton PP. Provisional drug overdose death counts. National Center for Health Statistics. 2022. URL: https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm#citation [accessed 2022-01-24]
18. Afshar M, Sharma B, Bhalla S, Thompson HM, Dligach D, Boley RA, et al. External validation of an opioid misuse machine learning classifier in hospitalized adult patients. Addict Sci Clin Pract 2021 Mar 17;16(1):19 [FREE Full text] [doi: 10.1186/s13722-021-00229-7] [Medline: 33731210]
19. Blanco C, Compton WM, Volkow ND. Opportunities for research on the treatment of substance use disorders in the context of COVID-19. JAMA Psychiatry 2020 Sep 01:357-358. [doi: 10.1001/jamapsychiatry.2020.3177] [Medline: 32870241]
20. Thompson H, Sharma B, Bhalla S, Boley R, McCluskey C, Dligach D, et al. Bias and fairness assessment of a natural language processing opioid misuse classifier: detection and mitigation of electronic health record data disadvantages across racial subgroups. J Am Med Inform Assoc 2021 Oct 12;28(11):2393-2403 [FREE Full text] [doi: 10.1093/jamia/ocab148] [Medline: 34383925]
21. Alsan M, Wanamaker M, Hardeman RR. The Tuskegee study of untreated syphilis: a case study in peripheral trauma with implications for health professionals. J Gen Intern Med 2020 Jan;35(1):322-325 [FREE Full text] [doi: 10.1007/s11606-019-05309-8] [Medline: 31646456]
22. Joudrey PJ, Adams ZM, Bach P, Van Buren S, Chaiton JA, Ehrenfeld L, et al. Methadone access for opioid use disorder during the COVID-19 pandemic within the United States and Canada. JAMA Netw Open 2021 Jul 01;4(7):e2118223 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.18223] [Medline: 34297070]
23. Friedman J, Beletsky L, Schriger DL. Overdose-related cardiac arrests observed by emergency medical services during the US COVID-19 epidemic. JAMA Psychiatry 2021 May 01;78(5):562-564 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4218] [Medline: 33270086]
24. US Preventive Services Task Force, Krist AH, Davidson KW, Mangione CM, Barry MJ, Cabana M, et al. Screening for unhealthy drug use: US Preventive Services Task Force recommendation statement. JAMA 2020 Jun 09;323(22):2301-2309. [doi: 10.1001/jama.2020.8020] [Medline: 32515821]
25. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, N3C Consortium. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. J Am Med Inform Assoc 2021 Mar 01;28(3):427-443 [FREE Full text] [doi: 10.1093/jamia/ocaa196] [Medline: 32805036]
26. Afshar M, Sharma B, Dligach D, Oguss M, Brown R, Chhabra N, et al. Substance Misuse Algorithm for Referral to Treatment Using Artificial Intelligence (SMART-AI): Multi-Modal Validation with Interpretation and Bias Assessment. SSRN Preprint posted online on September 21, 2021. [doi: 10.2139/ssrn.3922677]
27. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]
28. Albalawi R, Yeap TH, Benyoucef M. Using topic modeling methods for short-text data: a comparative analysis. Front Artif Intell 2020;3:42 [FREE Full text] [doi: 10.3389/frai.2020.00042] [Medline: 33733159]

29. Chakkarwar V, Tamane SC. Quick insight of research literature using topic modeling. In: Zhang YD, Mandal J, So-In C, Thakur N, editors. Smart Trends in Computing and Communications. Singapore: Springer; 2020:189-197.
30. Xu A, Qi T, Dong X. Analysis of the Douban online review of the MCU: based on LDA topic model. In: J Phys Conf Ser. 2020 Jan 01 Presented at: 2nd International Symposium on Big Data and Applied Statistics (ISBDAS2019); Sep 20-22, 2019; Dalian, China p. 012102. [doi: 10.1088/1742-6596/1437/1/012102]
31. Park JN, Rouhani S, Beletsky L, Vincent L, Saloner B, Sherman SG. Situating the continuum of overdose risk in the social determinants of health: a new conceptual framework. Milbank Q 2020 Sep;98(3):700-746 [FREE Full text] [doi: 10.1111/1468-0009.12470] [Medline: 32808709]
32. Kalani C, Venigalla T, Bailey J, Udeani G, Surani S. Sepsis patients in critical care units with obesity: is obesity protective? Cureus 2020 Feb 10;12(2):e6929 [FREE Full text] [doi: 10.7759/cureus.6929] [Medline: 32190482]
33. Thompson H, Hill K, Jadhav R, Webb T, Pollack M, Karnik N. The substance use intervention team: a preliminary analysis of a population-level strategy to address the opioid crisis at an academic health center. J Addict Med 2019;13(6):460-463 [FREE Full text] [doi: 10.1097/ADM.0000000000000520] [Medline: 31689260]
34. Thompson HM, Faig W, VanKim NA, Sharma B, Afshar M, Karnik NS. Differences in length of stay and discharge destination among patients with substance use disorders: The effect of Substance Use Intervention Team (SUIT) consultation service. PLoS One 2020;15(10):e0239761 [FREE Full text] [doi: 10.1371/journal.pone.0239761] [Medline: 33035229]
35. Han S, Zhang RF, Shi L, Richie R, Liu H, Tseng A, et al. Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing. J Biomed Inform 2022 Mar;127:103984. [doi: 10.1016/j.jbi.2021.103984] [Medline: 35007754]
36. Kurtz P, Bastos LSL, Dantas LF, Zampieri FG, Soares M, Hamacher S, et al. Evolving changes in mortality of 13,301 critically ill adult patients with COVID-19 over 8 months. Intensive Care Med 2021 May;47(5):538-548 [FREE Full text] [doi: 10.1007/s00134-021-06388-0] [Medline: 33852032]
37. Perkins GD, Ji C, Connolly BA, Couper K, Lall R, Baillie JK, RECOVERY-RS Collaborators. Effect of noninvasive respiratory strategies on intubation or mortality among patients with acute hypoxemic respiratory failure and COVID-19: The RECOVERY-RS randomized clinical trial. JAMA 2022 Feb 08;327(6):546-558 [FREE Full text] [doi: 10.1001/jama.2022.0028] [Medline: 35072713]
38. Weerakkody S, Arina P, Glenister J, Cottrell S, Boscaini-Gilroy G, Singer M, et al. Non-invasive respiratory support in the management of acute COVID-19 pneumonia: considerations for clinical practice and priorities for research. Lancet Respir Med 2022 Feb;10(2):199-213 [FREE Full text] [doi: 10.1016/S2213-2600(21)00414-8] [Medline: 34767767]
39. Goldman JE, Waye KM, Periera KA, Krieger MS, Yedinak JL, Marshall BDL. Perspectives on rapid fentanyl test strips as a harm reduction practice among young adults who use drugs: a qualitative study. Harm Reduct J 2019 Jan 08;16(1):3 [FREE Full text] [doi: 10.1186/s12954-018-0276-0] [Medline: 30621699]
40. Alegría M, NeMoyer A, Falgàs Bagué I, Wang Y, Alvarez K. Social determinants of mental health: where we are and where we need to go. Curr Psychiatry Rep 2018 Sep 17;20(11):95 [FREE Full text] [doi: 10.1007/s11920-018-0969-9] [Medline: 30221308]

Abbreviations

BERT: bidirectional encoder representations from transformers
cTAKES: Clinical Text and Knowledge Extraction System version 4.0
CUI: concept unique identifier
CV: coherence value
ED: emergency department
EHR: electronic health record
ICD-10: International Classification of Disease, Tenth Revision
ICU: intensive care unit
LDA: latent Dirichlet allocation
MOUD: medication for opioid use disorder
OR: odds ratio
OUD: opioid use disorder
RUMC: Rush University Medical Center
SMART-AI: Substance Misuse and Referral to Treatment Artificial Intelligence
SUD: substance use disorders
UAU: unhealthy alcohol use
UOU: unhealthy opioid use

Edited by A Mavragani, G Eysenbach; submitted 21.03.22; peer-reviewed by W Ceron, M Gong, L Zhang; comments to author 02.05.22; revised version received 23.05.22; accepted 18.10.22; published 08.12.22.

Please cite as:
Thompson HM, Sharma B, Smith DL, Bhalla S, Erondu I, Hazra A, Ilyas Y, Pachwicewicz P, Sheth NK, Chhabra N, Karnik NS, Afshar M
Machine Learning Techniques to Explore Clinical Presentations of COVID-19 Severity and to Test the Association With Unhealthy Opioid Use: Retrospective Cross-sectional Cohort Study
JMIR Public Health Surveill 2022;8(12):e38158
URL: https://publichealth.jmir.org/2022/12/e38158
doi:10.2196/38158
PMID:36265163

Review

The Relationship Between Physical Activity and Mobile Phone Addiction Among Adolescents and Young Adults: Systematic Review and Meta-analysis of Observational Studies

Wen Xiao[1*], MA; Jinlong Wu[2*], MA; Joanne Yip[2], PhD; Qiuqiong Shi[2], PhD; Li Peng[3], PhD; Qiwen Emma Lei[2], BA; Zhanbing Ren[1], PhD

[1]College of Physical Education, Shenzhen University, Shenzhen, China

[2]School of Fashion and Textiles, Hong Kong Polytechnic University, Hong Kong, China

[3]College of Physical Education, Southwest University, Chongqing, China

[*]these authors contributed equally

Corresponding Author:
Zhanbing Ren, PhD
College of Physical Education
Shenzhen University
No 3588 Nanhai Road
Nanshan District
Shenzhen, 518060
China
Phone: 86 0755 265347
Email: rzb@szu.edu.cn

Abstract

Background: Previous studies have reported a potential negative correlation between physical activity (PA) and mobile phone addiction (MPA) among adolescents and young adults. To date, the strength of this correlation has not been well characterized.

Objective: This review and meta-analysis aimed to synthesize available empirical studies to examine the correlations between PA and MPA among adolescents and young adults. We also explored several potential moderators, including time of data collection, country or region, and type of population, associated with the relationship between PA and MPA.

Methods: Four electronic databases (PubMed, Scopus, PsycINFO, and Web of Science) were searched from database inception to March 2022 to identify relevant studies. The pooled Pearson correlation coefficients and their corresponding 95% CIs for the relationship between PA and MPA were calculated using the inverse variance method. The methodological quality of the included cross-sectional studies was determined based on the Joanna Briggs Institute appraisal checklist. The study conformed to the PRISMA (Preferred Reporting Items for Systematic Review and Meta-analyses) guidelines.

Results: In total, 892 relevant articles were identified, of which 22 were selected based on the inclusion and exclusion criteria. The final meta-analysis included 17 of the 22 studies. Results of random effects modeling revealed a moderate correlation between PA and MPA among adolescents and young adults (summary r=–0.243, P<.001). Sensitivity and publication bias analyses further demonstrated the robustness of our results. All the included studies were scored as high quality with a low risk of bias. Subgroup analysis further indicated that none of the hypothesized moderators (time of data collection, country or region, and type of population) significantly affected the relationship between PA and MPA, as confirmed by the mixed effects analysis. In addition, in the data collection subgroups, medium effect sizes were obtained for data collected before COVID-19 (r=–0.333, P<.001) and data collected during COVID-19 (r=–0.207, P<.001). In subgroup analyses for country or region, the correlation coefficient for China and other developing regions showed a similarly moderate effect size (r=–0.201, P<.001 and r= –0.217, P<.001, respectively). However, the effect sizes for developed regions were not significant (r=–0.446, P=.39). In a subgroup analysis based on the type of population, we found that the effect size for young adults was moderate (r=–0.250, P<.001). However, that of adolescents was not significant (r=–0.129, P=.24).

Conclusions: Our results demonstrate a moderately negative relationship between PA and MPA among young adults. The strength of this relationship was not influenced by the time of data collection, country or region, or type of population.

(JMIR Public Health Surveill 2022;8(12):e41606) doi:10.2196/41606

KEYWORDS

mobile phone addiction; physical activity; adolescents; young adults; systematic review; phone addiction; association; correlation; phone use

Introduction

Mobile phone addiction (MPA) is defined as an addictive behavior in which individuals show uncontrollable use of mobile phones that severely impairs their physical, psychological, and social functions [1,2]. Generally, MPA is considered a negative behavior that is socially inappropriate, or even hazardous, in circumstances such as driving, walking, and unauthorized live streaming [3]. It is also categorized as a behavioral addiction (ie, a nonsubstance addiction) that can potentially cause physical, emotional, and financial harm [4,5].

Previous epidemiological surveys of MPA in different countries and regions in the past 5 years have revealed a high rate of MPA among adolescents and young adults. Recent surveys have also found that the rate of MPA among Brazilian adolescents aged 15 to 18 years was approximately 70.3% [6]. The rate of MPA among college students in Hainan province in China aged 18 to 26 years was 40.5% [7], the rate among Egyptian college students with average age of 18 to 21 years was 64.2% [8], and the rate among college students in a regional city in India with a mean age of 20.1 (SD 1.3) years was 39% to 44% [9].

Numerous investigations have demonstrated that MPA negatively affects mental health by causing anxiety [10] and depression [11,12], affecting sleep quality [13-15] and cognitive function [16], and causing muscle pain [17,18], thereby affecting work productivity and the quality of life of individuals. Thus, MPA is now considered an important worldwide public health topic [19]. MPA has been exacerbated by the spread of SARS-CoV-2 in the recent past and restrictions imposed on social gatherings. This has caused negative psychological effects (eg, anxiety, depression, frustration, fear, and stress) in many individuals [20,21]. As a consequence, the overuse of smartphones, social media, and video gaming has increased as people have found mobile phones to be a coping mechanism to alleviate negative emotions [22,23]. It has been shown that adolescents and young adults are more likely to use mobile phones excessively [24] due to their mental immaturity and lower ability to self-regulate compared to middle-aged and older adults [25,26].

A variety of factors that influence MPA have been explored to develop interventions for preventing MPA in young populations, including physical activity (PA). Data show that PA has broad health benefits, including prolonged life expectancy and better physical and psychological well-being [27]. The World Health Organization recommends that adolescents participate in at least 60 minutes of moderate to vigorous PA daily and 150 to 300 minutes of moderate to vigorous PA per week [28]. A previous meta-analysis showed that PA, including tai chi, basketball, badminton, dance, running, and bicycling, had positive effects on individuals with smartphone addiction [29].

Some cross-sectional studies have predicted that higher levels of PA may reduce rates of MPA among adolescents and young adults, suggesting that there might be a negative correlation between PA and MPA [30-32]. A study in China found a significant negative correlation between PA and MPA in adolescents (ie, people aged 10 to 19 years), which indicates that active participation in PA is a potential strategy to reduce MPA levels [33]. Similar findings were obtained in a study of young adults [30,34]. However, a weak relationship has been reported between PA and MPA among young adults aged 18 to 24 years in other research [35]. Physical inactivity (ie, sedentariness) has been demonstrated to increase the risk of MPA due to the prolonged use of mobile phones. There is evidence that sedentary behaviors and low PA levels are strong predictors of time spent using smartphones [36-38] in adolescents and adults.

To the best of our knowledge, no systematic review and meta-analysis has been conducted to examine the correlation between PA and MPA. Thus, an up-to-date literature review of previous findings on the relationship between PA and MPA is needed. This review identified three knowledge gaps. First, previous findings regarding the strength of the correlation coefficient between PA and MPA in adolescents and young adults are inconsistent. Only one, small-scale systematic review [39] reported a negative correlation between PA and MPA in adolescents. This finding cannot be explained without a quantitative analysis [39]. Second, during the COVID-19 pandemic, isolation policies reduced outdoor PA and increased psychological stress among young adults, which may have increased MPA. However, whether the correlation between PA and MPA was influenced by the pandemic is unclear. Third, as mentioned above, the prevalence of MPA differs across countries and regions. Nevertheless, the question of whether the correlation between PA and MPA is influenced by country or region has remained underexplored.

Therefore, this systematic review and meta-analysis is timely. We sought to examine the overall correlation between PA and MPA and address an important research topic. Furthermore, factors such as the time of data collection (ie, before or during COVID-19), country or region, and type of population (adolescents and young adults) are potential variables influencing the correlation between PA and MPA that we explored and examined with a subgroup analysis.

Methods

Protocol Registration

This systematic review and meta-analysis was conducted in line with the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) guidelines [40].

Search Strategy

We searched 4 electronic databases (PubMed, Scopus, PsycINFO, and Web of Science) from database inception until March 26, 2022, to identify relevant studies. A manual search was conducted of the retrieved publications to identify potentially missing studies. The search strategy consisted of 2

strings of keywords, including PA- and MPA-related terms. These included the following: ("cell phone" OR "cell phones" OR "cellular phone" OR "cellular phones" OR "cellular telephone" OR "cellular telephones" OR "mobile devices" OR "mobile phone" OR "smart phone" OR "smartphone") AND ("addiction" OR "dependence" OR "dependency" OR "abuse" OR "addicted to" OR "overuse" OR "problem use" OR "compensatory use") OR ("problematic smartphone use" OR "problematic smart phone use" OR "problematic mobile phone use" OR "problematic cell phone use" OR "problematic cellular phone use" OR "Nomophobia" OR "Phubbing" OR "fear of missing out" OR "FoMO" OR "smartphone separation anxiety" OR "smartphone use disorder" OR "compulsive mobile phone use") AND ("physical activity" OR "walk*" OR "exercise*" OR "physical activity*" OR "strength training" OR "resistance training" OR "resistance exercise*" OR "conditioning muscle" OR "training" OR "leisure training" OR "leisure activities" OR "physical fitness" OR "motor activity"). The detailed search strategy is presented in Multimedia Appendix 1. We manually performed a complementary Google search using the abovementioned keyword combinations to broaden the results on September 20, 2022. Secondary searches were performed by manually screening reference lists of included studies and tracking cited articles to ensure no relevant study was omitted.

The identified and retrieved studies were imported into EndNote X7 software (Thompson Reuters). Duplicates were excluded using the deduplication function in Endnote. This screening and processing was conducted by 2 reviewers, who independently read the titles and abstracts and assessed the studies against predetermined inclusion criteria. The full text of the included studies was also independently examined by the 2 reviewers. Inclusion checklists were completed for each study, along with details on the decision to exclude. The reference list of each included study and the articles cited were thoroughly reviewed to ensure that no relevant studies were missed. At all stages, any discrepancies in the results obtained were resolved through consensus or by involving a third reviewer.

Inclusion Criteria and Study Selection

Population

A study was deemed eligible if it included healthy adolescents or young adults aged between 11 and 24 years [41].

Exposure and Outcome

Data on PA were collected using measurement tools that included self-reported scales, questionnaires, and accelerometers. Data on different aspects of PA, such as steps taken; time spent each day engaging in light, moderate, and vigorous PA; and PA in different scenarios (ie, for leisure, with family, during active travel, or for work) were recorded. Measurements of MPA levels were collected using internationally used scales or questionnaires (eg, the MPA tendency scale, the mobile phone addiction tendency scale, or the smartphone addiction scale). The contents of the MPA measurement questionnaires or other questionnaires were required to include withdrawal, loss of control and escape, and other MPA symptoms. Studies that only provided data on the duration of mobile phone use were excluded.

Study Design

Quantitative observational (cross-sectional and cohort/longitudinal) studies were included.

Other Criteria

Studies were included if they were published in peer-reviewed journals and were written in English. If 2 studies were based on the same data set, the study published earlier was selected for inclusion in the review.

Exclusion Criteria

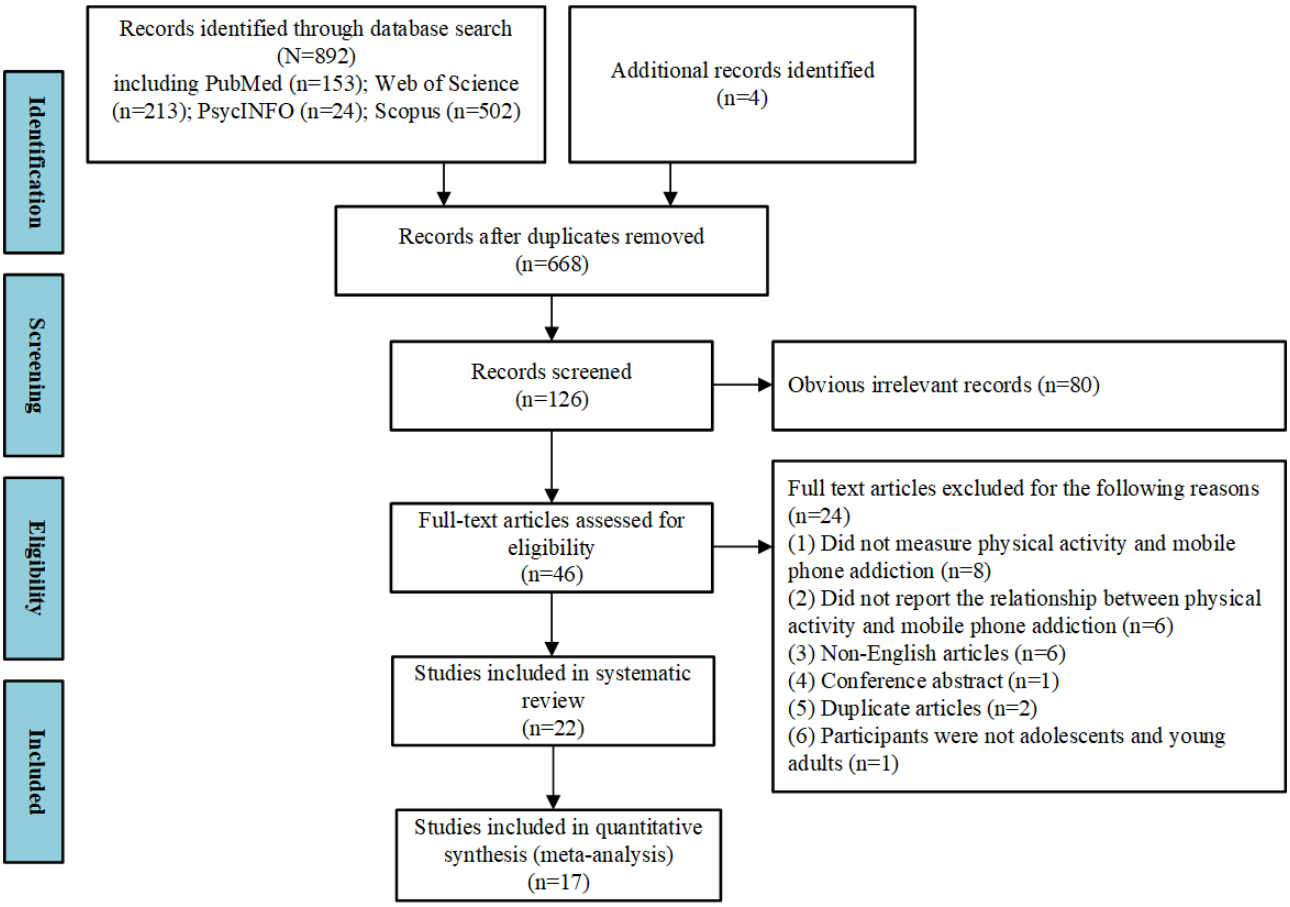
Case-control studies were excluded because they examined specific groups that were beyond the scope of this review. Furthermore, reviews, meta-analyses, commentaries, replies, clinical guidelines, conference abstracts, theses, and book chapters were also excluded.

Data Extraction and Synthesis

A total of 892 studies were identified by reading the titles and abstracts. Among these, 46 candidate studies were identified after reading their full text. At this stage, 24 studies were excluded based on the above criteria. The remaining 22 studies were deemed eligible and included in the systematic review. The final meta-analysis included 17 of the 22 studies (Figure 1 shows the details of the article screening process).

Two reviewers independently extracted data from the included articles and entered the data into a form tailored to the requirements of this review. The extracted data included (1) publication details (author, year, and country); (2) sample characteristics (sample size, sex of participants, and type of participant); (3) time of data collection; (4) measurements of PA and MPA; and (5) the main study outcome (ie, correlation coefficient).

Figure 1. Flow diagram of article screening process.



Methodological Quality Assessment

The Joanna Briggs Institute (JBI) appraisal checklist, which has 10 items, was used to examine the methodological quality of the included cross-sectional studies [42]. The studies were given a score of 0 to 2 for each item. Studies with an overall score higher than 70% were considered high quality with a low risk of bias. Details of the scoring criteria applied in the JBI appraisal checklist are presented in Multimedia Appendix 2.

Data Analysis

All statistical analyses were conducted with Comprehensive Meta-Analysis software (version 3; Biostat Inc).

All data were extracted from the included studies. The pooled Pearson correlation coefficients (with the corresponding 95% CIs) between PA and MPA were calculated with the inverse variance method. Subsequently, the Pearson correlation coefficients were transformed to Fisher *z* scores before the pooled estimate was obtained to calculate variance-stabilized correlation coefficients, as described previously [43] The effect sizes were interpreted in line with recent suggestions concerning correlations for psychometrics with *r*: small (*r*=0.10-0.20), medium (*r*=0.21-0.35), and large (*r*>0.35) [44].

The Cochran Q test and the I^2 statistic were employed to measure heterogeneity across studies. The Cochran Q determines the conformity to the normal distribution of effect sizes. A significant value (P<.10) indicates heterogeneity. I^2 is an estimate of the ratio of true heterogeneity in the observed variation. I^2>50% reflects statistically significant between-study heterogeneity [45]. For such studies, the random effects model was used to calculate the summary of the Pearson correlation coefficients with a P value <.05 or I^2>50% [44]. Otherwise, the fixed effects model was used [44].

To determine potential moderators of heterogeneity, subgroup analyses were carried out for country or region, population (college students and adolescents), and time of data collection (before or during COVID-19). All subgroup analyses were conducted with a mixed effects analysis. The random effects model was used to summarize the studies within the respective subgroups, and the fixed effects model was used to test for significant differences between the subgroups [46]. Full details of coding forms for the subgroups are provided in Multimedia Appendix 3.

To determine the influence of individual studies on the summary correlation coefficients and test the robustness of the correlations between PA and MPA, sensitivity analyses were conducted by sequentially omitting one study at time [11].

Funnel plots were established to determine the existence of potential publication bias. Additionally, the Begg rank correlation test and Egger linear regression test were performed to determine publication bias, with P<.05 indicating significant publication bias [47,48]. In the case of publication bias, the trim-and-fill method was used to adjust for funnel plot asymmetry [11].

Other statistical analyses performed included valid measures of the association between PA and MPA, measured with the correlation coefficient (*r*), standardized regression coefficient (b), unstandardized regression coefficient (β), odds ratio (OR), mean, and SD. To include as many eligible studies as possible, several data transformation steps were used. For studies that reported the mean and SD, the Cohen *d* effect size was calculated and converted to a correlation [49]. Studies that reported relevant ORs with 95% CIs were converted to Cohen *d* effect estimates and then to correlations [50]. The authors of the eligible studies were contacted if potentially relevant data were missing.

Results

Descriptive Characteristics

Table 1 presents a summary of the characteristics of the included studies. Overall, 23,365 participants aged between 15 and 26 years were included. Eighteen studies (numbers 1 to 22) were included in the systematic review, and 17 studies (numbers 1 to 17) were included in the meta-analysis. Moreover, 17 studies reported a correlation between PA and MPA. Considering the high heterogeneity among studies (Q=468.050, *P*<.001; I^2=96.582), the random effects model was used to estimate the effect size of summary *r* (*r*=–0.243; 95% CI –0.309 to –0.175; *P*<.001; Table 2 and Figure 2). This result showed that PA was moderately negatively correlated with MPA.

Table 1. Characteristics of the studies included in the review.

Study	Country	Size, n	Male, n	Population	Age (years)	Time period	MPA[a] measurement	PA[b] measurement	*r*
1. Kim et al, 2015 [30]	South Korea	110	67	College students	Mean 21.03 (SD 1.61)	2015	SAPS[c]	3D sensor pedometer	–0.798
2. Haug et al, 2015 [34]	Switzerland	1519	732	Adolescents	Range 16-21	Feb 2015 to Jun 2015	SAS-SV[d]	"Outside school: How many hours a week do you exercise or participate in sports that make you sweat or become out of breath?"	–0.019
3. Yang et al, 2019 [32]	China	608	158	College students	—[e]	Dec 2018 to Jan 2019	MPATS[f]	PARS-3[g]	–0.124
4. Haripriya et al, 2019 [51]	India	113	63	College students	Mean 22.15 (SD 1.69)	Apr 2019 to May 2019	SAPS	IPAQ-SF[h]	–0.335
5. Numanoğlu-Akbaş et al, 2020 [52]	Turkey	388	129	College students	Range 17-25	Jan 2019 to Jun 2019	SAS[i]	IPAQ-SF	–0.112
6. Zhong et al, 2021 [31]	China	394	115	College students	—	Jul 29, 2020	CSMDQ[j]	PARS-3	–0.190
7. Hosen et al, 2021 [53]	Bangladesh	601	344	College students	—	Oct 2020 to Nov 2020	SABAS[k]	Physical exercise questions (eg, at least 30 minutes daily walking, cycling, swimming, or other activities regularly)	–0.249
8. Li et al, 2021 [33]	China	2407	280	Adolescents	Mean 16.27 (SD 1.02)	Dec 2020 to Feb 2021	Self-rating questionnaire for adolescent problematic mobile phone use	PA questionnaire A[l]	–0.235
9. Buke et al, 2021 [54]	Turkey	300	166	College students	Mean 21.36 (SD 2.33)	Apr 2020	SAS-SV	IPAQ[m]	–0.262
10. Abbasi et al, 2021 [35]	Malaysia	250	145	College students	—	May 2020	SAS-SV	Physical activity questionnaire B[n]	–0.201
11. Islam et al, 2021 [55]	Bangladesh	5511	3254	College students	Mean 21.20 (SD 1.70)	Jul 2020	SABAS	Questions were asked regarding the engagement in infrequent activities (including home quarantine regular/frequent activities (ie, academic/other studies, social-media use, watching television, household chores, and professional activities)	–0.238
12. Ding et al, 2021 [56]	China	1724	740	College students	Mean 19.56 (SD 0.95)	Sep 2020	MPATS	PARS-3	–0.445
13. Halil, 2021 [57]	Pakistan	236	123	College students	—	2020 to 2021	SAS-SV	IPAQ-SF	–0.258
14. Guo et al, 2022 [58]	China	1433	704	College students	Mean 19.67 (SD 1.62)	Dec 2020 to Feb 2021	MPATS	PARS-3	–0.158

Study	Country	Size, n	Male, n	Population	Age (years)	Time period	MPA[a] measurement	PA[b] measurement	*r*
15. Saffari et al, 2022 [59]	Taiwan	391	0	College students	Mean 22.85	Aug 2021 to Sep 2021	SABAS	IPAQ-SF	–0.255
16. Lin et al, 2022 [60]	China	1787	628	College students	Range 18-22	Aug 2020 to Sep 2021	SAS	IPAQ-SF	–0.153
17. Chen et al, 2022 [61]	China	9406	3516	College students	Mean 19.58 (SD 1.07)	Mar 2022 to Apr 2022	MPAS[o]	IPAQ-L[p]	–0.060
18. Venkatesh et al, 2019 [62]	Saudi Arabia	205	101	College students	Mean 23.28	Jan 2016 to Mar 2016	SAS-SV	"Outside school: How many hours a week do you exercise or participate in sports that make you sweat or become out of breath?"	—
19. Xie et al, 2019 [63]	China	2134	917	College students	Mean 19.25 (SD 1.42)	Jun 2014 to Dec 2014	Self-rating questionnaire for adolescent problematic mobile phone use	During the past 7 days, on how many days were you physically active for a total of at least 60 minutes per day?	—
20. Pereira et al, 2020 [64]	Brazil	667	308	Adolescents	Range 13-18	—	SAS-SV	IPAQ-SF	—
21. Tao et al, 2020 [65]	China	4624	2057	College students	Mean 19.91 (SD 1.27)	May 2018 to Jun 2018	Self-rating questionnaire for adolescent problematic mobile phone use	IPAQ-SF	—
22. Zou et al, 2021 [66]	China	251	52	College students	Mean 19.01 (SD 0.85)	Apr 2019 to Jun 2019	Self-rating questionnaire for adolescent problematic mobile phone use	IPAQ-C[q]	—

[a]MPA: mobile phone addiction.

[b]PA: physical activity.

[c]SAPS: Smartphone Addiction Proneness Scale.

[d]SAS-SV: Smartphone Addiction Scale–Short Version.

[e]Not available.

[f]MPATS: Mobile Phone Addiction Tendency Scale.

[g]PARS-3: Physical Activity Rating Scale–3.

[h]IPAQ-SF: International Physical Activity Questionnaire–Short Form.

[i]SAS: Smartphone Addiction Scale.

[j]CSMDQ: College Students Mobile Phone Dependence Questionnaire.

[k]SABAS: Smartphone Application-Based Addiction Scale.

[l]Physical activity questionnaire A was derived from [67].

[m]IPAQ: International Physical Activity Questionnaires.

[n]Physical activity questionnaire B was derived from [68].

[o]MPAS: Mobile Phone Addiction Scale.

[p]IPAQ-L: International Physical Activity Questionnaire–Long Form.

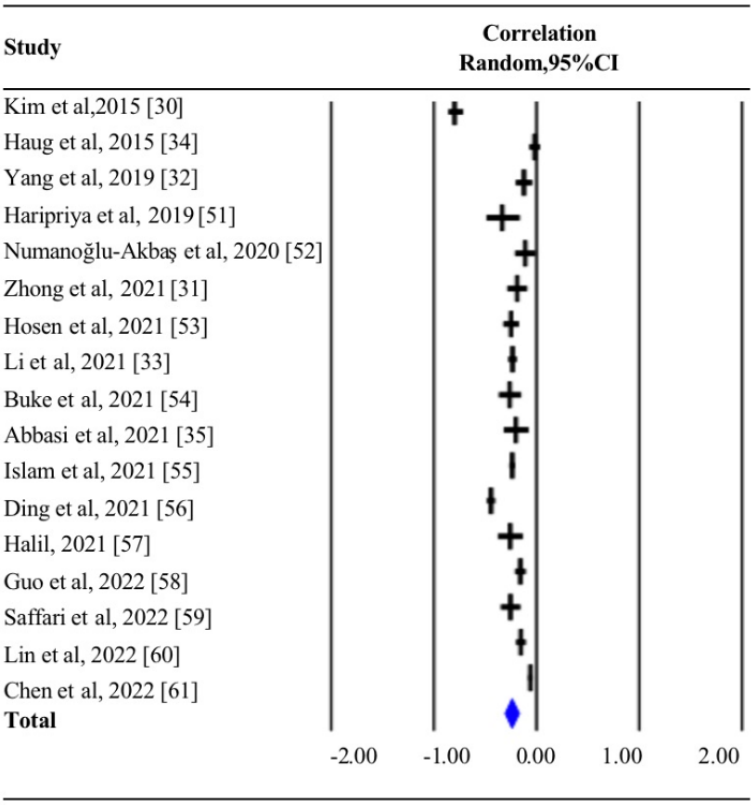
[q]IPAQ C: International Physical Activity Questionnaire–Chinese.

Table 2. Statistics for each study[a].

Study	*r* (total *r*=0.243)	95% CI (total 95% CI –0.309 to –0.175)	*z* (total *z*=–6.810)	*P* value (total *P*<.001)	Weight (total 100%)
Kim et al, 2015 [30]	–0.798	–0.190 to 0.184	–0.031	.98	4.48%
Haug et al, 2015 [34]	–0.019	–0.069 to 0.031	–0.740	.46	6.35%
Yang et al, 2019 [32]	–0.124	–0.202 to –0.045	–3.066	.002	6.06%
Haripriya et al, 2019 [51]	–0.335	–0.486 to –0.165	–3.753	<.001	4.60%
Numanoğlu-Akbaş et al, 2020 [52]	–0.112	–0.209 to –0.013	–2.207	.03	5.81%
Zhong et al, 2021 [31]	–0.190	–0.283 to –0.093	–3.803	<.001	5.82%
Hosen et al, 2021 [53]	–0.249	–0.323 to –0.172	–6.220	<.001	6.05%
Li et al, 2021 [33]	–0.235	–0.272 to –0.197	–11.742	<.001	6.42%
Buke et al, 2021 [54]	–0.262	–0.364 to –0.153	–4.623	<.001	5.62%
Abbasi et al, 2021 [35]	–0.201	–0.317 to –0.079	–3.203	<.001	5.46%
Islam et al, 2021 [55]	–0.238	–0.263 to –0.213	–18.009	<.001	6.50%
Ding et al, 2021 [56]	–0.445	–0.482 to –0.406	–19.848	<.001	6.37%
Halil, 2021 [57]	–0.258	–0.373 to –0.135	–4.029	<.001	5.41%
Guo et al, 2022 [58]	–0.158	–0.208 to –0.107	–6.025	<.001	6.34%
Saffari et al, 2022 [59]	–0.255	–0.345 to –0.160	–5.136	<.001	5.81%
Lin et al, 2022 [60]	–0.153	–0.198 to –0.107	–6.513	<.001	6.38%
Chen et al, 2022 [61]	–0.060	–0.080 to –0.040	–5.825	<.001	6.52%

[a]Heterogeneity: Q=468.050; *P*<.001; I^2=96.582.

Figure 2. Summary of pooled correlation between physical activity and mobile phone addiction. The blue diamond represents the overall pooled correlation for the random effects model [30-35,51-61].



Subgroup Analysis

As shown in Table 3, the summary correlation coefficient between PA and MPA did not change when stratified by time of data collection, country or region, or type of population (all P^b>.05). However, to allow comparison with other studies, we present model-implied effect sizes for each level of the moderator.

The time of data collection did not significantly moderate the effect sizes (between-subgroup P^b=.14). Notably, the summary correlation coefficient of the studies reporting on data collected before COVID-19 was slightly higher compared with that for data collected during COVID-19. Specifically, the effect sizes for data collection before COVID-19 were moderate, with a 95% CI that did not overlap with 0 (r=–0.333, 95% CI –0.466 to –0.187; k=4; P^a<.001), whereas the effect sizes for data collected during COVID-19 were also moderate, with 95% CIs that overlapped with 0 (r=–0.207, 95% CI –0.285 to –0.126, k=13; P^a<.001).

Similarly, we did not find significant moderator effect sizes for country or region (between-subgroup P^b=.71). The summary correlation coefficient for both China and other developing regions showed a similarly moderate effect size (China: r=–0.201, 95% CI –0.311 to –0.127; k=7; P^a<.001; other developing regions: r= –0.217, 95% CI –0.326 to –0.103; k=8; P^a<.001). However, the effect sizes for developed regions with a 95% CI that overlapped with 0 (r=–0.446, 95% CI –0.616 to 236, k=2; P^a=.24) were not significant.

In addition, there were no significant moderator effect sizes for the type of population (between-subgroup: P^b=.26). Specifically, we found that the effect sizes for young adults were moderate, with a 95% CI that did not overlap with 0 (r=0.250, 95% CI –0.325 to –0.173, k=15; P^a<.001). However, the effect sizes for adolescents were not significant, with a 95% CI that overlapped with 0 (r=–0.129, 95% CI –0.333 to 0.086, k=2, P^a=.24).

Table 3. Subgroup analyses of summary correlation between PA and MPA. P^a values for the within-subgroup effect sizes were calculated with the z test; P^b values for between-subgroup differences were calculated with the Q test; and P^c values for heterogeneity within subgroups were calculated with the Q test.

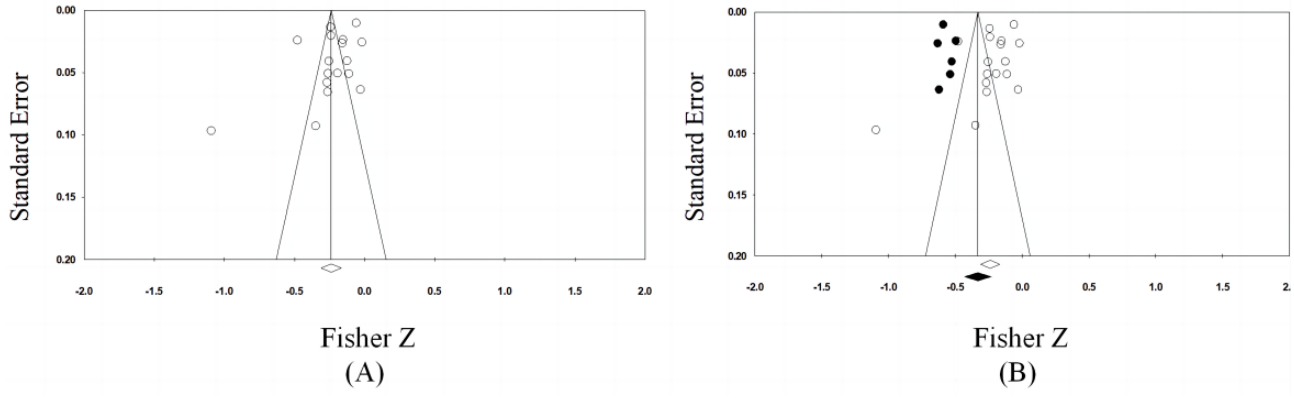
Moderator	Studies, n	Summary r (95% CI)	P^a value	Heterogeneity		P^b value
				I^2 (%)	P^c value	
Time of data collection						.14
Before COVID-19	4	–0.333 (–0.466 to –0.187)	<.001	97.555	<.001	
During COVID-19	13	–0.207 (–0.285 to –0.126)	<.001	96.438	<.001	
Country or region						.71
Developed regions	2	–0.446 (–0.616 to 0.236)	.39	99.133	<.001	
China	7	–0.201 (–0.311 to –0.127)	<.001	97.873	<.001	
Other developing regions	8	–0.217 (–0.326 to –0.103)	<.001	63.310	<.001	
Population						.26
Young adults	15	–0.250 (–0.325 to –0.173)	<.001	96.681	<.001	
Adolescents	2	–0.129 (–0.333 to 0.086)	.24	97.787	<.001	

Sensitivity Analyses and Publication Bias

In the analysis that removed studies one at a time, no evident outliers were identified. Thus, the correlation coefficient for removing each study was in the range of r=–0.195 to –0.248. This shows that no one study significantly skewed or changed the correlation coefficient or influenced the overall results of the meta-analysis. Therefore, the results were reliable.

Subjectively speaking, we could not determine the existence of publication bias from the funnel plots for the summary correlation coefficients, as shown in Figure 3A. Studies with a small sample size are unlikely to result in symmetrical distributions of scattered points. The Begg rank correlation tests and Egger linear regression tests showed no significant publication bias (P=.62 and P=.14, respectively). After the trim-and-fill analysis, the correlation between PA and MPA remained statistically significant (number to trim and fill 6, summary r=–0.319, 95% CI –0.40552 to –0.22771; Figure 3B). Therefore, the overall modeling results after correction remained unchanged. Thus, we concluded that there was no publication bias.

Figure 3. Funnel plots of (A) publication bias and (B) publication bias with trim and fill.



Methodological Quality Assessment

The methodological quality assessment results are shown in Table 4. Notably, the mean scores of the included studies and the other studies were 15.29 (SD 1.53) and 15.40 (SD 0.89), respectively. All included studies were of high quality with a low risk of bias.

Table 4. Methodological quality of the studies.

Number	Study	Joanna Briggs Institute appraisal checklist items										Total score (%)	Overall risk of bias
		1	2	3	4	5	6	7	8	9	10		
1	Kim et al, 2015 [30]	1	1	1	1	2	2	2	2	2	1	15 (75)	Low
2	Haug et al, 2015 [34]	2	0	1	1	2	2	2	2	2	2	16 (80)	Low
3	Yang et al, 2019 [32]	1	2	1	2	2	2	2	2	2	2	18 (80)	Low
4	Haripriya et al, 2019 [51]	2	0	1	1	2	2	2	2	2	1	15 (75)	Low
5	Numanoğlu-Akbaş et al, 2020 [52]	2	0	0	0	2	2	2	2	1	2	13 (65)	Mid
6	Zhong et al, 2021 [31]	2	2	1	1	2	2	0	2	2	2	16 (80)	Low
7	Hosen et al, 2021 [53]	1	1	1	1	1	2	2	2	1	2	14 (70)	Low
8	Li et al, 2021 [33]	2	1	1	1	2	2	2	2	2	2	17 (85)	Low
9	Buke et al, 2021 [54]	1	1	1	1	2	2	2	2	2	1	15 (75)	Low
10	Abbasi et al, 2021 [35]	2	1	1	1	2	2	0	1	1	1	15 (75)	Low
11	Islam et al, 2021 [55]	1	0	1	2	2	2	2	2	1	2	15 (75)	Low
12	Ding et al, 2021 [56]	2	1	1	1	2	2	0	2	2	1	14 (70)	Low
13	Halil, 2021 [57]	2	0	1	1	2	2	0	2	2	2	14 (70)	Low
14	Guo et al, 2022 [58]	2	2	1	1	2	2	2	2	2	2	18 (80)	Low
15	Saffari et al, 2022 [59]	2	0	1	2	2	2	2	2	2	2	17 (85)	Low
16	Lin et al, 2022 [60]	2	0	0	0	2	2	2	2	2	1	13 (65)	Mid
17	Chen et al, 2022 [61]	2	2	0	1	2	2	0	2	2	2	15 (75)	Low
18	Venkatesh et al, 2019 [62]	1	0	1	1	2	2	2	2	2	1	14 (70)	Low
19	Xie et al, 2019 [63]	1	0	1	2	2	2	2	2	2	2	16 (80)	Low
20	Pereira et al, 2020 [64]	1	0	1	2	2	2	2	2	2	2	16 (80)	Low
21	Tao et al, 2020 [65]	2	1	1	2	2	2	0	2	2	1	15 (75)	Low
22	Zou et al, 2021 [66]	2	0	1	2	2	2	2	2	2	1	16 (80)	Low

Discussion

Meta-analytic Findings

To the best of our knowledge, this is the first meta-analysis to explore pooled correlation coefficients of PA and MPA. Our analysis of 17 studies found a moderately negative correlation between PA and MPA, with a summary Pearson correlation coefficient of r=–0.243. This is in line with a previous review [40]. Sensitivity analyses did not find significant publication bias, indicating that the pooled analyses of the correlation coefficients provided reliable and convincing results. In addition, all the included studies were high quality with a low risk of bias. Subgroup analysis showed that none of the hypothesized moderators (data collection, country or region, and type of population) significantly influenced the relationship between PA and MPA, as confirmed with a mixed effects analysis.

The target subjects of research into MPA are adolescents and young adults, who are relatively less self-disciplined in controlling their frequency of mobile phone use and are more susceptible to smartphone use addiction compared to middle-aged or older adults [29]. Lack of self-control is an essential factor influencing MPA among adolescents and young adults. Previous studies found that self-control regulates the correlation between PA and MPA. Factors such as negative emotions (eg, anxiety [69,70] and loneliness [71,72]) and mental toughness [73,74] have been shown to affect the relationship between PA and MPA. We speculate that these factors may modulate the relationship between PA and MPA.

Results of magnetic resonance imaging studies suggest that MPA is associated with structural brain abnormalities, like other types of addiction dependence. For example, the insula cortex participates in the formation of addictive behaviors, because these behaviors may influence the decision-making process in terms of choosing immediate rewards that are always associated with physiological state while eliciting strong interoceptive signals [75]. Two recent studies reported changes in gray matter volume in this region (ie, the insula cortex) among people with MPA [66,76]. Exercise has been shown to improve brain health [77,78]. Therefore, we hypothesize that the relationship between PA and MPA might be influenced by the structure and function of the insula and even other brain regions.

Differences in Subgroup Analysis

Notably, the time of data collection did not significantly influence the relationship between PA and MPA. Moreover, a moderate negative relationship was found between PA and MPA among adolescents and young adults before and during the COVID-19 pandemic. According to the compensatory internet use theory, when people encounter psychosocial problems in the real world, they are likely to use the internet or smartphones as a coping mechanism to alleviate negative emotions [79]. The restrictions imposed on participation in social activities and gatherings during the COVID-19 pandemic increased anxiety, depression, and stress levels in people [80]. Therefore, they were more likely to overly rely on their mobile phones to cope with stress. Moreover, recent studies have shown that adolescents and young adults have had low PA levels during the COVID-19 pandemic [81,82]. The target subjects in our study were adolescents and young adults, who are more inclined to use social media for physical training. A previous study found that young-adult Spanish university students used social media apps to improve their high-intensity interval training [82], mind-body activities, and strength exercises. In other words, adolescents and young adults used social media to facilitate their participation in PA during the pandemic [82]. This reduced the time spent sitting for long periods of time and reduced leisure-time screen activities, subsequently reducing the risk of MPA in young people [30].

The present findings demonstrate that country or region do not have a significant moderating role on the relationship between PA and MPA. Notably, a medium-strength negative relationship between PA and MPA has been reported in China and other developing counties among adolescents and young adults. However, this correlation was not found in developed countries. This finding should be interpreted with caution, because it is based on 2 studies from developed countries. These 2 studies were carefully reviewed elsewhere [30,34]. In addition, we found that one of these studies reported a weak negative correlation between PA and MPA [34], whereas the other found a significant, strong negative correlation [30]. It should be noted that the 2 studies were published around the same time. The difference in the correlation results may be due to the type of PA measurement tools used. For instance, one of the studies used a pedometer sensor to measure the level of PA, which is more precise [30]. The influence of measurement errors associated with self-reported PA questionnaires also needs to be acknowledged. The majority of the reviewed studies used self-reported scales or questionnaires; thus, we suggest that accelerometry should be adopted in future studies to obtain more reliable data.

Further analysis revealed that population type did not significantly affect the relationship between PA and MPA. This may be explained by the widespread use of mobile phones. This is especially true for young people, as their ownership rate for smartphones is very high. Additionally, subgroup analysis revealed that there was no significant correlation between PA and MPA among adolescents (P=.26). We presume that this might be influenced by the degree of external restrictions on the use of mobile phones. Compared with adults (eg, college students), adolescents are subjected to more control and restriction measures on mobile phone use by their parents, schools, and even by the mobile phone apps themselves. Therefore, they are less likely to influence the correlation [83]. These findings, however, should be interpreted with caution, because only 2 studies on adolescents were analyzed.

Limitations and Strengths

In conclusion, our study indicates that a low PA level contributes to MPA behavior. This is because low PA encourages a sedentary lifestyle among young adults. The PA guidelines of the World Health Organization encourage individuals of different ages to participate in PA. Previous studies have shown that increasing the PA level of young adults can reduce MPA behavior. We recommend higher PA levels than those stipulated in the guidelines of the World Health Organization, because more PA could bring more mental health benefits. From a

practical perspective, the findings of this study may help to inform countermeasures to prevent MPA behavior among adolescents and young adults amid the COVID-19 pandemic and future public health crises.

All previous findings were objectively stated, analyzed, and interpreted using an appropriate research design. All original data were retained to provide a reference for future research. The repeatability and reproducibility of our analyses have been ensured. However, there are some limitations and potential sources of bias that need to be noted. First, only studies published in English were included in our meta-analysis. Second, the studies mainly provided cross-sectional data, which do not allow determination of causality in the relationship between PA and MPA. Third, we only analyzed a young population. Fourth, no study reported moderating variables between PA and MPA. Finally, although a sensitivity analysis was conducted, sources of bias were identified, and our results should thus be interpreted with caution. Further case-control and cohort studies are needed to test the benefits of PA on MPA in young adults.

Conclusion

Our findings demonstrate a moderate negative relationship between PA and MPA among young adults. The strength of the relationship between PA and MPA did not differ by time of data collection, country or region, or type of population.

Acknowledgments

This study received funding from the Research Foundation for Young Teachers of Shenzhen University (grant QNJS0274), the High-level Scientific Research Foundation for the Introduction of Talent of Shenzhen University (grant RC00228), the Natural Science Featured Innovation Projects in Ordinary Universities in Guangdong Province (grant 2021KTSCX297), and the Scientific Research Platform and Project of Colleges and Universities of the Education Department of Guangdong Province (grant 2022ZDZX2087).

Authors' Contributions

WX and ZR were responsible for conceptualization of the research, investigation, and forming the hypothesis. JW and WX conducted the systematic search, data extraction, and quality assessment and data analyses. JW and WX wrote the first draft of the manuscript. JY, QS, LP, QEL, and ZR reviewed and edited the initial draft and its revisions. All authors agree with the results and conclusions. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search strategy.
[DOCX File , 15 KB - publichealth_v8i12e41606_app1.docx]

Multimedia Appendix 2
Details of the scoring criteria in the JBI appraisal checklist. JBI: Joanna Briggs Institute.
[DOCX File , 14 KB - publichealth_v8i12e41606_app2.docx]

Multimedia Appendix 3
Full details of coding forms for the subgroups.
[DOCX File , 16 KB - publichealth_v8i12e41606_app3.docx]

References

1. Billieux J, Maurage P, Lopez-Fernandez O, Kuss DJ, Griffiths MD. Can disordered mobile phone use be considered a behavioral addiction? an update on current evidence and a comprehensive model for future research. Curr Addict Rep 2015 Jun 01;2(2):156-162 [FREE Full text] [doi: 10.1007/s40429-015-0054-y]
2. Sert H, Taskin Yilmaz F, Karakoc Kumsar A, Aygin D. Effect of technology addiction on academic success and fatigue among Turkish university students. Fatigue: Biomed Health Behav 2019 Feb 26;7(1):41-51. [doi: 10.1080/21641846.2019.1585598]
3. Battle DE. Diagnostic and Statistical Manual of Mental Disorders (DSM). Codas 2013;25(2):191-192 [FREE Full text] [doi: 10.1590/s2317-17822013000200017] [Medline: 24413388]
4. Zou Z, Wang H, d'Oleire Uquillas F, Wang X, Ding J, Chen H. Definition of substance and non-substance addiction. Adv Exp Med Biol 2017;1010:21-41. [doi: 10.1007/978-981-10-5562-1_2] [Medline: 29098666]
5. Amiri M, Dowran B, Salimi H, Zarghami MH. The problematic use of mobile phone and mental health: A review study in Iran. J Educ Health Promot 2020;9:290 [FREE Full text] [doi: 10.4103/jehp.jehp_268_20] [Medline: 33282995]

6. Nunes PPDB, Abdon APV, Brito CBD, Silva FVM, Santos ICA, Martins DDQ, et al. Factors related to smartphone addiction in adolescents from a region in Northeastern Brazil. Cien Saude Colet 2021 Jul;26(7):2749-2758 [FREE Full text] [doi: 10.1590/1413-81232021267.08872021] [Medline: 34231688]
7. Zou Y, Chen S, Dong Q, Wang H, Xie X, Ji S. Current situation of mobile phone addiction and its relationship with personality among college students in Hainan Province. China Journal of Health Psychology 2017;25:025-025. [doi: 10.13342/j.cnki.cjhp.2017.09.025]
8. Soliman Elserty N, Ahmed Helmy N, Mohmed Mounir K. Smartphone addiction and its relation to musculoskeletal pain in Egyptian physical therapy students. Eur J Physiother 2018 Dec 31;22(2):70-78. [doi: 10.1080/21679169.2018.1546337]
9. Jahagirdar V, Rama K, Soppari P, Kumar M. Mobile phones: vital addiction or lethal addiction? mobile phone usage patterns and assessment of mobile addiction among undergraduate medical students in Telangana, India. J Addict 2021;2021:8750650 [FREE Full text] [doi: 10.1155/2021/8750650] [Medline: 34721921]
10. Kim K, Yee J, Chung JE, Kim HJ, Han JM, Kim JH, et al. Smartphone addiction and anxiety in adolescents - a cross-sectional study. Am J Health Behav 2021 Sep 30;45(5):895-901. [doi: 10.5993/AJHB.45.5.9] [Medline: 34702436]
11. Li Y, Li G, Liu L, Wu H. Correlations between mobile phone addiction and anxiety, depression, impulsivity, and poor sleep quality among college students: A systematic review and meta-analysis. J Behav Addict 2020 Oct 12;9(3):551-571 [FREE Full text] [doi: 10.1556/2006.2020.00057] [Medline: 32903205]
12. Ivanova A, Gorbaniuk O, Błachnio A, Przepiórka A, Mraka N, Polishchuk V, et al. Mobile phone addiction, phubbing, and depression among men and women: a moderated mediation analysis. Psychiatr Q 2020 Sep;91(3):655-668 [FREE Full text] [doi: 10.1007/s11126-020-09723-8] [Medline: 32146681]
13. Matar Boumosleh J, Jaalouk D. Depression, anxiety, and smartphone addiction in university students- A cross sectional study. PLoS One 2017;12(8):e0182239 [FREE Full text] [doi: 10.1371/journal.pone.0182239] [Medline: 28777828]
14. Mohammadbeigi A, Absari R, Valizadeh F, Saadati M, Sharifimoghadam S, Ahmadi A, et al. Sleep quality in medical students; the impact of over-use of mobile cell-phone and social networks. J Res Health Sci 2016;16(1):46-50 [FREE Full text] [Medline: 27061997]
15. Lane H, Chang C, Huang C, Chang Y. An investigation into smartphone addiction with personality and sleep quality among university students. Int J Environ Res Public Health 2021 Jul 16;18(14):7588 [FREE Full text] [doi: 10.3390/ijerph18147588] [Medline: 34300037]
16. Zhang T, Gong N, Jia R, Li H, Ni X. Stroop effect in smartphone addiction among college students. Medicine (Baltimore) 2021 Jul 30;100(30):e26741 [FREE Full text] [doi: 10.1097/MD.0000000000026741] [Medline: 34397714]
17. Shah PP, Sheth MS. Correlation of smartphone use addiction with text neck syndrome and SMS thumb in physiotherapy students. Int J Community Med Public Health 2018 May 22;5(6):2512 [FREE Full text] [doi: 10.18203/2394-6040.ijcmph20182187]
18. Alotaibi MS, Fox M, Coman R, Ratan ZA, Hosseinzadeh H. Smartphone addiction prevalence and its association on academic performance, physical health, and mental well-being among university students in Umm Al-Qura University (UQU), Saudi Arabia. Int J Environ Res Public Health 2022 Mar 21;19(6):3710 [FREE Full text] [doi: 10.3390/ijerph19063710] [Medline: 35329397]
19. Public health implications of excessive use of the internet, computers, smartphones and similar electronic devices: Meeting report, Main Meeting Hall, Foundation for Promotion of Cancer Research, National Cancer Research Centre, Tokyo, Japan, 27-29 August 2014. World Health Organization. 2015. URL: https://apps.who.int/iris/bitstream/handle/10665/184264/9789241509367_eng.pdf?sequence=1&isAllowed=y [accessed 2022-12-05]
20. Banerjee D. The COVID-19 outbreak: crucial role the psychiatrists can play. Asian J Psychiatr 2020 Apr;50:102014 [FREE Full text] [doi: 10.1016/j.ajp.2020.102014] [Medline: 32240958]
21. Islam MS, Potenza MN, van Os J. Posttraumatic stress disorder during the COVID-19 pandemic: upcoming challenges in Bangladesh and preventive strategies. Int J Soc Psychiatry 2021 Mar;67(2):205-206. [doi: 10.1177/0020764020954469] [Medline: 32873127]
22. Brand M, Young KS, Laier C, Wölfling K, Potenza MN. Integrating psychological and neurobiological considerations regarding the development and maintenance of specific Internet-use disorders: An Interaction of Person-Affect-Cognition-Execution (I-PACE) model. Neurosci Biobehav Rev 2016 Dec;71:252-266 [FREE Full text] [doi: 10.1016/j.neubiorev.2016.08.033] [Medline: 27590829]
23. Elhai JD, Yang H, McKay D, Asmundson GJG. COVID-19 anxiety symptoms associated with problematic smartphone use severity in Chinese adults. J Affect Disord 2020 Sep 01;274:576-582 [FREE Full text] [doi: 10.1016/j.jad.2020.05.080] [Medline: 32663990]
24. Li L, Xu D, Chai J, Wang D, Li L, Zhang L, et al. Prevalence of Internet addiction disorder in Chinese university students: A comprehensive meta-analysis of observational studies. J Behav Addict 2018 Sep 01;7(3):610-623 [FREE Full text] [doi: 10.1556/2006.7.2018.53] [Medline: 30010411]
25. Lian S, Sun X, Niu G, Yang X, Zhou Z, Yang C. Mobile phone addiction and psychological distress among Chinese adolescents: The mediating role of rumination and moderating role of the capacity to be alone. J Affect Disord 2021 Jan 15;279:701-710 [FREE Full text] [doi: 10.1016/j.jad.2020.10.005] [Medline: 33197839]

26. Long J, Liu T, Liao Y, Qi C, He H, Chen S, et al. Prevalence and correlates of problematic smartphone use in a large random sample of Chinese undergraduates. BMC Psychiatry 2016 Nov 17;16(1):408 [FREE Full text] [doi: 10.1186/s12888-016-1083-3] [Medline: 27855666]
27. Kuan G, Abdullah N, Kueh YC, Ismail M, Shafei MN, Morris T. Co-curricular activities and motives for participating in physical activity among health sciences students at Universiti Sains Malaysia, Malaysia. Malays J Med Sci 2019 Jan;26(1):138-146 [FREE Full text] [doi: 10.21315/mjms2019.26.1.13] [Medline: 30914901]
28. WHO guidelines on physical activity and sedentary behaviour: web annex: evidence profiles. World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/336657 [accessed 2022-12-05]
29. Liu S, Xiao T, Yang L, Loprinzi PD. Exercise as an alternative approach for treating smartphone addiction: a systematic review and meta-analysis of random controlled trials. Int J Environ Res Public Health 2019 Oct 15;16(20):3912 [FREE Full text] [doi: 10.3390/ijerph16203912] [Medline: 31618879]
30. Kim S, Kim J, Jee Y. Relationship between smartphone addiction and physical activity in Chinese international students in Korea. J Behav Addict 2015 Sep;4(3):200-205 [FREE Full text] [doi: 10.1556/2006.4.2015.028] [Medline: 26551911]
31. Zhong W, Wang Y, Zhang G. The impact of physical activity on college students' mobile phone dependence: the mediating role of self-control. Int J Ment Health Addiction 2020 Jul 29;19(6):2144-2159. [doi: 10.1007/s11469-020-00308-x]
32. Yang G, Tan G, Li Y, Liu H, Wang S. Physical exercise decreases the mobile phone dependence of university students in China: the mediating role of self-control. Int J Environ Res Public Health 2019 Oct 24;16(21):4098 [FREE Full text] [doi: 10.3390/ijerph16214098] [Medline: 31652978]
33. Li Y, Sun Q, Sun M, Sun P, Sun Q, Xia X. Physical exercise and psychological distress: the mediating roles of problematic mobile phone use and learning burnout among adolescents. Int J Environ Res Public Health 2021 Sep 02;18(17):9261 [FREE Full text] [doi: 10.3390/ijerph18179261] [Medline: 34501851]
34. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. J Behav Addict 2015 Dec;4(4):299-307 [FREE Full text] [doi: 10.1556/2006.4.2015.037] [Medline: 26690625]
35. Abbasi GA, Jagaveeran M, Goh Y, Tariq B. The impact of type of content use on smartphone addiction and academic performance: Physical activity as moderator. Technol Soc 2021 Feb;64:101521. [doi: 10.1016/j.techsoc.2020.101521]
36. Fennell C, Barkley JE, Lepp A. The relationship between cell phone use, physical activity, and sedentary behavior in adults aged 18–80. Comput Hum Behav 2019 Jan;90:53-59. [doi: 10.1016/j.chb.2018.08.044]
37. Barkley JE, Lepp A. Mobile phone use among college students is a sedentary leisure behavior which may interfere with exercise. Comput Hum Behav 2016 Mar;56:29-33. [doi: 10.1016/j.chb.2015.11.001]
38. Xiang M, Lin L, Wang Z, Li J, Xu Z, Hu M. Sedentary behavior and problematic smartphone use in Chinese adolescents: the moderating role of self-control. Front Psychol 2019;10:3032 [FREE Full text] [doi: 10.3389/fpsyg.2019.03032] [Medline: 32038398]
39. Zagalaz-Sánchez ML, Cachón-Zagalaz J, Sánchez-Zafra M, Lara-Sánchez A. Mini review of the use of the mobile phone and its repercussion in the deficit of physical activity. Front Psychol 2019 Jun 6;10:1307 [FREE Full text] [doi: 10.3389/fpsyg.2019.01307] [Medline: 31244720]
40. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
41. Adhia A, Kernic MA, Hemenway D, Vavilala MS, Rivara FP. Intimate partner homicide of adolescents. JAMA Pediatr 2019 Jun 01;173(6):571-577 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.0621] [Medline: 30985886]
42. Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R. Chapter 7: Systematic reviews of etiology and risk. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: The Joanna Briggs Institute; 2017.
43. Zhang Y, Long G, Ding B, Sun G, Ouyang W, Liu M, et al. The impact of ambient temperature on the incidence of urolithiasis: a systematic review and meta-analysis. Scand J Work Environ Health 2020 Mar 01;46(2):117-126 [FREE Full text] [doi: 10.5271/sjweh.3866] [Medline: 31747456]
44. Polanin JR, Snilstveit B. Converting between effect sizes. Campbell Syst Rev 2016;12:1-13. [doi: 10.4073/CMPN.2016.3]
45. Heeke C, Kampisiou C, Niemeyer H, Knaevelsrud C. A systematic review and meta-analysis of correlates of prolonged grief disorder in adults exposed to violent loss. Eur J Psychotraumatol 2019;10(1):1583524 [FREE Full text] [doi: 10.1080/20008198.2019.1583524] [Medline: 30949303]
46. Fu Z, Burger H, Arjadi R, Bockting CLH. Effectiveness of digital psychological interventions for mental health problems in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Psychiatry 2020 Oct;7(10):851-864 [FREE Full text] [doi: 10.1016/S2215-0366(20)30256-X] [Medline: 32866459]
47. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994 Dec;50(4):1088-1101. [Medline: 7786990]
48. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]

49. Xin F, Chen S, Clark C, Hong J, Liu Y, Cai Y. Relationship between fundamental movement skills and physical activity in preschool-aged children: a systematic review. Int J Environ Res Public Health 2020 May 19;17(10):3566 [FREE Full text] [doi: 10.3390/ijerph17103566] [Medline: 32438736]
50. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Hoboken, NJ: John Wiley & Sons; 2021.
51. Haripriya S, Samuel SE, Megha M. Correlation between smartphone addiction, sleep quality and physical activity among young adults. J Clin Diagn Res 2019;13(10):5-9. [doi: 10.7860/jcdr/2019/42168.13212]
52. Numanoğlu-Akbaş A, Suner-Keklik S, Yakut H. Investigation of the relationship between smart phone addiction and physical activity in university students. Balt J Health Phys Act 2020 Dec 31;Supplement 1(1):63-73 [FREE Full text] [doi: 10.29359/bjhpa.2020.suppl.1.07]
53. Hosen I, Al Mamun F, Sikder MT, Abbasi AZ, Zou L, Guo T, et al. Prevalence and associated factors of problematic smartphone use during the COVID-19 pandemic: a Bangladeshi study. Risk Manag Healthc Policy 2021;14:3797-3805 [FREE Full text] [doi: 10.2147/RMHP.S325126] [Medline: 34548828]
54. Buke M, Egesoy H, Unver F. The effect of smartphone addiction on physical activity level in sports science undergraduates. J Bodyw Mov Ther 2021 Oct;28:530-534. [doi: 10.1016/j.jbmt.2021.09.003] [Medline: 34776190]
55. Islam MS, Sujan MSH, Tasnim R, Mohona RA, Ferdous MZ, Kamruzzaman S, et al. Problematic smartphone and social media use among Bangladeshi college and university students amid COVID-19: the role of psychological well-being and pandemic related factors. Front Psychiatry 2021;12:647386 [FREE Full text] [doi: 10.3389/fpsyt.2021.647386] [Medline: 33935834]
56. Ding Z, Yan J, Fu J. Internet and Mobile Phone Addiction Self-Control Mediate Physical Exercise and Subjective Well-Being in Young Adults Using IoT. Mob Inf Syst 2021 May 18;2021:1-6 [FREE Full text] [doi: 10.1155/2021/9923833]
57. Halil T. The risk of physical activity and smart phone addiction in sports high school students: an example of a state school. Pakistan J Medical Health Sci 2021;15(2):706-711 [FREE Full text]
58. Guo K, Ma QS, Yao SJ, Liu C, Hui Z, Jiang J, et al. The relationship between physical exercise and mobile phone addiction tendency of university students in China: a moderated mediation model. Front Psychol 2022;13:730886 [FREE Full text] [doi: 10.3389/fpsyg.2022.730886] [Medline: 35237204]
59. Saffari M, Chen J, Wu H, Fung XCC, Chang C, Chang Y, et al. Effects of weight-related self-stigma and smartphone addiction on female university students' physical activity levels. Int J Environ Res Public Health 2022 Feb 24;19(5):2631 [FREE Full text] [doi: 10.3390/ijerph19052631] [Medline: 35270328]
60. Lin B, Teo EW, Yan T. The impact of smartphone addiction on Chinese university students' physical activity: exploring the role of motivation and self-efficacy. Psychol Res Behav Manag 2022;15:2273-2290 [FREE Full text] [doi: 10.2147/PRBM.S375395] [Medline: 36039111]
61. Chen H, Wang C, Lu T, Tao B, Gao Y, Yan J. The relationship between physical activity and college students' mobile phone addiction: the chain-based mediating role of psychological capital and social adaptation. Int J Environ Res Public Health 2022 Jul 29;19(15):9286 [FREE Full text] [doi: 10.3390/ijerph19159286] [Medline: 35954644]
62. Venkatesh E, Al Jemal MY, Samani ASA. Smart phone usage and addiction among dental students in Saudi Arabia: a cross sectional study. Int J Adolesc Med Health 2017 Apr 06;31(1):20160133. [doi: 10.1515/ijamh-2016-0133] [Medline: 28384117]
63. Xie H, Tao S, Zhang Y, Tao F, Wu X. Impact of problematic mobile phone use and insufficient physical activity on depression symptoms: a college-based follow-up study. BMC Public Health 2019 Dec 05;19(1):1640 [FREE Full text] [doi: 10.1186/s12889-019-7873-z] [Medline: 31805915]
64. Pereira FS, Bevilacqua GG, Coimbra DR, Andrade A. Impact of problematic smartphone use on mental health of adolescent students: association with mood, symptoms of depression,. Cyberpsychol Behav Soc Netw 2020 Sep;23(9):619-626. [doi: 10.1089/cyber.2019.0257] [Medline: 32580574]
65. Tao S, Wu X, Yang Y, Tao F. The moderating effect of physical activity in the relation between problematic mobile phone use and depression among university students. J Affect Disord 2020 Aug 01;273:167-172. [doi: 10.1016/j.jad.2020.04.012] [Medline: 32421598]
66. Zou L, Wu X, Tao S, Yang Y, Zhang Q, Hong X, et al. Neural correlates of physical activity moderate the association between problematic mobile phone use and psychological symptoms. Front Behav Neurosci 2021;15:749194 [FREE Full text] [doi: 10.3389/fnbeh.2021.749194] [Medline: 35177970]
67. Yuan J, Liwei Z, Zhixiong M. Physical exercise and mental health: the effect of emotion regulation self-efficacy and emotion regulation strategy. Studies of Psychology and Behavior 2018;16(4):570-576 [FREE Full text]
68. Arneklev BJ, Grasmick HG, Tittle CR, Bursik RJ. Low self-control and imprudent behavior. J Quant Criminol 1993 Sep;9(3):225-247. [doi: 10.1007/bf01064461]
69. Yang N, Yang X. Anxiety and depression in graduating university students during the COVID-19 pandemic: a longitudinal study. Am J Transl Res 2022;14(4):2668-2676 [FREE Full text] [Medline: 35559421]
70. Zhang C, Hao J, Liu Y, Cui J, Yu H. Associations between online learning, smartphone addiction problems, and psychological symptoms in Chinese college students after the COVID-19 pandemic. Front Public Health 2022;10:881074. [doi: 10.3389/fpubh.2022.881074] [Medline: 35602144]

71. Pinto ADA, Oppong Asante K, Puga Barbosa RMDS, Nahas MV, Dias DT, Pelegrini A. Association between loneliness, physical activity, and participation in physical education among adolescents in Amazonas, Brazil. J Health Psychol 2021 Apr;26(5):650-658. [doi: 10.1177/1359105319833741] [Medline: 30841751]
72. Li X, Feng X, Xiao W, Zhou H. Loneliness and mobile phone addiction among Chinese college students: the mediating roles of boredom proneness and self-control. Psychol Res Behav Manag 2021;14:687-694 [FREE Full text] [doi: 10.2147/PRBM.S315879] [Medline: 34149289]
73. Lancaster MR, Callaghan P. The effect of exercise on resilience, its mediators and moderators, in a general population during the UK COVID-19 pandemic in 2020: a cross-sectional online study. BMC Public Health 2022 Apr 25;22(1):827 [FREE Full text] [doi: 10.1186/s12889-022-13070-7] [Medline: 35468747]
74. Li S, Cui G, Yin Y, Tang K, Chen L, Liu X. Prospective association between problematic mobile phone use and eating disorder symptoms and the mediating effect of resilience in Chinese college students: a 1-year longitudinal study. Front Public Health 2022;10:857246 [FREE Full text] [doi: 10.3389/fpubh.2022.857246] [Medline: 35570941]
75. Benarroch EE. Insular cortex: Functional complexity and clinical correlations. Neurology 2019 Nov 19;93(21):932-938. [doi: 10.1212/WNL.0000000000008525] [Medline: 31645470]
76. Horvath J, Mundinger C, Schmitgen MM, Wolf ND, Sambataro F, Hirjak D, et al. Structural and functional correlates of smartphone addiction. Addict Behav 2020 Jun;105:106334. [doi: 10.1016/j.addbeh.2020.106334] [Medline: 32062336]
77. Wrann CD, White JP, Salogiannnis J, Laznik-Bogoslavski D, Wu J, Ma D, et al. Exercise induces hippocampal BDNF through a PGC-1α/FNDC5 pathway. Cell Metab 2013 Nov 05;18(5):649-659 [FREE Full text] [doi: 10.1016/j.cmet.2013.09.008] [Medline: 24120943]
78. Urdinguio RG, Tejedor JR, Fernández-Sanjurjo M, Pérez RF, Peñarroya A, Ferrero C, et al. Physical exercise shapes the mouse brain epigenome. Mol Metab 2021 Dec;54:101398 [FREE Full text] [doi: 10.1016/j.molmet.2021.101398] [Medline: 34801767]
79. Kardefelt-Winther D. A conceptual and methodological critique of internet addiction research: Towards a model of compensatory internet use. Comput Human Behav 2014 Feb;31:351-354 [FREE Full text] [doi: 10.1016/j.chb.2013.10.059]
80. GebreEyesus FA, Tarekegn TT, Amlak BT, Shiferaw BZ, Emeria MS, Geleta OT, et al. Levels and predictors of anxiety, depression, and stress during COVID-19 pandemic among frontline healthcare providers in Gurage zonal public hospitals, Southwest Ethiopia, 2020: A multicenter cross-sectional study. PLoS One 2021;16(11):e0259906 [FREE Full text] [doi: 10.1371/journal.pone.0259906] [Medline: 34843502]
81. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, et al. Worldwide effect of COVID-19 on physical activity: a descriptive study. Ann Intern Med 2020 Nov 03;173(9):767-770 [FREE Full text] [doi: 10.7326/M20-2665] [Medline: 32598162]
82. Rodríguez-Larrad A, Mañas A, Labayen I, González-Gross M, Espin A, Aznar S, et al. Impact of COVID-19 confinement on physical activity and sedentary behaviour in Spanish university students: role of gender. Int J Environ Res Public Health 2021 Jan 06;18(2):369 [FREE Full text] [doi: 10.3390/ijerph18020369] [Medline: 33418907]
83. Alavi SS, Ghanizadeh M, Mohammadi MR, Mohammadi Kalhory S, Jannatifard F, Sepahbodi G. The survey of personal and national identity on cell phone addicts and non-addicts. Iran J Psychiatry 2018 Jan;13(1):15-21 [FREE Full text] [Medline: 29892313]

Abbreviations

JBI: Joanna Briggs Institute
MPA: mobile phone addiction
OR: odds ratio
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Review and Meta-analyses

Edited by A Mavragani, G Eysenbach; submitted 01.08.22; peer-reviewed by W Liang, E O'Loughlin; comments to author 09.09.22; revised version received 07.10.22; accepted 17.10.22; published 14.12.22.

Please cite as:
Xiao W, Wu J, Yip J, Shi Q, Peng L, Lei QE, Ren Z
The Relationship Between Physical Activity and Mobile Phone Addiction Among Adolescents and Young Adults: Systematic Review and Meta-analysis of Observational Studies
JMIR Public Health Surveill 2022;8(12):e41606
URL: https://publichealth.jmir.org/2022/12/e41606
doi:10.2196/41606
PMID:36515994

Review

The Practice and Potential Role of HIV Self-testing in China: Systematic Review and Meta-analysis

Xiangfei Xiu[1], MM; Yuanyuan Qin[2,3], MM; Yugang Bao[4], PhD; Yaokai Chen[3], MD, PhD; Hao Wu[2], MD; Xiaojie Huang[2*], MD, PhD; Lu Wang[1*], PhD

[1]National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

[2]Clinical and Research Center for Infectious Diseases, Beijing Youan Hospital, Capital Medical University, Beijing, China

[3]Division of Infectious Diseases, Chongqing Public Health Medical Center, Chongqing, China

[4]AIDS Healthcare Foundation, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Lu Wang, PhD
National Center for AIDS/STD Control and Prevention
Chinese Center for Disease Control and Prevention
155 Changbai Road
Changping District
Beijing, 102206
China
Phone: 86 13910584589
Email: wanglu64@chinaaids.cn

Abstract

Background: HIV self-testing (HIVST) is recommended by the World Health Organization as a valid approach to routine HIV testing services. The scale of HIVST use has gradually been expanded in China over the past 5 years. To take a closer look at the role of HIVST in China, we reviewed the promotion and application of HIVST within China.

Objective: The main objective of this study was to systematically analyze the proportion of past use and actual uptake of HIVST within China. Moreover, we aimed to quantify the effect of HIVST on HIV prevention and treatment.

Methods: In all, 5 medical databases and 2 registration systems, including PubMed, Web of Science, MEDLINE, WanFang, China National Knowledge Internet, ClinicalTrials.gov, and the Chinese Clinical Trial Registry were systematically searched for studies reporting the prevalence of HIVST use from January 1, 2010, to December 25, 2021. Meta-analyses of the pooled proportion estimates were carried out by the meta-package in R software (version 4.1.2). Statistical heterogeneity among the studies was estimated using Cochran Q test and the inconsistency index (I^2).

Results: A total of 50 studies were included in our systematic review. The estimated pooled prevalence of HIVST use in China was 29.9% (95% CI 22.5%-37.9%). Among individuals who have ever used HIVST, 47.5% (95% CI 37.2%-57.8%) were tested for HIV for the first time. The pooled reactive rate of HIVST was 4.2% (95% CI 3.1%-5.8%). When HIVST revealed a reactive result, 81.3% (95% CI 70.9%-91.6%) of individuals sought medical care.

Conclusions: In recent times, HIVST has become a valuable tool for HIV prevention in China. The widespread use of HIVST in non–men who have sex with men populations needs to be endorsed and promoted. The long-term applications of HIVST and the potential consequences of self-financing of HIVST in China have yet to be explored.

Trial Registration: PROSPERO CRD42022304846; https://tinyurl.com/54d9pxy8

(JMIR Public Health Surveill 2022;8(12):e41125) doi:10.2196/41125

KEYWORDS

HIV; self-testing; China; meta-analysis; prevalence

Introduction

Approximately 5.9 million people were unaware that they were living with HIV in 2021, according to preliminary the Joint United Nations Programme on HIV/AIDS 2021 epidemiological estimates [1]. To achieve the goal of "95-95-95" as defined by the Joint United Nations Programme on HIV/AIDS, much work remains to be done, especially during the currently prevailing period when HIV services are being severely disrupted by the impact of COVID-19; in a worst case scenario, 7.7 million HIV-related deaths might possibly be incurred [2]. HIV self-testing (HIVST) has been shown to be an effective tool to potentially supplement other testing modalities to reach the 95-95-95 targets advocated by the World Health Organization [3].

HIVST has been recommended by the World Health Organization as a viable and effective extension to routine HIV-testing services since 2016. It has been shown to be a reliable strategy to promote HIV testing and simultaneously protects the privacy of those tested. Those tested can be made aware of their HIV infection status promptly by simply interpreting the test results, using the specimen collected by themselves at their convenience [4]. Oral fluid–based and blood-based HIVST are both accurate and practicable testing approaches in the study setting [5]. Additionally, the first urine-based HIVST testing kit was approved for use in China in 2019 [6]. Different categories of HIVST testing kits may mediate the acceptance and expansion of the use of HIVST within health care facilities. Previous meta-analyses have investigated the effects of HIVST and its purported benefits in key populations and have shown that HIVST plays a key role in HIV prevention by increasing the frequency of testing [7-9]. However, critical post-HIVST patient support needs to be diligently studied and holistically understood and includes issues related to the confirmation of results and linkage to ongoing care [7].

The scale of HIVST use has been expanded gradually within China [10]. The Chinese government released official directives encouraging the use of HIVST during the "Thirteenth Five-Year Plan" (2017-2022) period [11]. Since then, China has taken a series of measures to promote the expansion of HIVST. The National Center for AIDS/STD Control and Prevention and the Chinese Center for Disease Control and Prevention have conducted pilot HIVST projects in many cities across China and has worked with community-based organizations to explore more effective HIV-testing strategies [11]. Additionally, the HIVST strategy has been included in the National Guideline for Detection of HIV/AIDS, published by the Chinese Center for Disease Control and Prevention, since 2020 [12]. However, the expansion of the HIVST strategy in China has created both opportunities and challenges. The standardization, implementation, sustainability, and linkage of HIVST are concerns that have regrettably persisted [11].

Robust data, and the scrupulous analysis thereof, are required to define the practice and role of HIVST in China over the past decade, which may provide critical evidence for future decision-making related to HIVST promotion and application. Despite the existence of studies that have reported on the prevalent rates of HIVST in different areas or regions within China, the estimated HIVST prevalence rate at the national level has rarely been formulated [13-16]. To evaluate the practicality and influence of HIVST on HIV prevention in China, we conducted a systematic review and meta-analysis of studies that tracked the history and overall influence of HIVST on HIV prevention and treatment.

Methods

The present systematic review and meta-analysis followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. This study was duly registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration number: CRD42022304846).

Search Strategy and Selection Criteria

We searched for HIVST-associated studies that were conducted in China and published from January 1, 2010, to December 25, 2021, in 5 databases and 2 registration systems (PubMed, Web of Science, MEDLINE, WanFang, China National Knowledge Internet, ClinicalTrials.gov, and the Chinese Clinical Trial Registry) and then further reviewed them. Our search syntax was based on the following core concepts: "HIV," "self-testing," and "China" (see Multimedia Appendix 1 for details regarding the search strategy).

To qualify for inclusion, a research article had to meet the following criteria: (1) peer-reviewed articles reporting on the performance of rapid diagnostic HIV tests by those who self-tested, and (2) available numerator and denominator data to confirm rate values. Studies investigating foreign residents living within China were excluded.

Selection Process and Data Extraction

Two reviewers (XX and YQ) independently screened and assessed the titles and abstracts of all articles identified via the search strategy. Each reviewer decided whether to exclude or include a study in accordance with a standardized form that defined the criteria for inclusion and exclusion. Data were independently extracted from individual studies by 3 reviewers (XX, YQ, and YB). If a dispute occurred, consensus was achieved through discussions with 3 other authors (HW, XH, and LW). EndNote reference management software (version X9.3.3; Clarivate) was used to filter duplicate studies. Microsoft Excel spreadsheet software (Microsoft Office 2016) was used to record the data extracted from eligible studies.

Data Analysis

In all, 6 indicators were analyzed in this work: proportion of those who previously self-tested, proportion of actual uptake of HIVST, proportion of self-testing as lifetime first HIV screening, proportion of results feedback, reactive rate of HIVST, and proportion of linkage to care. The proportion of those who previously self-tested was defined as the number of individuals who have personal experience of HIVST among all the investigated people. The proportion of self-testing as lifetime first HIV screening was defined as the number of individuals who used self-testing as their first-ever HIV screening among

people who have ever used HIVST. The proportion of results feedback was defined as the number of individuals who self-reported the results of HIVST or returned the images of HIVST to investigators or health institutions among those who self-tested for HIV. The reactive rate of HIVST was defined as the number of individuals receiving a positive reactive HIVST result among those who have self-tested for HIV. The proportion of linkage to care was defined as the number of patients who sought in-person confirmation of their HIV status (with or without accessing treatment) at local health facilities among individuals whose HIVST was reactive.

Meta-analyses of the pooled proportion estimates were carried out by the meta-package in R software (version 4.1.2; R Foundation for Statistical Computing). Statistical heterogeneity among the studies was estimated using Cochran Q test and the inconsistency index (I^2). Very low, low, moderate, and high degrees of heterogeneity were defined as I^2 of ≤25%, 25% to ≤50%, 50% to ≤75%, and ≥75%, respectively.

Quality Assessment

The tool developed by Hoy et al [17] was used to assess the quality of the included studies [18]. The quality of each study was assessed according to 10 items with a maximum score of 10 (one point for a “Yes”). A total score of 0-5, 6-8, and 9-10 was considered high, moderate, and low risk of bias, respectively.

Results

Study Characteristics and Quality Assessment

A total of 688 records was found based on the initial search, of which 213 were selected for full-text evaluation after the removal of duplicates and initial screening. Finally, 50 studies were included in the systematic review based on critical appraisal and were included in our systematic review (Figure 1). Among them, 30 articles were written in the English language, and 20 articles were published in Chinese-language journals. Of the 50 studies, 6 (12%) were found to show a low risk of bias, 36 (72%) were classified as having a moderate risk, and 8 (16%) were classified as having a high risk of bias. The characteristics of eligible studies are summarized in Table 1.

Figure 1. Flowchart presenting the selection of studies for inclusion in the systematic review and meta-analysis. HIVST: HIV self-testing.

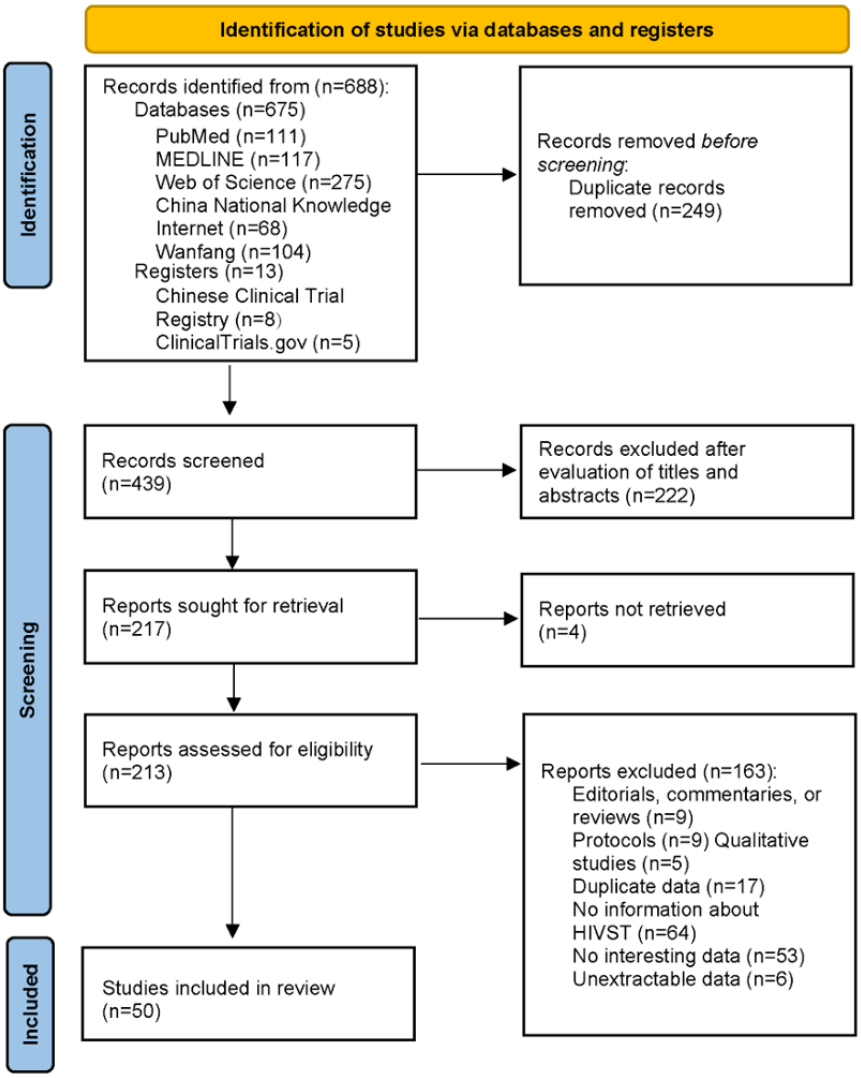


Table 1. Characteristics of eligible studies.

Author	Publication year	Project year	Study design	Population	Setting	Provides HIVST[a] kits	Fee for testing kits	Ways of supply	Quality score
Qi [19]	2013	2011-2012	Cohort	MSM[b] with HIV-negative status	Beijing	Yes	Free	On-site	8
Han et al [20]	2014	2013	Cross-sectional	MSM	Guangdong and Chongqing	No	N/A[c]	N/A	7
Tao et al [21]	2014	2012	Cross-sectional	MSM with HIV-negative or unknown status	National	Yes	US $10 deposit	Postal	7
Wong et al [13]	2015	2013	Cross-sectional	MSM	Hong Kong	No	N/A	N/A	8
Yan et al [14]	2015	2013-2014	Cross-sectional	MSM with HIV uninfected or unknown status	Jiangsu	No	N/A	N/A	8
Zhong et al [22]	2017	2015	Cross-sectional	MSM	Guangdong	Yes	US $23 deposit	Postal	4
Qin et al [23]	2017	2015	Cross-sectional	MSM	National	No	N/A	N/A	9
Ren et al [24]	2017	2016	Cross-sectional	MSM with HIV-negative or unknown status having ever taken an HIV self-test	Beijing	No	N/A	N/A	8
Ren et al [15]	2017	2016	Cross-sectional	MSM with HIV-negative or unknown status	Beijing	No	N/A	N/A	7
Zhou et al [25]	2017	2016-2017	Cross-sectional	MSM	Guangdong	Yes	¥100 (US $13.94) deposit	Postal	6
Jin et al [26]	2017	2015-2016	Cross-sectional	MSM never tested for HIV	Unknown	Yes	Free	Postal	6
Jin [27]	2017	2016	Cross-sectional	MSM with HIV-positive status	Guangdong	No	N/A	N/A	5
Tang et al [28]	2018	2016	Cross-sectional	MSM	Guangdong and Shandong	No	N/A	N/A	9
Tang et al [29]	2018	2016-2017	RCT[d]	MSM HIV-negative or unknown status	Guangdong or Shandong	Yes	Free	Postal	8
Wei et al [30]	2018	2017	Qualitative participant observation study	MSM who self-report being HIV negative or unknown status	Jiangsu	Yes	Free	On-site	6
Wei et al [31]	2018	2017	Cross-sectional	MSM with HIV-negative status	Jiangsu	No	N/A	N/A	8
Xue [32]	2018	Unknown	Cross-sectional	MSM who self-report being HIV negative or unknown status	Guangdong and Shandong	No	N/A	N/A	4
Cheng et al [33]	2019	Unknown	RCT	Unknown	Unknown	Unknown	Unknown	Unknown	6
Fan et al [34]	2019	2017-2018	Cross-sectional	Students at 5 universities	Sichuan	Yes	Unknown	Postal or on-site	6
Tang et al [35]	2019	2016-2017	Cohort	MSM who self-report being HIV negative or unknown status	Guangdong and Shandong	No	N/A	N/A	8
Wei et al [36]	2019	2015-2017	Cross-sectional	MSM	Guangdong	No	N/A	N/A	9

Author	Publication year	Project year	Study design	Population	Setting	Provides HIVST[a] kits	Fee for testing kits	Ways of supply	Quality score
Jin et al [37]	2019	2017	Cross-sectional	MSM with HIV-negative or unknown status	14 provinces in China	Yes	A US $5 deposit and a nonrefundable shipping fee of US $2 to US $3	Postal	6
Yang et al [38]	2019	Unknown	Cross-sectional	MSM who have ever tested for HIV	Nationwide	No	N/A	N/A	5
Zhu [39]	2019	Unknown	RCT	HIV-negative MSM	Anhui	Yes	Free	Postal or on-site	7
Mao [40]	2019	2017-2018	Cross-sectional	Unknown	Unknown	Yes	¥50 (US $6.97) deposit	Unknown	8
Zhao [41]	2019	2018-2019	Cohort	HIV-negative MSM	Anhui	Yes	Free	On-site	6
Liu et al [16]	2020	2017	Cross-sectional	MSM with HIV uninfected or unknown status	Chongqing	Yes	Free	On-site	9
Luo et al [42]	2020	2014-2016	Cross-sectional	Traceable sexual partners of newly diagnosed HIV-positive MSM	Zhejiang	Yes	Free	Sexual partners	7
Wang et al [43]	2020	2019	Cross-sectional	Female sex workers	7 provinces in China	No	N/A	N/A	9
Zhang et al [44]	2020	2018	Cross-sectional	MSM	National	Yes	Free	Postal or on-site	7
Luo et al [45]	2020	2019	Cross-sectional	MSM with HIV-negative or unknown status	National	No	N/A	N/A	7
Zhao [46]	2020	2017-2019	Cross-sectional	MSM	National	Yes	¥50 (US $6.97) deposit	Postal	7
Huang et al [47]	2020	2019	Cross-sectional	MSM	Guangdong	Yes	Unknown	Postal	7
Chan et al [48]	2021	2017	Cross-sectional	MSM	Hong Kong	Yes	Free	Postal or on-site	6
Cheng et al [49]	2021	2016	RCT	HIV-negative MSM	Guangdong	No	N/A	N/A	8
Hong et al [50]	2021	2019	Cross-sectional	MSM	Zhejiang	No	N/A	N/A	8
Lau et al [51]	2021	Before 2018 (unknown)	Cross-sectional	Male clients of female sex workers	Hong Kong	No	N/A	N/A	7
Li et al [52]	2021	2020	Cross-sectional	MSM who have ever tested for HIV	Jiangsu	No	N/A	N/A	7
Li et al [53]	2021	2017-2019	Cross-sectional	MSM	National	Yes	US $7 deposit	Postal or partner distribution	8
Ni et al [54]	2021	2019-2020	RCT	MSM	Unknown	Yes	US $15 deposit	Postal or partner distribution	3
Wu et al [55]	2021	2018-2019	Cross-sectional	MSM	Guangdong	Yes	US $15 deposit	Postal or partner distribution	6
Shan et al [56]	2021	2019	Cross-sectional	MSM	Beijing	No	N/A	N/A	5
Wu et al [57]	2021	2013,2014,2015, 2016,2018	Cross-sectional	MSM	National	No	N/A	N/A	3

Author	Publication year	Project year	Study design	Population	Setting	Provides HIVST[a] kits	Fee for testing kits	Ways of supply	Quality score
Zhou et al [58]	2021	2017-2019	Cross-sectional	MSM	Guangdong	Yes	¥100 (US $13.94) deposit	Postal or partner distribution	6
Li et al [59]	2021	2020	Cross-sectional	MSM	Beijing	Yes	Unknown	Postal	6
Chu [60]	2021	2018	RCT	HIV-negative MSM	National	Yes	Free	Postal	5
Zhao et al [61]	2021	2019	RCT	MSM without syphilis or unknown status	National	Yes	Free	Postal	5
Jin et al [62]	2021	2018-2019	Cross-sectional	MSM with HIV-negative status	4 provinces in China	No	N/A	N/A	6
Bao et al [63]	2021	2018-2019	Cross-sectional	MSM who self-report being HIV negative or unknown status	Shanghai	No	N/A	N/A	6
Bao et al [64]	2021	2019-2020	Cross-sectional	MSM who self-report being HIV negative or unknown status	Shanghai	No	N/A	N/A	6

[a]HIVST: HIV self-testing.

[b]MSM: men who have sex with men.

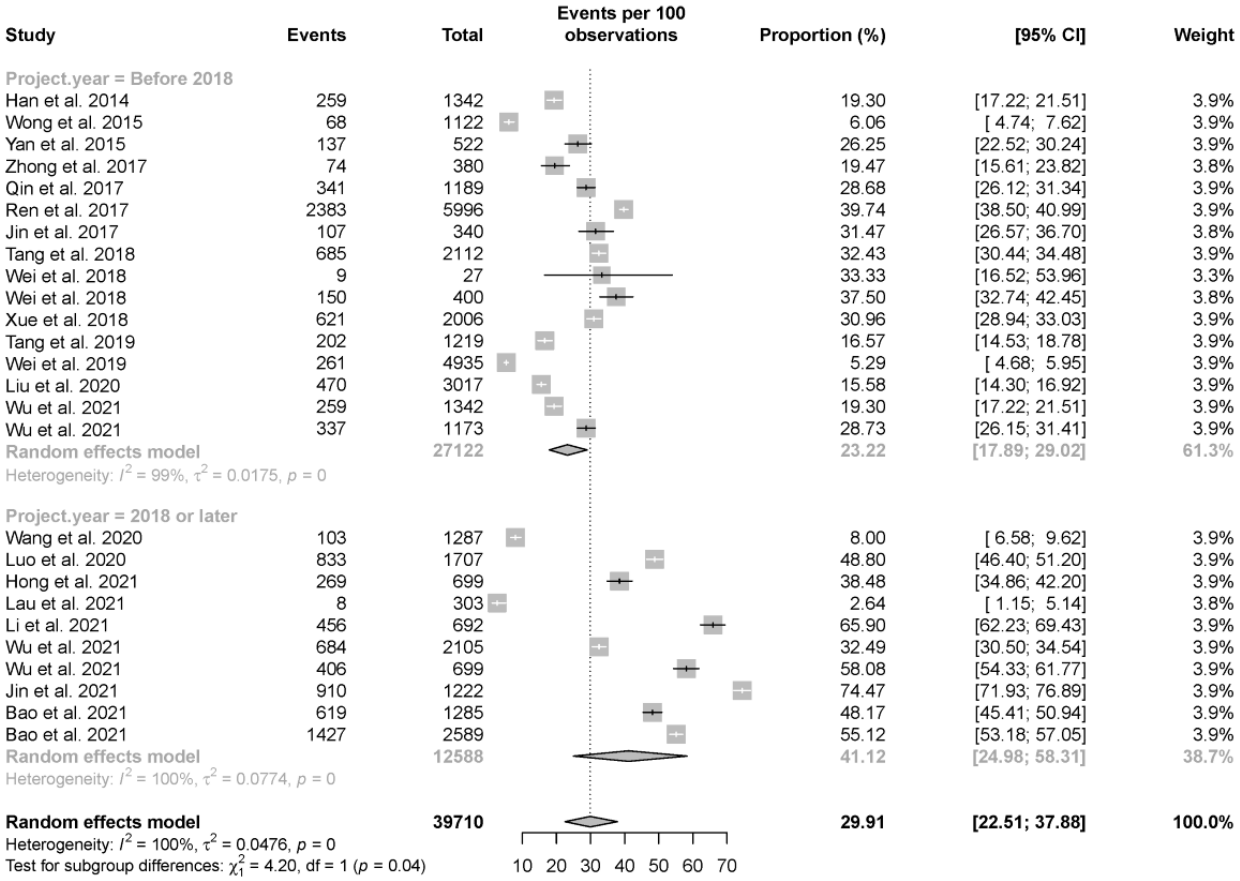
[c]N/A: not applicable.

[d]RCT: randomized controlled trial.

Proportion of Previously Self-tested

A total of 23 studies investigated the proportion of those who previously self-tested. Only 2 studies reported on the prevalence of HIVST used in the past in HIV high-risk populations other than men who have sex with men (MSM) [43,51], with the lowest proportion of those who previously self-tested being the male clients of female sex workers (2.6%), as reported by Lau et al [51]. The highest prevalence of HIVST use in the past was observed in the study by Jin et al [62], that is, 74.5% in HIV-negative MSM. The estimated pooled prevalence of HIVST use in China was 29.91% (95% CI 22.51%-37.88%), with a higher estimated pooled prevalence in 2018 or later (41.12%, 95% CI 24.98%-58.31%) than before 2018 (23.22%, 95% CI 17.89%-29.02%; Figure 2).

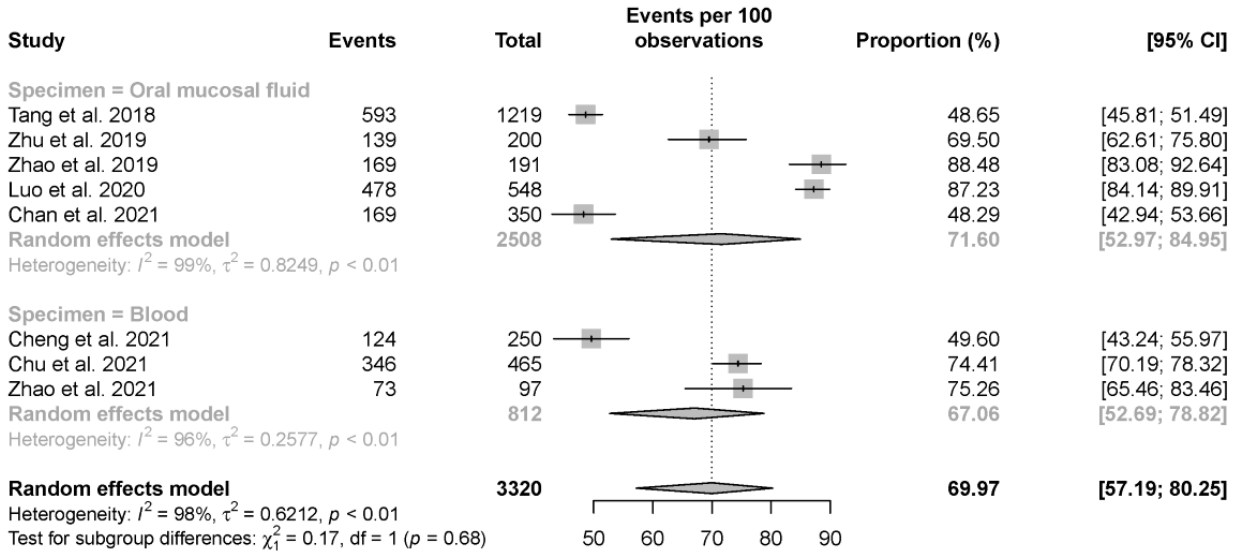
Figure 2. Pooled proportion of HIV self-testing use. Forest plot shows the estimated proportion of previously self-tested individuals before or after 2018.



Actual Uptake of HIVST

In all, 8 studies investigated the proportion of HIVST uptake after distributing HIVST kits to their study populations. The proportion of actual uptake of HIVST ranged from 48.29% to 88.48% across studies. It was shown that the pooled estimate of the proportion of HIVST was 69.97% (95% CI 51.19%-80.25%). Subgroup analysis revealed that the proportion of actual oral mucosal fluid–based HIVST was 71.6% (95% CI 52.97%-84.95%). The proportion of actual blood-based HIVST was 67.06% (95% CI 57.69%-78.82%). Heterogeneity among the studies was found to be statistically significant (I^2=98%; P<.001; Figure 3). To analyze the source of heterogeneity, we conducted a meta-regression to evaluate the impact that the categories of HIVST kits had on the substantial heterogeneity between the studies reporting actual HIVST uptake. However, we found no statistically significant difference in uptake of HIVST using either oral mucosal fluid– or blood-based HIVST kits (P=.45).

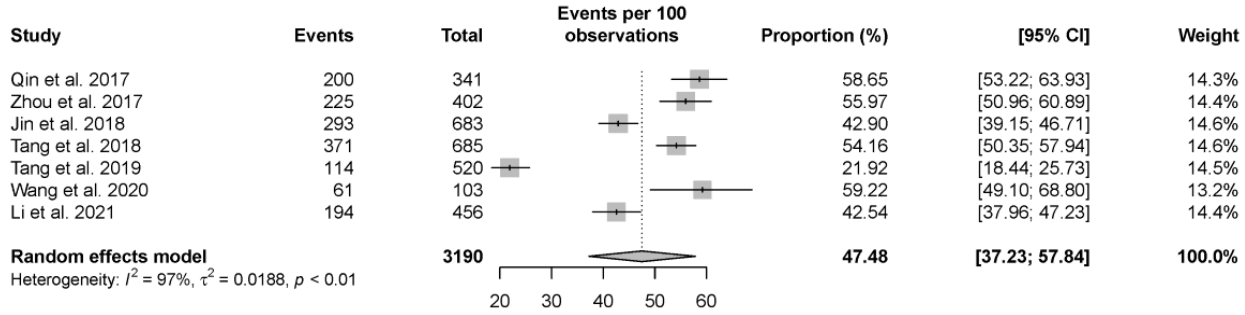
Figure 3. Pooled proportion of actual HIV self-testing uptake. Forest plot shows the estimated proportion of actual HIV self-testing uptake.



Self-testing as Lifetime First HIV Screening

In all, 7 studies examined individuals who used HIVST as their first-ever HIV-screening experience. The pooled proportion of individuals using HIVST as their lifetime first HIV screening was 47.48% (95% CI 37.23%-57.84%) but had high heterogeneity (I^2=97%; P<.001; Figure 4).

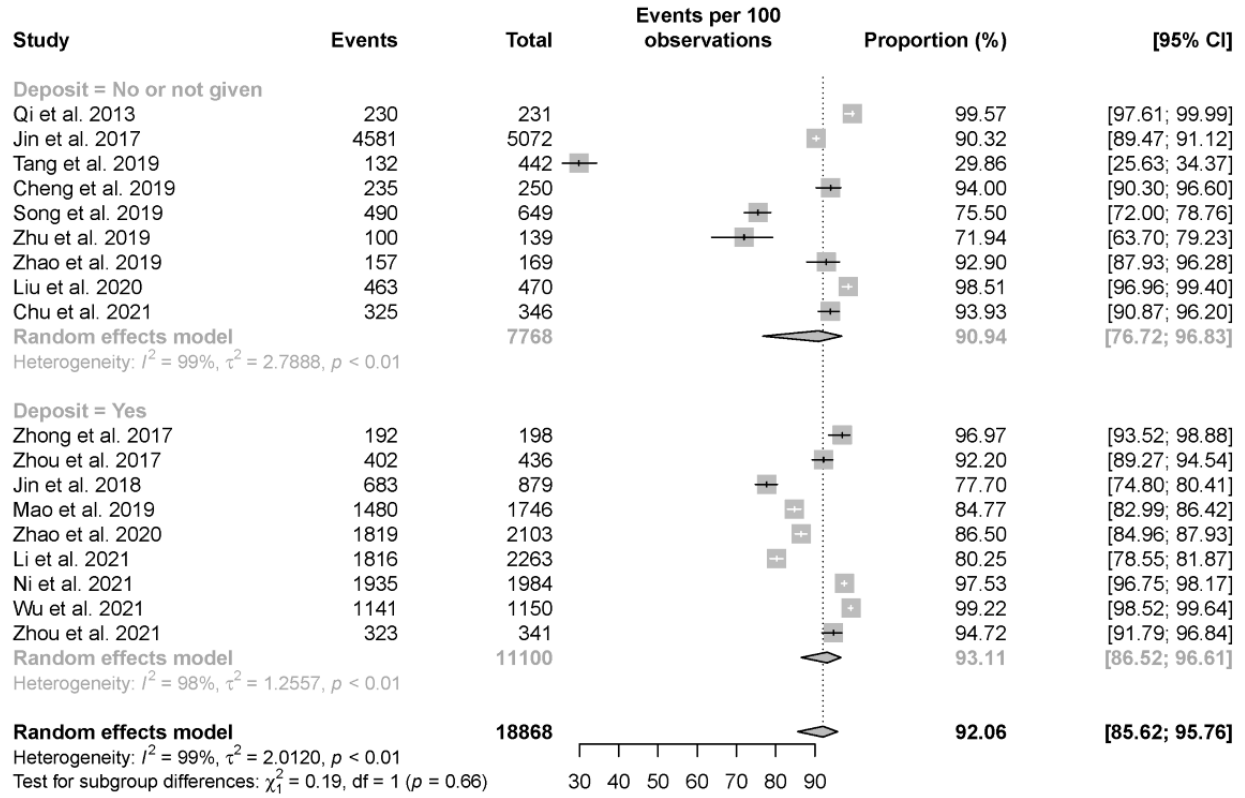
Figure 4. Pooled proportion of individuals using HIV self-testing (HIVST) as their first ever HIV test. Forest plot shows the proportion of individuals who tested for HIV for the first time via HIVST.



Results Feedback

Figure 5 presents the proportion of results feedback among those self-tested for HIV among 18 studies. The overall estimate of the feedback rate was 92.1% (95% CI 85.6%-95.8%) among those who self-tested for HIV. Cochran Q testing indicated substantial heterogeneity among the studies (I^2=99%; P<.001; Figure 5).

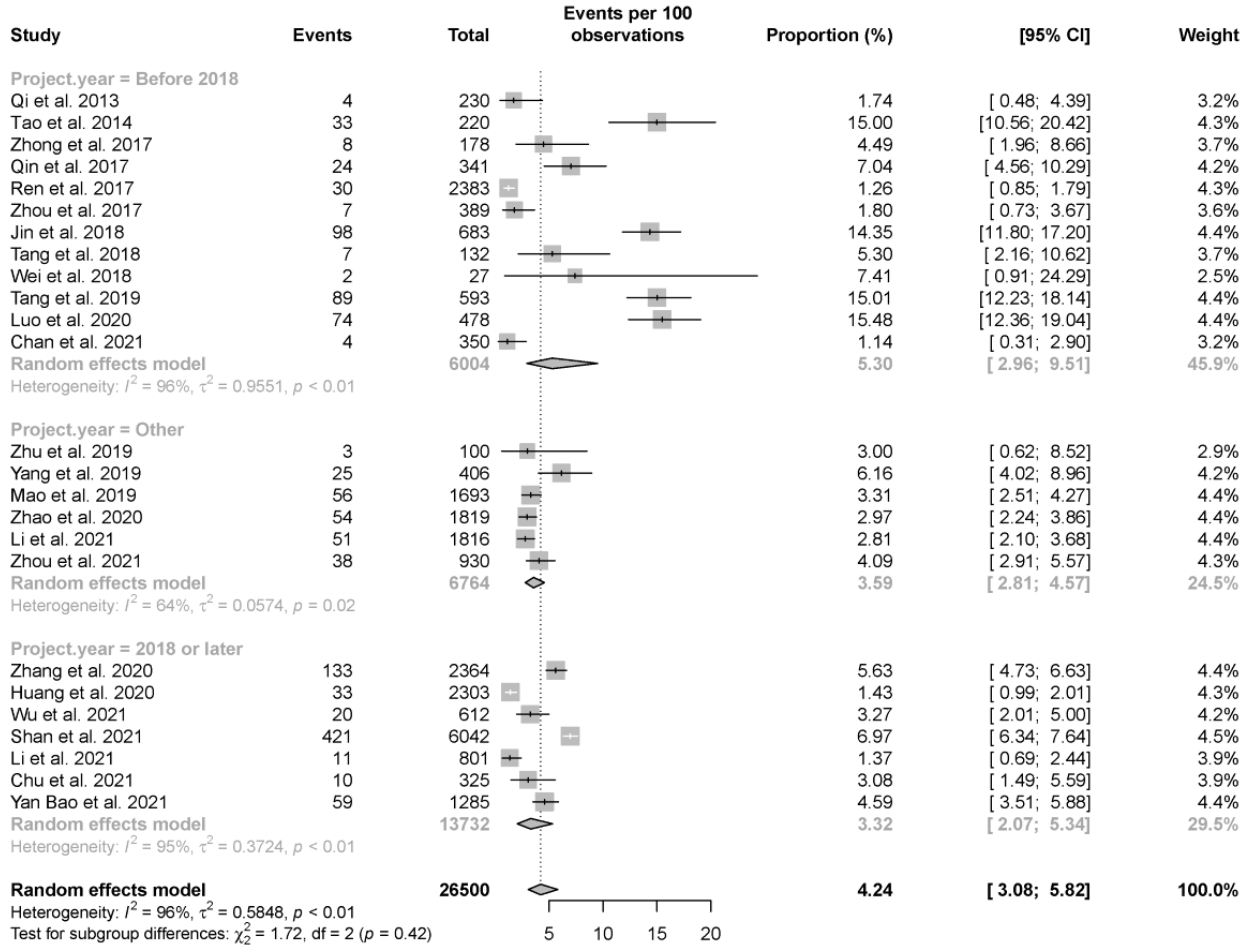
Figure 5. Pooled proportion of results feedback in those who self-tested for HIV. Forest plot shows estimated feedback rate among those self-tested for HIV.



Reactive Rate of HIVST

In all, 25 studies showed the reactive rate of HIVST. The pooled reactive rate of HIVST was 4.24% (95% CI 3.08%-5.82%). Interestingly, the pooled reactive rate of HIVST before 2018 was 5.3% (95% CI 2.96%-9.51%), which is higher than that in 2018 or later (3.32%, 95% CI 2.07%-5.34%). The included studies showed high heterogeneity (I^2=96%; P<.001; Figure 6).

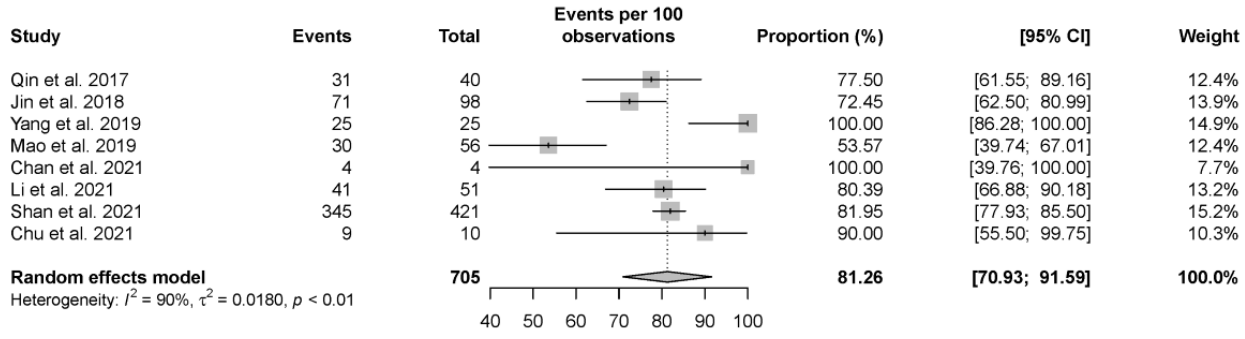
Figure 6. Pooled reactive rate of HIV self-testing (HIVST). Forest plot shows the reactive rate of HIVST before and after 2018. Project year "Other" refers to studies where the year that the project was conducted is unknown or cannot be distinguished.



Linkage to Care

Among the 25 studies reporting the reactive rate of HIVST, 8 studies reported the incidence of linkage to care in HIV-positive self-tested individuals. The pooled proportion of linkage to care among HIV-positive individuals was 81.26% (95% CI 70.93%-91.59%; Figure 7).

Figure 7. Pooled proportion of linkage to care. Forest plot shows the incidence of linkage to care among individuals whose HIV self-testing presented as reactive.



Sensitivity Analysis

Although only studies with moderate and low risk of bias were included, the pooled proportion of HIVST used previously (29.49%, 95%CI 19.99%-40%), the pooled proportion of actual uptake of HIVST (68.29%, 95% CI 51.02%-81.66%), the pooled proportion of individuals using HIVST as their lifetime first HIV screening (47.48%, 95% CI 37.23%-57.84%), the pooled feedback rate (91.38%, 95% CI 82.51%-95.97%), the pooled reactive rate of HIVST (4.21%, 95% CI 2.91%-6.09%), and the pooled proportion of linkage to care (74.68%, 95% CI 62.44%-86.92%) were similar to results obtained when including all studies (Multimedia Appendix 2).

Publication Bias

Funnel plots for the outcomes are shown in Multimedia Appendix 3. The *P* values obtained from the Egger test for asymmetry of the funnel plot were not significant for the pooled

proportion of HIVST used previously (*P*=.59), the pooled proportion of actual uptake of HIVST (*P*=.15), the pooled proportion of individuals using HIVST as their lifetime first HIV screening (*P*=.56), the pooled feedback rate (*P*=.28), and the pooled proportion of linkage to care (*P*=.58). Using the Egger test, a significant publication bias was observed in studies concerning the reactive rate of HIVST (*P*=.003), which is concerning since this tends to undermine, to a degree, the validity of the method that was used to measure the reactive rate of HIVST.

Discussion

Principal Findings

HIV testing is a key entry point for HIV prevention and treatment programs. Previous studies have shown that HIVST may enhance HIV testing uptake, increase HIV testing frequency, and limit potential harm among key populations [9,43]. According to Figueroa et al [65], when using HIV rapid diagnostic tests with high accuracy, those who self-test may obtain results similar to those obtained by health care workers in health care settings.

The potential factors influencing the use of HIVST include the demographic characteristics of population, age, level of education, and marital status [16,66,67]. In China, only 8% of female sex workers had ever tested for HIV using self-test kits [43]. HIVST uptake among MSM is much more prevalent than in other populations [15,31]. The higher uptake of HIVST in MSM may be attributable to HIV awareness raising by MSM-associated social organizations. For instance, crowdsourcing has been an effective strategy for enhancing HIVST uptake in MSM, especially in low- and middle-income countries [29]. Furthermore, the rate of HIVST use among MSM who use pre-exposure prophylaxis (PrEP), especially among those on daily PrEP, was observed to be high in a multicenter trial in China [68]. Thus, more extensive and specific applications of HIVST, such as using HIVST during the follow-up of PrEP users, may generate improved data for HIVST uptake and use. All forms of sex work are illegal in China, which restricts female sex workers and their male clients from actively seeking relevant and valuable knowledge regarding HIV. Therefore, inventive strategies are required to enhance the uptake of HIVST in key populations who are at high risk of HIV infection but possess limited knowledge of HIV prevention. In the 23 studies reporting the proportion of those who previously self-tested, only 2 studies [14,23] reported on the number of oral fluid–based HIVST used by self-testers (13.87% and 29%, respectively). Although it has been reported in the past that oral fluid–based HIVST is valid, acceptable, and accurate [5], the practical application of this testing method remains a challenge in China. Notably, oral fluid–based HIVST kit users in China were more likely to make errors during the oral HIVST testing procedure. Data gleaned from statistical meta-regression suggest that the category of the testing kit may not be the main or only reason for the actual difference in uptake between oral and blood-based HIVST kits.

According to our meta-analysis, it is estimated that approximately half (47.48%) of all individuals who were tested for HIV in China used HIVST as their first-ever test. HIVST has the potential to reach high-risk populations who have never been tested for HIV or who refuse voluntary counseling and testing. Qin et al [23], in contrast, observed that MSM who find sexual partners through the internet prefer medical facility–based testing rather than self-testing as their first-ever HIV test. This finding might be attributable to the fact that MSM who use the internet may have legitimate concerns regarding the accuracy of self-test kits after coming across relevant information on the internet, given their regular and routine access to internet-based information, in this instance, related to the overall quality and validity of HIVST kits on the internet [23]. HIVST with digital support may be an effective way to engage with those who self-test for HIV for the first time and some hard-to-reach populations, according to McGuire et al [69]. HIVST along with web-based counseling may be an effective strategy to increase the prevalence of HIV testing and reduce sexual risk behaviors [70], especially during the COVID-19 pandemic, when accessing HIV testing may have structural, regulatory, and psychological public health–related barriers [71,72]. Therefore, raising awareness of the accuracy and reliability of HIVST in a web-based digital manner might be desirable. In addition, monetary incentives combined with peer referral in the MSM population is also an effective manner to encourage first-time testing [73]. However, it is essential that other novel and inventive approaches for the promotion of first-time HIV testing are further explored.

The overall result feedback rate in those who self-test for HIV is known to be relatively high. Jin et al [37] compared the behaviors between those who did and those who did not submit their results after self-testing and observed, curiously, that individuals at lower HIV risk were more reluctant to submit their test results [37]. The lower feedback rate among individuals at lower HIV risk may lead to an inaccurate estimation of HIV prevalence in these populations. In this meta-analysis, we found that the cost of participants acquiring self-testing kits did not play a significant role in influencing results feedback.

The pooled reactive rate of HIVST among MSM in our study is similar to the overall national prevalence of HIV among MSM from 2001 to 2018, as estimated by Dong et al [74] (4.6%, 95% CI 3.2%-6.5% vs 5.7%, 95% CI 5.4%-6.1%). One recent study suggested that the lower reactive rate of HIVST is likely to be associated with a wider participant base in the study, and participants are not necessarily restricted to potential high-risk individuals only [3]. Our findings support the aforementioned speculation—that the reactive rate in China has declined steadily along with the widespread use of HIVST after 2018. However, beyond that, the reduced reactive rate may also reflect a demonstration of the significance of HIVST for HIV prevention.

Choko et al [75] indicated that financial incentives and partner-delivered approaches may likely increase male linkage into posttest HIV care. Among the studies providing self-test kits (excluding 2 studies with small sample size [48,60]; n≤10), those providing self-test kits distributed by sexual partners [53] showed a higher linkage-to-care rate among HIV-positive patients (81.9%), whereas for those studies only distributing testing kits by post [37,40], the linkage-to-care rate was from 53.6% to 72.4%. Our review observed that sexual partners may

play a critical role in accessing timely medical intervention for their partners with HIV-positive self-tests.

Limitations

We acknowledge several limitations to our study. First, most of the data used in our study were derived from the MSM population, and our analysis is based exclusively on this data. Therefore, the results of our study provide only limited outcomes and knowledge with respect to the application of HIVST in other high-risk populations. Second, data associated with the use of HIVST over extended periods are limited. One recent longitudinal study observed that HIVST adherence reduced to a paltry 10% during 1-year of follow-up [76]. Thus, the potential role of HIVST use over prolonged periods requires further exploration. Third, HIVST kits provided by research sponsors were available for free or on a refundable basis in the included studies. The effect of actual cost of access to HIVST kits in the real world in China was, therefore, not assessed or commented upon in our discussion.

Conclusion

In summary, HIVST has evolved over recent years into an important pillar of HIV prevention in China. However, the use of HIVST in non-MSM populations requires sustained upscaling. The long-term applications of HIVST and the effects of self-financing of HIVST in China have yet to be explored.

Acknowledgments

This work was supported by the Fund for Capital Health Development Research (2022-2-2185;2022-1G-3011), the Beijing Talent Project in the New Millennium (2020A35), and the National Science and Technology Major Project of China during the Thirteenth Five-Year Plan Period (grant 2017ZX10201101).

Authors' Contributions

XX conceptualized and led the design of the study. YC and XH contributed to the development of the study methods. XX, YQ, and YB assessed the quality and selected the studies. XX and YQ led the statistical analysis and drafting of the manuscript. HW, XH, and LW reviewed the manuscript and provided relevant feedback on the contents of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 17 KB - publichealth_v8i12e41125_app1.docx]

Multimedia Appendix 2
Sensitivity Analysis.
[DOCX File , 654 KB - publichealth_v8i12e41125_app2.docx]

Multimedia Appendix 3
Funnel plots for the outcomes.
[DOCX File , 118 KB - publichealth_v8i12e41125_app3.docx]

References

1. Joint United Nations Programme on HIV/AIDS. 2021. URL: https://www.unaids.org/en/resources/fact-sheet [accessed 2022-11-28]
2. HIV: key facts. World Health Organization. 2021 Nov 09. URL: https://www.who.int/news-room/fact-sheets/detail/hiv-aids [accessed 2022-11-25]
3. Muchedzi A, Mpofu M, Mudzengerere FH, Bateganya M, Mavimba T, Satti H, et al. High HIV Positivity Rates Following Large-Scale HIV Self-Testing Implementation in Zimbabwe, 2018-2020. Front Public Health 2021 Mar 23;9:606376 [FREE Full text] [doi: 10.3389/fpubh.2021.606376] [Medline: 33834013]
4. Ten threats to global health in 2019. World Health Organization. 2019. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2022-11-25]
5. Tonen-Wolyec S, Sarassoro A, Muwonga Masidi J, Twite Banza E, Nsiku Dikumbwa G, Maseke Matondo DM, et al. Field evaluation of capillary blood and oral-fluid HIV self-tests in the Democratic Republic of the Congo. PLoS One 2020;15(10):e0239607 [FREE Full text] [doi: 10.1371/journal.pone.0239607] [Medline: 33017442]
6. Lv Y, Li G, Hu M, Xu C, Lu H, Chen L, et al. Anonymous linkage between college students and human immunodeficiency virus (HIV) facilities: systematic evaluation of urine self-collection for HIV testing initiative in China. Clin Infect Dis 2021 Sep 07;73(5):e1108-e1115. [doi: 10.1093/cid/ciaa1816] [Medline: 33294913]

7. Johnson CC, Kennedy C, Fonner V, Siegfried N, Figueroa C, Dalal S, et al. Examining the effects of HIV self-testing compared to standard HIV testing services: a systematic review and meta-analysis. J Int AIDS Soc 2017 May 15;20(1):21594 [FREE Full text] [doi: 10.7448/IAS.20.1.21594] [Medline: 28530049]
8. Zhang C, Li X, Brecht M, Koniak-Griffin D. Can self-testing increase HIV testing among men who have sex with men: a systematic review and meta-analysis. PLoS One 2017 Nov 30;12(11):e0188890 [FREE Full text] [doi: 10.1371/journal.pone.0188890] [Medline: 29190791]
9. Witzel TC, Eshun-Wilson I, Jamil MS, Tilouche N, Figueroa C, Johnson CC, et al. Comparing the effects of HIV self-testing to standard HIV testing for key populations: a systematic review and meta-analysis. BMC Med 2020 Dec 03;18(1):381 [FREE Full text] [doi: 10.1186/s12916-020-01835-z] [Medline: 33267890]
10. Tucker JD, Ong J, Conserve D, Pan S, Tang W. Scaling up HIV self-testing in China and Africa. J Virus Erad 2017 Jul 01;3(3):167 [FREE Full text] [Medline: 28758026]
11. Tang W, Wu D. Opportunities and challenges for HIV self-testing in China. Lancet HIV 2018 Nov;5(11):e611-e612 [FREE Full text] [doi: 10.1016/S2352-3018(18)30244-3] [Medline: 30213725]
12. The National Guideline for Detection of HIV/AIDS (2020). Chinese Center for Disease Control and Prevention. 2020 Mar. URL: https://ncaids.chinacdc.cn/zxzx/zxdteff/202005/W020200522484711502629.pdf [accessed 2022-11-25]
13. Wong HTH, Tam HY, Chan DPC, Lee SS. Usage and acceptability of HIV self-testing in men who have sex with men in Hong Kong. AIDS Behav 2015 Mar 22;19(3):505-515. [doi: 10.1007/s10461-014-0881-1] [Medline: 25145608]
14. Yan H, Yang H, Raymond HF, Li J, Shi L, Huan X, et al. Experiences and correlates of HIV self-testing among men who have sex with men in Jiangsu province, China. AIDS Behav 2015 Mar;19(3):485-491 [FREE Full text] [doi: 10.1007/s10461-014-0968-8] [Medline: 25480598]
15. Ren XL, Wu ZY, Mi GD, McGoogan JM, Rou KM, Zhao Y. Uptake of HIV Self-testing among men who have sex with men in Beijing, China: a cross-sectional study. Biomed Environ Sci 2017 Jun;30(6):407-417 [FREE Full text] [doi: 10.3967/bes2017.054] [Medline: 28705264]
16. Liu Y, Wu G, Lu R, Ou R, Hu L, Yin Y, et al. Facilitators and barriers associated with uptake of HIV self-testing among men who have sex with men in Chongqing, China: a cross-sectional survey. Int J Environ Res Public Health 2020 Mar 03;17(5):1634 [FREE Full text] [doi: 10.3390/ijerph17051634] [Medline: 32138263]
17. Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. J Clin Epidemiol 2012 Sep;65(9):934-939. [doi: 10.1016/j.jclinepi.2011.11.014] [Medline: 22742910]
18. Mogire RM, Mutua A, Kimita W, Kamau A, Bejon P, Pettifor JM, et al. Prevalence of vitamin D deficiency in Africa: a systematic review and meta-analysis. Lancet Glob Health 2020 Jan;8(1):e134-e142 [FREE Full text] [doi: 10.1016/S2214-109X(19)30457-7] [Medline: 31786117]
19. Qi X. HIV infection and testing of regular sex partners among men who have sex with men in Beijing [dissertation]. Article in Chinese. Hebei Medical University. 2013. URL: https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201302&filename=1013287632.nh [accessed 2022-11-28]
20. Han L, Bien CH, Wei C, Muessig KE, Yang M, Liu F, et al. HIV self-testing among online MSM in China: implications for expanding HIV testing among key populations. J Acquir Immune Defic Syndr 2014 Oct 01;67(2):216-221 [FREE Full text] [doi: 10.1097/QAI.0000000000000278] [Medline: 24991972]
21. Tao J, Li M, Qian H, Wang L, Zhang Z, Ding H, et al. Home-based HIV testing for men who have sex with men in China: a novel community-based partnership to complement government programs. PLoS One 2014 Jul 22;9(7):e102812 [FREE Full text] [doi: 10.1371/journal.pone.0102812] [Medline: 25051160]
22. Zhong F, Tang W, Cheng W, Lin P, Wu Q, Cai Y, et al. Acceptability and feasibility of a social entrepreneurship testing model to promote HIV self-testing and linkage to care among men who have sex with men. HIV Med 2017 May;18(5):376-382 [FREE Full text] [doi: 10.1111/hiv.12437] [Medline: 27601301]
23. Qin Y, Tang W, Nowacki A, Mollan K, Reifeis SA, Hudgens MG, et al. Benefits and potential harms of human immunodeficiency virus self-testing among men who have sex with men in China: an implementation perspective. Sex Transm Dis 2017 Apr;44(4):233-238 [FREE Full text] [doi: 10.1097/OLQ.0000000000000581] [Medline: 28282650]
24. Ren XL, Wu ZY, Mi GD, McGoogan JM, Rou KM, Zhao Y, et al. HIV care-seeking behaviour after HIV self-testing among men who have sex with men in Beijing, China: a cross-sectional study. Infect Dis Poverty 2017 Jun 28;6(1):112 [FREE Full text] [doi: 10.1186/s40249-017-0326-y] [Medline: 28655340]
25. Zhou Y, Du L, Dai WC, Yu Y, He X, Li Y. Result analysis of internet-based HIV self-testing among MSM in Zhuhai city. Article in Chinese. Chinese Journal of AIDS & STD 2017 Dec 14;23(12):1128-1130. [doi: 10.13419/j.cnki.aids.2017.12.14]
26. Jin X, Xiao D, Xiu X, Ding ZW, Bao YG. Risk factors of men who have sex with men seeking for internet based HIV self-testing. Article in Chinese. Chinese Journal of AIDS & STD 2017;23(11):1043-1046. [doi: 10.13419/j.cnki.aids.2017.11.19]
27. Jin W. Willingness and its correlations of promoting HIV testing of sexual partners by self-testing kits among HIV-positive MSM in Guangzhou [dissertation]. Article in Chinese. Guangdong Pharmaceutical University. 2017. URL: https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201801&filename=1017734470.nh [accessed 2022-11-28]

28. Tang W, Liu C, Cao B, Pan SW, Zhang Y, Ong J, SESH Study Group. Receiving HIV serostatus disclosure from partners before sex: results from an online survey of Chinese men who have sex with men. AIDS Behav 2018 Dec;22(12):3826-3835 [FREE Full text] [doi: 10.1007/s10461-018-2062-0] [Medline: 29470809]
29. Tang W, Wei C, Cao B, Wu D, Li KT, Lu H, et al. Crowdsourcing to expand HIV testing among men who have sex with men in China: a closed cohort stepped wedge cluster randomized controlled trial. PLoS Med 2018 Aug;15(8):e1002645 [FREE Full text] [doi: 10.1371/journal.pmed.1002645] [Medline: 30153265]
30. Wei C, Yan L, Li J, Su X, Lippman S, Yan H. Which user errors matter during HIV self-testing? a qualitative participant observation study of men who have sex with men (MSM) in China. BMC Public Health 2018 Sep 10;18(1):1108 [FREE Full text] [doi: 10.1186/s12889-018-6007-3] [Medline: 30200905]
31. Wei C, Yan L, Lippman SA, Su X, Li J, Chen L, et al. Prevalence and correlates of point-of-sex human immunodeficiency virus self-testing among human immunodeficiency virus-negative men who have sex with men in China. Sex Transm Dis 2018 Dec;45(12):818-822 [FREE Full text] [doi: 10.1097/OLQ.0000000000000885] [Medline: 29944646]
32. Xue L. Reliability and validity of HIV testing self-efficacy scale and anticipated HIV stigma scale and their impact on HIV testing among MSM [dissertation]. Article in Chinese. Shandong University. 2018. URL: https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201901&filename=1018106479.nh [accessed 2022-11-28]
33. Cheng W, Friedman M, Egan J, Hampton M, Zhong F, Xu H. O11.5 Impact of HIV self-testing on the promotion of HIV testing among men who have sex with men in China: a randomized controlled trial. Sex Transm Infect 2019 Jul 16;95:A64. [doi: 10.1136/sextrans-2019-sti.168]
34. Fan S, Chen R, Chen H, Feng C, Jin X, Zou H. Provision of online and offline HIV self-testing kits for students attending universitiesxperience from Sichuan, China. Sex Transm Infect 2019 Jul 15;95:A195. [doi: 10.1136/sextrans-2019-sti.493]
35. Tang W, Huang W, Lu H, Cao B, Wu D, Ong J, et al. What happens after HIV self-testing? results from a longitudinal cohort of Chinese men who have sex with men. BMC Infect Dis 2019 Sep 14;19(1):807 [FREE Full text] [doi: 10.1186/s12879-019-4455-8] [Medline: 31521123]
36. Wei L, Chen L, Zhang H, Yang Z, Liu S, Tan W, et al. Relationship between gay app use and HIV testing among men who have sex with men in Shenzhen, China: a serial cross-sectional study. BMJ Open 2019 Aug 24;9(8):e028933 [FREE Full text] [doi: 10.1136/bmjopen-2019-028933] [Medline: 31446409]
37. Jin X, Xu J, Smith MK, Xiao D, Rapheal ER, Xiu X, et al. An internet-based self-testing model (Easy Test): cross-sectional survey targeting men who have sex with men who never tested for HIV in 14 provinces of China. J Med Internet Res 2019 May 15;21(5):e11854 [FREE Full text] [doi: 10.2196/11854] [Medline: 31094339]
38. Yang F, Tang W, Wang C. P034 HIV self-testing and potential linkage to care among men who have sex with men in China: a cross-sectional online survey. Sex Transm Infect 2019 Jul 15;95:A95. [doi: 10.1136/sextrans-2019-sti.242]
39. Zhu XF. Demand and obstacles of rapid HIV self-test among men who have sex with men: a qualitative study [dissertation]. Article in Chinese. Anhui Medical University. 2019. URL: https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201902&filename=1019175534.nh [accessed 2022-11-28]
40. Mao X. The surveillance of HIV epidemic and promotion of HIV test among men who have sex with men in China based on HIV self-testing strategy [dissertation]. Article in Chinese. China Medical University. 2019. URL: https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CDFDLAST2022&filename=1019891767.nh [accessed 2022-11-28]
41. Zhao Y. Oral HIV self-testing and counseling service model among men who have sex with men [dissertation]. Article in Chinese. Anhui Medical University. 2019. URL: https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201902&filename=1019175531.nh [accessed 2022-11-28]
42. Luo M, Hann K, Zhang G, Pan X, Ma Q, Jiang J, et al. HIV testing uptake and yield among sexual partners of HIV-positive men who have sex with men in Zhejiang Province, China, 2014-2016: a cross-sectional pilot study of a choice-based partner tracing and testing package. PLoS One 2020 Jun 04;15(6):e0232268 [FREE Full text] [doi: 10.1371/journal.pone.0232268] [Medline: 32497114]
43. Wang C, Wang YJ, Tucker JD, Xiong MZ, Fu HF, Smith MK, et al. Correlates of HIV self-testing among female sex workers in China: implications for expanding HIV screening. Infect Dis Poverty 2020 Oct 22;9(1):147 [FREE Full text] [doi: 10.1186/s40249-020-00765-5] [Medline: 33092641]
44. Zhang W, Hu Q, Tang W, Jin X, Mao X, Lu T, et al. HIV self-testing programs to men who have sex with men delivered by social media key opinion leaders and community-based organizations are both effective and complementary: a national pragmatic study in China. J Acquir Immune Defic Syndr 2020 Aug 15;84(5):453-462. [doi: 10.1097/QAI.0000000000002375] [Medline: 32692103]
45. Luo QQ, Luo YC, Ren XL. Comparison of HIV infection risk between 15 to 24 year-old student men who have sex with men and non-student men who have sex with men: a cross-sectional study. Article in Chinese. Zhonghua Yu Fang Yi Xue Za Zhi 2020 Dec 06;54(12):1435-1440. [doi: 10.3760/cma.j.cn112150-20191217-00938] [Medline: 33333663]
46. Zhao N. Application of self-testing method to implement HIV and HAV/HBV/HCV combined testing among MSM in China: a stepped-wedge cluster randomized controlled trial [dissertation]. Article in Chinese. China Medical University. 2020. URL: https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CDFDLAST2022&filename=1020383079.nh [accessed 2022-11-28]

47. Huang SZ, Dai WC, Li XF, Jiang XH, Tang WM, Zhou Y. Cost-effectiveness of HIV self-testing strategy in men who have sex with men. Article in Chinese. Zhonghua Liu Xing Bing Xue Za Zhi 2020 May 10;41(5):753-757. [doi: 10.3760/cma.j.cn112338-20191231-00926] [Medline: 32447920]
48. Chan PS, Chidgey A, Lau J, Ip M, Lau JTF, Wang Z. Effectiveness of a novel HIV self-testing service with online real-time counseling support (HIVST-Online) in increasing HIV testing rate and repeated HIV testing among men who have sex with men in Hong Kong: results of a pilot implementation project. Int J Environ Res Public Health 2021 Jan 15;18(2):729 [FREE Full text] [doi: 10.3390/ijerph18020729] [Medline: 33467770]
49. Cheng W, Egan JE, Zhong F, Xu H, Stall R, Friedman MR. Effectiveness of HIV self-testing on regular HIV testing among ever-tested men who have sex men in China: a pragmatic randomized controlled trial. AIDS Behav 2021 Nov;25(11):3494-3502. [doi: 10.1007/s10461-021-03161-8] [Medline: 33506306]
50. Hong H, Shi HB, Jiang HB, Dong HJ, Shen YL. Prevalence and associated factors of HIV self-testing among men who have sex with men in Ningbo, China: a cross-sectional study. AIDS Res Ther 2021 Apr 20;18(1):14 [FREE Full text] [doi: 10.1186/s12981-021-00339-x] [Medline: 33879191]
51. Lau CYK, Wang Z, Fang Y, Ip M, Wong KM, Chidgey A, et al. Prevalence of and factors associated with behavioral intention to take up home-based HIV self-testing among male clients of female sex workers in China - an application of the Theory of Planned Behavior. AIDS Care 2021 Aug 28;33(8):1088-1097. [doi: 10.1080/09540121.2020.1785996] [Medline: 32597197]
52. Li J, Marley G, Zhang Y, Chen Y, Tang W, Rongbin Y, et al. Determinants of recent HIV self-testing uptake among men who have sex with men in Jiangsu Province, China: an online cross-sectional survey. Front Public Health 2021 Nov 01;9:736440 [FREE Full text] [doi: 10.3389/fpubh.2021.736440] [Medline: 34790641]
53. Li S, Zhang J, Mao X, Lu T, Gao Y, Zhang W, et al. Feasibility of indirect secondary distribution of HIV self-test kits via WeChat among men who have sex with men: national cross-sectional study in China. J Med Internet Res 2021 Oct 26;23(10):e28508 [FREE Full text] [doi: 10.2196/28508] [Medline: 34698651]
54. Ni Y, Lu Y, Zhou Y, Tang W. S08.3b Using social media and social network to expand HIV self-testing in china. Sex Transm Infect 2021 Jul 06;97(SUPPL 1):A12. [doi: 10.1136/sextrans-2021-sti.40]
55. Wu D, Zhou Y, Yang N, Huang S, He X, Tucker J, et al. Social media-based secondary distribution of human immunodeficiency virus/syphilis self-testing among Chinese men who have sex with men. Clin Infect Dis 2021 Oct 05;73(7):e2251-e2257 [FREE Full text] [doi: 10.1093/cid/ciaa825] [Medline: 32588883]
56. Shan D, Xiong R, Xiao D, Li J, Shi YJ, Pan L. A survey on current situation of HIV self-testing kits use. Article in Chinese. Chinese Journal of AIDS & STD 2021 Feb 13;27(02):166-169. [doi: 10.13419/j.cnki.aids.2021.02.13]
57. Wu D, Cheng WB, Xiong Y, Lu Y, Ni YX, Tang WM. Changing trends of HIV testing and HIV self-testing in men who have sex with men on a gay social networking app. Article in Chinese. Zhonghua Liu Xing Bing Xue Za Zhi 2021 Feb 10;42(2):269-272. [doi: 10.3760/cma.j.cn112338-20200624-00882] [Medline: 33626614]
58. Zhou Y, Wu D, Tang WM, Li XF, Huang SZ, Liu YW, et al. The roles of two HIV self-testing models in promoting HIV-testing among men who have sex with men. Article in Chinese. Zhonghua Liu Xing Bing Xue Za Zhi 2021 Feb 10;42(2):263-268 [FREE Full text] [doi: 10.3760/cma.j.cn112338-20200629-00893] [Medline: 33626613]
59. Li J, Xin R, Li J, Lu H. Analysis on the status and influencing factors of Internet Plus-based HIV self-testing from wrine among MSM in Beijing. Article in Chinese. Int J Virol 2021 Jun;28(03):214-217 [FREE Full text] [doi: 10.3760/cma.j.issn.1673-4092.2021.03.009]
60. Chu ZX. The effect of HIV self-testing on condom use and serosorting behaviors among HIV negative men who have sex with men: a stepped-wedge randomized controlled trial [dissertation]. Article in Chinese. China Medical University. 2021. URL: https://kns.cnki.net/kcms/detail/detail.aspx?doi=10.27652/d.cnki.gzyku.2021.000087&dbcode=CDFD [accessed 2022-11-28]
61. Zhao PZ, Cheng WB, Tang WM, Zheng HP, Yang B, Wang C. Evaluating the effectiveness of HIV/syphilis joint self-testing in promoting syphilis testing among men who have sex with men. Article in Chinese. Zhonghua Liu Xing Bing Xue Za Zhi 2021 Feb 10;42(2):273-277. [doi: 10.3760/cma.j.cn112338-20200624-00880] [Medline: 33626615]
62. Jin X, Wang HY, Zhang J, Chu ZX, Hu ZL, Bao RT, et al. HIV self-testing reagent use in pre-exposure prophylaxis and related factors in men who have sex with men. Article in Chinese. Zhonghua Liu Xing Bing Xue Za Zhi 2021 Feb 10;42(2):278-283. [doi: 10.3760/cma.j.cn112338-20200420-00603] [Medline: 33626616]
63. Bao Y, Shen B, Gu KK, Yang Y, Wang HT, Han L. Analysis of the HIV self-testingits influencing factors among men who have sex with men in Shanghai. Article in Chinese. Chinese Journal of AIDS & STD 2021 Jul 10;27(7):715-718. [doi: 10.13419/j.cnki.aids.2021.07.10]
64. Bao Y, Xu YC, Shen B, Gu KK, Yang Y, Wang HT. Factors asociated with the self use of oral fluid HIV rapid test among men who have sex with men in Shanghai. Chinese Preventive Medicine. 2021 2021 Apr 11;22(4):297-300. [doi: 10.16506/j.1009-6639.2021.04.011]
65. Figueroa C, Johnson C, Ford N, Sands A, Dalal S, Meurant R, et al. Reliability of HIV rapid diagnostic tests for self-testing compared with testing by health-care workers: a systematic review and meta-analysis. Lancet HIV 2018 Jun;5(6):e277-e290 [FREE Full text] [doi: 10.1016/S2352-3018(18)30044-4] [Medline: 29703707]

66. Kurth AE, Cleland CM, Chhun N, Sidle JE, Were E, Naanyu V, et al. Accuracy and acceptability of oral fluid HIV self-testing in a general adult population in Kenya. AIDS Behav 2016 Apr;20(4):870-879 [FREE Full text] [doi: 10.1007/s10461-015-1213-9] [Medline: 26438487]
67. Wang X, Wu Z, Tang Z, Nong Q, Li Y. Promoting HIV testing with home self-test kit among men who have sex with men in China: a feasibility study. Lancet 2015 Oct 01;386:S68. [doi: 10.1016/s0140-6736(15)00649-2]
68. Zhang J, Tucker J, Tang W, Wang H, Chu Z, Hu Q, et al. Internet-based HIV self-testing among men who have sex with men through pre-exposure prophylaxis: 3-month prospective cohort analysis from China. J Med Internet Res 2021 Aug 27;23(8):e23978 [FREE Full text] [doi: 10.2196/23978] [Medline: 34448709]
69. McGuire M, de Waal A, Karellis A, Janssen R, Engel N, Sampath R, et al. HIV self-testing with digital supports as the new paradigm: a systematic review of global evidence (2010-2021). EClinicalMedicine 2021 Sep;39:101059 [FREE Full text] [doi: 10.1016/j.eclinm.2021.101059] [Medline: 34430835]
70. Wang Z, Lau JTF, Ip M, Ho SPY, Mo PKH, Latkin C, et al. A randomized controlled trial evaluating efficacy of promoting a home-based HIV self-testing with online counseling on increasing HIV testing among men who have sex with men. AIDS Behav 2018 Jan;22(1):190-201. [doi: 10.1007/s10461-017-1887-2] [Medline: 28831616]
71. Zhang K, Chen S, Chan PS, Fang Y, Cao H, Chen H, et al. Changes in HIV testing utilization among Chinese men who have sex with men during the COVID-19 pandemic in Shenzhen, China: an observational prospective cohort study. Front Med (Lausanne) 2022 Jun 02;9:842121 [FREE Full text] [doi: 10.3389/fmed.2022.842121] [Medline: 35755039]
72. Zhang KC, Fang Y, Cao H, Chen H, Hu T, Chen YQ, et al. The impacts of the COVID-19 pandemic on HIV testing utilization among men who have sex with men in China: cross-sectional online survey. JMIR Public Health Surveill 2022 May 25;8(5):e30070 [FREE Full text] [doi: 10.2196/30070] [Medline: 35486811]
73. Zhou Y, Lu Y, Ni Y, Wu D, He X, Ong JJ, et al. Monetary incentives and peer referral in promoting secondary distribution of HIV self-testing among men who have sex with men in China: a randomized controlled trial. PLoS Med 2022 Feb;19(2):e1003928 [FREE Full text] [doi: 10.1371/journal.pmed.1003928] [Medline: 35157727]
74. Dong MJ, Peng B, Liu ZF, Ye QN, Liu H, Lu XL, et al. The prevalence of HIV among MSM in China: a large-scale systematic analysis. BMC Infect Dis 2019 Nov 27;19(1):1000 [FREE Full text] [doi: 10.1186/s12879-019-4559-1] [Medline: 31775654]
75. Choko AT, Corbett EL, Stallard N, Maheswaran H, Lepine A, Johnson CC, et al. HIV self-testing alone or with additional interventions, including financial incentives, and linkage to care or prevention among male partners of antenatal care clinic attendees in Malawi: an adaptive multi-arm, multi-stage cluster randomised trial. PLoS Med 2019 Jan;16(1):e1002719 [FREE Full text] [doi: 10.1371/journal.pmed.1002719] [Medline: 30601823]
76. Yan X, Su H, Zhang B, Li Y, Zhang L, Jia Z. Adherence of HIV self-testing among men who have sex with men in China: longitudinal study. J Med Internet Res 2020 Sep 17;22(9):e19627 [FREE Full text] [doi: 10.2196/19627] [Medline: 32940619]

Abbreviations

HIVST: HIV self-testing
MSM: Men who have sex with men
PrEP: pre-exposure prophylaxis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by Y Khader; submitted 16.07.22; peer-reviewed by Z Wang, X Jin; comments to author 12.08.22; revised version received 15.10.22; accepted 18.11.22; published 02.12.22.

Please cite as:
Xiu X, Qin Y, Bao Y, Chen Y, Wu H, Huang X, Wang L
The Practice and Potential Role of HIV Self-testing in China: Systematic Review and Meta-analysis
JMIR Public Health Surveill 2022;8(12):e41125
URL: https://publichealth.jmir.org/2022/12/e41125
doi:10.2196/41125
PMID:36459393

Review

HIV-Specific Reported Outcome Measures: Systematic Review of Psychometric Properties

Ziqi Wang[1*], PhD; Yaxin Zhu[1*], PhD; Xiyu Duan[1], PhD; Hao Kang[2], MPH; Bo Qu[1,3], PhD

[1]Institute for International Health Professions Education and Research, China Medical University, Shenyang, China

[2]Administration Department of Nosocomial Infection, Southwest Hospital, Third Military Medical University, Chongqing, China

[3]School of Public Health, China Medical University, Shenyang, China

[*]these authors contributed equally

Corresponding Author:
Bo Qu, PhD
School of Public Health
China Medical University
No.77 Puhe Road
Shenyang North New Area
Shenyang, 110122
China
Phone: 86 18900910198
Email: qubo6666@163.com

Abstract

Background: The management of people living with HIV and AIDS is multidimensional and complex. Using patient-reported outcome measures (PROMs) has been increasingly recognized to be the key factor for providing patient-centered health care to meet the lifelong needs of people living with HIV and AIDS from diagnosis to death. However, there is currently no consensus on a PROM recommended for health care providers and researchers to assess health outcomes in people living with HIV and AIDS.

Objective: The purpose of this systematic review was to summarize and categorize the available validated HIV-specific PROMs in adults living with HIV and AIDS and to assess these PROMs using the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) methodology.

Methods: This systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. A literature search of 3 recommended databases (PubMed, Embase, and PsychINFO) was conducted on January 15, 2021. Studies were included if they assessed any psychometric property of HIV-specific PROMs in adults living with HIV and AIDS and met the eligibility criteria. The PROMs were assessed for 9 psychometric properties, evaluated in each included study following the COSMIN methodology by assessing the following: the methodological quality assessed using the COSMIN risk of bias checklist; overall rating of results; level of evidence assessed using the modified Grading of Recommendations, Assessment, Development, and Evaluation approach; and level of recommendation.

Results: A total of 88 PROMs classified into 8 categories, assessing the psychometric properties of PROMs for adults living with HIV and AIDS, were identified in 152 studies including 79,213 people living with HIV and AIDS. The psychometric properties of most included PROMs were rated with insufficient evidence. The PROMs that received class A recommendation were the Poz Quality of Life, HIV Symptom Index or Symptoms Distress Module of the Adult AIDS Clinical Trial Group, and People Living with HIV Resilience Scale. In addition, because of a lack of evidence, recommendations regarding use could not be made for most of the remaining assessed PROMs (received class B recommendation).

Conclusions: This systematic review recommends 3 PROMs to assess health outcomes in adults living with HIV and AIDS. However, all these PROMs have some shortcomings. In addition, most of the included PROMs do not have sufficient evidence for assessing their psychometric properties and require a more comprehensive validation of the psychometric properties in the future to provide more scientific evidence. Thus, our findings may provide a reference for the selection of high-quality HIV-specific PROMs by health care providers and researchers for clinical practice and research.

(JMIR Public Health Surveill 2022;8(12):e39015) doi:10.2196/39015

KEYWORDS

HIV; AIDS; people living with HIV and AIDS; patient-reported outcome measures; psychometric properties

Introduction

Background

According to the statistics from the Joint United Nations Program on HIV/AIDS, 28.2 million individuals were accessing antiretroviral therapy (ART) as of mid-2021 [1]. Although effective treatment via ART has improved the life expectancy of people living with HIV and AIDS [2], this population still faces substantial challenges brought by HIV [3-6]. Therefore, Lazarus et al [7] proposed the *Fourth 90* target to ensure that 90% of people living with HIV and AIDS with viral suppression have a good health-related quality of life (HRQoL) after the World Health Organization proposed the *90-90-90* targets. They proposed that HRQoL in people living with HIV and AIDS should be considered as important as viral suppression [8]. For people living with HIV and AIDS, the focus should be shifted toward improving HIV-related care [9].

The management of people living with HIV and AIDS is multidimensional and complex. To overcome the obstacles to achieving the *Fourth 90* [10], patient-centered care that can meet the lifelong needs of people living with HIV and AIDS from diagnosis to death is the key requirement [9]. The collection and use of patient-reported outcome (PRO) data is one of the most effective approaches for ensuring that the care reflects the needs and priorities of people living with HIV and AIDS [9]. Compared with clinician-reported outcomes, PROs present a more comprehensive method for assessing the subjective perceptions of people living with HIV and AIDS of their own health that cannot be observed or are not easily observed directly and have been shown to accurately predict health outcomes among this population [11,12]. Furthermore, there is sufficient evidence that PROs can be used to improve the care quality and health outcomes in people living with HIV and AIDS, such as by improving patient-physician communication [13], clinical decision-making [14], and symptom recognition [15].

Why Did This Systematic Review Only Include HIV-Specific PRO Measures?

Patient-reported outcome measures (PROMs) are the actual tool developed for collecting PRO data. There are 2 types of PROMs: generic (designed for use in any population and cover general aspects of outcome measures) and disease specific (designed for use in people with a condition and measure specific aspects of an outcome of importance). Many generic and HIV-specific PROMs have been validated in people living with HIV and AIDS. The advantage of a generic PROM is that it enables researchers to compare the health outcomes of people living with HIV and AIDS with those of other populations based on the same measurements [16]. However, unlike generic PROMs, HIV-specific PROMs do not have a significant ceiling and floor effect and do not overestimate health outcomes in people living with HIV and AIDS [17,18]. Furthermore, HIV-specific PROMs are more closely associated with HIV than are generic PROMs. In addition, they have the sensitivity for detecting and quantifying minor changes and specificity needed for HIV-specific domains, such as HIV-related stigma, comorbidities, and ART-related treatment [19]. Some related reviews have recommended a strategy to combine generic and HIV-specific PROMs to supplement HIV-specific health care outcomes that cannot be obtained with generic PROMs alone [20,21]. Clayson et al [20] suggested that the right combination of generic and HIV-specific PROMs can improve the comprehensiveness of assessment content, such that it includes not only the 3 core domains that generic PROMs focus on, that is, physical function, social or role function, and mental health or emotional well-being, but also the items or domains addressing issues relevant to HIV or AIDS and its treatment. Considering that many HIV-specific PROMs were developed before the widespread use of ART, they may not be able to detect the impact of current treatment on people living with HIV and AIDS and serve as an assessment tool for the long-term management of people living with HIV and AIDS [9]. In addition, many poorly designed PROMs lack a standardized development process. Therefore, it is necessary to summarize the existing HIV-specific PROMs and assess their psychometric properties.

Previous Studies

With the rapid development of this field, many HIV-specific PROMs have been developed. After a preliminary literature search in MEDLINE using a comprehensive search strategy (Table S1 in Multimedia Appendix 1), we found some relevant reviews. Wen et al [19] recently conducted a systematic review on a similar topic; however, they only aimed at identifying and assessing the psychometric properties of HRQoL in people living with HIV and AIDS. Engler et al [22] identified 117 different HIV-specific PROMs in 2016; however, they did not quantitatively assess the psychometric properties of these PROMs. Cooper [16] reported an overview of the available reviews and summarized the PROMs with <40 items for measuring HRQoL in people living with HIV and AIDS in 2017. Earlier, several researchers conducted nonsystematic reviews of some PROMs in specific contexts [20,23,24]. Although many previous reviews have summarized the content of some existing HIV-specific PROMs, few have comprehensively reported the psychometric properties of these PROMs and given recommendations for the use of these PROMs.

As accurate and reliable PROMs are a precondition for obtaining robust results, PROMs with good psychometric properties are indispensable for research [25]. *Lancet HIV* also suggested in the special issue of "HIV outcomes beyond viral suppression" that the psychometric properties of the existing PROMs should be assessed in line with the existing guidelines, such as the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines [9]. The COSMIN guidelines provide a consecutive procedure to help health care providers and researchers improve the selection of the most suitable PROMs in research and clinical practice [26]. Therefore, we conducted a systematic review to identify studies assessing the psychometric properties of HIV-specific PROMs

validated in a population of adults living with HIV and AIDS and categorized these PROMs based on the type of outcome measure. We further assessed the methodological quality and level of evidence of these PROMs in association with their psychometric properties.

Objective

The purpose of this systematic review was to summarize and categorize the available and validated HIV-specific PROMs for adults living with HIV and AIDS. This systematic review also aimed to use the COSMIN methodology to assess the psychometric properties of these PROMs and make an evidence-based and completely transparent recommendation for the use of these PROMs.

Methods

Overview

This systematic review was conducted and reported according to the COSMIN guidelines [27] and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [28]. It included only a secondary data analysis of publicly available content not involving human participants. Therefore, ethics approval was not required for this review.

Search Strategy

Three literature databases (MEDLINE, Embase, and PsycINFO) were searched on January 15, 2021. Two important web databases, PROQOLID and PROMIS, which contain a large number of PROMs and cover a wide range of populations and therapeutic areas, were also searched for PROMs. These 2 databases were developed by the Mapi Research Trust in France and the National Institutes of Health in the United States to facilitate the selection process of PROMs and are now used by many clinical investigators. The reference lists of relevant reviews in the preliminary literature search and the included studies were further examined for relevant publications. The search strategy used three COSMIN-guided search terms in reference to the search for constructs developed by Terwee et al [29]: (1) construct of interest, (2) condition of interest, and (3) psychometric properties (Table S2 in Multimedia Appendix 1). A comprehensive search strategy was developed under the guidance of a senior health research librarian.

Study Selection

The eligibility criteria of the studies were as follows: (1) the study validated HIV-specific PROMs for adults living with HIV or AIDS and assessed at least one of the 9 psychometric properties defined by the COSMIN guidelines: content validity, structural validity, internal consistency, cross-cultural validity or measurement invariance, reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness [30]; (2) the study was published in English in a peer-reviewed journal; and (3) the study applied self-administered PROMs for patients.

Studies were excluded if (1) they used the PROM mainly for outcome measures rather than for assessing the 9 psychometric properties; (2) they developed and used PROMs for screening or diagnostic purposes only; (3) they were not an original investigation, such as reviews, letters, and editorials; (4) they included generic PROMs or other disease-specific PROMs not related to or only partially related to HIV (such as the 36-Item Short-Form Health Survey Questionnaire); and (5) they provided indirect evidence of psychometric properties (such as studies using a PROM in a validation study of another instrument [30]).

The retrieved literature was imported into the EndNote software (version X9; Clarivate Plc), and duplications were automatically removed. A 2-stage screening process was used to select eligible studies. First, the titles and abstracts were screened based on the predetermined selection criteria (stage I). Subsequently, the full texts of articles deemed relevant or possibly relevant were obtained and further assessed for eligibility (stage II). Two independent researchers (ZW and YZ) determined study eligibility, and any disagreement was settled by consensus or discussion with a third researcher (BQ).

Data Exclusion

For the eligible studies, data were independently extracted by the same 2 researchers (ZW and YZ) using a standardized form, and completeness and correctness were confirmed. Any discrepancy was resolved via a discussion with the third researcher (BQ). The extracted data included the characteristics of PROMs (name of the PROM[abbreviation], year of PROM development, targeted concept, recall period, number of items, each domain and the number of items in each domain, response options and score range, and original language), characteristics of the included studies (first author [year of publication], the total number of patients [N], age, gender, patient description, years diagnosis, severity of disease, recruitment context, country of research, and effective response rate of the questionnaire), and results of the included studies (COSMIN risk of bias information, evidence of the 9 psychometric properties, and COSMIN summary and rating).

Data Analysis

According to the suggestions mentioned in the COSMIN guidelines, each PROM was assessed via a 4-step process [27]. First, the methodological quality for every psychometric property in each study was assessed using the COSMIN risk of bias checklist based on a four-point response, “very good,” “adequate,” “doubtful,” or “inadequate,” and an overall rating of the psychometric property was determined based on the item with the worst rating [30]. Second, the results for every psychometric property in each study were rated based on the updated criteria for good psychometric properties [27], and each result was graded as positive (+), negative (–), or indeterminate (?). Third, the overall results for each psychometric property of a PROM were rated as sufficient (+), insufficient (–), inconsistent (±), or indeterminate (?), and the level of evidence for each psychometric property of a PROM was rated as “high,” “moderate,” “low,” or “very low” by following the Grading of Recommendations, Assessment, Development, and Evaluation approach, which considered the initial level of evidence to be high, with subsequent downgrading based on the score for 4 criteria: risk of bias, inconsistencies, imprecision, and indirectness. Finally, a table summarizing the findings was constructed and used to make recommendations for the selection of the most suitable PROMs.

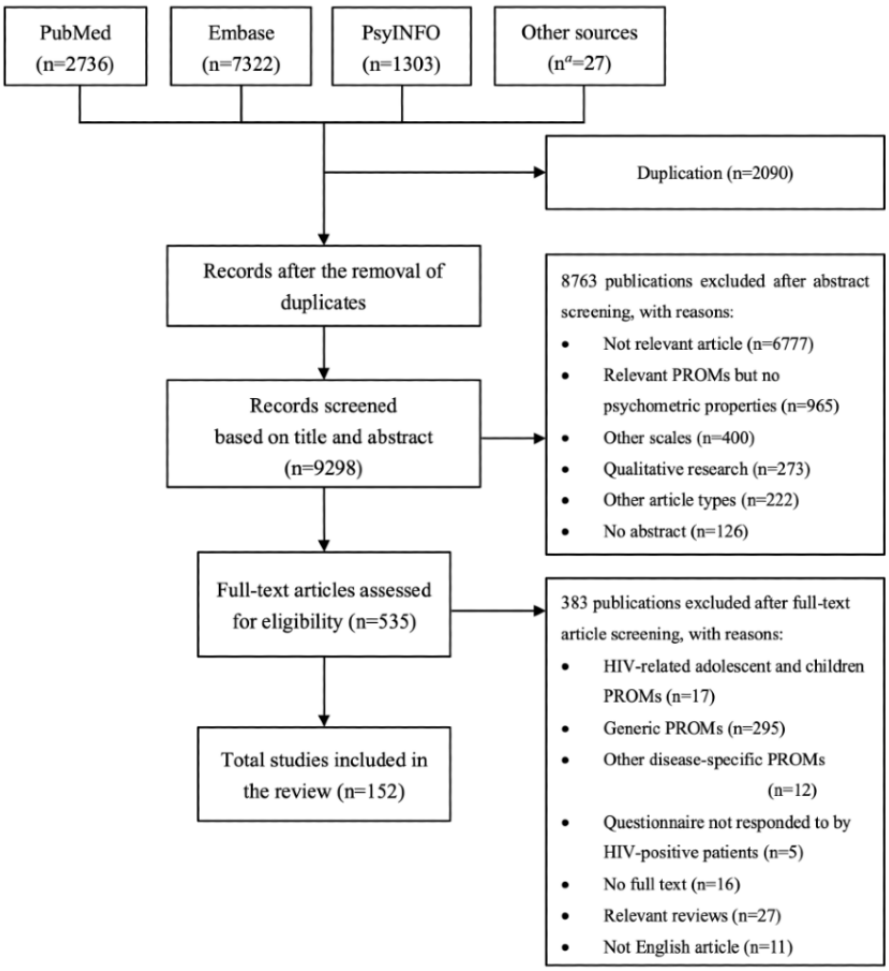
All assessments were conducted independently by 3 researchers (ZW, HK, and XD), and any disagreement was settled via consensus or discussion with a fourth researcher (YZ). The Cohen κ coefficient was calculated using the SPSS software (version 24.0; IBM) to evaluate the interrater agreement for title and abstract screening, study selection, and data extraction.

Results

Search Results

A total of 11,361 articles were identified in the literature search, and another 27 articles were identified through reference and citation searches. Of these, 2090 were excluded because of duplication. After screening the titles and abstracts, 535 articles were found to be potentially relevant, and their full text was reviewed for further assessment. Of these, 152 articles were finally included [31-182]. The PRISMA flow diagram and the reasons for exclusion are presented in Figure 1. The average Cohen κ coefficients for the title and abstract screening, study selection, and data extraction were 0.85, 0.82, and 0.89, respectively, indicating that the 2 researchers reached a "substantial agreement" as defined by Landis and Koch [183] in 1991.

Figure 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. PROM: patient-reported outcome measure. [a]These studies were identified through further research of the reference lists of relevant reviews in the preliminary literature search and the included studies.



Characteristics of the Included PROMs

Table 1 lists the characteristics of the included PROMs, with details of the subscales provided in Table S3 in Multimedia Appendix 1. A total of 88 PROMs were reported in the 152 included studies, and these PROMs can be divided into 8 categories (improved based on the initial taxonomy developed by Engler et al [22]): HRQoL (24/88, 27% of PROMs) [31-102], symptoms (10/88, 11% of PROMs) [103-120], stigma (15/88, 17% of PROMs) [121-142], psychological (8/88, 9% of PROMs) [143-151], body and facial appearance (5/88, 6% of PROMs) [152-156], treatment (17/88, 19% of PROMs) [153-173], social support (3/88, 3% of PROMs) [174-176], and self-management and self-care (6/88, 7% of PROMs) [177-182]. All the included PROMs were tools self-administered by people living with HIV and AIDS either in a clinical or research context. Of these 88 PROMs, 22 (25%) PROMs were developed before 2000, 31 (35%) between 2000 and 2009, and 35 (40%) after 2010. The recall period for PROMs ranged from "past 7 days" to "last 12 months." The number of items varied between 4 and 165. The original language for most PROMs was English, and the response option format for most PROMs was the 5-point Likert scale.

Table 1. Characteristics of the included patient-reported outcome measures (PROMs)[a].

PROM; year of development	Targeted concept	Recall period	Total no. of items	Response options	Score range	Original language
HRQoL[b]						
MOS-HIV[c] [31-51]; 1996	HRQoL	Past 4 weeks	35	Multiple response options	Raw scores for each scale were transformed to a scale of 0 to 100	English
MOS-HIV-17 [53]; 2000	HRQoL	Past 4 weeks	17	Multiple response options	Raw scores for each scale were transformed to a scale of 0 to 100	English
MOS-HIV-29 [52]; 2012	HRQoL	Past 30 days	29	Multiple response options	Raw scores for each scale were transformed to a scale of 0 to100	Luganda
HIV Overview of Problems Evaluation System [54,55]; 1992	HRQoL	__[d]	165	5-point Likert scale (0-4)	Summary of scales: physical scale, medical interaction, psychosocial scale, sexual scale, and significant others or partners	English
HIV-Related Quality of Life Questions [56]; 1993	HRQoL	Past month	34	Multiple response options	—	English
AIDS Health Assessment Questionnaire [57]; 1997	HRQoL	Different recall periods per dimensions	116	Multiple response options	Raw scores were transformed to a scale of 0 to 100	English
HIV-PARSE[e] [58]; 1994	HRQoL	Different recall periods per dimensions	30	Multiple response options	Perceived Health Index (25 items)	English
HIV-PARSE-Brief [59]; 1995	HRQoL	Different recall periods per dimensions	21	Multiple response options	Perceived Health Index (13 items)	English
HRQoL [60]; 1995	HRQoL	Past 4 weeks	64	Multiple response options	A physical health dimension and a Mental health dimension	English
Functional Assessment of HIV Infection [61-65]; 1996	HRQoL	Past 7 days	44	5-point Likert scale (0-4)	Sum of all item scores (0-176)	English
General Health Self-Assessment [66]; 1997	HRQoL	Past 4 weeks	49	Multiple response options	The subscales are scored as summated and transformed on a scale of 0 to 100	English
HIV Quality of Life 31-item scale [67]; 1997	HRQoL	—	31	Dichotomous: yes or no	Simple summation of dichotomous response options	French
HAT-QoL[f]-42 [68,69]; 1997	HRQoL	Past 4 weeks	42	5-point Likert scale (1-5)	All subscales are coded to range from 0 to 100	English
HAT-QoL-30 [35]; 1999	HRQoL	Past 4 weeks	30	5-point Likert scale (1-5)	All subscales are coded to range from 0 to 100	English
HAT-QoL-34 [42,70,71]; 2008	HRQoL	Past 4 weeks	34	5-point Likert scale (1-5)	All subscales are coded to range from 0 to 100	English
MQoL[g] for patients with HIV or AIDS [34,72-75]; 1997	HRQoL	—	40	7-point Likert scale (1-7)	Each subscale ranged from 4 to 28; mental health score + (2 × physical functioning score) = overall index for MQoL (12-84)	English

PROM; year of development	Targeted concept	Recall period	Total no. of items	Response options	Score range	Original language
Living with HIV Scale [76]; 1998	HRQoL	—	32	5-point Likert scale (0-4)	Sum of all item scores: 0-128; subscale scores range: 0-24	English
WHOQOL-HIV[h] [77-83]; 2004	HRQoL	The last 2 weeks	120 (30 facets)	5-point Likert scale (1-5)	Facet scores range: 4-20	English
WHOQOL-HIV-BREF[i] [84-95]; 2012	HRQoL	The last 2 weeks	31	5-point Likert scale (1-5)	Facet scores range: 4-20	English
Instituto Superiore di Sanità Quality of Life [96]; 2006	HRQoL	Past 4 weeks	62	5-point Likert scales (1-5)	All subscales are coded to range from 0 to 100	Italian
Symptom Quality of Life Adherence [97]; 2009	HRQoL	Past 4 weeks	26	HRQoL: 5-point Likert scales (1-5), symptoms: yes or no, and adherence: 10 cm VAS[j]	HRQoL: standardized sum (0-100), symptoms: summed, and adherence: score 0-10 VAS	French
PROQOL-HIV[k]-43 [98-100]; 2012	HRQoL	Past 2 weeks	43	5-point Likert scale (0-4)	Sum of the 8 subscales and coded as a total score range from 0 to 100	English
PROQOL-HIV-38 [101]; 2016	HRQoL	Past 2 weeks	38	5-point Likert scale (0-4)	Four subscale scores are summed of item responses, coded to range from 0 to 100	French
Poz Quality of Life [102]; 2018	HRQoL	—	13	5-point Likert scale (1-5)	Items were averaged to create the total score and scores for each subscale	English
Symptoms						
Riverside Symptom Checklist [103]; 1993	HIV-related symptoms	Past 3 months	28	5-point Likert scale (0-4)	The subscales are scored as summated and transformed on a scale from 0 to 100	English
HIV Symptom Index [104]; 1994	HIV-related symptoms	Past 2 weeks	12	4-point Likert scales (0-3)	Scores range: 0-24	English
HIV Assessment Tool [105]; 1994	HIV-related symptoms	—	34 in each exploratory factor analysis	100-mm linear scale	Items were averaged to create the total score (0-100)	English
SSC-HIV[l] [106,107]; 1999	HIV-related symptoms	—	26	4-point Likert scales (0-3)	The items within a factor are summed for a subscale score	English
SSC-HIV-rev [108]; 2001	HIV-related symptoms	—	72	4-point Likert scales (0-3)	The items within a factor are summed for a subscale score	English
HIV Cost and Services Utilization Study Symptom Measure [109]; 2000	HIV-related symptoms	Preceding 6 months	13 for male and 14 for female respondents	5-point Likert scale (1-5)	The subscales are scored as summated and transformed on a scale of 0 to 100	English
HIV Symptom Index or Symptoms Distress Module of the ACTG[m] [110-112]; 2001	HIV-related symptoms	Past 4 weeks	20	5-point Likert scale (0-4)	Score range: 0-80	English
HIV-Related Fatigue Scale [113-115]; 2002	HIV-related fatigue	Past week	56	Multiple response options	All subscales are coded to range from 1 to 10	English
HIV Disability Questionnaire [116-119]; 2013	HIV-related disability	Past week	69	Disability presence scores: yes or no; disability severity scores: 5-point Likert scale (0-4); episodic scores: yes or no; 7-point Likert scale (0-6) [116]	Each method of calculating scores was to sum the scores and transform them into scores out of 100	English

PROM; year of development	Targeted concept	Recall period	Total no. of items	Response options	Score range	Original language
Istituto Superiore di Sanità-HIV symptoms scale [120]; 2016	HIV-related symptoms	Past 4 weeks	22	5-point Likert scale (no score for each option)	—	Italian
Stigma						
HSS^n-40 [121,122]; 2001	HIV-related stigma	—	40	4-point Likert scale (1-4)	Score range: 40-160	English
HSS-32 [123,124]; 2007	HIV-related stigma	—	32	4-point Likert scale (1-4)	Score range: 32-128	English
HSS-12 [125-127]; 2010	HIV-related stigma	—	12	4-point Likert scale (1-4)	Score range: 12-48	Swedish
HSS-39 [128]; 2014	HIV-related stigma	—	39	4-point Likert scale (1-4)	Score range: 39-156	Swedish
HSS-30 [129]; 2015	HIV-related stigma	—	30	4-point Likert scale (1-4)	Score range: 30-120	Spanish
HSS-10 [130]; 2020	HIV-related stigma	—	10	5-point Likert scale (1-5)	Score range: 10-50	Japanese
HIV or AIDS stigma instrument–People living with AIDS [131,132]; 2007	HIV-related stigma	Two recall periods: past 3 months and ever since HIV diagnosis	33	4-point Likert scale (0-3)	Score range: 0-3	English
Internalized HIV Stigma Measure [133]; 2008	Internalized HIV-related stigma	—	28	Transformed linearly to a range of 0 to 100	The subscales are scored as summated and transformed on a scale of 0 to 100	English
Internalized AIDS-Related Stigma Scale [134-136]; 2009	Internalized HIV-related stigma	—	6	Binary response: 1=agree and 0=disagree	Total scores range of endorsed stigma items: 0-6	English
Internalized Stigma in Those With HIV or AIDS [137]; 2011	Internalized HIV-related stigma	Ever since HIV diagnosis	10	5-point Likert scale (1-5)	Score range: 10-50	English
HIV- and Abuse-Related Shame Inventory [138]; 2012	HIV- and abuse-related shame	Past month	31	5-point Likert scale (0-4)	Score range: 0-124	English
Self, Experienced, and Perceived HIV or AIDS Stigma Scales [139]; 2012	HIV-related stigma	—	22	4-point Likert scale (1-4)	Score range: 22-88	English
HIV Stigma Mechanisms [140]; 2013	HIV stigma mechanisms	—	24	5-point Likert scales (1-5)	Items were averaged to create composite scores	English
HIV or AIDS Stigma Assessment for Latino Gay Men, Bisexual Men, and Transgender Women Living With HIV [141]; 2013	HIV-related stigma	—	36	4-point Likert scale (1-4)	Score range: 36-144	English
Van Rie HIV or AIDS-Related Stigma Scale-Revised for use in the United States [142]; 2015	HIV-related stigma	—	15	4-point Likert scale (0-3)	Items were averaged to create composite scores and subscales scores	English
Psychological						
The Mental Adjustment to HIV scale [143]; 1994	Mental adjustment	—	40	4-point Likert scale (1-4)	The subscales are scored as summated	English
HIV or AIDS Stress Scale [144]; 2002	Stress and coping	Past month	23	5-point Likert scale (0-4)	Score range: 0-92	English

PROM; year of development	Targeted concept	Recall period	Total no. of items	Response options	Score range	Original language
Perceived Stress Scale Among People Living With HIV or AIDS [145]; 2008	HIV-related stress	Past month	35	5-point Likert scale (1-5)	Score range: 35-175	Simplified Chinese
Screenphiv [146,147]; 2012	Psychological issues related to HIV	—	63	VAS (0-100 mm)	Items were averaged to create composite scores	Spanish
Impact on Self-Concept Scale [148]; 2013	Impact of HIV on self-concept	—	10	6-point Likert scale (1-6)	Items were averaged to create composite scores	English
Impact of HIV [149]; 2015	Challenges of HIV survivorship	—	38	5-point Likert scale (1-5)	Score range: 38-190	English
HIV Meaningfulness Scale [150]; 2015	HIV meaningfulness	—	4	7-point Likert scale	Score range: 1-28	English
People Living with HIV Resilience Scale [151]; 2019	Resilience	Past 12 months	10	Positively affected: "+1," not affected: "0," and negatively affected: "–1"	Score range: (–10 to 10)	English
Body and facial appearance						
Body Image in Patients With HIV or AIDS [152]; 2005	Perceived body image	—	12	5-point VAS	Score range: 12-60	English
Owen Clinic Lipodystrophy Scale [153]; 2006	Body change	—	12	Dichotomous: (yes or no)	—	English
ACTG-ABCD[o] [154]; 2006	Body change and distress	Part 3: past 4 weeks	27	Part 1: dichotomous: (yes or no); part 2 and part 3: 5-point Likert scale (1-5)	Sum of all item scores in part 3 (20 items)	English
ACTG-ABCD Short form [155]; 2014	Body change and distress	Past 4 weeks	18	5-point Likert scale (1-5)	Sum of all item scores	English
Facial Appearance Inventory [156]; 2016	Appearance	Past 4 weeks	10	7-point Likert scale (1-7)	Score range: 24-168; final score is linearly transformed to 0-100	English
Treatment						
Medication Attribution Scale [157]; 1998	Attributions about ART[p] (its limitations on functioning, etc)	—	10	11-point Likert scale (0-10)	Sum of all item scores	English
HIVTSQ[q] [158]; 2001	Satisfaction with ART	Past 4 weeks	9	7-point Likert scale (0-6)	Total treatment satisfaction is the sum of the 9 item scores	English
HIV Treatment Satisfaction Questionnaire status version [159]; 2006	Satisfaction with ART	Past few weeks	10	7-point Likert scale (0-6)	Total treatment satisfaction is the sum of the 10 item scores	English
Treatment-Related Empowerment Scale [160]; 2001	Empowerment (involvement in treatment decision-making)	—	10	5-point Likert scale (1-5)	Sum of all item scores	English
Subcutaneous Injection Survey [161]; 2002	Satisfaction with ART–subcutaneous injection	—	15	5-point Likert scale (1-5)	Score range: 20-100	English
Quality of care through the patient's eyes [162]; 2003	Quality of care	—	27	Importance and performance were measured using a 4-point Likert scale (1-4)	$[Qij = Iij \times Pij]^r$	English

PROM; year of development	Targeted concept	Recall period	Total no. of items	Response options	Score range	Original language
Attitudes Toward HIV Health Care Provider scale [163]; 2004	Attitudes toward health care providers	—	19	6-point Likert scale (1-6)	Sum of all item scores	English
Antiretroviral General Adherence Scale [164]; 2006	Ease and ability to adhere to ART	Past 30 days	5	6-point Likert scale (1-6)	Sum of all item scores or a proportion (by dividing this score by the total possible score)	English
Health Care Relationship Trust Scale [165]; 2006	Trust toward health care providers	—	15	5-point Likert scale (0-4)	Sum of item scores and the mean of item scores	English
HIV Medication Readiness Scale [166]; 2007	Readiness to adhere to ART	—	10	5-point Likert scale (0-4)	Score range: 0-40	English and French
SECope [167]; 2007	Coping with the side effects of ART	—	20	5-point Likert scale (0-4)	Score range: 0-80	English
HIV Treatment Optimism Scale [168]; 2009	Optimism about ART	—	19	7-point Likert scale (1-7)	Score range: 19-133	English
HIV Medication Taking Self-Efficacy Scale [169]; 2010	Self-efficacy to adhere to ART	—	26	11-point Likert scale (0-10)	Sum of all item scores	English
Brief Estimate of Health Knowledge and Action-HIV version [170]; 2010	ART-related health literacy	—	8	Part I: 4-point Likert scale (0-3); Part II: 6-point Likert scale (0-5)	Sum of all item scores	English
HIV Treatment Readiness Measure [171]; 2011	Factors affecting the readiness for ART	Alcohol and drug use subscale in the past 3 months	38	5-point Likert scale (1-5)	Sum of all item scores and the mean of all item scores	English
HIV Treatment Regimen Fatigue Scale [172]; 2015	Regimen fatigue	—	22	–3 to 3 (excluding 0)	Sum of all item scores	English
HIV Engagement in and Continuity of Care Scale [173]; 2017	Engagement in care	—	26	5-point Likert scale	—	English
Social support						
Social Support Inventory [174]; 1999	Received social support	—	14/17	Satisfaction: 5-point Likert scale (1-5); want: yes or no; have: yes, no, or not applicable	Nine subscales: 0-5	English
Unsupportive Social Interactions Inventory-HIV version [175]; 1999	Unsupportive social interactions	—	24	4-point Likert scale (0-4)	An overall score, the Unsupportive Social Interactions Inventory-18, is based on 3 of its subscales	English
Perceived Social Support for HIV [176]; 2014	Perceived social support	—	12	5-point Likert scale (1-5)	Sum of all item scores; Score range: 12-60	Spanish
Self-management and self-care						
HIV Treatment Adherence Self-Efficacy Scale [177]; 2007	Self-efficacy to adhere to HIV care	Past 1 month	12	11-point Likert scale (0-10)	Item scores were averaged for each respondent	English
Perceived HIV Self-Management Scale [178]; 2011	Self-efficacy for HIV self-management	—	8	6-point Likert scale (1-6)	Sum of all item scores	English
HIV Self-Management Scale (Women) [179]; 2012	HIV Self-Management Scale (Women)	—	20	4-point Likert scale (0-3)	Subscale score range: 0-3	English

PROM; year of development	Targeted concept	Recall period	Total no. of items	Response options	Score range	Original language
HIV Intention Measure [180]; 2012	Intention to adhere to HIV care	—	14	6-point Likert scale (1-6)	—	English
HIV Exercise Stereotypes Scale [181]; 2016	Stereotypes related to exercise in people living with HIV	—	14	6-point Likert scale (1-6)	Three subscale scores are computed as the mean of item responses	French
HIV Symptom Management Self-Efficacy for Women Scale [182]; 2011	Self-efficacy for HIV symptom management	—	9	11-point Likert scale (0-10)	The final score is calculated as the mean of the 9 item scores	English

[a]Each version of a PROM is considered a separate PROM.

[b]HRQoL: health-related quality of life.

[c]MOS-HIV: Medical Outcomes Study-HIV Health Survey.

[d]—: not reported.

[e]HIV-PARSE: HIV Patient–Reported Status and Experience.

[f]HAT-QoL: HIV or AIDS-Targeted Quality of Life Instrument.

[g]MQoL: Multidimensional Quality of Life.

[h]WHOQOL-HIV: World Health Organization Quality of Life-HIV.

[i]WHOQOL-HIV-BREF: World Health Organization Quality of Life-HIV-Bref instrument.

[j]VAS: visual analog scale.

[k]PROQOL-HIV: Patient-Reported Outcome Quality of Life-HIV Questionnaire.

[l]SSC-HIV: Sign and Symptom Checklist for HIV.

[m]ACTG: Adult AIDS Clinical Trial Group.

[n]HSS: HIV Stigma Scale.

[o]ACTG-ABCD: Adult AIDS Clinical Trial Group's Assessment of Body Change and Distress.

[p]ART: antiretroviral therapy.

[q]HIVTSQ: HIV Treatment Satisfaction Questionnaire.

[r]The quality improvement score (Q) on a health service (j) by an individual patient (i) is equal to the importance score (I) multiplied by the (perceived) performance score (P).

Characteristics of the Included Records

As 3 studies [34,35,42] included the assessment of 2 PROMs, 155 records were included. Table S4 in Multimedia Appendix 1 shows the characteristics of the 155 included records. Of these 155 records, 31 (20%) records were reported before 2000, 46 (29.7%) records were reported between 2000 and 2009, and 78 (50.3%) records were reported after 2010. The total sample size of these records was 79,213 (range 20-5521). There were more men than women in 83.2% (129/155) of the records, and 1.3% (2/155) of records did not indicate gender data. Most records gave the mean (SD) or median (IQR) age data for samples (range 16-84 years), and 8.4% (13/155) of records indicated no age data. There were 70.3% (109/155) records from high-income countries (64/155, 41.3% records from the United States), 20.6% (32/155) records from low- and medium-income countries (9/155, 5.8% records from China), and 9% (14/155) of records from multiple countries. Table S4 in Multimedia Appendix 1 also summarizes the years since diagnosis, the severity of the disease, recruitment context, and effective response rate.

Methodological Quality Assessment

The methodological quality for each psychometric property of every record is summarized in Table S5 in Multimedia Appendix 1 based on the COSMIN risk of bias checklist. As there is no generally accepted "golden standard" for assessing health outcomes in adults living with HIV and AIDS, the criterion validity of all studies was not considered. Most records assessed internal consistency (146/155, 94.2% of records) and structural validity (96/155, 61.9% of records), and most of them were rated as "very good" or "adequate." Although 79.4% (123/155) of records assessed the hypotheses testing for construct validity, most were rated as "doubtful" or "inadequate." As for the remaining psychometric properties, only a few records assessed them, and most of them were rated as "doubtful" or "inadequate."

Overall Results and the Level of Evidence

Table S6 in Multimedia Appendix 1 shows the results of each psychometric property of each record. The overall results and the level of evidence are presented in Table S7 in Multimedia Appendix 1. There are only few studies on PROMs, except for some well-known PROMs; accordingly, there is little evidence for psychometric properties.

Of the 88 PROMs, PROM development was assessed in 18% (16/88) PROMs, and original content validity was assessed in 3% (3/88) PROMs. However, no PROM exhibited "sufficient" high-quality evidence for content validity. Subsequently, we found that 16% (14/88) of the PROMs had "sufficient" high-quality evidence of structural validity; however, most

others had "indeterminate" moderate-quality evidence. The internal consistency for a PROM can be assessed only if it has at least low-quality evidence for "sufficient" structural validity; otherwise, the internal consistency will be considered "indeterminate" [30]. Therefore, although 83% (73/88) of PROMs presented high-quality evidence for internal consistency, only 16% (14/88) demonstrated "sufficient" results. Evidence supporting hypotheses testing for construct validity was available for 81% (71/88) of the PROMs. Furthermore, reliability was assessed in 30% (26/88) PROMs, but no PROM presented "sufficient" high-quality evidence. The responsiveness of 8% (7/88) of PROMs was evaluated as "sufficient," but only 2% (2/88) PROMs (Functional Assessment of HIV Infection [61-66] and HIV Medication Readiness Scale [166]) showed high-quality evidence. Cross-cultural validity or measurement invariance was assessed in only 6% (5/88) of PROMs with low or very low quality [82,111,122,124,127]. Finally, only 1% (1/88) of PROMs assessed measurement error with "indeterminate" low-quality evidence [118].

Recommendations

The following recommendations are presented according to the COSMIN guidelines (Table 2):

- Class A: The PROMs with evidence for "sufficient" content validity (any level) and at least low-quality evidence for "sufficient" internal consistency included the following: Poz Quality of Life (PozQoL) [102], HIV Symptom Index or Symptoms Distress Module of the Adult AIDS Clinical Trial Group (HIV-SI or SDM) [110-112], and People Living with HIV Resilience Scale (PLHIV-RS) [151]. These may be recommended for use, and the results obtained may be credible.
- Class B: The remaining PROMs have the potential to be recommended for use; however, further research is required to assess their quality (PROMs not included in class A or C).
- Class C: The PROMs with high-quality evidence for an "insufficient" psychometric property included the following: Multidimensional Quality of Life for patients With HIV and AIDS [72-75], Patient-Reported Outcome Quality of Life-HIV Questionnaire-38 [101], HIV-Related Fatigue Scale [113-115], HIV Stigma Scale-10 [130], HIV or AIDS Stress Scale [144], Screenphiv [146,147], SECope [167], and HIV Exercise Stereotypes Scale [181]. They may not be recommended for use.

Although 3 PROMs have been recommended, they all have some shortcomings, reducing the strength of the recommendation for their routine use. Furthermore, although PozQoL [102] and PLHIV-RS [151] achieved class A, they were developed and assessed based on a single validation study. In addition, some items in HIV-SI or SDM have significant differential item functioning between different cultural groups [111], indicating low-quality evidence for "insufficient" cross-cultural validity.

Table 2. Summary of findings[a].

PROM[b]	Content validity		Structural validity		Internal consistency[c]		CCV or MI[d]		Reliability		Measurement error		HTCV[e,f]		Responsiveness		Class[g]
	Results	LoE[h]	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	
MOS-HIV[i] [31-51]			+	M	±	M			±	M			+	H			B
MOS-HIV-17 [53]					?	H							–	L			B
MOS-HIV-29 [52]			?	M	?	M							+	L	+	L	B
HOPES[j] [54,55]					?	H							+	VL	+	VL	B
HIV-QoL[k] [56]					?	H											B
AIDS-HAQ[l] [57]					?	H							+	H	+	L	B
HIV-PARSE[m] [58]					?	H											B
HIV-PARSE-Brief [59]					?	H											B
HRQOL[n] [60]			–	VL	?	H							+	L			B
FAHI[o] [61-65]	±	M	–	M	?	H							+	H	+	H	B
GHSA[p] [66]			?	M	?	H							+	L			B
HIV-QL31[q] [67]	?	VL	–	M	?	H							–	L			B
HAT-QoL[r]-42 [68,69]					?	H							–	M			B
HAT-QoL-30 [35]					?	H							+	L			B
HAT-QoL-34 [42,70,71]			?	M	?	H			–	M			+	H			B
MQoL-HIV[s] [34,72-75]	?	M	?	M	?	H			–	H			+	H	+	M	C
LWHIVS[t] [76]	±	VL	?	M	?	H							–	L			B
WHOQOL-HIV[u] [77-83]			+	M	?	L	?	VL	+	M			+	H			B
WHOQOL-HIV-BREF [84-95]			±	L	?	H			?	M			+	L			B
ISSQoL[v] [96]	+	L			?	H							+	L			B
HIV-SQUAD[w] [97]			?	M	?	H							+	L			B
PROQOL[x]-HIV-43 [98-100]	+	VL	?	M	?	H			+	L			+	L			B
PROQOL-HIV-38 [101]			–	H	?	H							+	VL			C
PozQol[y] [102]	+	L	+	H	+	H			+	M			+	M			A
RSC[z] [103]					?	H							+	H			B
HSI[aa] [104]					?	H			+	L			–	L	+	VL	B
HAT[ab] [105]			?	VL	?	VL			+	L							B
SSC-HIV[ac] [106,107]			+	M	+	H											B
SSC-HIV-rev [108]			?	M	?	H							–	L			B

PROM[b]	Content validity		Structural validity		Internal consistency[c]		CCV or MI[d]		Reliability		Measurement error		HTCV[e,f]		Responsiveness		Class[g]
	Results	LoE[h]	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	
HCSUS-SM[ad] [109]					?	H											B
HIV-SI or SDM[ae] [110-112]	+	L	+	H	+	H	–	L					+	L			A
HRFS[af] [113-115]	?	VL			?	H			–	H			+	M			C
HDQ[ag] [116-119]			+	M	+	H			+	M	?	L	±	M			B
ISS-HIV-SS[ah] [120]			?	M	?	H							+	H			B
HSS[ai]-40 [121,122]			?	M	?	H	–	L	+	M			+	H			B
HSS-32 [123,124]			±	M	?	H	–	VL					+	L			B
HSS-12 [125-127]			+	H	+	H	–	L					–	M			B
HSS-39 [128]			?	VL	?	H							+	VL			B
HSS-30 [129]	?	M	+	H	+	H							+	H			B
HSS-10 [130]			+	H	–	H							+	L			C
HASI[aj]-P [131,132]			?	H	?	H							+	H			B
IHSM[ak] [133]			?	M	?	H							–	L			B
IA-RSS[al] [134-136]			+	H	+	H			–	M			±	M			B
ISAT[am] [137]			?	M	?	H							+	H			B
HARSI[an] [138]			?	M	?	H			–	VL							B
SEP-HASS[ao] [139]					?	H							+	VL			B
HIV-SM[ap] [140]					?	M							–	VL			B
HA-SAL-GBT[aq] [141]			?	VL	?	H							–	L			B
VR-HARSSR[ar] [142]			?	L	?	M							+	M			B
MAH[as] [143]			?	VL	?	H											B
SS-HIV[at] [144]			–	H	?	H			+	M			+	L			C
PSSHIV[au] [145]	±	VL	?	M	?	H			+	L			+	L			B
Screenphiv [146,147]	?	L	+	H	–	H							+	H			C
ISCS[av] [148]			?	M	?	H							+	L			B
IHIV[aw] [149]			+	H	+	H											B
HIVMS[ax] [150]			?	M	?	H			+	L			+	VL			B
PLHIV-RS[ay] [151]	+	L	+	H	+	H							+	L			A
BIS[az] [152]			?	M	?	L			+	VL							B
OCLS[ba] [153]			?	M	?	VL											B

PROM[b]	Content validity		Structural validity		Internal consistency[c]		CCV or MI[d]		Reliability		Measurement error		HTCV[e,f]		Responsiveness		Class[g]
	Results	LoE[h]	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	
ACTG-ABCD[bb] [154]					?	H							–	L			B
ACTG-ABCD-SF[bc] [155]			?	M	?	H							+	L			B
FAI[bd] [156]	?	L			?	VL							+	VL			B
MAS[be] [157]					?	M							–	M			B
HIVTSQ[bf] [158]			?	M	?	H							–	L			B
HIVTSQ status version [159]			+	H	+	H							–	L			B
TES[bg] [160]					?	L							–	VL			B
SIS[bh] [161]			?	M	?	H							–	L			B
QUOTE-HIV[bi] [162]					?	L											B
AHHCP[bj] [163]			?	M	?	H							+	H			B
AGAS[bk] [164]			?	M	?	H							+	L			B
HCR[bl] [165]	±	VL	?	L	?	M			–	VL			–	VL			B
HMRS[bm] [166]			?	M	?	H			+	VL			+	L	+	H	B
SECope [167]	±	VL	+	H	–	H			–	M			–	L			C
HTOS[bn] [168]			?	M	?	H											B
HIV-MT-SES[bo] [169]			+	H	+	H			–	L			+	L			B
BEHKA-HIV[bp] [170]			?	M	?	H											B
HTRM[bq] [171]			?	M	?	H			–	M							B
HTRFS[br] [172]			?	L	?	M							+	VL			B
HECCS[bs] [173]			+	M	+	H							+	H			B
SSI[bt] [174]	±	VL	?	M	?	H							+	VL			B
USII-HIV[bu] [175]			?	M	?	H							–	L			B
PSS-HIV[bv] [176]					?	L							+	VL			B
HIV-ASES[bw] [177]			+	H	+	H			?	L			+	L			B
PHIVSMS[bx] [178]					?	H							+	L			B
HIV-SMS-W[by] [179]	+	VL	+	H	+	H			?	VL							B
HIV-IM[bz] [180]	±	L	?	L	?	H							+	L			B
HIVESS[ca] [181]			–	H	?	H							+	VL			C

PROM[b]	Content validity		Structural validity		Internal consistency[c]		CCV or MI[d]		Reliability		Measurement error		HTCV[e,f]		Responsiveness		Class[g]
	Results	LoE[h]	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	Results	LoE	
HSM-SEWS[cb] [182]			?	L	?	M			?	VL							B

[a]As there is no generally accepted "golden standard" for assessing health outcomes in adults living with HIV and AIDS, the criterion validity of all studies was not considered. Overall results of PROMs are rated as +: sufficient; ?: indeterminate; ±: inconsistent; and –: insufficient. LoE is rated as H: high, M: moderate, L: low; VL: very low. Blank cells indicate that the data are not available.

[b]PROM: patient-reported outcome measure.

[c]Internal consistency can be rated as "sufficient" if there is at least low evidence for "sufficient" structural validity, and Cronbach α values≥.70 for each unidimensional scale or subscale; the evidence for "sufficient" structural validity may come from different studies, and the "at least low evidence" was defined by grading the evidence according to the Grading of Recommendations, Assessment, Development, and Evaluation approach.

[d]CCV or MI: cross-cultural validity or measurement invariance.

[e]HTCV: hypotheses testing for construct validity.

[f]The results of all included records should be taken together, and it should then be decided if 75% of the results are in accordance with the hypotheses. Only assessed measurement properties are shown.

[g]Class A represents evidence for sufficient content validity (any level) and at least low-quality evidence for sufficient internal consistency (PROMs can be recommended for use); class B, PROMs categorized not in class A or C; and class C, high-quality evidence for an insufficient measurement property; PROMs with class B recommendation require further evaluation to assess their quality before recommendation for use; PROMs with class C recommendation are not recommended for use.

[h]LoE: level of evidence (using the Grading of Recommendations, Assessment, Development, and Evaluations assessment tool).

[i]MOS-HIV: Medical Outcomes Study-HIV Health Survey.

[j]HOPES: HIV Overview of Problems Evaluation System.

[k]HIV-QoL: HIV-Related Quality of Life Questions.

[l]AIDS-HAQ: AIDS Health Assessment Questionnaire.

[m]HIV-PARSE: HIV Patient–Reported Status and Experience.

[n]HRQoL: health-related quality of life.

[o]FAHI: Functional Assessment of HIV Infection.

[p]GHSA: General Health Self-Assessment.

[q]HIV-QL31: HIV Quality of Life 31-item scale.

[r]HAT-QoL: HIV or AIDS-Targeted QoL Instrument.

[s]MQoL-HIV: Multidimensional QoL for patients with HIV or AIDS.

[t]LWHIVS: Living with HIV Scale.

[u]WHOQOL-HIV: World Health Organization Quality of Life-HIV.

[v]ISSQoL: Instituto Superiore di Sanità Quality of Life.

[w]HIV-SQUAD: Symptom Quality of Life Adherence.

[x]PROQOL-HIV: Patient-Reported Outcome Quality of Life-HIV Questionnaire.

[y]PozQol: Poz Quality of Life.

[z]RSC: Riverside Symptom Checklist.

[aa]HSI: HIV Symptom Index.

[ab]HAT: HIV Assessment Tool.

[ac]SSC-HIV: Sign and Symptom Checklist for HIV.

[ad]HCSUS-SM: HIV Cost and Services Utilization Study Symptom Measure.

[ae]HIV-SI or SDM: HIV Symptom Index or Symptoms Distress Module of the Adult AIDS Clinical Trial Group.

[af]HRFS: HIV-Related Fatigue Scale.

[ag]HDQ: HIV Disability Questionnaire.

[ah]ISS-HIV-SS: Istituto Superiore di Sanità-HIV symptoms scale.

[ai]HSS-40: HIV Stigma Scale.

[aj]HASI-P: HIV or AIDS Stigma Instrument-PLWA.

[ak]IHSM: Internalized HIV Stigma Measure.

[al]IA-RSS: Internalized AIDS-Related Stigma Scale.

[am]ISAT: Internalized Stigma in Those With HIV or AIDS.

[an]HARSI: HIV- and Abuse-Related Shame Inventory.

[ao]SEP-HASS: Self, Experienced, and Perceived HIV or AIDS Stigma Scales.

[ap]HIV-SM: HIV stigma mechanisms.
[aq]HA-SAL-GBT: HIV or AIDS Stigma Assessment for Latino Gay Men, Bisexual Men and Transgender Women Living With HIV.
[ar]VR-HARSSR: Van Rie HIV or AIDS-Related Stigma Scale-Revised for use in the United States.
[as]MAH: Mental Adjustment to HIV scale.
[at]SS-HIV: HIV or AIDS Stress Scale.
[au]PSSHIV: Perceived Stress Scale Among People Living With HIV and AIDS.
[av]ISCS: Impact on Self-Concept Scale.
[aw]IHIV: Impact of HIV.
[ax]HIVMS: HIV Meaningfulness Scale.
[ay]PLHIV-RS: People Living with HIV Resilience Scale.
[az]BIS: Body Image in Patients With HIV or AIDS.
[ba]OCLS: Owen Clinic Lipodystrophy Scale.
[bb]ACTG-ABCD: Adult AIDS Clinical Trial Group's Assessment of Body Change and Distress.
[bc]ACTG-ABCD-SF: ACTG-ABCD Short Form.
[bd]FAI: Facial Appearance Inventory.
[be]MAS: Medication Attribution Scale.
[bf]HIVTSQ: HIV Treatment Satisfaction Questionnaire.
[bg]TES: Treatment-Related Empowerment Scale.
[bh]SIS: Subcutaneous Injection Survey.
[bi]QUOTE-HIV: quality of care through the patient's eyes.
[bj]AHHCP: Attitudes Toward HIV Health Care Provider scale.
[bk]AGAS: Antiretroviral General Adherence Scale.
[bl]HCR: Health Care Relationship Trust Scale.
[bm]HMRS: HIV Medication Readiness Scale.
[bn]HTOS: HIV Treatment Optimism Scale.
[bo]HIV-MT-SES: HIV Medication Taking Self-Efficacy Scale.
[bp]BEHKA-HIV: Brief Estimate of Health Knowledge and Action-HIV version.
[bq]HTRM: HIV Treatment Readiness Measure.
[br]HTRFS: HIV Treatment Regimen Fatigue Scale.
[bs]HECCS: HIV Engagement in and Continuity of Care Scale.
[bt]SSI: Social Support Inventory.
[bu]USII-HIV: Unsupportive Social Interactions Inventory-HIV version.
[bv]PSS-HIV: Perceived Social Support for HIV.
[bw]HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale.
[bx]PHIVSMS: Perceived HIV Self-Management Scale.
[by]HIV-SMS-W: HIV Self-Management Scale (Women).
[bz]HIV-IM: HIV Intention Measure.
[ca]HIVESS: HIV Exercise Stereotypes Scale.
[cb]HSM-SEWS: HIV Symptom Management Self-Efficacy for Women Scale.

Discussion

Principal Findings

From the 152 included studies, we identified 88 PROMs in 8 categories for adults living with HIV, and the psychometric properties of the majority of the included PROMs were rated with insufficient evidence. The principal finding of this review was the lack of comprehensively validated HIV-specific PROMs for the assessment of health outcomes in adults living with HIV and AIDS. Although 3 available PROMs (PozQoL, HIV-SI or SDM, and PLHIV-RS) have been recommended based on the COSMIN guidelines, they all have some shortcomings. In addition, because of limited evidence, recommendations regarding the use of most of the remaining assessed PROMs (class B recommendation) cannot be made. These findings emphasize on the need for a more comprehensive validation of the psychometric properties of the existing PROMs. Furthermore, our findings indicate the need for a robust and rapid validation of PROMs through the use of electronic PROMs (ePROMs) and modern measurement theories (such as Item Response Theory).

Taxonomy of HIV-Specific PROMs

This systematic review updated the review reported by Engler [22] and provided improvisations on the inclusion and exclusion criteria, such that many unvalidated PROMs were excluded because if we include these PROMs, we cannot summarize the overall status of their psychometric properties. In addition, using the 12 categories reported by inductive content analysis in the review of Engler [22] as reference, this review reported 8 integrated categories (Table 1). The 2 categories of "ART and adherence-related views and experiences" and

"healthcare-related views and experiences" in the study by Engler et al [22] were integrated into "treatment," and "psychological challenges" and "psychological resources" were integrated into the category "psychological"; the PROMs in the "sexual and reproductive health" category were excluded because they did not meet the inclusion criteria for our study. Finally, the "Disability" category was integrated with "Symptoms." The new taxonomy proposed in this review should be helpful for health care providers and researchers in selecting PROM.

In addition, although some of the PROMs included cognitive function or symptoms to some extent (such as "cognitive functioning" of Medical Outcomes Study-HIV Health Survey and "cognitive symptoms" of HIV Disability Questionnaire), no PROM specifically designed to measure cognitive concerns was included in the analysis. However, considering the high prevalence of HIV-associated neurocognitive disorders and HIV-associated dementia in people living with HIV and AIDS, it is important to assess their cognition via PROMs [184]. Askari [185,186] conducted a series of studies to progressively simplify the item pool and developed a PROM (the Communicating Cognitive Concerns Questionnaire) aimed at assessing the cognitive abilities of people living with HIV and AIDS. The main cognitive dimensions measured by this PROM included memory, concentration, executive function, language, emotions, and motivation. Although the Communicating Cognitive Concerns Questionnaire did not correlate strongly with cognitive test performance in people living with HIV and AIDS, it reflected the real-life concerns of people living with HIV and AIDS in terms of their mood, work, and work productivity. Although the related PROMs were not included in this review, we will further explore these cognitive concerns as an independent PROM category in future studies.

Psychometric Properties

Overview

A thorough validation process is important for ensuring the applicability of a PROM to individual patient care [187]. However, in this review, most included PROMs were short of evidence for many psychometric properties, such as content validity, measurement error, cross-cultural validity or measurement invariance, and responsiveness. Therefore, it was difficult to assess the quality of these PROMs.

Content Validity

On the basis of the most up-to-date COSMIN methodology [26], content validity is the most important psychometric property, and the current guidance suggests that it is very important for patients to participate in development and validation studies [25]. As suggested by Selby and Velikova [188], and public involvement should appear as a core feature in PROM design and application. In addition, Wilson [189] believed that the perception of patients was essential for providing better insights into how a disease affects HRQoL. However, they were short of evidence in terms of patient and public involvement in the development process of the included PROMs. To determine whether a PROM was well designed, it should be confirmed that the PROM is relevant, comprehensible, and comprehensive from a patient perspective and for their context of use [190]. In addition, PROMs should be able to record the experience of people living with HIV and AIDS and how HIV affects their lives so as to make a study more relevant and have better content validity [191].

Internal Structure

Internal consistency was the most frequently reported psychometric property. However, many studies used internal consistency as the only indicator of reliability, which was definitely not enough. Besides, structural validity is also one of the most important psychometric properties [192]. The premise for assessing internal consistency is at least "low" evidence for "sufficient" structural validity, and this evidence may come from different studies [27]. However, only exploratory factor analysis was conducted in many studies for the assessment of structural validity instead of confirmatory factor analysis. Accordingly, this property can only obtain the rating of "indeterminate," further affecting the assessment of internal consistency. In addition, the assessment of structural validity in most studies included in this review was based on classical test theory. Only 2 studies used Rasch analysis to assess the extent of interval level measurement and implementation of unidimensionality in this review [62,67]. However, no guidance has been provided in the COSMIN guidelines with regard to relying on only Rasch analysis without classical test theory statistics to assess the structural validity of PROMs. Therefore, Recchioni [193] suggested that it is necessary to provide additional guidance for the study that only uses Rasch analysis, especially in the development of new PROMs.

A PROM developed in one particular context may not be suitable for another. Therefore, it is necessary to use the same PROM for direct comparisons between different populations. No positive results for cross-cultural validity or measurement invariance were reported in this review [82,111,122,124,127], showing that the validity and transferability of the included PROMs between different geographies, cultural contexts, and risk populations were still unclear. Many researchers directly use the existing PROMs through simple translations and ignore cross-cultural adaptation [194]. However, there are great differences in the understanding of some concepts among people of different cultures, global regions, genders, ages, and socioeconomic strata [195]. The use of PROMs in different contexts is not simply dependent on translating items but should be processed based on a 7-key-step process for comprehensive cross-cultural adaptation [196].

Remaining Psychometric Properties

Measurement error was also important for interpreting PROs. Minimal important change is best calculated from multiple studies and using multiple anchors with an anchor-based longitudinal approach [197]. In this review, only 1 study reported the smallest detectable change ranging from 7.3 to 15.0 points without minimal important change. Therefore, measurement error was assessed as "indeterminate" [118]. Moreover, only few studies assessed responsiveness. However, responsiveness was vital to assess the effectiveness of a clinical intervention designed to improve the health outcomes of people living with HIV and AIDS. This identifies several gaps for

future research in the area of HIV. Without such information, it is impossible to understand whether changes in the levels of health outcomes of people living with HIV and AIDS are meaningful and matter to health care providers and researchers.

Clinical Implications

Despite a 64% reduction in HIV-related deaths in 2020 compared with the peak reported in 2004, a total of 680,000 people living with HIV and AIDS still died from HIV-related illnesses in 2020. This was largely due to the unique physical and psychosocial symptoms [1]. These symptoms seriously affect the physical function and clinical outcomes of people living with HIV and AIDS [4,198-200]. PRO data can be used in a variety of ways to improve care and health outcomes at a patient, institution, and population level [201-204]. Considering the particularity of people living with HIV and AIDS on subjective and privacy issues, PROs should be the primary outcome or end point. Many regulatory agencies and guidelines also recommend the inclusion of PROMs as the primary or secondary end points in clinical trials [205,206]. In addition, the development of the current ART regimen aims at simplifying the form of administration to meet the needs of long-term ART and maintain viral suppression with minimal toxicity [207]. Therefore, PRO data are becoming increasingly important for determining which ART regimen to use [208]. Therefore, a reliable, valid, and sensitive PROM is invaluable to health care providers and researchers.

In this systematic review, only 3 available PROMs (PozQoL, HIV-SI or SDM, and PLHIV-RS) were recommended based on the COSMIN guidelines, wherein PozQoL was used to assess HIV-related HRQoL, HIV-SI or SDM was used to assess HIV-related symptoms, and PLHIV-RS was used to assess HIV-related resilience. Health care providers can adopt these 3 PROMs for different application purposes. With regard to PROMs that received class B recommendation, although these PROMs are not recommended in this systematic review, researchers can select the PROMs with relatively good results for psychometric properties and use them according to the research purpose or further validate them for use in their context. For administrators, selecting validated PROMs can aid in the development of continuous quality improvement reports to understand health care providers' performance against the measurement framework and standard key performance indicators [209]. On the basis of the data collected through validated PROMs, policy makers can further evaluate system performance by comparing outcomes over time and support health care policy decision-making [210]. In summary, this review will help health care providers, administrators, policy makers, and researchers to choose suitable PROMs in different contexts, which in turn will promote the systematic use of these PROMs, identify areas that need to be improved from a patient perspective, and improve the quality of assessment for intervention.

Limitations

Our study has some limitations. First, although this systematic review additionally searched 2 important web-based databases of PROMs (PROQOLID and PROMIS) that are considered to be an important source of gray literature, we did not search dissertations, non-English literature, and other gray literature. This may have caused some relevant studies to be left out of our analysis, and these studies may help provide some evidence to support or refute our findings. Furthermore, evidence on the validation of PROMs can be deduced from the results of some studies. However, it was not the primary purpose of these studies; therefore, these studies were not included. Furthermore, some other PROMs were not included because they are still under study. Moreover, this systematic review may have ignored PROMs that only assessed a certain domain related to specific comorbidities, such as PROMs specifically designed to measure cognitive concerns. Considering the importance of evaluating these comorbidities in people living with HIV and AIDS, we will conduct further research on these PROMs. Furthermore, because no generally accepted "golden standard" measure for adults living with HIV and AIDS currently exists, the criterion validity of the included PROMs was not assessed. In addition, an insufficient number of studies reporting PROM development and content validity were included in this systematic review. Although we excluded many qualitative studies during the title and abstract screening stage, none of these studies researched on content validity. However, this is the same as the other relevant reviews [16,19] that also searched for insufficient studies reporting on the content validity of HIV-related PROMs.

One another limitation of this review is that the selection of studies, scoring of methodological quality, and grading of evidence were subjective in nature. However, this systematic review strictly followed the steps of the COSMIN guidelines, and the processes mentioned earlier involved multiple researchers. We believe that this could resolve discrepancies and reduce variability in interpretation, thereby minimizing the chance of errors. Furthermore, given that the negative results of many PROMs are less likely to be published, the possibility of publication bias cannot be eliminated. Moreover, some included studies may have reported on only some psychometric properties; accordingly, there may be a selective reporting bias. Finally, quantitative pooled summary or meta-analyses were not performed because of the possible large heterogeneity. These limitations may help to explain why concrete recommendations for the use of some PROMs were not made because there were few included studies for some PROMs, and not all psychometric properties were assessed in these studies.

Future Work

Although there are a large number of PROMs in each category, it would be necessary to validate the existing PROMs, or even develop new PROMs in some categories, because not enough validated PROMs are available. Considering the shortcomings of the 3 class A PROMs, efforts in future research should focus on validation as well as class B PROMs. It should be noted that multiple personnel such as patients themselves, their family members, health care providers, and researchers should participate in the development and validation of all PROMs [211]. In the future research on PROMs, researchers should follow the suggestions of the COSMIN guidelines to ensure the complete reporting of research details and accurate interpretation of results [27].

For the existing PROMs, research should focus on the validation of content validity and measurement error to determine the suitability of a PROM for use in the care of people living with HIV and AIDS. Moreover, these PROMs should be applied to different regions or populations to assess their cross-cultural validity or measurement invariance and explore the comparability of the results. In addition, future research should use more longitudinal or experimental study designs to assess the responsiveness of PROMs [9].

With the gradual aging of people living with HIV and AIDS, new and adjusted PROMs should focus on exploring the impact of aging on people living with HIV and AIDS, such as complex complications [212], polypharmacy [213], menopause in older women [214], low social support [215], cognitive impairment [216], and special symptoms of early exposure to HIV [9]. PROMs for children will be summarized in our future research.

In the past decades, researchers have mainly used interviewer-administered surveys and self-administered paper questionnaires to collect data [217]. However, several limitations of these methods have been found in the actual application process. ePROMs are becoming increasingly popular in recent years, greatly saving labor and time costs, minimizing errors, and realizing complex survey management [9]. Despite the fact that ePROMs are rapidly developing, future research should pay attention to evaluating the equivalence between electronic questionnaires and paper questionnaires [218]. Some researchers have used the most advanced technologies to integrate ePROMs into electronic hospital records or routine HIV care, allowing health care providers to easily and conveniently assess the qualitative and quantitative health outcomes of people living with HIV and AIDS. In addition, there are independent apps and software used in clinical practice and research.

Moreover, with the development of computer adaptive tests (CATs) in recent years, future research can develop and improve the item bank for people living with HIV and AIDS and use the CAT technology to dynamically select items for administration based on the respondent's previous answers for finally assessing their PROs [219-221]. However, the item bank of the CAT instrument requires a large number of unidimensional scales, posing a great challenge to the content validity of each PROM and its subconstructs. At the same time, the development of a CAT item bank can promote the improvement of the existing HIV-specific PROMs and the development of new HIV-specific PROMs, further promoting the vigorous development of research in related fields in the future.

Conclusions

This systematic review provides a detailed assessment of the psychometric properties of the existing HIV-specific PROMs for adults living with HIV and AIDS. Class A rating of PROMs was achieved for PozQoL, HIV-SI or SDM, and PLHIV-RS. However, all of these have a few shortcomings. Therefore, this study believes that future studies should conduct a more comprehensive validation of the psychometric properties of the existing PROMs to provide sufficient assessment evidence. These findings may provide a reference for the selection of high-quality HIV-specific PROMs by health care providers and researchers for clinical practice and research.

Acknowledgments

The authors would like to thank all the authors and research teams who developed all the PROMs included in this review. Furthermore, they would like to acknowledge the COSMIN group for their advice on the study design and application of the COSMIN tools. In addition, they thank the senior health research librarian (Liyuan Cui from the library of China Medical University) for her guidance on the search strategy for this review. Finally, they would like to thank MogoEdit for its English editing during the preparation of this manuscript. This work was supported by the National Natural Science Foundation of China (grant 71473269 and 71673301), the Distinguished Professor Program of Liaoning Province, and the "Bai-Qian-Wan" Talents program of Liaoning Province. The funding agencies had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; and decision to submit the manuscript for publication.

Authors' Contributions

ZW, YZ, and BQ conceptualized and designed the study. ZW and HK performed literature search, screening, and selection. ZW, HK, and XD extracted the data. ZW, HK, and YZ performed quality appraisal and statistical analysis. ZW, HK, XD, YZ, and BQ contributed to COSMIN evaluation. YZ and BQ supervised the study. ZW, YZ, and BQ drafted the manuscript. ZW, YZ, and BQ critically revised the manuscript for important intellectual content. ZW, YZ, and BQ provided administrative, technical, or material support. All the authors critically reviewed the manuscript and approved the final version before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Literature search strategy for existing review, literature search strategy for HIV-specific patient-reported outcome measures, subscales of the included PROMs, characteristics of the included records, methodological quality assessment of the included

records, results and ratings of each psychometric property of each record, the overall results and the level of evidence, PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 abstract checklist, and PRISMA 2020 main checklist. [DOCX File , 258 KB - publichealth_v8i12e39015_app1.docx]

References

1. Global HIV and AIDS statistics — Fact sheet. Joint United Nations Programme on HIV/AIDS. 2021 Dec 1. URL: https://www.unaids.org/en/resources/fact-sheet [accessed 2021-12-18]
2. Antiretroviral Therapy Cohort Collaboration. Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. Lancet HIV 2017 Aug;4(8):e349-e356 [FREE Full text] [doi: 10.1016/S2352-3018(17)30066-8] [Medline: 28501495]
3. Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, Kirtley S, WHO–UNAIDS Network for HIV Isolation Characterisation. Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis. Lancet Infect Dis 2019 Feb;19(2):143-155. [doi: 10.1016/S1473-3099(18)30647-9] [Medline: 30509777]
4. Remien RH, Stirratt MJ, Nguyen N, Robbins RN, Pala AN, Mellins CA. Mental health and HIV/AIDS: the need for an integrated response. AIDS 2019 Jul 15;33(9):1411-1420 [FREE Full text] [doi: 10.1097/QAD.0000000000002227] [Medline: 30950883]
5. Sued O, Figueroa MI, Cahn P. Clinical challenges in HIV/AIDS: hints for advancing prevention and patient management strategies. Adv Drug Deliv Rev 2016 Aug 01;103:5-19. [doi: 10.1016/j.addr.2016.04.016] [Medline: 27117711]
6. Back D, Marzolini C. The challenge of HIV treatment in an era of polypharmacy. J Int AIDS Soc 2020 Feb;23(2):e25449 [FREE Full text] [doi: 10.1002/jia2.25449] [Medline: 32011104]
7. Lazarus JV, Safreed-Harmon K, Barton SE, Costagliola D, Dedes N, Del Amo Valero J, et al. Beyond viral suppression of HIV - the new quality of life frontier. BMC Med 2016 Jun 22;14(1):94 [FREE Full text] [doi: 10.1186/s12916-016-0640-4] [Medline: 27334606]
8. Serwadda DM. Beyond HIV viral suppression: an African perspective. Lancet HIV 2019 Dec;6(12):e812-e814. [doi: 10.1016/S2352-3018(19)30349-2] [Medline: 31776100]
9. Kall M, Marcellin F, Harding R, Lazarus JV, Carrieri P. Patient-reported outcomes to enhance person-centred HIV care. Lancet HIV 2020 Jan;7(1):e59-e68. [doi: 10.1016/S2352-3018(19)30345-5] [Medline: 31776101]
10. Yang Z, Zhu Z, Wen H, Han S, Xing W, Dong T, et al. Psychometric properties of self-reported measures of health-related quality of life in people living with HIV: a systematic review protocol. JBI Evid Synth 2021 May 14;19(10):2829-2838. [doi: 10.11124/JBIES-20-00190] [Medline: 33993150]
11. Jerant A, Tancredi DJ, Franks P. Mortality prediction by quality-adjusted life year compatible health measures: findings in a nationally representative US sample. Med Care 2011 May;49(5):443-450. [doi: 10.1097/MLR.0b013e318206c231] [Medline: 21368679]
12. DeSalvo KB, Jones TM, Peabody J, McDonald J, Fihn S, Fan V, et al. Health care expenditure prediction with a single item, self-rated health measure. Med Care 2009 Apr;47(4):440-447. [doi: 10.1097/MLR.0b013e318190b716] [Medline: 19238099]
13. Herrmann S, Power B, Rashidi A, Cypher M, Mastaglia F, Grace A, et al. Supporting patient-clinician interaction in chronic HIV care: design and development of a patient-reported outcomes software application. J Med Internet Res 2021 Jul 30;23(7):e27861 [FREE Full text] [doi: 10.2196/27861] [Medline: 34328442]
14. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. BMJ Qual Saf 2014 Jun;23(6):508-518. [doi: 10.1136/bmjqs-2013-002524] [Medline: 24505110]
15. Etkind SN, Daveson BA, Kwok W, Witt J, Bausewein C, Higginson IJ, et al. Capture, transfer, and feedback of patient-centered outcomes data in palliative care populations: does it make a difference? A systematic review. J Pain Symptom Manage 2015 Mar;49(3):611-624 [FREE Full text] [doi: 10.1016/j.jpainsymman.2014.07.010] [Medline: 25135657]
16. Cooper V, Clatworthy J, Harding R, Whetham J, Emerge Consortium. Measuring quality of life among people living with HIV: a systematic review of reviews. Health Qual Life Outcomes 2017 Nov 15;15(1):220 [FREE Full text] [doi: 10.1186/s12955-017-0778-6] [Medline: 29141645]
17. Huang IC, Frangakis C, Atkinson MJ, Willke RJ, Leite WL, Vogel WB, et al. Addressing ceiling effects in health status measures: a comparison of techniques applied to measures for people with HIV disease. Health Serv Res 2008 Feb;43(1 Pt 1):327-339 [FREE Full text] [doi: 10.1111/j.1475-6773.2007.00745.x] [Medline: 18211533]
18. Seguiti C, Salvo PF, Di Stasio E, Lamonica S, Fedele AL, Manfrida S, et al. Health-related quality of life (HRQoL) from HIV patients' perspective: comparison of patient-reported outcome (PRO) measures among people living with HIV (PLWH) and other chronic clinical conditions. J Patient Rep Outcomes 2022 Mar 26;6(1):27 [FREE Full text] [doi: 10.1186/s41687-022-00423-4] [Medline: 35347476]
19. Wen H, Yang Z, Zhu Z, Han S, Zhang L, Hu Y. Psychometric properties of self-reported measures of health-related quality of life in people living with HIV: a systematic review. Health Qual Life Outcomes 2022 Jan 10;20(1):5 [FREE Full text] [doi: 10.1186/s12955-021-01910-w] [Medline: 35012574]

20. Clayson DJ, Wild DJ, Quarterman P, Duprat-Lomon I, Kubin M, Coons SJ. A comparative review of health-related quality-of-life measures for use in HIV/AIDS clinical trials. Pharmacoeconomics 2006;24(8):751-765. [doi: 10.2165/00019053-200624080-00003] [Medline: 16898846]
21. Wu AW, Hanson KA, Harding G, Haider S, Tawadrous M, Khachatryan A, et al. Responsiveness of the MOS-HIV and EQ-5D in HIV-infected adults receiving antiretroviral therapies. Health Qual Life Outcomes 2013 Mar 12;11:42 [FREE Full text] [doi: 10.1186/1477-7525-11-42] [Medline: 23497257]
22. Engler K, Lessard D, Lebouché B. A review of HIV-specific patient-reported outcome measures. Patient 2017 Apr;10(2):187-202. [doi: 10.1007/s40271-016-0195-7] [Medline: 27637488]
23. Gakhar H, Kamali A, Holodniy M. Health-related quality of life assessment after antiretroviral therapy: a review of the literature. Drugs 2013 May;73(7):651-672 [FREE Full text] [doi: 10.1007/s40265-013-0040-4] [Medline: 23591907]
24. Robberstad B, Olsen JA. The health related quality of life of people living with HIV/AIDS in sub-Saharan Africa - a literature review and focus group study. Cost Eff Resour Alloc 2010 Apr 16;8:5 [FREE Full text] [doi: 10.1186/1478-7547-8-5] [Medline: 20398367]
25. Terwee CB, Prinsen CA, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018 May;27(5):1159-1170 [FREE Full text] [doi: 10.1007/s11136-018-1829-0] [Medline: 29550964]
26. Mokkink LB, de Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res 2018 May;27(5):1171-1179 [FREE Full text] [doi: 10.1007/s11136-017-1765-4] [Medline: 29260445]
27. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May;27(5):1147-1157 [FREE Full text] [doi: 10.1007/s11136-018-1798-3] [Medline: 29435801]
28. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. J Clin Epidemiol 2021 Jun;134:178-189 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.03.001] [Medline: 33789819]
29. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009 Oct;18(8):1115-1123 [FREE Full text] [doi: 10.1007/s11136-009-9528-5] [Medline: 19711195]
30. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs) user manual. COSMIN. 2018 Feb. URL: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf [accessed 2021-03-07]
31. Carretero MD, Burgess AP, Soler P, Soler M, Catalán J. Reliability and validity of an HIV-specific health-related quality-of-life measure for use with injecting drug users. AIDS 1996 Dec;10(14):1699-1705. [doi: 10.1097/00002030-199612000-00015] [Medline: 8970691]
32. Murri R, Ammassari A, Fantoni M, Scoppettuolo G, Cingolani A, De Luca A, et al. Disease-related factors associated with health-related quality of life in people with nonadvanced HIV disease assessed using an Italian version of the MOS-HIV Health Survey. J Acquir Immune Defic Syndr Hum Retrovirol 1997 Dec 15;16(5):350-356. [doi: 10.1097/00042560-199712150-00007] [Medline: 9420313]
33. Revicki DA, Sorensen S, Wu AW. Reliability and validity of physical and mental health summary scores from the Medical Outcomes Study HIV Health Survey. Med Care 1998 Feb;36(2):126-137. [doi: 10.1097/00005650-199802000-00003] [Medline: 9475468]
34. Badia X, Podzamczer D, Garcia M, López-Lavid C, Consiglio E. A randomized study comparing instruments for measuring health-related quality of life in HIV-infected patients. Spanish MOS-HIV and MQOL-HIV Validation Group. Medical Outcomes Study HIV Health Survey. AIDS 1999 Sep 10;13(13):1727-1735. [doi: 10.1097/00002030-199909100-00017] [Medline: 10509575]
35. Holmes WC, Shea JA. Two approaches to measuring quality of life in the HIV/AIDS population: HAT-QoL and MOS-HIV. Qual Life Res 1999 Sep;8(6):515-527. [doi: 10.1023/a:1008931006866] [Medline: 10548867]
36. Scott-Lennox JA, Wu AW, Boyer JG, Ware JE. Reliability and validity of French, German, Italian, Dutch, and UK English translations of the Medical Outcomes Study HIV Health Survey. Med Care 1999 Sep;37(9):908-925. [doi: 10.1097/00005650-199909000-00007] [Medline: 10493469]
37. Delate T, Coons SJ. The use of 2 health-related quality-of-life measures in a sample of persons infected with human immunodeficiency virus. Clin Infect Dis 2001 Feb 01;32(3):E47-E52. [doi: 10.1086/318492] [Medline: 11170970]
38. Paton NI, Chapman CA, Chan S, Tan KM, Leo Y, Aboulhab J, et al. Validation of the Medical Outcomes Study HIV Health Survey as a measure of quality of life in HIV-infected patients in Singapore. Int J STD AIDS 2002 Jul;13(7):456-461. [doi: 10.1258/09564620260079608] [Medline: 12171664]
39. Schifano P, Borgia P, Wu AW, Spadea T, Milanese G, Perucci CA. Validity and reliability of the Italian translation of the MOS-HIV health survey in persons with AIDS. Qual Life Res 2003 Dec;12(8):1137-1146. [doi: 10.1023/a:1026151931248] [Medline: 14651431]

40. Ichikawa M, Natpratan C. Quality of life among people living with HIV/AIDS in northern Thailand: MOS-HIV Health Survey. Qual Life Res 2004 Apr;13(3):601-610. [doi: 10.1023/B:QURE.0000021319.73865.5a] [Medline: 15130024]
41. Lau JT, Tsui HY, Patrick LC, Rita CW, Molassiotis A. Validation of a Chinese version of the Medical Outcomes Study HIV Health Survey (MOS-HIV) among Chinese people living with HIV/AIDS in Hong Kong. Qual Life Res 2006 Aug;15(6):1079-1089. [doi: 10.1007/s11136-005-5914-9] [Medline: 16900288]
42. Taylor TN, Dolezal C, Tross S, Holmes WC. Reliability and validity of two HIV/AIDS-specific quality of life instruments adapted for use in HIV-positive Zimbabweans. AIDS Care 2009 May;21(5):598-607 [FREE Full text] [doi: 10.1080/09540120802302574] [Medline: 19444668]
43. Henderson WA, Schlenk EA, Kim KH, Hadigan CM, Martino AC, Sereika SM, et al. Validation of the MOS-HIV as a measure of health-related quality of life in persons living with HIV and liver disease. AIDS Care 2010 Apr;22(4):483-490 [FREE Full text] [doi: 10.1080/09540120903207292] [Medline: 20140792]
44. Stasinopoulou PG, Tzavara C, Dimitrakaki C, Georgiou O, Baraboutis IG, Skoutelis A, et al. Reliability and validity of the Greek translation of the MOS-HIV health survey in HIV-infected individuals. Qual Life Res 2010 Mar;19(2):199-205. [doi: 10.1007/s11136-009-9573-0] [Medline: 20066565]
45. Chariyalertsak S, Wansom T, Kawichai S, Ruangyuttikarna C, Kemerer VF, Wu AW. Reliability and validity of Thai versions of the MOS-HIV and SF-12 quality of life questionnaires in people living with HIV/AIDS. Health Qual Life Outcomes 2011 Mar 15;9:15 [FREE Full text] [doi: 10.1186/1477-7525-9-15] [Medline: 21406088]
46. Hsiung PC, Fang CT, Lee KL, Sheng WH, Wu CY, Wang JD, et al. Validation of the medical outcomes study HIV (MOS-HIV) health survey among HIV-infected patients in Taiwan. Qual Life Res 2011 Mar;20(2):281-286. [doi: 10.1007/s11136-010-9733-2] [Medline: 20803317]
47. Epino HM, Rich ML, Kaigamba F, Hakizamungu M, Socci AR, Bagiruwigize E, et al. Reliability and construct validity of three health-related self-report scales in HIV-positive adults in rural Rwanda. AIDS Care 2012;24(12):1576-1583. [doi: 10.1080/09540121.2012.661840] [Medline: 22428702]
48. Huang ZJ, Tian M, Dai SY, Ye DQ. Feasibility, reliability and validity of the Chinese simplified version of the MOS-HIV health survey among AIDS patients in China. Qual Life Res 2013 Mar;22(2):403-407. [doi: 10.1007/s11136-012-0148-0] [Medline: 22392524]
49. Marzieh A, Peyman J, Seyyed Mohammad Taghi A, Parvin Afsar K. Reliability and validity of Persian version of medical outcome study-HIV health survey in Iranian people living with HIV. J Tradit Chin Med 2017 Oct;37(5):695-701 [FREE Full text] [Medline: 32188232]
50. Liu J, Zhu Y, Qu B. Reliability and validity of the Chinese version of the Medical Outcomes Study HIV Health Survey (MOS-HIV) in people living with HIV/AIDS (PLWHA) in China. PLoS One 2018 Jul 25;13(7):e0201177 [FREE Full text] [doi: 10.1371/journal.pone.0201177] [Medline: 30044881]
51. Shim EJ, Ha H, Lee SH, Kim NJ, Kim ES, Bang JH, et al. Psychometric properties of the Korean version of the medical outcomes study HIV health survey: results from a multicenter survey in Korea. Health Qual Life Outcomes 2018 May 15;16(1):92 [FREE Full text] [doi: 10.1186/s12955-018-0919-6] [Medline: 29764440]
52. Stangl AL, Bunnell R, Wamai N, Masaba H, Mermin J. Measuring quality of life in rural Uganda: reliability and validity of summary scores from the medical outcomes study HIV health survey (MOS-HIV). Qual Life Res 2012 Nov;21(9):1655-1663. [doi: 10.1007/s11136-011-0075-5] [Medline: 22198741]
53. McDonnell KA, Gielen AC, Wu AW, O'Campo P, Faden R. Measuring health related quality of life among women living with HIV. Qual Life Res 2000;9(8):931-940. [doi: 10.1023/a:1008909919456] [Medline: 11284212]
54. Schag CA, Ganz PA, Kahn B, Petersen L. Assessing the needs and quality of life of patients with HIV infection: development of the HIV Overview of Problems-Evaluation System (HOPES). Qual Life Res 1992 Dec;1(6):397-413. [doi: 10.1007/BF00704434] [Medline: 1299472]
55. De Boer JB, Sprangers MA, Aaronson NK, Lange JM, van Dam FS. A study of the reliability, validity and responsiveness of the HIV overview of problems evaluation system (HOPES) in assessing the quality of life of patients with AIDS and symptomatic HIV infection. Qual Life Res 1996 Jun;5(3):339-347. [doi: 10.1007/BF00433918] [Medline: 8763802]
56. Cleary PD, Fowler Jr FJ, Weissman J, Massagli MP, Wilson I, Seage 3rd GR, et al. Health-related quality of life in persons with acquired immune deficiency syndrome. Med Care 1993 Jul;31(7):569-580. [doi: 10.1097/00005650-199307000-00001] [Medline: 8326772]
57. Lubeck DP, Fries JF. Assessment of quality of life in early stage HIV-infected persons: data from the AIDS Time-oriented Health Outcome Study (ATHOS). Qual Life Res 1997 Aug;6(6):494-506. [doi: 10.1023/a:1018404014821] [Medline: 9330550]
58. Bozzette SA, Hays RD, Berry SH, Kanouse DE. A Perceived Health Index for use in persons with advanced HIV disease: derivation, reliability, and validity. Med Care 1994 Jul;32(7):716-731. [doi: 10.1097/00005650-199407000-00005] [Medline: 8028406]
59. Bozzette SA, Hays RD, Berry SH, Kanouse DE, Wu AW. Derivation and properties of a brief health status assessment instrument for use in HIV disease. J Acquir Immune Defic Syndr Hum Retrovirol 1995 Mar 01;8(3):253-265. [doi: 10.1097/00042560-199503010-00006] [Medline: 7859137]

60. Hays RD, Cunningham WE, Ettl MK, Beck CK, Shapiro MF. Health related quality of life in HIV disease. Assessment 1995 Dec;2(4):363-380. [doi: 10.1177/1073191195002004006]
61. Cella DF, McCain NL, Peterman AH, Mo F, Wolen D. Development and validation of the Functional Assessment of Human Immunodeficiency Virus Infection (FAHI) quality of life instrument. Qual Life Res 1996 Aug;5(4):450-463. [doi: 10.1007/BF00449920] [Medline: 8840825]
62. Peterman AH, Cella D, Mo F, McCain N. Psychometric validation of the revised Functional Assessment of Human Immunodeficiency Virus Infection (FAHI) quality of life instrument. Qual Life Res 1997 Aug;6(6):572-584. [doi: 10.1023/a:1018416317546] [Medline: 9330556]
63. Viala-Danten M, Dubois D, Gilet H, Martin S, Peeters K, Cella D. Psychometric evaluation of the functional assessment of HIV Infection (FAHI) questionnaire and its usefulness in clinical trials. Qual Life Res 2010 Oct;19(8):1215-1227. [doi: 10.1007/s11136-010-9674-9] [Medline: 20509050]
64. Byrne S, Petry NM. Reliability and validity of the Functional Assessment of Human Immunodeficiency Virus Infection (FAHI) in patients with drug and alcohol use disorders. AIDS Care 2013;25(1):118-125 [FREE Full text] [doi: 10.1080/09540121.2012.687811] [Medline: 22646736]
65. Nyongesa MK, Sigilai A, Hassan AS, Thoya J, Odhiambo R, Van de Vijver FJ, et al. A mixed methods approach to adapting and evaluating the functional assessment of HIV infection (FAHI), Swahili version, for use with low literacy populations. PLoS One 2017 Apr 5;12(4):e0175021 [FREE Full text] [doi: 10.1371/journal.pone.0175021] [Medline: 28380073]
66. Lenderking WR, Testa MA, Katzenstein D, Hammer S. Measuring quality of life in early HIV disease: the modular approach. Qual Life Res 1997 Aug;6(6):515-530. [doi: 10.1023/a:1018408115729] [Medline: 9330552]
67. Leplège A, Rude N, Ecosse E, Ceinos R, Dohin E, Pouchot J. Measuring quality of life from the point of view of HIV-positive subjects: the HIV-QL31. Qual Life Res 1997 Aug;6(6):585-594. [doi: 10.1023/a:1018468301617] [Medline: 9330557]
68. Holmes WC, Shea JA. Performance of a new, HIV/AIDS-targeted quality of life (HAT-QoL) instrument in asymptomatic seropositive individuals. Qual Life Res 1997 Aug;6(6):561-571. [doi: 10.1023/a:1018464200708] [Medline: 9330555]
69. Holmes WC, Shea JA. A new HIV/AIDS-targeted quality of life (HAT-QoL) instrument: development, reliability, and validity. Med Care 1998 Feb;36(2):138-154. [doi: 10.1097/00005650-199802000-00004] [Medline: 9475469]
70. Holmes WC, Ruocco JE. Test-retest evaluation of HAT-QoL and SF-36 in an HIV-seropositive sample. AIDS Care 2008 Oct;20(9):1084-1092. [doi: 10.1080/09540120701797437] [Medline: 18608069]
71. Mafirakureva N, Dzingirai B, Postma MJ, van Hulst M, Khoza S. Health-related quality of life in HIV/AIDS patients on antiretroviral therapy at a tertiary care facility in Zimbabwe. AIDS Care 2016 Jul;28(7):904-912. [doi: 10.1080/09540121.2016.1173639] [Medline: 27098289]
72. Smith KW, Avis NE, Mayer KH, Swislow L. Use of the MQoL-HIV with asymptomatic HIV-positive patients. Qual Life Res 1997 Aug;6(6):555-560. [doi: 10.1023/a:1018412216638] [Medline: 9330554]
73. Kemmler G, Schmied B, Shetty-Lee A, Zangerle R, Hinterhuber H, Schüssler G, et al. Quality of life of HIV-infected patients: psychometric properties and validation of the German version of the MQOL-HIV. Qual Life Res 2003 Dec;12(8):1037-1050. [doi: 10.1023/a:1026114004548] [Medline: 14651421]
74. Remple VP, Hilton BA, Ratner PA, Burdge DR. Psychometric assessment of the Multidimensional Quality of Life Questionnaire for Persons with HIV/AIDS (MQOL-HIV) in a sample of HIV-infected women. Qual Life Res 2004 Jun;13(5):947-957. [doi: 10.1023/B:QURE.0000025584.77779.e1] [Medline: 15233508]
75. Watanabe M, Nishimura K, Inoue T, Kimura S, Oka S, QoL Research Group Of The AIDS Clinical Centre And Eight Regional AIDS Treatment Hospitals In Japan. A discriminative study of health-related quality of life assessment in HIV-1-infected persons living in Japan using the Multidimensional Quality of Life Questionnaire for persons with HIV/AIDS. Int J STD AIDS 2004 Feb;15(2):107-115. [doi: 10.1258/095646204322764307] [Medline: 15006073]
76. Holzemer WL, Gygax Spicer J, Skodol Wilson H, Kemppainen JK, Coleman C. Validation of the quality of life scale: living with HIV. J Adv Nurs 1998 Sep;28(3):622-630. [doi: 10.1046/j.1365-2648.1998.00800.x] [Medline: 9756232]
77. Starace F, Cafaro L, Abrescia N, Chirianni A, Izzo C, Rucci P, et al. Quality of life assessment in HIV-positive persons: application and validation of the WHOQOL-HIV, Italian version. AIDS Care 2002 Jun;14(3):405-415. [doi: 10.1080/09540120220123793a] [Medline: 12042086]
78. WHOQOL HIV Group. WHOQOL-HIV for quality of life assessment among people living with HIV and AIDS: results from the field test. AIDS Care 2004 Oct;16(7):882-889. [doi: 10.1080/09540120412331290194] [Medline: 15385243]
79. Zimpel RR, Fleck MP. Quality of life in HIV-positive Brazilians: application and validation of the WHOQOL-HIV, Brazilian version. AIDS Care 2007 Aug;19(7):923-930. [doi: 10.1080/09540120701213765] [Medline: 17712697]
80. Canavarro MC, Pereira M, Simoes MR, Pintassilgo AL. Quality of life assessment in HIV-infection: validation of the European Portuguese version of WHOQOL-HIV. AIDS Care 2011 Feb;23(2):187-194. [doi: 10.1080/09540121.2010.498870] [Medline: 21259131]
81. Mweemba P, Zeller R, Ludwick R, Gosnell D, Michelo C. Validation of the World Health Organization Quality of Life HIV instrument in a Zambian sample. J Assoc Nurses AIDS Care 2011;22(1):53-66. [doi: 10.1016/j.jana.2010.04.006] [Medline: 20619690]
82. Olsen M, Jensen NK, Tesfaye M, Holm L. Conceptual equivalence of WHOQOL-HIV among people living with HIV in Ethiopia. Qual Life Res 2013 Mar;22(2):361-367. [doi: 10.1007/s11136-012-0141-7] [Medline: 22367635]

83. Reychler G, Caty G, Vincent A, Billo S, Yombi J. Validation of the French version of the World Health Organization quality of life HIV instrument. PLoS One 2013 Sep 3;8(9):e73180 [FREE Full text] [doi: 10.1371/journal.pone.0073180] [Medline: 24019904]
84. Saddki N, Noor MM, Norbanee TH, Rusli MA, Norzila Z, Zaharah S, et al. Validity and reliability of the Malay version of WHOQOL-HIV BREF in patients with HIV infection. AIDS Care 2009 Oct;21(10):1271-1278. [doi: 10.1080/09540120902803216] [Medline: 20024703]
85. Hsiung PC, Fang CT, Wu CH, Sheng WH, Chen SC, Wang JD, et al. Validation of the WHOQOL-HIV BREF among HIV-infected patients in Taiwan. AIDS Care 2011 Aug;23(8):1035-1042. [doi: 10.1080/09540121.2010.543881] [Medline: 21500023]
86. O'Connell KA, Skevington SM. An international quality of life instrument to assess wellbeing in adults who are HIV-positive: a short form of the WHOQOL-HIV (31 items). AIDS Behav 2012 Feb 23;16(2):452-460. [doi: 10.1007/s10461-010-9863-0] [Medline: 21181253]
87. Tran BX. Quality of life outcomes of antiretroviral treatment for HIV/AIDS patients in Vietnam. PLoS One 2012 Jul 20;7(7):e41062 [FREE Full text] [doi: 10.1371/journal.pone.0041062] [Medline: 22911742]
88. Canavarro MC, Pereira M. Factor structure and psychometric properties of the European Portuguese version of a questionnaire to assess quality of life in HIV-infected adults: the WHOQOL-HIV-Bref. AIDS Care 2012 Nov 22;24(6):799-807. [doi: 10.1080/09540121.2011.630362] [Medline: 22107018]
89. Pereira M, Martins A, Alves S, Canavarro MC. Assessing quality of life in middle-aged and older adults with HIV: psychometric testing of the WHOQOL-HIV-Bref. Qual Life Res 2014 Nov;23(9):2473-2479. [doi: 10.1007/s11136-014-0707-7] [Medline: 24791929]
90. Meemon N, Paek SC, Yenchai D, Wan TT. Application of the WHOQOL-HIV-BREF questionnaire in HIV-infected Thai patients: reliability and validity of the instrument. J Assoc Nurses AIDS Care 2016;27(5):698-708. [doi: 10.1016/j.jana.2016.04.007] [Medline: 27220328]
91. Salehi M, Niroumand S, Erfanian MR, Sajjadi RB, Dadgarmoghaddam M. Validation of Persian version of WHOQOL-HIV BREF questionnaire in Islamic Republic of Iran. East Mediterr Health J 2016 Dec 12;22(9):647-653 [FREE Full text] [doi: 10.26719/2016.22.9.647] [Medline: 27966765]
92. Tesfaye M, Olsen MF, Medhin G, Friis H, Hanlon C, Holm L. Adaptation and validation of the short version WHOQOL-HIV in Ethiopia. Int J Ment Health Syst 2016;10:29 [FREE Full text] [doi: 10.1186/s13033-016-0062-x] [Medline: 27064377]
93. Zhu Y, Liu J, Qu B. Psychometric properties of the Chinese version of the WHOQOL-HIV BREF to assess quality of life among people living with HIV/AIDS: a cross-sectional study. BMJ Open 2017 Aug 21;7(8):e016382 [FREE Full text] [doi: 10.1136/bmjopen-2017-016382] [Medline: 28827253]
94. Fuster-RuizdeApodaca MJ, Laguía A, Safreed-Harmon K, Lazarus JV, Cenoz S, Del Amo J. Assessing quality of life in people with HIV in Spain: psychometric testing of the Spanish version of WHOQOL-HIV-BREF. Health Qual Life Outcomes 2019 Aug 19;17(1):144 [FREE Full text] [doi: 10.1186/s12955-019-1208-8] [Medline: 31426799]
95. Barger D, Hessamfar M, Neau D, Vareil MO, Lazaro E, Duffau P, et al. Assessing the psychometric properties of the French WHOQOL-HIV BREF within the ANRS CO3 Aquitaine Cohort's QuAliV ancillary study. Health Qual Life Outcomes 2020 Jul 10;18(1):220 [FREE Full text] [doi: 10.1186/s12955-020-01451-8] [Medline: 32650781]
96. Bucciardini R, Murri R, Guarinieri M, Starace F, Martini M, Vatrella A, et al. ISSQoL: a new questionnaire for evaluating the quality of life of people living with HIV in the HAART era. Qual Life Res 2006 Apr;15(3):377-390. [doi: 10.1007/s11136-005-3212-1] [Medline: 16547775]
97. Spire B, Arnould B, Barbier F, Durant J, Gilquin J, Landman R, et al. Simplification and first validation of a short battery of patient questionnaires for clinical management of HIV-infected patients: the HIV-SQUAD (Symptom Quality of life Adherence) Questionnaire. HIV Clin Trials 2009 Jan 6;10(4):215-232. [doi: 10.1310/hct1004-215] [Medline: 19723610]
98. Duracinsky M, Herrmann S, Berzins B, Armstrong AR, Kohli R, Le Coeur S, et al. The development of PROQOL-HIV: an international instrument to assess the health-related quality of life of persons living with HIV/AIDS. J Acquir Immune Defic Syndr 2012 Apr 15;59(5):498-505. [doi: 10.1097/QAI.0b013e318245cafe] [Medline: 22205438]
99. Duracinsky M, Lalanne C, Le Coeur S, Herrmann S, Berzins B, Armstrong AR, et al. Psychometric validation of the PROQOL-HIV questionnaire, a new health-related quality of life instrument-specific to HIV disease. J Acquir Immune Defic Syndr 2012 Apr 15;59(5):506-515. [doi: 10.1097/QAI.0b013e31824be3f2] [Medline: 22293550]
100. Duracinsky M, Lalanne C, Goujard C, Herrmann S, Cheung-Lung C, Brosseau JP, et al. Electronic versus paper-based assessment of health-related quality of life specific to HIV disease: reliability study of the PROQOL-HIV questionnaire. J Med Internet Res 2014 Apr 25;16(4):e115 [FREE Full text] [doi: 10.2196/jmir.3330] [Medline: 24769643]
101. Lalanne C, Chassany O, Carrieri P, Marcellin F, Armstrong AR, Lert F, et al. A reduced factor structure for the PROQOL-HIV questionnaire provided reliable indicators of health-related quality of life. J Clin Epidemiol 2016 Apr;72:116-125. [doi: 10.1016/j.jclinepi.2015.10.009] [Medline: 26548542]
102. Brown G, Mikołajczak G, Lyons A, Power J, Drummond F, Cogle A, et al. Development and validation of PozQoL: a scale to assess quality of life of PLHIV. BMC Public Health 2018 Apr 20;18(1):527 [FREE Full text] [doi: 10.1186/s12889-018-5433-6] [Medline: 29678156]

103. Burgess AP, Irving G, Riccio M. The reliability and validity of a symptom checklist for use in HIV infection: a preliminary analysis. Int J STD AIDS 1993 Jun 25;4(6):333-338. [doi: 10.1177/095646249300400606] [Medline: 8305574]
104. Whalen CC, Antani M, Carey J, Landefeld CS. An index of symptoms for infection with human immunodeficiency virus: reliability and validity. J Clin Epidemiol 1994 May;47(5):537-546. [doi: 10.1016/0895-4356(94)90300-x] [Medline: 7730879]
105. Nokes KM, Wheeler K, Kendrew J. Development of an HIV assessment tool. Image J Nurs Sch 1994;26(2):133-138. [doi: 10.1111/j.1547-5069.1994.tb00932.x] [Medline: 8063320]
106. Sousa KH, Tann SS, Kwok OM. Reconsidering the assessment of symptom status in HIV/AIDS care. J Assoc Nurses AIDS Care 2006;17(2):36-46. [doi: 10.1016/j.jana.2006.01.004] [Medline: 16800166]
107. Holzemer WL, Henry SB, Nokes KM, Corless IB, Brown MA, Powell-Cope GM, et al. Validation of the sign and symptom check-list for persons with HIV disease (SSC-HIV). J Adv Nurs 1999 Nov;30(5):1041-1049. [doi: 10.1046/j.1365-2648.1999.01204.x] [Medline: 10564402]
108. Holzemer WL, Hudson A, Kirksey KM, Hamilton MJ, Bakken S. The revised sign and symptom check-list for HIV (SSC-HIVrev). J Assoc Nurses AIDS Care 2001;12(5):60-70. [doi: 10.1016/s1055-3290(06)60263-x] [Medline: 11565239]
109. Mathews WC, McCutchan JA, Asch S, Turner BJ, Gifford AL, Kuromiya K, et al. National estimates of HIV-related symptom prevalence from the HIV cost and services utilization study. Med Care 2000 Jul;38(7):750-762. [doi: 10.1097/00005650-200007000-00007] [Medline: 10901358]
110. Justice AC, Holmes W, Gifford AL, Rabeneck L, Zackin R, Sinclair G, Adult AIDS Clinical Trials Unit Outcomes Committee. Development and validation of a self-completed HIV symptom index. J Clin Epidemiol 2001 Dec;54 Suppl 1:S77-S90. [doi: 10.1016/s0895-4356(01)00449-8] [Medline: 11750213]
111. Regnault A, Marfatia S, Louie M, Mear I, Meunier J, Viala-Danten M. Satisfactory cross-cultural validity of the ACTG symptom distress module in HIV-1-infected antiretroviral-naive patients. Clin Trials 2009 Dec;6(6):574-584. [doi: 10.1177/1740774509352515] [Medline: 19933717]
112. Marc LG, Wang MM, Testa MA. Psychometric evaluation of the HIV symptom distress scale. AIDS Care 2012 Mar 12;24(11):1432-1441 [FREE Full text] [doi: 10.1080/09540121.2012.656567] [Medline: 22409246]
113. Barroso J, Lynn MR. Psychometric properties of the HIV-Related Fatigue Scale. J Assoc Nurses AIDS Care 2002;13(1):66-75. [doi: 10.1016/S1055-3290(06)60242-2] [Medline: 11828861]
114. Pence BW, Barroso J, Leserman J, Harmon JL, Salahuddin N. Measuring fatigue in people living with HIV/AIDS: psychometric characteristics of the HIV-related fatigue scale. AIDS Care 2008 Aug;20(7):829-837 [FREE Full text] [doi: 10.1080/09540120701694063] [Medline: 18608084]
115. Li SY, Wu HS, Barroso J. The development and psychometric analysis of the Chinese HIV-Related Fatigue Scale. J Clin Nurs 2016 Apr;25(7-8):1025-1034. [doi: 10.1111/jocn.13151] [Medline: 26879371]
116. O'Brien KK, Solomon P, Bayoumi AM. Measuring disability experienced by adults living with HIV: assessing construct validity of the HIV Disability Questionnaire using confirmatory factor analysis. BMJ Open 2014 Sep 01;4(8):e005456 [FREE Full text] [doi: 10.1136/bmjopen-2014-005456] [Medline: 25180054]
117. O'Brien KK, Solomon P, Bergin C, O'Dea S, Stratford P, Iku N, et al. Reliability and validity of a new HIV-specific questionnaire with adults living with HIV in Canada and Ireland: the HIV Disability Questionnaire (HDQ). Health Qual Life Outcomes 2015 Aug 12;13:124 [FREE Full text] [doi: 10.1186/s12955-015-0310-9] [Medline: 26263898]
118. Brown DA, Simmons B, Boffito M, Aubry R, Nwokolo N, Harding R, et al. Evaluation of the psychometric properties of the HIV Disability Questionnaire among adults living with HIV in the United Kingdom: a cross-sectional self-report measurement study. PLoS One 2019 Jul 10;14(7):e0213222 [FREE Full text] [doi: 10.1371/journal.pone.0213222] [Medline: 31291243]
119. O'Brien KK, Kietrys D, Galantino ML, Parrott JS, Davis T, Tran Q, et al. Reliability and validity of the HIV disability questionnaire (HDQ) with adults living with HIV in the United States. J Int Assoc Provid AIDS Care 2019;18:2325958219888461 [FREE Full text] [doi: 10.1177/2325958219888461] [Medline: 31769326]
120. Bucciardini R, Pugliese K, Francisci D, Costantini A, Schiaroli E, Cognigni M, et al. Validation of a self-reported HIV symptoms list: the ISS-HIV symptoms scale. AIDS Res Ther 2016 Apr 9;13:18 [FREE Full text] [doi: 10.1186/s12981-016-0102-2] [Medline: 27064450]
121. Berger BE, Ferrans CE, Lashley FR. Measuring stigma in people with HIV: psychometric assessment of the HIV stigma scale. Res Nurs Health 2001 Dec;24(6):518-529. [doi: 10.1002/nur.10011] [Medline: 11746080]
122. Rao D, Pryor JB, Gaddist BW, Mayer R. Stigma, secrecy, and discrimination: ethnic/racial differences in the concerns of people living with HIV/AIDS. AIDS Behav 2008 Mar;12(2):265-271. [doi: 10.1007/s10461-007-9268-x] [Medline: 17588146]
123. Bunn JY, Solomon SE, Miller C, Forehand R. Measurement of stigma in people with HIV: a reexamination of the HIV Stigma Scale. AIDS Educ Prev 2007 Jun;19(3):198-208. [doi: 10.1521/aeap.2007.19.3.198] [Medline: 17563274]
124. Reinius M, Rao D, Manhart LE, Wiklander M, Svedhem V, Pryor J, et al. Differential item functioning for items in Berger's HIV Stigma Scale: an analysis of cohorts from the Indian, Swedish, and US contexts. Qual Life Res 2018 Jun;27(6):1647-1659 [FREE Full text] [doi: 10.1007/s11136-018-1841-4] [Medline: 29574526]

125. Rongkavilit C, Wright K, Chen X, Naar-King S, Chuenyam T, Phanuphak P. HIV stigma, disclosure and psychosocial distress among Thai youth living with HIV. Int J STD AIDS 2010 Feb;21(2):126-132. [doi: 10.1258/ijsa.2009.008488] [Medline: 20089999]
126. Reinius M, Wettergren L, Wiklander M, Svedhem V, Ekström AM, Eriksson LE. Development of a 12-item short version of the HIV stigma scale. Health Qual Life Outcomes 2017 May 30;15(1):115 [FREE Full text] [doi: 10.1186/s12955-017-0691-z] [Medline: 28558805]
127. Luz PM, Torres TS, Almeida-Brasil CC, Marins LM, Bezerra DR, Veloso VG, et al. Translation and validation of the Short HIV Stigma scale in Brazilian Portuguese. Health Qual Life Outcomes 2020 Oct 02;18(1):322 [FREE Full text] [doi: 10.1186/s12955-020-01571-1] [Medline: 33008400]
128. Lindberg MH, Wettergren L, Wiklander M, Svedhem-Johansson V, Eriksson LE. Psychometric evaluation of the HIV stigma scale in a Swedish context. PLoS One 2014 Dec 18;9(12):e114867 [FREE Full text] [doi: 10.1371/journal.pone.0114867] [Medline: 25522127]
129. Fuster-RuizdeApodaca MJ, Molero F, Holgado FP, Ubillos S. Adaptation of the HIV stigma scale in Spaniards with HIV. Span J Psychol 2015 Sep 15;18:E66. [doi: 10.1017/S1138741615000694] [Medline: 26369905]
130. Kagiura F, Fujii T, Kihana N, Maruyama E, Shimoji Y, Kakehashi M. Brief HIV stigma scale for Japanese people living with HIV: validation and restructuring using questionnaire survey data. AIDS Care 2020;32(sup1):1-9. [doi: 10.1080/09540121.2019.1683809] [Medline: 31658827]
131. Holzemer WL, Uys LR, Chirwa ML, Greeff M, Makoae LN, Kohi TW, et al. Validation of the HIV/AIDS stigma instrument - PLWA (HASI-P). AIDS Care 2007 Sep;19(8):1002-1012. [doi: 10.1080/09540120701245999] [Medline: 17851997]
132. Maluccio JA, Wu F, Rokon RB, Rawat R, Kadiyala S. Assessing the impact of food assistance on stigma among people living with HIV in Uganda using the HIV/AIDS stigma instrument-PLWA (HASI-P). AIDS Behav 2017 Mar;21(3):766-782. [doi: 10.1007/s10461-016-1476-9] [Medline: 27372803]
133. Sayles JN, Hays RD, Sarkisian CA, Mahajan AP, Spritzer KL, Cunningham WE. Development and psychometric assessment of a multidimensional measure of internalized HIV stigma in a sample of HIV-positive adults. AIDS Behav 2008 Sep;12(5):748-758 [FREE Full text] [doi: 10.1007/s10461-008-9375-3] [Medline: 18389363]
134. Kalichman SC, Simbayi LC, Cloete A, Mthembu PP, Mkhonta RN, Ginindza T. Measuring AIDS stigmas in people living with HIV/AIDS: the Internalized AIDS-Related Stigma Scale. AIDS Care 2009 Jan;21(1):87-93. [doi: 10.1080/09540120802032627] [Medline: 19085224]
135. Tsai AC, Weiser SD, Steward WT, Mukiibi NF, Kawuma A, Kembabazi A, et al. Evidence for the reliability and validity of the internalized AIDS-related stigma scale in rural Uganda. AIDS Behav 2013 Jan;17(1):427-433 [FREE Full text] [doi: 10.1007/s10461-012-0281-3] [Medline: 22869104]
136. Geibel S, Gottert A, Friedland BA, Jeremiah K, McClair TL, Mallouris C, PLHIV Stigma Index 2.0 Study Group in Cambodia, the Dominican Republic, and Uganda. Internalized stigma among people living with HIV: assessing the Internalized AIDS-Related Stigma Scale in four countries. AIDS 2020 Sep 01;34 Suppl 1:S33-S41. [doi: 10.1097/QAD.0000000000002649] [Medline: 32881792]
137. Phillips KD, Moneyham L, Tavakoli A. Development of an instrument to measure internalized stigma in those with HIV/AIDS. Issues Ment Health Nurs 2011;32(6):359-366. [doi: 10.3109/01612840.2011.575533] [Medline: 21692574]
138. Neufeld SA, Sikkema KJ, Lee RS, Kochman A, Hansen NB. The development and psychometric properties of the HIV and Abuse Related Shame Inventory (HARSI). AIDS Behav 2012 May;16(4):1063-1074 [FREE Full text] [doi: 10.1007/s10461-011-0086-9] [Medline: 22065235]
139. Zelaya CE, Sivaram S, Johnson SC, Srikrishnan AK, Suniti S, Celentano DD. Measurement of self, experienced, and perceived HIV/AIDS stigma using parallel scales in Chennai, India. AIDS Care 2012;24(7):846-855. [doi: 10.1080/09540121.2011.647674] [Medline: 22272891]
140. Earnshaw VA, Smith LR, Chaudoir SR, Amico KR, Copenhaver MM. HIV stigma mechanisms and well-being among PLWH: a test of the HIV stigma framework. AIDS Behav 2013 Jun;17(5):1785-1795 [FREE Full text] [doi: 10.1007/s10461-013-0437-9] [Medline: 23456594]
141. Molina Y, Ramirez-Valles J. HIV/AIDS stigma: measurement and relationships to psycho-behavioral factors in Latino gay/bisexual men and transgender women. AIDS Care 2013;25(12):1559-1568 [FREE Full text] [doi: 10.1080/09540121.2013.793268] [Medline: 23668809]
142. Kipp AM, Audet CM, Earnshaw VA, Owens J, McGowan CC, Wallston KA. Re-validation of the Van Rie HIV/AIDS-related stigma scale for use with people living with HIV in the United States. PLoS One 2015 Mar 4;10(3):e0118836 [FREE Full text] [doi: 10.1371/journal.pone.0118836] [Medline: 25738884]
143. Ross MW, Hunter CE, Condon J, Collins P, Begley K. The Mental Adjustment to HIV scale: measurement and dimensions of response to AIDS/HIV disease. AIDS Care 1994;6(4):407-411. [doi: 10.1080/09540129408258655] [Medline: 7833358]
144. Niu L, Qiu Y, Luo D, Chen X, Wang M, Pakenham KI, et al. Cross-culture validation of the HIV/AIDS stress scale: the development of a revised Chinese version. PLoS One 2016 Apr 4;11(4):e0152990 [FREE Full text] [doi: 10.1371/journal.pone.0152990] [Medline: 27043134]
145. Su X, Lau JT, Mak WW, Chen L, Feng T, Chen X, et al. Development of the perceived stress scale for people living with HIV/AIDS in China. AIDS Patient Care STDS 2008 Dec;22(12):989-998. [doi: 10.1089/apc.2008.0095] [Medline: 19072105]

146. Remor E, Fuster MJ, Ballester-Arnal R, Gómez-Martínez S, Fumaz CR, González-Garcia M, et al. Development of a new instrument for the assessment of psychological predictors of well-being and quality of life in people with HIV or AIDS. AIDS Behav 2012 Nov;16(8):2414-2423. [doi: 10.1007/s10461-012-0230-1] [Medline: 22692820]
147. Remor E, Fuster-RuizdeApodaca MJ, Ballester-Arnal R, Gómez-Martínez S, Fumaz CR, González-Garcia M, et al. Field psychometric testing of the instrument for assessment of psychological predictors of well-being and quality of life in people with HIV or AIDS. AIDS Behav 2016 Jun;20(6):1360-1369. [doi: 10.1007/s10461-015-1253-1] [Medline: 26584813]
148. Golub SA, Rendina HJ, Gamarel KE. Identity-related growth and loss in a sample of HIV-positive gay and bisexual men: initial scale development and psychometric evaluation. AIDS Behav 2013 Feb;17(2):748-759 [FREE Full text] [doi: 10.1007/s10461-012-0338-3] [Medline: 23086425]
149. Buscher AL, Kallen MA, Suarez-Almazor ME, Giordano TP. Development of an "Impact of HIV" instrument for HIV survivors. J Assoc Nurses AIDS Care 2015;26(6):720-731 [FREE Full text] [doi: 10.1016/j.jana.2015.08.002] [Medline: 26324524]
150. Audet CM, Wagner LJ, Wallston KA. Finding meaning in life while living with HIV: validation of a novel HIV meaningfulness scale among HIV-infected participants living in Tennessee. BMC Psychol 2015 May 2;3(1):15 [FREE Full text] [doi: 10.1186/s40359-015-0070-7] [Medline: 25945254]
151. Gottert A, Friedland B, Geibel S, Nyblade L, Baral SD, Kentutsi S, et al. The people living with HIV (PLHIV) resilience scale: development and validation in three countries in the context of the PLHIV stigma index. AIDS Behav 2019 Sep;23(Suppl 2):172-182 [FREE Full text] [doi: 10.1007/s10461-019-02594-6] [Medline: 31350712]
152. Martinez SM, Kemper CA, Diamond C, Wagner G, California Collaborative Treatment Group. Body image in patients with HIV/AIDS: assessment of a new psychometric measure and its medical correlates. AIDS Patient Care STDS 2005 Mar;19(3):150-156. [doi: 10.1089/apc.2005.19.150] [Medline: 15798382]
153. Lee D, Patel P, Sachs J, Basinger S, Mathews WC, Barber RE. Psychometric properties of a lipodystrophy scale. AIDS Patient Care STDS 2006 Jan;20(1):30-35. [doi: 10.1089/apc.2006.20.30] [Medline: 16426153]
154. Guaraldi G, Orlando G, Murri R, Vandelli M, De Paola M, Beghetto B, Esposito, et al. Quality of life and body image in the assessment of psychological impact of lipodystrophy: validation of the Italian version of assessment of body change and distress questionnaire. Qual Life Res 2006 Feb;15(1):173-178. [doi: 10.1007/s11136-005-8342-y] [Medline: 16411042]
155. Blashill AJ, Wilson JM, Baker JS, Mayer KH, Safren SA. Assessing appearance-related disturbances in HIV-infected men who have sex with men (MSM): psychometrics of the body change and distress questionnaire-short form (ABCD-SF). AIDS Behav 2014 Jun;18(6):1075-1084 [FREE Full text] [doi: 10.1007/s10461-013-0620-z] [Medline: 24057934]
156. Wu AW, Wansom T, Huang IC, CoFrancesco Jr J, Conant MA, Sarwer DB. The facial appearance inventory: development and preliminary evidence for reliability and validity in people with HIV and lipoatrophy. Aesthet Surg J 2016 Jul;36(7):842-851. [doi: 10.1093/asj/sjw010] [Medline: 26931304]
157. Aversa S, Kimberlin C, Segal R. The Medication Attribution Scale: perceived effects of antiretrovirals and quality of life. Qual Life Res 1998 Apr;7(3):205-214. [doi: 10.1023/a:1024913425993] [Medline: 9584550]
158. Woodcock A, Bradley C. Validation of the HIV treatment satisfaction questionnaire (HIVTSQ). Qual Life Res 2001;10(6):517-531. [doi: 10.1023/a:1013050904635] [Medline: 11789552]
159. Woodcock A, Bradley C. Validation of the revised 10-item HIV Treatment Satisfaction Questionnaire status version and new change version. Value Health 2006;9(5):320-333 [FREE Full text] [doi: 10.1111/j.1524-4733.2006.00121.x] [Medline: 16961550]
160. Webb DG, Horne R, Pinching AJ. Treatment-related empowerment: preliminary evaluation of a new measure in patients with advanced HIV disease. Int J STD AIDS 2001 Feb;12(2):103-107. [doi: 10.1258/0956462011916875] [Medline: 11236098]
161. Green J, Kleinman L, Ciesla G, Huang J, Wintfeld N, Revicki D. Subcutaneous injection survey: psychometric evaluation of a treatment satisfaction instrument associated with a novel HIV medication. HIV Clin Trials 2002;3(5):387-395. [doi: 10.1310/TT4M-WKFF-JTJ7-VQU6] [Medline: 12407488]
162. Hekkink CF, Sixma HJ, Wigersma L, Yzermans CJ, Van Der Meer JT, Bindels PJ, et al. QUOTE-HIV: an instrument for assessing quality of HIV care from the patients' perspective. Qual Saf Health Care 2003 Jun;12(3):188-193 [FREE Full text] [doi: 10.1136/qhc.12.3.188] [Medline: 12792008]
163. Bodenlos JS, Grothe KB, Kendra K, Whitehead D, Copeland AL, Brantley PJ. Attitudes toward HIV Health Care Providers scale: development and validation. AIDS Patient Care STDS 2004 Dec;18(12):714-720. [doi: 10.1089/apc.2004.18.714] [Medline: 15659882]
164. Holstad MM, Foster V, Diiorio C, McCarty F, Teplinskiy I. An examination of the psychometric properties of the Antiretroviral General Adherence Scale (AGAS) in two samples of HIV-infected individuals. J Assoc Nurses AIDS Care 2010;21(2):162-172 [FREE Full text] [doi: 10.1016/j.jana.2009.08.002] [Medline: 19804994]
165. Bova C, Fennie KP, Watrous E, Dieckhaus K, Williams AB. The health care relationship (HCR) trust scale: development and psychometric evaluation. Res Nurs Health 2006 Oct;29(5):477-488. [doi: 10.1002/nur.20158] [Medline: 16977644]
166. Balfour L, Tasca GA, Kowal J, Corace K, Cooper CL, Angel JB, et al. Development and validation of the HIV Medication Readiness Scale. Assessment 2007 Dec;14(4):408-416. [doi: 10.1177/1073191107304295] [Medline: 17986658]

167. Johnson MO, Neilands TB. Coping with HIV treatment side effects: conceptualization, measurement, and linkages. AIDS Behav 2007 Jul;11(4):575-585 [FREE Full text] [doi: 10.1007/s10461-007-9229-4] [Medline: 17436074]
168. Brennan DJ, Welles SL, Miner MH, Ross MW, Mayer KH, Rosser BR, Positive Connections Team. Development of a treatment optimism scale for HIV-positive gay and bisexual men. AIDS Care 2009 Sep;21(9):1090-1097 [FREE Full text] [doi: 10.1080/09540120802705859] [Medline: 20024767]
169. Erlen JA, Cha ES, Kim KH, Caruthers D, Sereika SM. The HIV Medication Taking Self-efficacy Scale: psychometric evaluation. J Adv Nurs 2010 Nov;66(11):2560-2572 [FREE Full text] [doi: 10.1111/j.1365-2648.2010.05400.x] [Medline: 20722799]
170. Osborn CY, Davis TC, Bailey SC, Wolf MS. Health literacy in the context of HIV treatment: introducing the Brief Estimate of Health Knowledge and Action (BEHKA)-HIV version. AIDS Behav 2010 Feb;14(1):181-188. [doi: 10.1007/s10461-008-9484-z] [Medline: 19023653]
171. Fernández MI, Hosek S, Warren JC, Jacobs RJ, Hernandez N, Martinez J, Adolescent Medicine Trials Network (ATN). Development of an easy to use tool to assess HIV treatment readiness in adolescent clinical care settings. AIDS Care 2011 Nov;23(11):1492-1499 [FREE Full text] [doi: 10.1080/09540121.2011.565020] [Medline: 22022853]
172. Claborn KR, Miller MB, Meier E. Initial validation of the HIV treatment regimen fatigue scale for adults prescribed antiretroviral therapy. J Assoc Nurses AIDS Care 2015;26(4):308-315 [FREE Full text] [doi: 10.1016/j.jana.2015.01.005] [Medline: 26066687]
173. Logie CH, Kennedy VL, Tharao W, Ahmed U, Loutfy MR. Engagement in and continuity of HIV care among African and Caribbean Black women living with HIV in Ontario, Canada. Int J STD AIDS 2017 Sep;28(10):969-974. [doi: 10.1177/0956462416683626] [Medline: 27956646]
174. Renwick R, Halpen T, Rudman D, Friedland J. Description and validation of a measure of received support specific to HIV. Psychol Rep 1999 Apr;84(2):663-673. [doi: 10.2466/pr0.1999.84.2.663] [Medline: 10335081]
175. Ingram KM, Jones DA, Fass RJ, Neidig JL, Song YS. Social support and unsupportive social interactions: their association with depression among people living with HIV. AIDS Care 1999 Jun;11(3):313-329. [doi: 10.1080/09540129947947] [Medline: 10474630]
176. Cortes A, Hunt N, McHale S. Development of the scale of perceived social support in HIV (PSS-HIV). AIDS Behav 2014 Dec;18(12):2274-2284. [doi: 10.1007/s10461-014-0902-0] [Medline: 25245475]
177. Johnson MO, Neilands TB, Dilworth SE, Morin SF, Remien RH, Chesney MA. The role of self-efficacy in HIV treatment adherence: validation of the HIV Treatment Adherence Self-Efficacy Scale (HIV-ASES). J Behav Med 2007 Oct;30(5):359-370 [FREE Full text] [doi: 10.1007/s10865-007-9118-3] [Medline: 17588200]
178. Wallston KA, Osborn CY, Wagner LJ, Hilker KA. The Perceived Medical Condition Self-Management Scale applied to persons with HIV/AIDS. J Health Psychol 2011 Jan;16(1):109-115 [FREE Full text] [doi: 10.1177/1359105310367832] [Medline: 20656769]
179. Webel AR, Asher A, Cuca Y, Okonsky JG, Kaihura A, Dawson Rose C, et al. Measuring HIV self-management in women living with HIV/AIDS: a psychometric evaluation study of the HIV Self-management Scale. J Acquir Immune Defic Syndr 2012 Jul 01;60(3):e72-e81 [FREE Full text] [doi: 10.1097/QAI.0b013e318256623d] [Medline: 22569267]
180. Nelsen A, Trautner BW, Petersen NJ, Gupta S, Rodriguez-Barradas M, Giordano TP, et al. Development and validation of a measure for intention to adhere to HIV treatment. AIDS Patient Care STDS 2012 Jun;26(6):329-334 [FREE Full text] [doi: 10.1089/apc.2011.0318] [Medline: 22680281]
181. Gray L, Falzon C, Bergamaschi A, Schuft L, Durant J, Rosenthal E, et al. Exercise stereotypes and health-related outcomes in French people living with HIV: development and validation of an HIV Exercise Stereotypes Scale (HIVESS). Health Qual Life Outcomes 2016 Nov 14;14(1):157 [FREE Full text] [doi: 10.1186/s12955-016-0562-z] [Medline: 27842557]
182. Webel AR, Okonsky J. Psychometric properties of a Symptom Management Self-Efficacy Scale for women living with HIV/AIDS. J Pain Symptom Manage 2011 Mar;41(3):549-557 [FREE Full text] [doi: 10.1016/j.jpainsymman.2010.05.013] [Medline: 21145198]
183. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
184. Rosca EC, Tadger P, Cornea A, Tudor R, Oancea C, Simu M. International HIV dementia scale for HIV-associated neurocognitive disorders: a systematic review and meta-analysis. Diagnostics (Basel) 2021 Jun 20;11(6):1124 [FREE Full text] [doi: 10.3390/diagnostics11061124] [Medline: 34202994]
185. Askari S, Fellows L, Brouillette MJ, Moriello C, Duracinsky M, Mayo NE. Development of an item pool reflecting cognitive concerns expressed by people with HIV. Am J Occup Ther 2018;72(2):7202205070p1-7202205070p9. [doi: 10.5014/ajot.2018.023945] [Medline: 29426385]
186. Askari S, Fellows LK, Brouillette MJ, Mayo NE. Development and validation of a voice-of-the-patient measure of cognitive concerns experienced by people living with HIV. Qual Life Res 2021 Mar;30(3):921-930. [doi: 10.1007/s11136-020-02679-z] [Medline: 33104940]
187. Nobels-Janssen E, van der Wees PJ, Verhagen WI, Westert GP, Bartels RH, Boogaarts JD. Patient-reported outcome measures in subarachnoid hemorrhage: a systematic review. Neurology 2019 Jun 04;92(23):1096-1112 [FREE Full text] [doi: 10.1212/WNL.0000000000007618] [Medline: 31076533]

188. Selby P, Velikova G. Taking patient reported outcomes centre stage in cancer research - why has it taken so long? Res Involv Engagem 2018 Jul 19;4:25 [FREE Full text] [doi: 10.1186/s40900-018-0109-z] [Medline: 30038798]
189. Wilson R. Patient led PROMs must take centre stage in cancer research. Res Involv Engagem 2018 Feb 26;4(1):7. [doi: 10.1186/s40900-018-0092-4]
190. Venda Nova C, Zakrzewska JM, Baker SR, Ni Riordain R. Patient reported outcome measures in trigeminal neuralgia - a systematic review of psychometric performance. Eur J Pain 2021 Aug;25(7):1449-1461. [doi: 10.1002/ejp.1779] [Medline: 33934425]
191. Delvin NJ, Appleby J. Getting the most out of PROMs: Putting health outcomes at the heart of NHS decision-making. The King's Fund. 2010 Mar 11. URL: http://www.kingsfund.org.uk/publications/proms.html [accessed 2021-12-18]
192. Gutiérrez-Sánchez D, Gómez-García R, Cuesta-Vargas AI, Pérez-Cruzado D. The suffering measurement instruments in palliative care: a systematic review of psychometric properties. Int J Nurs Stud 2020 Oct;110:103704. [doi: 10.1016/j.ijnurstu.2020.103704] [Medline: 32717488]
193. Recchioni A, Aiyegbusi OL, Cruz-Rivera S, Rauz S, Slade A. A systematic review assessing the quality of patient reported outcomes measures in dry eye diseases. PLoS One 2021 Aug 9;16(8):e0253857 [FREE Full text] [doi: 10.1371/journal.pone.0253857] [Medline: 34370748]
194. Prakash V, Shah S, Hariohm K. Cross-cultural adaptation of patient-reported outcome measures: a solution or a problem? Ann Phys Rehabil Med 2019 May;62(3):174-177 [FREE Full text] [doi: 10.1016/j.rehab.2019.01.006] [Medline: 30753895]
195. Lawrence V, Murray J, Banerjee S, Turner S, Sangha K, Byng R, et al. Concepts and causation of depression: a cross-cultural study of the beliefs of older adults. Gerontologist 2006 Feb;46(1):23-32. [doi: 10.1093/geront/46.1.23] [Medline: 16452281]
196. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011 Apr;17(2):268-274. [doi: 10.1111/j.1365-2753.2010.01434.x] [Medline: 20874835]
197. de Vet HC, Terwee CB. The minimal detectable change should not replace the minimal important difference. J Clin Epidemiol 2010 Jul;63(7):804-806. [doi: 10.1016/j.jclinepi.2009.12.015] [Medline: 20399609]
198. Currier JS, Havlir DV. CROI 2019: complications and coinfections in HIV infection. Top Antivir Med 2019 Apr;27(1):34-40 [FREE Full text] [Medline: 31137001]
199. Srinivasa S, Grinspoon SK. Metabolic and body composition effects of newer antiretrovirals in HIV-infected patients. Eur J Endocrinol 2014 May;170(5):R185-R202. [doi: 10.1530/EJE-13-0967] [Medline: 24523497]
200. Lampe FC, Harding R, Smith CJ, Phillips AN, Johnson M, Sherr L. Physical and psychological symptoms and risk of virologic rebound among patients with virologic suppression on antiretroviral therapy. J Acquir Immune Defic Syndr 2010 Aug;54(5):500-505. [doi: 10.1097/QAI.0b013e3181ce6afe] [Medline: 20150819]
201. Kjær AS, Rasmussen TA, Hjollund NH, Rodkjaer LO, Storgaard M. Patient-reported outcomes in daily clinical practise in HIV outpatient care. Int J Infect Dis 2018 Apr;69:108-114 [FREE Full text] [doi: 10.1016/j.ijid.2018.02.015] [Medline: 29476900]
202. Monroe AK, Jabour SM, Peña S, Keruly JC, Moore RD, Chander G, et al. A qualitative study examining the benefits and challenges of incorporating patient-reported outcome substance use and mental health questionnaires into clinical practice to improve outcomes on the HIV care continuum. BMC Health Serv Res 2018 Jun 07;18(1):419 [FREE Full text] [doi: 10.1186/s12913-018-3203-x] [Medline: 29879962]
203. Rosen S, Maskew M, Brennan AT, Fox MP, Vezi L, Ehrenkranz PD, et al. Improved simplified clinical algorithm for identifying patients eligible for immediate initiation of antiretroviral therapy for HIV (SLATE II): protocol for a randomized evaluation. Trials 2018 Oct 11;19(1):548 [FREE Full text] [doi: 10.1186/s13063-018-2928-5] [Medline: 30305142]
204. Lowther K, Selman L, Simms V, Gikaara N, Ahmed A, Ali Z, et al. Nurse-led palliative care for HIV-positive patients taking antiretroviral therapy in Kenya: a randomised controlled trial. Lancet HIV 2015 Aug;2(8):e328-e334. [doi: 10.1016/S2352-3018(15)00111-3] [Medline: 26423375]
205. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health 2007;10 Suppl 2:S125-S137 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00275.x] [Medline: 17995471]
206. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 2013 Feb 27;309(8):814-822. [doi: 10.1001/jama.2013.879] [Medline: 23443445]
207. Vitoria M, Rangaraj A, Ford N, Doherty M. Current and future priorities for the development of optimal HIV drugs. Curr Opin HIV AIDS 2019 Mar;14(2):143-149. [doi: 10.1097/COH.0000000000000527] [Medline: 30562177]
208. Simpson KN, Hanson KA, Harding G, Haider S, Tawadrous M, Khachatryan A, et al. Patient reported outcome instruments used in clinical trials of HIV-infected adults on NNRTI-based therapy: a 10-year review. Health Qual Life Outcomes 2013 Oct 03;11:164 [FREE Full text] [doi: 10.1186/1477-7525-11-164] [Medline: 24090055]
209. Marshall DA, Jin X, Pittman LB, Smith CJ. The use of patient-reported outcome measures in hip and knee arthroplasty in Alberta. J Patient Rep Outcomes 2021 Oct 12;5(Suppl 2):87 [FREE Full text] [doi: 10.1186/s41687-021-00362-6] [Medline: 34636973]

210. Sawatzky R, Kwon JY, Barclay R, Chauhan C, Frank L, van den Hout WB, Response Shift – in Sync Working Group. Implications of response shift for micro-, meso-, and macro-level healthcare decision-making using results of patient-reported outcome measures. Qual Life Res 2021 Dec;30(12):3343-3357 [FREE Full text] [doi: 10.1007/s11136-021-02766-9] [Medline: 33651278]
211. Weldring T, Smith SM. Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). Health Serv Insights 2013 Aug 4;6:61-68 [FREE Full text] [doi: 10.4137/HSI.S11093] [Medline: 25114561]
212. Antela A, Rivero A, Llibre JM, Moreno S, RET Group. Redefining therapeutic success in HIV patients: an expert view. J Antimicrob Chemother 2021 Sep 15;76(10):2501-2518 [FREE Full text] [doi: 10.1093/jac/dkab168] [Medline: 34077524]
213. Althoff KN, Smit M, Reiss P, Justice AC. HIV and ageing: improving quantity and quality of life. Curr Opin HIV AIDS 2016 Sep;11(5):527-536 [FREE Full text] [doi: 10.1097/COH.0000000000000305] [Medline: 27367780]
214. King EM, Prior JC, Pick N, van Schalkwyk J, Kestler M, Tkachuk S, et al. Menopausal hormone therapy for women living with HIV. Lancet HIV 2021 Sep;8(9):e591-e598. [doi: 10.1016/S2352-3018(21)00148-X] [Medline: 34384545]
215. Brown MJ, Adeagbo O. HIV and aging: double stigma. Curr Epidemiol Rep 2021;8(2):72-78 [FREE Full text] [doi: 10.1007/s40471-021-00265-6] [Medline: 33728256]
216. Cabrera DM, Diaz MM, Grimshaw A, Salvatierra J, Garcia PJ, Hsieh E. Aging with HIV in Latin America and the Caribbean: a systematic review. Curr HIV/AIDS Rep 2021 Feb;18(1):1-47 [FREE Full text] [doi: 10.1007/s11904-020-00538-7] [Medline: 33400168]
217. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes 2015 Oct 07;13:167 [FREE Full text] [doi: 10.1186/s12955-015-0362-x] [Medline: 26446159]
218. Byrom B, Doll H, Muehlhausen W, Flood E, Cassedy C, McDowell B, et al. Measurement equivalence of patient-reported outcome measure response scale types collected using bring your own device compared to paper and a provisioned device: results of a randomized equivalence trial. Value Health 2018 May;21(5):581-589 [FREE Full text] [doi: 10.1016/j.jval.2017.10.008] [Medline: 29753356]
219. Edwards TC, Fredericksen RJ, Crane HM, Crane PK, Kitahata MM, Mathews WC, et al. Content validity of patient-reported outcomes measurement information system (PROMIS) items in the context of HIV clinical care. Qual Life Res 2016 Feb;25(2):293-302 [FREE Full text] [doi: 10.1007/s11136-015-1096-2] [Medline: 26245710]
220. Gibbons LE, Fredericksen R, Merrill JO, McCaul ME, Chander G, Hutton H, et al. Suitability of the PROMIS alcohol use short form for screening in a HIV clinical care setting. Drug Alcohol Depend 2016 Jul 01;164:113-119 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.04.038] [Medline: 27209223]
221. Gibbons LE, Fredericksen R, Batey DS, Dant L, Edwards TC, Mayer KH, Centers for AIDS Research Network of Integrated Clinical Systems (CNICS). Validity assessment of the PROMIS fatigue domain among people living with HIV. AIDS Res Ther 2017 Apr 11;14:21 [FREE Full text] [doi: 10.1186/s12981-017-0146-y] [Medline: 28400850]

Abbreviations

ART: antiretroviral therapy
CAT: computer adaptive test
COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments
ePROM: electronic patient-reported outcome measure
HIV-SI or SDM: HIV Symptom Index or Symptoms Distress Module of the Adult AIDS Clinical Trial Group
HRQoL: health-related quality of life
PLHIV-RS: People Living with HIV Resilience Scale
PozQoL: Poz Quality of Life
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO: patient-reported outcome
PROM: patient-reported outcome measure

Edited by H Bradley; submitted 25.04.22; peer-reviewed by D Finitsis, MJ Brouillette, A Olekalns; comments to author 26.06.22; revised version received 03.08.22; accepted 11.10.22; published 08.12.22.

Please cite as:
Wang Z, Zhu Y, Duan X, Kang H, Qu B
HIV-Specific Reported Outcome Measures: Systematic Review of Psychometric Properties
JMIR Public Health Surveill 2022;8(12):e39015
URL: https://publichealth.jmir.org/2022/12/e39015
doi:10.2196/39015
PMID:36222289

Original Paper

The Use of HIV Pre- and Postexposure Prophylaxis Among a Web-Based Sample of HIV-Negative and Unknown Status Cisgender and Transgender Sexual Minority Men: Cross-sectional Study

Steven A John[1], MPH, PhD; K Marie Sizemore[2], PhD; Ruben H Jimenez[2], BA; S Scott Jones[2], MS; Andrew E Petroll[1], MD, MS; H Jonathon Rendina[2], MPH, PhD

[1]Health Intervention Sciences Group / Center for AIDS Intervention Research, Department of Psychiatry and Behavioral Medicine, Medical College of Wisconsin, Milwaukee, WI, United States

[2]Department of Psychology, Hunter College of the City University of New York, New York, NY, United States

Corresponding Author:
Steven A John, MPH, PhD
Health Intervention Sciences Group / Center for AIDS Intervention Research
Department of Psychiatry and Behavioral Medicine
Medical College of Wisconsin
2071 N Summit Ave
Milwaukee, WI, 53202
United States
Phone: 1 414 955 7744
Email: sjohn@mcw.edu

Abstract

Background: HIV disproportionately affects sexual minority men (SMM) in the United States.

Objective: We sought to determine past HIV postexposure prophylaxis (PEP) use and current and prior pre-exposure prophylaxis (PrEP) use among a web-based sample of cisgender and transgender men who have sex with men.

Methods: In 2019, HIV-negative and unknown status SMM (n=63,015) were recruited via geosocial networking apps, social media, and other web-based venues to participate in a brief eligibility screening survey. Individuals were asked about past PEP use and current and prior PrEP use. We examined associations of demographics, socioeconomic indicators, and recent club drug use with PEP and PrEP use, as well as the association between past PEP use and current and prior PrEP use using generalized linear models and multinomial logistic regression. Statistical significance was considered at $P<.001$, given the large sample size; 99.9% CIs are reported.

Results: Prior PEP use was reported by 11.28% (7108/63,015) of the participants, with current or prior PrEP use reported by 21.95% (13,832/63,015) and 8.12% (5118/63,015), respectively. Nearly half (3268/7108, 46%) of the past PEP users were current PrEP users, and another 39.9% (2836/7108) of the participants who reported past PEP use also reported prior PrEP use. In multivariable analysis, past PEP use was associated with current (relative risk ratio [RRR] 23.53, 99.9% CI 14.03-39.46) and prior PrEP use (RRR 52.14, 99.9% CI 29.39-92.50). Compared with White men, Black men had higher prevalence of past PEP use and current PrEP use, Latino men had higher prevalence of PEP use but no significant difference in PrEP use, and those identifying as another race or ethnicity reported higher prevalence of past PEP use and lower current PrEP use. Past PEP use and current PrEP use were highest in the Northeast, with participants in the Midwest and South reporting significantly lower PEP and PrEP use. A significant interaction of Black race by past PEP use with current PrEP use was found (RRR 0.57, 99.9% CI 0.37-0.87), indicating that Black men who previously used PEP were less likely to report current PrEP use. Participants who reported recent club drug use were significantly more likely to report past PEP use and current or prior PrEP use than those without recent club drug use.

Conclusions: PrEP use continues to be the predominant HIV prevention strategy for SMM compared with PEP use. Higher rates of past PEP use and current PrEP use among Black SMM are noteworthy, given the disproportionate burden of HIV. Nonetheless, understanding why Black men who previously used PEP are less likely to report current PrEP use is an important avenue for future research.

(JMIR Public Health Surveill 2022;8(12):e31237) doi:10.2196/31237

KEYWORDS

HIV; pre-exposure prophylaxis; postexposure prophylaxis; sexual minority men; men who have sex with men

Introduction

Background

HIV disproportionately affects cisgender and transgender men who have sex with men—referred to herein as sexual minority men (SMM)—in the United States [1,2]. Despite decreasing HIV incidence nationally, cisgender SMM accounted for 68% of the sexually transmitted HIV incidence in 2019 [1]. HIV prevalence is estimated at 3.2% among transgender men, and 55.2% of the HIV-negative or unknown status transgender SMM could benefit from biomedical HIV prevention [3]. In 2019, the *Ending the HIV Epidemic: A Plan for America* was announced, with priorities that include expanding biomedical HIV prevention [4]. HIV pre-exposure prophylaxis (PrEP) is a method of biomedical HIV prevention that includes taking antiretrovirals once daily—as approved by the US Food and Drug Administration [5] with supporting guidelines from the Centers for Disease Control and Prevention [6]—or alternative dosing strategies (eg, 2-1-1) found to be highly effective in preventing HIV [7-10]. Nonetheless, engagement in anal sex can be unanticipated—both consensual and nonconsensual—and alternative options are needed after such encounters with an HIV-positive or unknown status partner.

Nonoccupational HIV postexposure prophylaxis (PEP) is a highly effective method of secondary HIV prevention [11-17] and can be administered within 72 hours after exposure or potential exposure to HIV. PEP, a 28-day strategy that includes taking a 3- or 4-drug regimen of HIV antiretrovirals after exposure [18,19], has been recommended as a strategy for HIV prevention by the Centers for Disease Control and Prevention since 2005 [20]. By contrast, PrEP is a 2-drug combination using emtricitabine with either tenofovir disoproxil fumarate or tenofovir alafenamide [21]. When indicated, individuals who complete the 28-day PEP regimen are then recommended to transition immediately to PrEP [6,22]. Strategies to support successful PEP-to-PrEP transition are beginning to be implemented in clinical settings, with initial data indicating high success [23].

Study Hypothesis

Many individuals who use PEP after potential sexual exposure to HIV are appropriate candidates to initiate PrEP upon completion of PEP [6,22], but research is limited on PEP uptake among SMM. As such, we sought to determine lifetime use of PEP among a large nationwide sample of SMM recruited on the internet. We hypothesized that SMM who had prior experience with PEP would report higher rates of PrEP use than SMM who had not used PEP. Given the dearth of data on PEP use among SMM, we also explored prior PEP use by demographic characteristics, health insurance status, socioeconomic status, and club drug use; in addition, we examined the effect of these factors on PrEP uptake.

Methods

Participants and Procedures

Participants were recruited via geosocial networking apps, social media, and other web-based venues targeting SMM between January 1, 2019, and December 31, 2019, to participate in a brief (5-10 minutes) screening survey used to determine eligibility for multiple paid research studies. Only individuals aged ≥13 years were eligible to take the screening survey. To be eligible for this analysis, participants were required to (1) identify as male (inclusive of transgender men), (2) report a male sexual partner in the past 6 months or a main partner who identified as male, (3) self-report HIV-negative or unknown status, and (4) reside in the United States, including Puerto Rico and other territories. On the basis of the recruitment procedures, advertisements, and venues targeted to men, women (inclusive of transgender women) were excluded from the analysis. Cisgender and transgender SMM were the focus of our analysis, given the disproportionate burden of HIV incidence in the United States [1,2]. All adolescent SMM were included in this analysis; children and adolescents are included in current PEP guidelines based on supporting safety data collected among young people [18], and PrEP is approved for use among minors weighing ≥35 kg [24]. Fraudulent responses were minimized by excluding any information of eligibility criteria in study advertisements and referral mechanisms and offering no incentive for completion of this screening survey. Potential duplicate responses were identified by corresponding birth month and year, zip code, HIV status, race, and ethnicity. Flagged cases were further screened by examining other demographic variables and metadata (eg, device and browser information) before being considered for removal, as recommended previously [25].

Ethical Considerations

An alteration of informed consent and assent was approved for this study, wherein participants agreed to participate after reading an informational letter describing the study procedures, risks, and benefits; parental permission was waived for all minors. No incentive was provided for participation in this screening survey. Surveys were conducted using Qualtrics, which provides Health Insurance Portability and Accountability Act privacy protection standards, and contact information was collected separately from survey data to reduce the risk harm in the case of loss of confidentiality. All study procedures were approved by the institutional review board of the City University of New York (319487).

Measures

Participants were asked to report their age, sexual orientation, gender, race and ethnicity, health insurance status, and location of residence in the United States. Age was categorized for analysis using thresholds used in the US HIV Surveillance Report [1]. Gender was determined using a 2-step approach:

participants indicated their sex assigned at birth with male and female response categories, and current gender identity was indicated by their response to the question "What is your current gender identity?" The response categories were male, female, and transgender. We regret the exclusion of additional gender identities in our response options, including but not limited to genderqueer, nonbinary, and 2 spirit. Individuals who reported being assigned female sex at birth and currently identified as male or transgender were coded as transgender men. Individuals were asked to indicate their race and ethnicity, and participants in the multiracial category either indicated >1 race or selected a multiracial category. Participants were also asked about their perceived socioeconomic status using the MacArthur Scale of Subjective Social Status [26], which measures participants' perceived socioeconomic rank compared with others, with 1 being the lowest and 10 being the highest. Individuals were coded as having used club drugs if they reported using any of the following substances in the past 90 days: crack and cocaine; 3,4-methylenedioxy-methamphetamine (MMDA); gamma-hydroxybutyrate (GHB); ketamine; or methamphetamine [27]. Participants were provided a brief introduction regarding PrEP and asked the following question about PrEP use: "Have you ever been prescribed HIV medications (e.g., Truvada) for use as PrEP (pre-exposure prophylaxis)?" The response options included (1) *Yes, I am currently on PrEP*; (2) *Yes, but I am no longer taking PrEP*; and (3) *No, I've never taken PrEP* [28,29]. Similarly, participants were provided a brief introduction regarding PEP and asked the following question: "Have you ever been prescribed PEP?" The response options included (1) *Yes, within the past 6 months*; (2) *Yes, more than 6 months ago*; and (3) *No, never*. PEP use was coded into past lifetime use (*yes* or *no*).

Data Analysis

Descriptive statistics were reported using frequency measures. For the past PEP use outcome, bivariate analyses were conducted using generalized linear models with log link function and Poisson distribution to produce prevalence ratios. We then examined associations between demographics and club drug use on ever using PEP using fully adjusted generalized linear models with log link function and Poisson distribution. For the current and prior PrEP use outcomes, bivariate analyses were conducted using multinomial logistic regression, which produced relative risk ratios (RRRs). We then examined associations among demographics, club drug use, and past PEP use with current and prior PrEP use using fully adjusted multinomial logistic regression; *never used PrEP* was the referent in the past PrEP use multinomial model. We removed insurance status and the socioeconomic status score from all adjusted models to reduce overadjustment bias [30] because of their role as hypothesized intermediate variables in the causal pathways between race and ethnicity (via racism) and PEP or PrEP use; insurance status and socioeconomic status score were thus removed to improve theoretical model precision. Interactions between race and ethnicity and past PEP use with PrEP use were explored by adding two interaction terms to the PrEP models: (1) Black, non-Hispanic×past PEP use and (2) Latino or Hispanic×past PEP use. Statistical significance was tested at α=.001 because of the large sample size, and unadjusted and adjusted prevalence ratios are reported with 99.9% CIs.

Results

Participant Characteristics

Recruitment activities resulted in 160,581 unique link clicks, with 120,274 (74.9%) participants agreeing to participate in the survey. Among those who agreed, 76.1% (91,526/120,274) completed the survey or provided data sufficient for analysis. Of these, 3.87% (3538/91,526) were ineligible by gender, 10.94% (10,011/91,526) did not report a recent male sexual partner or a main partner who identified as male, and 19.91% (18,219/91,526) self-reported living with HIV; individuals could be considered ineligible by ≥1 of the criteria. Thus, of the 91,526 SMM who agreed to participate and provided data sufficient for analysis, 63,015 (68.85%) were eligible for this analysis. The average age of respondents was 33.1 (SD 12.0) years (median 30, range 13-80; Table 1). Most of the participants identified as gay (45,251/63,015, 71.81%) or bisexual (15,129/63,015, 24%), and nearly all (62,446/63,015, 99.1%) identified as cisgender men. Past PEP use was reported by 11.28% (7108/63,015) of the participants, and 21.95% (13,832/63,015) and 8.12% (5118/63,015) reported current and prior PrEP use, respectively. Nearly half (3268/7108, 46%) of the past PEP users were current PrEP users, and another 39.9% (2836/7108) of the participants who reported past PEP use also reported prior PrEP use. Refer to Tables 1 and 2 for full sample characteristics.

In bivariate analyses, significant differences in past PEP use prevalence were found by age, sexual orientation, US region, race and ethnicity, health insurance status, and recent club drug use (Table 1). In addition, significant differences in PrEP uptake were found by age, sexual orientation, gender, US region, race and ethnicity, health insurance status, recent club drug use, and past PEP use (Table 2). Socioeconomic status was significant in both models, but effect sizes did not indicate a meaningful effect (Tables 1 and 2).

Table 1. Demographics, socioeconomic status indicators, club drug use, and current use of pre-exposure prophylaxis (PrEP) and their bivariate associations with previous postexposure prophylaxis (PEP) use (N=63,015).

	Values	Past PEP use		
		Values (n=7108)	Prevalence ratio (99.9% CI)	*P* value
Categorical variables, n (%)[a]				
Age (years; mean 33.12, SD 11.86; median 30, range 13-80)				
13 to 24	16,641 (26.41)	1160 (16.32)	N/A[b]	N/A
25 to 34	23,432 (37.18)	3200 (45.02)	*1.96*[c] (1.76-2.18)	<.001
35 to 44	11,502 (18.25)	1609 (22.64)	*2.01* (1.78-2.26)	<.001
≥45	11,440 (18.15)	1139 (16.05)	*1.43* (1.25-1.63)	<.001
Sexual orientation identity				
Gay	45,251 (71.81)	5875 (82.65)	N/A	N/A
Bisexual	15,129 (24)	929 (13.07)	*0.47* (0.42-0.53)	<.001
Queer	1758 (2.79)	266 (3.74)	*1.17* (0.96-1.41)	.008
Straight	877 (1.39)	38 (0.53)	*0.33* (0.20-0.56)	<.001
Gender				
Cisgender man	62,446 (99.1)	7041 (99.06)	N/A	N/A
Transgender man	569 (0.9)	67 (0.94)	1.04 (0.72-1.53)	.71
Region				
Northeast	12,823 (20.35)	1867 (26.27)	N/A	N/A
Midwest	11,359 (18.03)	1071 (15.07)	*0.65* (0.58-0.73)	<.001
South	21,087 (33.46)	1936 (27.24)	*0.63* (0.57-0.70)	<.001
West	17,418 (27.64)	2209 (31.08)	*0.87* (0.79-0.96)	<.001
US possession	255 (0.4)	16 (0.23)	*0.43* (0.19-0.96)	<.001
Military overseas	29 (0.05)	3 (0.04)	0.71 (0.12-4.30)	.53
Unknown	44 (0.07)	6 (0.08)	0.94 (0.27-3.27)	.86
Race and ethnicity				
Black, non-Hispanic	6628 (10.52)	774 (10.89)	*1.17* (1.04-1.33)	<.001
Latino or Hispanic	11,092 (17.6)	1474 (20.74)	*1.34* (1.21-1.47)	<.001
Multiracial	6385 (10.13)	909 (12.79)	0.61 (0.23-1.58)	.09
White, non-Hispanic	35,046 (55.61)	3485 (49.03)	N/A	N/A
Another	3864 (6.13)	466 (6.56)	*1.36* (1.23-1.50)	<.001
Health insurance status				
Has private health insurance	39,071 (62)	4352 (61.23)	*1.12* (1.01-1.23)	<.001
Has public health insurance (eg, Medicaid)	11,151 (17.7)	1481 (20.84)	*1.33* (1.18-1.50)	<.001
Uninsured	12,793 (20.3)	1257 (17.68)	N/A	N/A
Any club drug use (past 90 days)[d]				
No	50,411 (80)	5222 (73.47)	N/A	N/A
Yes	12,604 (20)	1886 (26.53)	*1.45* (1.33-1.57)	<.001
PrEP use status				
Never used	44,065 (69.93)	1003 (14.11)	N/A	N/A
Prior use	5118 (8.12)	2837 (39.91)	*24.35* (21.80-27.20)	<.001
Current use	13,832 (21.95)	3268 (45.98)	*10.38* (9.26-11.64)	<.001

	Values	Past PEP use		
		Values (n=7108)	Prevalence ratio (99.9% CI)	*P* value
Continuous variable, mean (SD)				
Socioeconomic status ladder (range 1-10)	6.66 (25.33)	7.69 (38.93)	*1.00* (1.00-1.00)	<.001

[a]Percentages may not add up to 100 because of rounding.

[b]N/A: not applicable.

[c]Italicized prevalence ratio values are significant at *P*<.001.

[d]Club drugs include crack and cocaine; 3,4-methylenedioxy-methamphetamine (MMDA); gamma-hydroxybutyrate (GHB); ketamine; and methamphetamine.

Table 2. Demographics, socioeconomic status indicators, club drug use, and past postexposure prophylaxis (PEP) use and their bivariate associations with current and prior pre-exposure prophylaxis (PrEP) use compared with never used PrEP (N=63,015).

	Current PrEP use			Prior PrEP use		
	Values (n=13,832)	RRR[a] (99.9% CI)	*P* value	Values (n=5118)	RRR (99.9% CI)	*P* value
Categorical variables, n (%)[b]						
Age (years)						
13 to 24	1911 (13.82)	N/A[c]	N/A	959 (18.74)	N/A	N/A
25 to 34	5807 (41.98)	*2.75*[d] (2.50-3.03)	<.001	2414 (47.17)	*2.28* (2.00-2.60)	<.001
35 to 44	3273 (23.66)	*3.28* (2.95-3.65)	<.001	1044 (20.4)	*2.09* (1.79-2.44)	<.001
≥45	2841 (20.54)	*2.59* (2.33-2.89)	<.001	701 (13.7)	*1.23* (1.08-1.51)	<.001
Sexual orientation identity						
Gay	11,792 (85.25)	N/A	N/A	4178 (81.63)	N/A	N/A
Bisexual	1533 (11.08)	*0.30* (0.27-0.33)	<.001	711 (13.89)	*0.39* (0.34-0.44)	<.001
Queer	479 (3.46)	1.11 (0.93-1.34)	.06	211 (4.12)	1.39 (1.07-1.79)	<.001
Straight	28 (0.2)	*0.08* (0.04-0.16)	<.001	18 (0.35)	*0.15* (0.07-0.33)	<.001
Gender						
Cisgender man	13,748 (99.39)	N/A	N/A	5069 (99.04)	N/A	N/A
Transgender man	84 (0.61)	*0.61* (0.41-0.91)	<.001	49 (0.96)	0.97 (0.59-1.59)	.83
Region						
Northeast	3294 (23.81)	N/A	N/A	1562 (30.52)	N/A	N/A
Midwest	2307 (16.68)	*0.71* (0.64-0.78)	<.001	1240 (24.23)	*0.68* (0.58-0.79)	<.001
South	4054 (29.31)	*0.66* (0.60-0.72)	<.001	833 (16.28)	*0.63* (0.55-0.72)	<.001
West	4131 (29.87)	*0.89* (0.81-0.97)	<.001	1466 (28.64)	0.89 (0.78-1.02)	.004
US possession	35 (0.25)	*0.42* (0.23-0.77)	<.001	10 (0.2)	*0.32* (0.11-0.93)	<.001
Military overseas	6 (0.04)	0.76 (0.16-3.50)	.55	3 (0.06)	1.00 (0.13-7.71)	.99
Unknown	5 (0.04)	0.36 (0.07-1.74)	.03	4 (0.08)	0.76 (0.13-4.35)	.61
Race and ethnicity						
Black, non-Hispanic	1311 (9.48)	*0.84* (0.75-0.94)	<.001	540 (8.1)	1.07 (0.90-1.26)	.21
Latino or Hispanic	2359 (17.05)	0.94 (0.86-1.03)	.02	1050 (9.5)	*1.29* (1.13-1.46)	<.001
Multiracial	1329 (9.61)	*0.21* (0.06-0.71)	<.001	630 (9.9)	0.25 (0.05-1.33)	.006
White, non-Hispanic	7996 (57.81)	N/A	N/A	2596 (7.4)	N/A	N/A
Another	837 (6.05)	0.94 (0.86-1.03)	.02	302 (7.8)	*1.24* (1.09-1.42)	<.001
Health insurance status						
Has private health insurance	10,118 (73.15)	*2.79* (2.52-3.08)	<.001	2932 (7.5)	0.96 (0.85-1.08)	.28
Has public health insurance (eg, Medicaid)	2293 (16.58)	*2.09* (1.85-2.36)	<.001	993 (8.9)	1.08 (0.93-1.25)	.10
Uninsured	1421 (10.27)	N/A	N/A	1193 (9.3)	N/A	N/A
Any club drug use (past 90 days)[e]						
No	10,512 (76)	N/A	N/A	3719 (7.4)	N/A	N/A
Yes	3320 (24)	*1.45* (1.34-1.57)	<.001	1399 (11.1)	*1.73* (1.55-1.93)	<.001
Past PEP use						
No	10,564 (76.37)	N/A	N/A	2281 (4.1)	N/A	N/A
Yes	3268 (23.63)	*13.28* (11.73-15.04)	<.001	2837 (39.9)	*53.40* (46.42-61.43)	<.001

	Current PrEP use			Prior PrEP use		
	Values (n=13,832)	RRR[a] (99.9% CI)	*P* value	Values (n=5118)	RRR (99.9% CI)	*P* value
Continuous variable, mean (SD)						
Socioeconomic status ladder (range 1-10)	7.47 (32.65)	*1.001* (1.000-1.003)	<.001	6.90 (31.02)	*1.001* (0.999-1.003)	.09

[a]RRR: relative risk ratio.

[b]Percentages may not add up to 100 because of rounding.

[c]N/A: not applicable.

[d]Italicized relative risk ratio values are significant at *P*<.001.

[e]Club drugs include crack and cocaine; 3,4-methylenedioxy-methamphetamine (MMDA); gamma-hydroxybutyrate (GHB); ketamine; and methamphetamine.

Multivariable Analyses

In multivariable analyses (Tables 2 and 3), past PEP use was associated with current (RRR 23.53, 99.9% CI 14.03-39.46) and prior PrEP use (RRR 52.14, 99.9% CI 29.39-92.50). Compared with White men, Black men had higher prevalence of past PEP use and current PrEP use, Latino men had higher prevalence of PEP use but no significant difference in PrEP use, and those identifying as another race or ethnicity reported higher prevalence of past PEP use and lower current PrEP use. Compared with White men, multiracial men had no significant difference in PEP or PrEP use. Past PEP use and current PrEP use were highest in the Northeast, with participants in the Midwest and South reporting significantly lower PEP and PrEP use. Men living in the West had significantly lower prevalence of past PEP use compared with men in the Northeast, but no significant difference in PrEP use was observed. Individuals living in a US possession also had significantly lower prevalence of past PEP use, as well as lower likelihood of current PrEP use. A significant interaction of Black race by past PEP use with current PrEP use was found (RRR 0.57, 99.9% CI 0.37-0.87), indicating that Black men who previously used PEP were less likely to report current PrEP use. Participants who reported recent club drug use were significantly more likely to report past PEP use and current or prior PrEP use than those without recent use. Refer to Tables 2 and 3 for full multivariable results.

Table 3. Results from generalized linear models with log link function and Poisson distribution predicting past postexposure prophylaxis (PEP) use and multinomial logistic regression comparing current and prior pre-exposure prophylaxis (PrEP) use with never used PrEP (N=63,015).

Categorical variables	Past PEP use (referent: never used)		PrEP use (referent: never used)			
	PR[a] (99.9% CI)	*P* value	Current use		Prior use	
			RRR[b] (99.9% CI)	*P* value	RRR (99.9% CI)	*P* value
Age (years)						
13 to 24	N/A[c]	N/A	N/A	N/A	N/A	N/A
25 to 34	*1.96*[d] (1.74-2.22)	<.001	*2.43* (2.20-2.69)	<.001	*1.23* (1.02-1.50)	<.001
35 to 44	*2.09* (1.83-2.40)	<.001	*2.91* (2.60-3.26)	<.001	*1.72* (1.44-2.06)	<.001
≥45	*1.59* (1.37-1.85)	<.001	*2.46* (2.19-2.76)	<.001	*1.88* (1.61-2.18)	<.001
Sexual orientation identity						
Gay	N/A	N/A	N/A	N/A	N/A	N/A
Bisexual	*0.46* (0.41-0.52)	<.001	*0.32* (0.29-3.36)	<.001	*0.45* (0.38-0.52)	<.001
Queer	1.13 (0.89-1.43)	.10	*1.22* (0.99-1.49)	.001	1.34 (0.99-1.82)	.001
Straight	*0.30* (0.17-0.52)	<.001	*0.08* (0.04-0.16)	<.001	*0.15* (0.06-0.35)	<.001
Gender						
Cisgender man	N/A	N/A	N/A	N/A	N/A	N/A
Transgender man	1.19 (0.75-1.89)	.21	0.71 (0.45-1.10)	.009	0.85 (0.46-1.57)	.39
Region						
Northeast	N/A	N/A	N/A	N/A	N/A	N/A
Midwest	*0.66* (0.58-0.76)	<.001	*0.78* (0.70-0.88)	<.001	0.85 (0.71-1.01)	.002
South	*0.61* (0.54-0.69)	<.001	*0.72* (0.66-0.79)	<.001	*0.77* (0.66-0.90)	<.001
West	*0.81* (0.72-0.91)	<.001	0.92 (0.83-1.02)	.006	0.93 (0.80-1.09)	.14
US possession	*0.35* (0.15-0.82)	<.001	*0.45* (0.24-0.86)	<.001	0.38 (0.12-1.23)	.007
Military overseas	0.70 (0.09-5.21)	.56	0.91 (0.18-4.72)	.85	1.36 (0.13-14.60)	.67
Unknown	1.14 (0.25-5.09)	.78	0.40 (0.07-2.18)	.07	0.75 (0.09-6.12)	.65
Race and ethnicity						
Black, non-Hispanic	*1.36* (1.18-1.57)	<.001	*1.60* (1.06-2.42)	<.001	1.29 (0.85-1.97)	.002
Latino or Hispanic	*1.41* (1.26-1.58)	<.001	1.08 (0.79-1.48)	.41	1.07 (0.78-1.47)	.49
Multiracial	1.10 (0.37-3.21)	.78	0.47 (0.14-1.64)	.47	0.37 (0.06-2.28)	.07
White, non-Hispanic	N/A	N/A	N/A	N/A	N/A	N/A
Another	*1.38* (1.23-1.56)	<.001	*0.90* (0.81-1.00)	<.001	0.98 (0.83-1.14)	.60
Any club drug use (past 90 days)[e]						
No	N/A	N/A	N/A	N/A	N/A	N/A
Yes	*1.40* (1.27-1.55)	<.001	*1.29* (1.18-1.40)	<.001	*1.47* (1.29-1.67)	<.001
Past PEP use						
No	N/A	N/A	N/A	N/A	N/A	N/A
Yes	N/A	N/A	*23.53* (14.03-39.46)	<.001	*52.14* (29.39-92.50)	<.001
Black, non-Hispanic×past PEP use	N/A	N/A	*0.57* (0.37-0.87)	<.001	0.87 (0.54-1.40)	.34
Latino or Hispanic×past PEP use	N/A	N/A	0.84 (0.61-1.16)	.07	1.09 (0.76-1.56)	.44

[a]PR*:* prevalence ratio.

[b]RRR*:* relative risk ratio.

[c]N/A: not applicable.

[d]Italicized prevalence ratio and relative risk ratio values are significant at $P<.001$.

[e]Club drugs include crack and cocaine; 3,4-methylenedioxy-methamphetamine (MMDA); gamma-hydroxybutyrate (GHB); ketamine; and methamphetamine.

Discussion

Principal Findings

We sought to determine lifetime use of PEP among SMM and hypothesized that SMM who had prior experience with PEP would report higher rates of PrEP use than SMM who had not used PEP. Specifically, we found that 11.28% (7108/63,015) of the participants reported past PEP use, but PrEP use was the more commonly used method of HIV prevention. As hypothesized, we found that men with a history of PEP use were more likely to report current PrEP use. When considering both current and prior PrEP use, 85.9% (6106/7108) of the past PEP users had also used PrEP currently or previously. As such, PEP use could be a gateway to PrEP use as a PEP-to-PrEP pathway to biomedical HIV prevention, supported by current guidelines and recommendations [6,22] as well as current PrEP and PEP implementation strategies [23].

We find it plausible that individuals who have previous experience taking antiretrovirals for PEP could have fewer barriers to taking PrEP. SMM frequently cite concerns about potential side effects with taking PrEP [31,32], concerns that could potentially diminish after the experience of taking PEP. Moreover, PEP is frequently obtained in urgent scenarios, given the short time interval to initiation, offering a cue to action for ongoing HIV prevention. PEP users are also most often put in contact with providers who could become their prescribers of PrEP. Further research is needed to explore the PEP-to-PrEP pathway to biomedical HIV prevention, including reasons for uptake of, or declining, PrEP, but our findings illustrate that nearly half (3268/7108, 46%) of the past PEP users are currently taking PrEP. Moreover, our findings about lifetime PEP uptake are higher than prior reports of PEP use more broadly, where a pooled estimate of PEP use was 5.8% in high-income countries in a systematic review [33]; yet, our nationwide findings find concordance with increasing uptake over time, including similar rates of PEP use reported among young SMM (ie, 11.5% [34]) and young SMM of color (ie, 15.3% [35]) in New York City—a high-resource area for HIV prevention.

Although our cross-sectional analysis is limited in our ability to distinguish temporality between past PEP use and prior PrEP use, our findings illustrate the potential need for further research in this area. Individuals who had previously used PEP had a >50-fold likelihood of prior PrEP use. Further research is needed to identify how PEP and PrEP can be used interchangeably to support individuals' HIV prevention goals. Specifically, PrEP use is intended to be flexible based on potential vulnerability to HIV infection, where individuals can discontinue daily PrEP during breaks in sexual behavior or in combination with other HIV prevention strategies, including mutual monogamy with a recently tested HIV-negative partner or a partner with an undetectable viral load (ie, HIV positive with sustained viral suppression). Research is robust on reasons for discontinuing PrEP use, such as lower perceived risk and challenges with cost and access [36-40]. Moreover, gaps in PrEP use are normalized and encouraged when biomedical HIV prevention is not necessary because many individuals report changes in sexual behavior and perceived HIV risk over time [41-43]. Advancements in 2-1-1 PrEP dosing also present new opportunities where unanticipated sexual behavior may result in condomless anal sex without PrEP protection—necessitating the potential need for PEP before PrEP reinitiation. Thus, PEP adds to the HIV prevention toolbox in combination with PrEP, but a study of how PEP is used among individuals who discontinued PrEP is needed.

PEP seems to potentially have a small role in combating disparities in HIV incidence, where Black SMM, Latino SMM, and SMM identifying as another race or ethnicity reported higher prevalence of past PEP use than White SMM. Disparities in PrEP uptake are well documented, with fewer Black and Latino SMM using PrEP compared with White SMM [44,45], despite accounting for 37% and 21% of HIV incidence among gay and bisexual men, respectively [46]. In crude statistics, our findings also indicate that fewer Black and Latino SMM are using PrEP compared with White SMM; yet, the magnitude of this difference is smaller within this web-based sample than within the aggregated commercial pharmacy data reported by AIDSVu [47]. HIV incidence decreased 15% among White SMM between 2014 and 2018, but HIV incidence remained stable for Black and Latino SMM [46], likely resulting from inequitable access and barriers to HIV treatment and PrEP. PEP is unique in its use because it can be dispensed in a single prescription, including all pills for the 28-day regimen, avoiding some of the barriers to PrEP uptake and persistence that include quarterly visits to a provider and ongoing navigation of insurance and copay assistance programs [39,48]. As such, PEP is especially important as a mechanism of HIV prevention because of notable gaps in, and barriers to, PrEP use among SMM.

PEP users in our web-based sample of SMM had a similar profile, by age and sexual orientation, as PrEP users as also reported in other samples. We found SMM aged <25 years to have lower prevalence of past PEP use than older SMM aged 25 to 44 years, similar to disparities in PrEP uptake and persistence [45,49]; yet, this is expected in lifetime use statistics, given that older people have had more time to access these interventions, especially as the length of time that PEP and PrEP have been available is increasing. Nonetheless, specific barriers to PrEP use among young SMM include privacy and insurance issues, including the challenges of living with parents and being on the parents' insurance plan, high cost of PrEP, and perceived adherence challenges [50-52]. Moreover, we found that those who identified as bisexual had lower prevalence of past PEP use than those who identified as gay, aligning with disparities in PrEP uptake where bisexual men were less likely to take PrEP than their gay counterparts in other research [45]. PrEP stigma is pervasive and a known barrier to PrEP uptake [53], compounded with homonegativity and the enduring effects of

early advertising of PrEP specifically targeted to men who have sex with men [53,54]. As such, structural interventions are needed to make PEP more accessible to younger SMM and to prioritize bisexual SMM in HIV prevention efforts, given the suboptimal biomedical HIV prevention uptake to date.

Our large sample provided an opportunity to compare PEP and PrEP uptake between cisgender and transgender SMM. Specifically, we found no difference in past PEP use between cisgender and transgender SMM, but fewer transgender SMM reported current PrEP use compared with cisgender SMM by a large magnitude (84/569, 14.8%, vs 13,748/62,446, 22%, respectively). Prior research found that nearly two-thirds of transgender men who have sex with men met clinical guidelines for PrEP in 2017; yet, uptake was reported by only 21.8% of transgender SMM [55]. Our findings here from 2019 found lower rates of both current and prior PrEP use among transgender SMM perhaps because of sampling strategies. We focused exclusively on web-based recruitment, whereas Reisner et al [55] also recruited via social networks, engagement with community-based organizations, and outreach at a Philadelphia-based transgender health-focused conference. Similarly, 26.1% of the transgender men recruited on the internet from October 2017 to May 2018 ever reported PrEP use [40]; yet, these findings were not disaggregated by current or prior PrEP use and are similar to our study's 23.4% (133/569) who reported ever being prescribed PrEP. Further efforts are needed to target barriers to PrEP uptake, such as reducing potential misconceptions about interactions with gender-affirming therapy, establishing trusting relationships between medical institutions and transgender patients, and reducing PrEP stigma negatively affecting PrEP knowledge and attitudes as thematically organized by a systematic review of the literature [56].

Finally, we found that SMM who had recently engaged in club drug use were more likely to report past PEP use and current or prior PrEP use in concordance with prior research [57]. There is substantial evidence that club drug use, including the use of methamphetamine and other stimulants, is strongly associated with condomless anal sex as well as HIV and sexually transmitted infection acquisition among SMM [27,57-60]. Moreover, researchers have identified altered rectal cytokines among SMM who used stimulants [61]. Researchers suggest the confluence of condomless anal sex and dysregulated rectal immune functioning as an important potential driver of HIV transmission among SMM who use stimulants [62]. As such, SMM who use club drugs are a priority population for biomedical HIV prevention. Our findings regarding greater engagement in biomedical HIV prevention among club drug users is promising because current PEP and PrEP implementation efforts are reaching SMM at heightened vulnerability to HIV via substance use. Importantly, our findings align with previous reports about PEP use among young SMM in New York City, where researchers found that young SMM who used methamphetamine had >6 times higher odds of past PEP use.

Limitations

Our research is not without limitations. First, we recruited a convenience sample on the internet without incentivizing participation, which may have resulted in biased enrollment and introduced selection bias, potentially limiting the generalizability of the findings. Second, there is a potential for recall bias, especially related to lifetime past PEP use. Third, social desirability bias cannot be ruled out, which may have resulted in, for example, higher endorsement of PEP and PrEP use and lower reports of substance use. Finally, we conducted a cross-sectional analysis describing PEP and PrEP use with potential issues related to temporality, especially regarding past PEP and prior PrEP use. Additional longitudinal and qualitative research is needed to better understand PEP use and its potential impact on PrEP uptake or discontinuation.

Conclusions

PrEP use was the predominate HIV prevention strategy reported in our web-based sample of SMM compared with PEP; yet, our findings indicate that PEP use could be a gateway to PrEP use because nearly half (3268/7108, 46%) of the current PrEP users reported prior use of PEP. Advertising and prescribing PEP could also support efforts to increase PrEP uptake and sustain HIV prevention during breaks or interruptions in daily or intermittent PrEP use. Further research is needed to better understand and support this phenomenon to maximize the use of currently available biomedical HIV prevention tools.

Acknowledgments

Data collection for this study was conducted at Hunter College of the City University of New York, and the affiliations reflect the authors' institutions at the time of manuscript submission. During the time of data collection for this study, several studies were contributing to the costs of advertising and screening for the survey, with data collection being supported by grants from the National Institute of Allergy and Infectious Diseases, National Institute of Mental Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development, and National Institute on Drug Abuse (UG3/UH3-AI133674, HJR; R01-MH114735, HJR; R01-DA041262, Tyrel J Starks; R34-DA043422, Tyrel J Starks; R01-DA045613, Tyrel J Starks; and U19-HD089875, Sylvie Naar). Additional funding support was provided by the National Institute of Mental Health (K01-MH118939, SAJ). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank their staff and colleagues at the Health Intervention Sciences Group/Center for AIDS Intervention Research and PRIDE Health Research Consortium, as well as the participants who volunteered their time.

Conflicts of Interest

AEP receives research funding from Gilead Sciences, Inc.

References

1. HIV surveillance report. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2019-08-09]
2. 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]
3. Reisner SL, Moore CS, Asquith A, Pardee DJ, Mayer KH. The pre-exposure prophylaxis cascade in at-risk transgender men who have sex with men in the United States. LGBT Health 2021;8(2):116-124 [FREE Full text] [doi: 10.1089/lgbt.2020.0232] [Medline: 33567245]
4. Fauci AS, Redfield RR, Sigounas G, Weahkee MD, Giroir BP. Ending the HIV epidemic: a plan for the United States. JAMA 2019 Mar 05;321(9):844-845. [doi: 10.1001/jama.2019.1343] [Medline: 30730529]
5. FDA approves first medication to reduce HIV risk 2012. US Food and Drug Administration. 2012 Jul 16. URL: https://wayback.archive-it.org/7993/20170406045106/https://www.fda.gov/ForConsumers/ConsumerUpdates/ucm311821.htm [accessed 2018-09-06]
6. Preexposure prophylaxis for the prevention of HIV infection in the United States - 2017 update. Centers for Disease Control and Prevention. 2018 Mar. URL: https://www.cdc.gov/hiv/pdf/risk/prep/cdc-hiv-prep-guidelines-2017.pdf [accessed 2018-04-04]
7. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, iPrEx Study Team. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
8. Molina J, Capitant C, Spire B, Pialoux G, Cotte L, Charreau I, ANRS IPERGAY Study Group. On-demand preexposure prophylaxis in men at high risk for HIV-1 infection. N Engl J Med 2015 Dec 03;373(23):2237-2246. [doi: 10.1056/NEJMoa1506273] [Medline: 26624850]
9. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. Lancet 2016 Jan 02;387(10013):53-60 [FREE Full text] [doi: 10.1016/S0140-6736(15)00056-2] [Medline: 26364263]
10. Molina J, Charreau I, Spire B, Cotte L, Chas J, Capitant C, ANRS IPERGAY Study Group. Efficacy, safety, and effect on sexual behaviour of on-demand pre-exposure prophylaxis for HIV in men who have sex with men: an observational cohort study. Lancet HIV 2017 Sep;4(9):e402-e410. [doi: 10.1016/S2352-3018(17)30089-9] [Medline: 28747274]
11. Beymer MR, Weiss RE, Bolan RK, Kofron RM, Flynn RP, Pieribone DL, et al. Differentiating nonoccupational postexposure prophylaxis seroconverters and non-seroconverters in a community-based clinic in Los Angeles, California. Open Forum Infect Dis 2017;4(2):ofx061 [FREE Full text] [doi: 10.1093/ofid/ofx061] [Medline: 28596981]
12. Donnell D, Mimiaga MJ, Mayer K, Chesney M, Koblin B, Coates T. Use of non-occupational post-exposure prophylaxis does not lead to an increase in high risk sex behaviors in men who have sex with men participating in the EXPLORE trial. AIDS Behav 2010 Oct;14(5):1182-1189 [FREE Full text] [doi: 10.1007/s10461-010-9712-1] [Medline: 20490908]
13. Foster R, McAllister J, Read TR, Pierce AB, Richardson R, McNulty A, et al. Single-tablet emtricitabine-rilpivirine-tenofovir as HIV postexposure prophylaxis in men who have sex with men. Clin Infect Dis 2015 Oct 15;61(8):1336-1341. [doi: 10.1093/cid/civ511] [Medline: 26123937]
14. Jain S, Oldenburg CE, Mimiaga MJ, Mayer KH. Subsequent HIV infection among men who have sex with men who used non-occupational post-exposure prophylaxis at a Boston community health center: 1997-2013. AIDS Patient Care STDS 2015 Jan;29(1):20-25 [FREE Full text] [doi: 10.1089/apc.2014.0154] [Medline: 25369451]
15. McAllister J, Read P, McNulty A, Tong WW, Ingersoll A, Carr A. Raltegravir-emtricitabine-tenofovir as HIV nonoccupational post-exposure prophylaxis in men who have sex with men: safety, tolerability and adherence. HIV Med 2014 Jan;15(1):13-22 [FREE Full text] [doi: 10.1111/hiv.12075] [Medline: 24007390]
16. Schechter M, do Lago RF, Mendelsohn AB, Moreira RI, Moulton LH, Harrison LH, Praca Onze Study Team. Behavioral impact, acceptability, and HIV incidence among homosexual men with access to postexposure chemoprophylaxis for HIV. J Acquir Immune Defic Syndr 2004 Apr 15;35(5):519-525. [doi: 10.1097/00126334-200404150-00010] [Medline: 15021317]
17. Sonder GJ, Prins JM, Regez RM, Brinkman K, Mulder J, Veenstra J, et al. Comparison of two HIV postexposure prophylaxis regimens among men who have sex with men in Amsterdam: adverse effects do not influence compliance. Sex Transm Dis 2010 Nov;37(11):681-686. [doi: 10.1097/OLQ.0b013e3181e2f999] [Medline: 20644499]
18. Updated guidelines for antiretroviral postexposure prophylaxis after sexual, injection drug use, or other nonoccupational exposure to HIV—United States, 2016. Centers for Disease Control and Prevention. 2016. URL: https://www.cdc.gov/hiv/pdf/programresources/cdc-hiv-npep-guidelines.pdf [accessed 2017-02-14]
19. Mayer KH, Jones D, Oldenburg C, Jain S, Gelman M, Zaslow S, et al. Optimal hiv postexposure prophylaxis regimen completion with single tablet daily elvitegravir/cobicistat/tenofovir disoproxil fumarate/emtricitabine compared with more

frequent dosing regimens. J Acquir Immune Defic Syndr 2017 Aug 15;75(5):535-539 [FREE Full text] [doi: 10.1097/QAI.0000000000001440] [Medline: 28696345]
20. Smith DK, Grohskopf LA, Black RJ, Auerbach JD, Veronese F, Struble KA, U.S. Department of HealthHuman Services. Antiretroviral postexposure prophylaxis after sexual, injection-drug use, or other nonoccupational exposure to HIV in the United States: recommendations from the U.S. Department of Health and Human Services. MMWR Recomm Rep 2005 Jan 21;54(RR-2):1-20 [FREE Full text] [Medline: 15660015]
21. Mayer KH, Molina J, Thompson MA, Anderson PL, Mounzer KC, De Wet JJ, et al. Emtricitabine and tenofovir alafenamide vs emtricitabine and tenofovir disoproxil fumarate for HIV pre-exposure prophylaxis (DISCOVER): primary results from a randomised, double-blind, multicentre, active-controlled, phase 3, non-inferiority trial. Lancet 2020 Jul 25;396(10246):239-254. [doi: 10.1016/S0140-6736(20)31065-5] [Medline: 32711800]
22. Saag MS, Benson CA, Gandhi RT, Hoy JF, Landovitz RJ, Mugavero MJ, et al. Antiretroviral drugs for treatment and prevention of HIV infection in adults: 2018 recommendations of the international antiviral society-USA panel. JAMA 2018 Jul 24;320(4):379-396 [FREE Full text] [doi: 10.1001/jama.2018.8431] [Medline: 30043070]
23. O'Byrne P, Orser L, Vandyk A. Immediate PrEP after PEP: results from an observational nurse-led PEP2PrEP study. J Int Assoc Provid AIDS Care 2020;19:2325958220939763 [FREE Full text] [doi: 10.1177/2325958220939763] [Medline: 32856549]
24. Tanner MR, Miele P, Carter W, Valentine SS, Dunville R, Kapogiannis BG, et al. Preexposure prophylaxis for prevention of HIV acquisition among adolescents: clinical considerations, 2020. MMWR Recomm Rep 2020 Apr 24;69(3):1-12 [FREE Full text] [doi: 10.15585/mmwr.rr6903a1] [Medline: 32324724]
25. Teitcher JE, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. J Law Med Ethics 2015;43(1):116-133 [FREE Full text] [doi: 10.1111/jlme.12200] [Medline: 25846043]
26. Adler NE, Epel ES, Castellazzo G, Ickovics JR. Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. Health Psychol 2000 Nov;19(6):586-592. [doi: 10.1037//0278-6133.19.6.586] [Medline: 11129362]
27. John SA, Parsons JT, Rendina HJ, Grov C. Club drug users had higher odds of reporting a bacterial STI compared with non-club drug users: results from a cross-sectional analysis of gay and bisexual men on HIV pre-exposure prophylaxis. Sex Transm Infect 2019 Dec;95(8):626-628 [FREE Full text] [doi: 10.1136/sextrans-2018-053591] [Medline: 30126949]
28. Parsons JT, Rendina HJ, Lassiter JM, Whitfield TH, Starks TJ, Grov C. Uptake of HIV pre-exposure prophylaxis (PrEP) in a national cohort of gay and bisexual men in the United States. J Acquir Immune Defic Syndr 2017 Mar 01;74(3):285-292 [FREE Full text] [doi: 10.1097/QAI.0000000000001251] [Medline: 28187084]
29. John SA, Rendina HJ, Starks TJ, Grov C, Parsons JT. Decisional balance and contemplation ladder to support interventions for HIV pre-exposure prophylaxis uptake and persistence. AIDS Patient Care STDS 2019 Feb;33(2):67-78 [FREE Full text] [doi: 10.1089/apc.2018.0136] [Medline: 30653348]
30. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology 2009 Jul;20(4):488-495 [FREE Full text] [doi: 10.1097/EDE.0b013e3181a819a1] [Medline: 19525685]
31. Bauermeister JA, Meanley S, Pingel E, Soler JH, Harper GW. PrEP awareness and perceived barriers among single young men who have sex with men. Curr HIV Res 2013 Oct;11(7):520-527 [FREE Full text] [doi: 10.2174/1570162x12666140129100411] [Medline: 24476355]
32. Golub SA, Gamarel KE, Rendina HJ, Surace A, Lelutiu-Weinberger CL. From efficacy to effectiveness: facilitators and barriers to PrEP acceptability and motivations for adherence among MSM and transgender women in New York City. AIDS Patient Care STDS 2013 Apr;27(4):248-254 [FREE Full text] [doi: 10.1089/apc.2012.0419] [Medline: 23565928]
33. Wang Z, Yuan T, Fan S, Qian H, Li P, Zhan Y, et al. HIV nonoccupational postexposure prophylaxis among men who have sex with men: a systematic review and meta-analysis of global data. AIDS Patient Care STDS 2020 May;34(5):193-204. [doi: 10.1089/apc.2019.0313] [Medline: 32396477]
34. Kaplun E, Martino RJ, Krause KD, Briganti M, D'Avanzo PA, Halkitis PN. Post-exposure prophylaxis and methamphetamine use among young sexual minority men: the p18 cohort study. Int J Environ Res Public Health 2022 Jan 09;19(2):712 [FREE Full text] [doi: 10.3390/ijerph19020712] [Medline: 35055534]
35. Koblin BA, Usher D, Nandi V, Tieu H, Bravo E, Lucy D, et al. Post-exposure prophylaxis awareness, knowledge, access and use among three populations in New York City, 2016-17. AIDS Behav 2018 Aug;22(8):2718-2732. [doi: 10.1007/s10461-018-2175-5] [Medline: 29858737]
36. Arnold T, Brinkley-Rubinstein L, Chan PA, Perez-Brumer A, Bologna ES, Beauchamps L, et al. Social, structural, behavioral and clinical factors influencing retention in Pre-Exposure Prophylaxis (PrEP) care in Mississippi. PLoS One 2017;12(2):e0172354 [FREE Full text] [doi: 10.1371/journal.pone.0172354] [Medline: 28222118]
37. D'Angelo AB, Lopez-Rios J, Flynn AW, Holloway IW, Pantalone DW, Grov C. Insurance- and medical provider-related barriers and facilitators to staying on PrEP: results from a qualitative study. Transl Behav Med 2021 Mar 16;11(2):573-581 [FREE Full text] [doi: 10.1093/tbm/ibz191] [Medline: 32065637]

38. Hong C, Horvath KJ, Stephenson R, Nelson KM, Petroll AE, Walsh JL, et al. PrEP use and persistence among young sexual minority men 17-24 years old during the COVID-19 pandemic. AIDS Behav 2022 Mar;26(3):631-638 [FREE Full text] [doi: 10.1007/s10461-021-03423-5] [Medline: 34387777]
39. Whitfield TH, John SA, Rendina HJ, Grov C, Parsons JT. Why I quit pre-exposure prophylaxis (PrEP)? A mixed-method study exploring reasons for PrEP discontinuation and potential re-initiation among gay and bisexual men. AIDS Behav 2018 Nov;22(11):3566-3575 [FREE Full text] [doi: 10.1007/s10461-018-2045-1] [Medline: 29404756]
40. Zarwell M, John SA, Westmoreland D, Mirzayi C, Pantalone DW, Golub S, et al. PrEP uptake and discontinuation among a U.S. national sample of transgender men and women. AIDS Behav 2021 Apr;25(4):1063-1071 [FREE Full text] [doi: 10.1007/s10461-020-03064-0] [Medline: 33057893]
41. Elsesser SA, Oldenburg CE, Biello KB, Mimiaga MJ, Safren SA, Egan JE, et al. Seasons of risk: anticipated behavior on vacation and interest in episodic antiretroviral pre-exposure prophylaxis (PrEP) among a large national sample of U.S. men who have sex with men (MSM). AIDS Behav 2016 Jul;20(7):1400-1407 [FREE Full text] [doi: 10.1007/s10461-015-1238-0] [Medline: 26538056]
42. Namey E, Agot K, Ahmed K, Odhiambo J, Skhosana J, Guest G, et al. When and why women might suspend PrEP use according to perceived seasons of risk: implications for PrEP-specific risk-reduction counselling. Cult Health Sex 2016 Sep;18(9):1081-1091 [FREE Full text] [doi: 10.1080/13691058.2016.1164899] [Medline: 27093238]
43. Underhill K, Guthrie KM, Colleran C, Calabrese SK, Operario D, Mayer KH. Temporal fluctuations in behavior, perceived HIV risk, and willingness to use pre-exposure prophylaxis (PrEP). Arch Sex Behav 2018 Oct;47(7):2109-2121 [FREE Full text] [doi: 10.1007/s10508-017-1100-8] [Medline: 29327091]
44. Kanny D, Jeffries WL, Chapin-Bardales J, Denning P, Cha S, Finlayson T, National HIV Behavioral Surveillance Study Group. Racial/ethnic disparities in HIV preexposure prophylaxis among men who have sex with men - 23 urban areas, 2017. MMWR Morb Mortal Wkly Rep 2019 Sep 20;68(37):801-806 [FREE Full text] [doi: 10.15585/mmwr.mm6837a2] [Medline: 31536484]
45. Whitfield TH, Parsons JT, Rendina HJ. Rates of pre-exposure prophylaxis use and discontinuation among a large U.S. National sample of sexual minority men and adolescents. Arch Sex Behav 2020 Jan;49(1):103-112 [FREE Full text] [doi: 10.1007/s10508-019-01602-z] [Medline: 31845148]
46. HIV and gay and bisexual men. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/group/msm/index.html [accessed 2021-02-18]
47. Sullivan P, Whitby S, Hipp P, Juhasz M, DuBose S, McGuinness P. Trends in PrEP inequity by racecensus region, United States, 2012-2021. In: Proceedings of the 24th International AIDS Conference. 2022 Presented at: 24th International AIDS Conference; Jul 29-Aug 2, 2022; Montreal, CA.
48. McMahan VM, McCanta L, Tran D, Herrera L, Viquez L, Swanson F, et al. Administrative and related barriers to covering the costs of preexposure prophylaxis at a safety-net clinic in Seattle, Washington. Clin Infect Dis 2020 Dec 03;71(9):2509-2512. [doi: 10.1093/cid/ciaa187] [Medline: 32152636]
49. Huang YA, Tao G, Smith DK, Hoover KW. Persistence with human immunodeficiency virus pre-exposure prophylaxis in the United States, 2012-2017. Clin Infect Dis 2021 Feb 01;72(3):379-385. [doi: 10.1093/cid/ciaa037] [Medline: 33527117]
50. Holloway IW, Tan D, Gildner JL, Beougher SC, Pulsipher C, Montoya JA, et al. Facilitators and barriers to pre-exposure prophylaxis willingness among young men who have sex with men who use geosocial networking applications in California. AIDS Patient Care STDS 2017 Dec;31(12):517-527 [FREE Full text] [doi: 10.1089/apc.2017.0082] [Medline: 29211513]
51. Marks SJ, Merchant RC, Clark MA, Liu T, Rosenberger JG, Bauermeister J, et al. Potential healthcare insurance and provider barriers to pre-exposure prophylaxis utilization among young men who have sex with men. AIDS Patient Care STDS 2017 Nov;31(11):470-478 [FREE Full text] [doi: 10.1089/apc.2017.0171] [Medline: 29087744]
52. Zapata JP, Dang M, Quinn KG, Horvath KJ, Stephenson R, Dickson-Gomez J, et al. COVID-19-related disruptions to HIV testing and prevention among young sexual minority men 17-24 years old: a qualitative study using synchronous online focus groups, April-September 2020. Arch Sex Behav 2022 Jan;51(1):303-314 [FREE Full text] [doi: 10.1007/s10508-021-02166-7] [Medline: 34773214]
53. Golub SA. PrEP stigma: implicit and explicit drivers of disparity. Curr HIV/AIDS Rep 2018 Apr;15(2):190-197 [FREE Full text] [doi: 10.1007/s11904-018-0385-0] [Medline: 29460223]
54. Garcia J, Parker C, Parker RG, Wilson PA, Philbin M, Hirsch JS. Psychosocial implications of homophobia and HIV stigma in social support networks: insights for high-impact HIV prevention among Black men who have sex with men. Health Educ Behav 2016 Apr;43(2):217-225 [FREE Full text] [doi: 10.1177/1090198115599398] [Medline: 27037286]
55. Reisner SL, Moore CS, Asquith A, Pardee DJ, Sarvet A, Mayer G, et al. High risk and low uptake of pre-exposure prophylaxis to prevent HIV acquisition in a national online sample of transgender men who have sex with men in the United States. J Int AIDS Soc 2019 Sep;22(9):e25391 [FREE Full text] [doi: 10.1002/jia2.25391] [Medline: 31536171]
56. Dang M, Scheim AI, Teti M, Quinn KG, Zarwell M, Petroll AE, et al. Barriers and facilitators to HIV pre-exposure prophylaxis uptake, adherence, and persistence among transgender populations in the United States: a systematic review. AIDS Patient Care STDS 2022 Jun;36(6):236-248. [doi: 10.1089/apc.2021.0236] [Medline: 35687813]

57. Hammoud MA, Jin F, Maher L, Bourne A, Haire B, Saxton P, et al. Biomedical HIV protection among gay and bisexual men who use crystal methamphetamine. AIDS Behav 2020 May;24(5):1400-1413. [doi: 10.1007/s10461-019-02739-7] [Medline: 31758350]
58. Achterbergh RC, de Vries HJ, Boyd A, Davidovich U, Drückler S, Hoornenborg E, et al. Identification and characterization of latent classes based on drug use among men who have sex with men at risk of sexually transmitted infections in Amsterdam, the Netherlands. Addiction 2020 Jan;115(1):121-133. [doi: 10.1111/add.14774] [Medline: 31400174]
59. Grov C, Westmoreland D, Morrison C, Carrico AW, Nash D. The crisis we are not talking about: one-in-three annual HIV seroconversions among sexual and gender minorities were persistent methamphetamine users. J Acquir Immune Defic Syndr 2020 Nov 01;85(3):272-279 [FREE Full text] [doi: 10.1097/QAI.0000000000002461] [Medline: 32740370]
60. Ostrow DG, Plankey MW, Cox C, Li X, Shoptaw S, Jacobson LP, et al. Specific sex drug combinations contribute to the majority of recent HIV seroconversions among MSM in the MACS. J Acquir Immune Defic Syndr 2009 Jul 01;51(3):349-355 [FREE Full text] [doi: 10.1097/QAI.0b013e3181a24b20] [Medline: 19387357]
61. Tapia GR, Glynn TR, Miller C, Manuzak JA, Broedlow CA, Mcgaugh A, et al. Syndemics and preexposure prophylaxis are independently associated with rectal immune dysregulation in sexual minority men. AIDS 2021 Jul 01;35(8):1295-1300 [FREE Full text] [doi: 10.1097/QAD.0000000000002866] [Medline: 33710016]
62. Viamonte M, Ghanooni D, Reynolds JM, Grov C, Carrico AW. Running with scissors: a systematic review of substance use and the pre-exposure prophylaxis care continuum among sexual minority men. Curr HIV/AIDS Rep 2022 Aug;19(4):235-250 [FREE Full text] [doi: 10.1007/s11904-022-00608-y] [Medline: 35701713]

Abbreviations

GHB: gamma-hydroxybutyrate
MMDA: 3,4-methylenedioxy-methamphetamine
PEP: postexposure prophylaxis
PrEP: pre-exposure prophylaxis
RRR: relative risk ratio
SMM: sexual minority men

Edited by T Sanchez; submitted 15.06.21; peer-reviewed by J Jones, P Serrano; comments to author 07.09.21; revised version received 08.12.21; accepted 28.10.22; published 16.12.22.

Please cite as:
John SA, Sizemore KM, Jimenez RH, Jones SS, Petroll AE, Rendina HJ
The Use of HIV Pre- and Postexposure Prophylaxis Among a Web-Based Sample of HIV-Negative and Unknown Status Cisgender and Transgender Sexual Minority Men: Cross-sectional Study
JMIR Public Health Surveill 2022;8(12):e31237
URL: https://publichealth.jmir.org/2022/12/e31237
doi:10.2196/31237
PMID:36306518

Original Paper

The Risk Factors for Cervical Cytological Abnormalities Among Women Infected With Non-16/18 High-Risk Human Papillomavirus: Cross-sectional Study

Ting Xiao[1*], BS; Chun-Quan Ou[1*], PhD; Jun Yang[2*], MD; Chunhua Wang[3], MD; Mei Yang[3], MD; Tingyu Yu[3], MM; Liang Shen[3], MD; Xiaohan Xu[1], BS; Hui Xing[3], MD

[1]State Key Laboratory of Organ Failure Research, Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou, China

[2]School of Public Health, Guangzhou Medical University, Guangzhou, China

[3]Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science, Xiangyang, China

[*]these authors contributed equally

Corresponding Author:
Hui Xing, MD
Xiangyang Central Hospital
Affiliated Hospital of Hubei University of Arts and Science
No 136, Jingzhou Street
Xiangcheng District
Xiangyang, 441021
China
Phone: 86 13508668281
Email: huixinghx123@163.com

Abstract

Background: High-risk human papillomavirus (hrHPV) infection is a necessary cause of almost all cervical cancers. Relative to hrHPV 16/18 infection, non-16/18 hrHPV infection is of less concern. However, the increasing prevalence of non-16/18 hrHPV infections has become an important public health issue. The early identification and treatment of cervical cytological abnormalities in women infected with non-16/18 hrHPV reduces the incidence of cervical cancer. To date, no study has examined the risk factors for cytological abnormalities in this high-risk population.

Objective: This population-based, cross-sectional study aimed to identify the risk factors for cervical cytological abnormalities in women infected with non-16/18 hrHPV.

Methods: A total of 314,587 women from the general population were recruited for cervical cancer screening at 136 primary care hospitals in Xiangyang, China. Of these, 311,604 women underwent HPV genotyping, and 17,523 non-16/18 hrHPV–positive women were referred for cytological screening according to the screening program. A logistic regression model was used to assess the risk factors for cytological abnormalities among these non-16/18 hrHPV–positive women. A separate analysis was performed to determine the factors influencing high-grade cytological abnormalities.

Results: The non-16/18 hrHPV infection rate was 5.88% (18,323/311,604), which was 3-fold higher than that of hrHPV 16/18 (6068/311,604, 1.95%). Among the non-16/18 hrHPV–positive women who underwent ThinPrep cytologic test, the overall prevalence rates of cervical cytological abnormalities and high-grade cytological abnormalities were 13.46% (2359/17,523) and 1.18% (206/17,523), respectively. Multivariate logistic regression analysis revealed that women with middle or high school educational attainment were at a higher risk of having cytological abnormalities than those who received primary education (odds ratio [OR] 1.31, 95% CI 1.17-1.45; *P*<.001, and OR 1.32, 95% CI 1.14-1.53; *P*<.001, respectively). Living in rural areas (OR 2.58, 95% CI 2.29-2.90; *P*<.001), gravidity ≥3 (OR 2.77, 95% CI 1.19-6.45; *P*=.02), cervix abnormalities detected in pelvic examination (OR 1.22, 95% CI 1.11-1.34; *P*<.001), and having a cervical cancer screening 3 years ago (OR 0.79, 95% CI 0.62-1.00; *P*=.048) were associated with cytological abnormalities. The risk factors for high-grade cytological abnormalities included middle school education (OR 1.45, 95% CI 1.07-1.98; *P*=.02), living in rural regions (OR 1.52, 95% CI 1.10-2.10; *P*=.01), and cervix abnormality (OR 1.72, 95% CI 1.30-2.26; *P*<.001).

Conclusions: The dominant epidemic of non-16/18 hrHPV infection is revealed in Chinese women. Multiple risk factors for cervical cytological abnormalities have been identified in women infected with non-16/18 hrHPV. These findings can provide

important information for clinically actionable decisions for the screening, early diagnosis, intervention, and prevention of cervical cancer in non-16/18 hrHPV–positive women.

(JMIR Public Health Surveill 2022;8(12):e38628) doi:10.2196/38628

KEYWORDS

non-16/18 high-risk human papillomavirus; cervical cytological abnormalities; risk factors; logistic regression; cervical cancer; screening; rural; pelvic examination; education; gravidity

Introduction

Globally, cervical cancer is one of the most serious threats to the lives of women. Cervical cancer ranks fourth in terms of both incidence and mortality among women, with an estimated 604,000 new cases and 342,000 deaths globally in 2020 [1]. In China, cervical cancer is a major public health concern because of its high incidence and heavy economic burden [2]. In 2020, it was estimated that there were approximately 110,000 new cases and 59,000 deaths from cervical cancer in China. It is the sixth most frequently diagnosed cancer and the seventh leading cause of cancer-related deaths among Chinese women [3].

Cervical cancer is the most preventable and treatable form of cancer via human papillomavirus (HPV) vaccination, early diagnosis, and effective management. Persistent infection with high-risk HPV (hrHPV) is a necessary but not sufficient cause of almost all cervical cancers [4,5]. There are 14 hrHPV genotypes (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68) that can be detected by real-time polymerase chain reaction assays [6], which are classified as hrHPV 16/18 and non-16/18 hrHPV by current diagnostic paradigms. The majority of cervical cancers are from infection with hrHPV 16, followed by hrHPV 18 [7]. Therefore, hrHPV 16/18 have been recognized as dominant risk factors for cervical cancer and are the focus of medical research, clinical diagnosis, and intervention. As a result, the prevalence of hrHPV 16/18 has significantly decreased over the years [8]. Researchers and the public are relatively less concerned about non-16/18 hrHPV because these infections are considered to be less prevalent and less risky than type 16/18 infections. However, recent studies have reported an increasing prevalence of non-16/18 hrHPV [9,10]. For example, a recent population-based study in China reported a prevalence of 2.2% and 15.3% for hrHPV 16/18 and non-16/18 hrHPV, respectively [11]. The prevalence of non-16/18 hrHPV infection is also a strong predictor of the persistence and progression of cervical diseases [12-15].

Women with cytological abnormalities in the cervix have a relatively high risk of cervical cancer [16]. Early identification and treatment of cervical abnormalities in the early stages or precursor phases of the neoplasm increases the likelihood of lesion regression and reduces the incidence of cervical cancer [17,18]. According to the guidelines of the American Society for Colposcopy and Cervical Pathology [19] and the Chinese Society for Colposcopy and Cervical Pathology [20], women infected with hrHPV 16/18 were directly subjected to colposcopy without cytological screening. Only the women with positive hrHPV genotypes were referred for ThinPrep cytologic test (TCT) followed by colposcopy among those with TCT-positive results. Therefore, following the detection of a non-16/18 hrHPV infection, cytological screening is a useful tool for the selection of women at risk of cervical cancer while reducing the colposcopy burden. A meta-analysis showed that cytological testing in women infected with non-16/18 hrHPV had an overall sensitivity of 69.6% and specificity of 90.2% for detecting cervical intraepithelial neoplasia or worse [21]. However, some women infected with non-16/18 hrHPV may not undergo cytological screening because of inadequate perception of the hazards associated with non-16/18 hrHPV infection or the lack of free screening programs, especially in resource-limited countries. Therefore, identifying the risk factors for cytological abnormalities among those with non-16/18 hrHPV infections will provide important information for impelling those at high risk to undergo screening and ultimately guide clinically actionable decisions for early diagnosis, monitoring, and intervention.

Nevertheless, no previous study has investigated the risk factors for abnormal cytological outcomes in individuals with non-16/18 hrHPV infections [8]. The majority of the previous studies were conducted on the whole population without considering HPV test results, and the factors under study and the conclusions were inconsistent. For example, an observational study in China showed that the risk of cytological abnormalities was associated with HPV genotype [22]. A population-based study in Nigeria showed that demographic characteristics, menopause, gravidity, parity, marital status, and education were associated with cytological abnormalities [23]. Moreover, some previous studies did not find an effect of age on cytological abnormalities in all women or those positive for HPV [24,25]. However, in some studies, the risk of cytological abnormalities significantly increased with age [22,26]. Two studies focused on individuals infected with HPV, among whom education level, years of sexually active life, and parity were risk factors for cytological abnormalities [27,28]. Besides these factors, recent studies have shown an association between cervical cancer and vaginal microbial infection [29,30]. Cervical cancer symptoms, such as bleeding after sex, abnormal vaginal discharge, and pelvic discomfort, may affect the timely diagnosis of cervical cancer [31]. The effects of these factors on cervical cytological abnormalities in individuals with non-16/18 hrHPV infections remain unknown. In particular, the potential impacts of some important factors, including vaginal microbial infection and pelvic examination (PE), on cervical cytological abnormalities have not been investigated previously.

This large population-based study of cervical cancer screening in Chinese women aimed to identify risk factors for cervical cytological abnormalities as well as high-grade cytological abnormalities among women with non-16/18 hrHPV infections, which would provide important information for the screening,

early diagnosis, management, and prevention of cervical cancer in the target population (ie, non-16/18 hrHPV–positive women).

Methods

Population

The cervical cancer screening program was conducted at 136 primary care hospitals in Xiangyang, China. Participants aged ≥30 years were recruited through media publicity and government notices between January 2017 and February 2018. Women who had received HPV vaccination, were pregnant, had no sexual history, had a hysterectomy, or had a history of pelvic radiotherapy were excluded. All participants were interviewed using questionnaires and underwent PE, vaginal microenvironment test, and HPV genotyping. Women infected with hrHPV 16/18 were directly subjected to colposcopy, whereas women positive for other hrHPV genotypes were referred for TCT, followed by colposcopy in women with TCT-positive results. Histopathological diagnosis was performed if the colposcopy was abnormal or if abnormalities were suspected. A technical manual was developed to regulate the screening process, and the medical staff were trained before the project began.

Questionnaires

The questionnaire, designed by gynecological oncologists, included age, educational level, residential type (rural or urban), whether the patient is in menopause, age at menopause, family history of cancer, gravidity, parity, contraceptive methods, personal history of other cancers, cervical cancer screening history, and presence of postcoital bleeding and abnormal leucorrhea (Multimedia Appendix 1). Professionally trained clinical staff distributed the questionnaires to the participants and collected data via face-to-face interviews. All data were inputted using the double-entry method.

PE and Vagina Microenvironment Test

All recruited women underwent routine PE and vaginal microenvironment test. The purpose of the PE was not only to assess pain, bleeding, and vaginal secretions but also to screen for cervical cancer and reproductive tract infections. The PE involved the visual inspection of the vulva, internal speculum examination of the vagina and cervix, and bimanual palpation of the adnexa and uterus. Vaginal secretions were collected with high-vaginal swabs and observed under a microscope to evaluate the vaginal microecosystem, including *Trichomonas vaginalis*, Candida, and Gardnerella [30].

HPV Genotyping

HPV genotyping was performed using the Cobas HPV test with the Cobas 4800 (Roche Molecular Systems) system, which is approved by the US Food and Drug Administration [6]. Specimens were collected using a cervical brush and sent to the laboratory for professional examination. The Cobas HPV test can provide individual results for hrHPV 16 and hrHPV 18 and simultaneously provide the pooled results for the other 12 non-16/18 hrHPV genotypes (31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68).

TCT Procedure

Women with non-16/18 hrHPV genotypes underwent TCT. The results were reported using the 2001 Bethesda System terminology [32], including negative for intraepithelial lesion or malignancy (NILM); low-grade squamous intraepithelial lesion (LSIL) or high-grade squamous intraepithelial lesion (HSIL); atypical squamous cells of undetermined significance (ASC-US) or atypical squamous cells not possible excluding HSIL (ASC-H); atypical glandular cells (AGC); and squamous cell carcinoma. NILM was considered normal, whereas the others (TCT result worse than ASC-US [ASC-US+]) were considered abnormal.

Ethics Approval

Ethical approval was obtained from the Ethics Review Committee of Xiangyang Central Hospital, and all procedures followed the ethical standards specified by the institution (approval 2017–004). Written informed consent was obtained from all participants. All examinations complied with the routine medical requirements, and there were provisions for patient safety.

Statistical Analysis

The enrolled participants were divided into 2 groups based on the TCT results: NILM and ASC-US+. ASC-US+ was considered to be a cervical cytological abnormality. Participants' characteristics were summarized as counts and percentages, and the chi-square or Fisher exact tests were used to compare whether there were statistical differences in the characteristics between the 2 groups.

Based on the literature and clinical knowledge about the risk factors for cytological abnormalities or cervical cancer, we considered 16 factors that may be associated with cervical cancer. Univariate logistic regression was used to quantify the effect of each factor on the TCT results. Multivariate logistic regression was subsequently performed for all included variables. The generalized variance inflation factor (GVIF) for each variable was calculated to estimate the existence of multicollinearity, and the variable with the largest $GVIF^{[1/(2\times df)]}$ was removed at each step until the $GVIF^{[1/(2\times df)]}$ for all remaining variables was less than 2.24 (ie, $5^{1/2}$) [33]. Odds ratios (ORs) and their 95% CIs were also calculated. Missing data were not inputted in this study because the rate was low, with 3.41% (597/17,523) of participants having missing values for at least one variable under study.

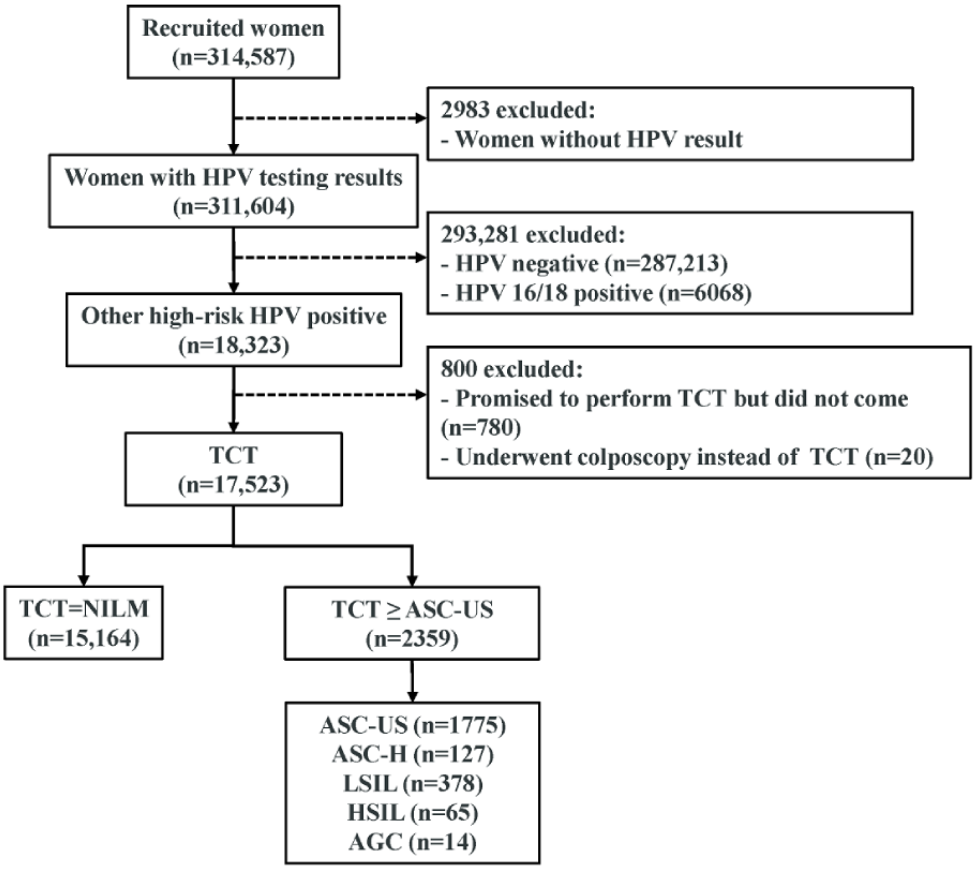
Since high-grade cytological abnormalities closely associated with cervical cancer require more attention, we specifically identified potential risk factors for high-grade cytological abnormalities (ASC-H, HSIL, AGC, and squamous cell carcinoma) [34] using univariate logistic regression. In the multivariate logistic regression analysis, only variables with $P<.10$ were considered independent variables due to the small sample size. All statistical analyses were performed using R statistical software (version 4.1.1; R Foundation for Statistical Computing). Two-sided statistical tests were used in all analyses, and $P<.05$ was considered statistically significant.

Results

Study Subjects

Figure 1 shows the flow of the identification and selection of participants in the study. A total of 311,604 participants in the study underwent HPV genotyping, among which 6068 (1.95%) were infected with hrHPV 16/18, and 18,323 (5.88%) were infected with non-16/18 hrHPV. Of the 18,323 non-16/18 hrHPV–positive participants, 780 (4.26%) promised to receive TCT but did not come back until the end of the program; 20 (0.11%) did not comply with the screening process and underwent colposcopy directly rather than TCT first. As a result, 17,523 participants who were infected with non-16/18 hrHPV and underwent TCT were included in the final analysis of factors associated with cervical abnormalities. The TCT results illustrated that, among them, 15,164 participants (86.54%) had NILM and 2359 (13.46%) had cytologically abnormal findings (ASC-US+). Of the 2359 cytologically abnormal findings, ASC-US was the primary abnormality in TCT (n=1775, 75.25%), followed by LSIL (n=378, 16.02%), ASC-H (n=127, 5.38%), HSIL (n=65, 2.76%), and AGC (n=14, 0.59%).

Figure 1. Flow diagram of the identification and selection of study subjects. AGC: atypical glandular cells; ASC-H: atypical squamous cells not possible excluding high-grade squamous intraepithelial lesion; ASC-US: atypical squamous cells of undetermined significance; HPV: human papillomavirus; HSIL: high-grade squamous intraepithelial lesion; LSIL: low-grade squamous intraepithelial lesion; NILM: negative for intraepithelial lesion or malignancy; TCT: ThinPrep cytologic test.



Participant Characteristics

Table 1 presents the demographic characteristics and diagnosis-related variables among participants with non-16/18 hrHPV infections. We stratified the participants into 4 groups based on age, and the ages of participants were mainly concentrated in the 40-60 years age range (12,594/17,523, 71.87%). Women with ASC-US+ were relatively less educated than women with NILM (P<.001), although 89.78% (15,733/17,523) of participants in both groups had only primary or secondary education. Individuals from rural areas represented the largest proportion of participants with ASC-US+ (1924/2359, 81.56%), with only 62.91% (9540/15,164) of participants with NILM coming from rural areas. A higher proportion of ASC-US+ were participants whose gravidity and parity were ≥3 (1259/2359, 53.37% vs 7219/15,164, 47.6% and 372/2359, 15.77% vs 2147/15,164, 14.16%, respectively). Participants in the ASC-US+ group was less likely to have undergone cervical screening within 3 years or >3 years ago than those in the NILM group (373/2357, 15.83% vs 2674/15,157, 17.64% and 89/2357, 3.78% vs 757/15,157, 4.99%, respectively). Cervix abnormalities detected in PE were more common in participants with ASC-US+ than in those with NILM (1020/2346, 43.48% vs 5678/15,067, 37.69%, respectively). There were no statistically significant differences in other factors between the 2 groups.

Table 1. Demographic characteristics and diagnosis-related variables for participants with non-16/18 high-risk human papillomavirus infection.

Characteristics	Overall, n (%)	Groups		*P* value
		NILM[a], n (%)	ASC-US+[b], n (%)	
Age (years; overall: n=17,523; NILM: n=15,164; ASC-US+: n=2359)				.84
<40	2670 (15.24)	2302 (15.18)	368 (15.6)	
40-50	5817 (33.2)	5028 (33.16)	789 (33.45)	
50-60	6777 (38.67)	5884 (38.8)	893 (37.85)	
≥60	2259 (12.89)	1950 (12.86)	309 (13.1)	
BMI[c,d] (overall: n=17,359; NILM: n=15,023; ASC-US+: n=2336)				.44
Normal	12,551 (72.3)	10,858 (72.28)	1693 (72.47)	
Underweight	699 (4.03)	595 (3.96)	104 (4.45)	
Overweight	4109 (23.67)	3570 (23.76)	539 (23.07)	
Education (overall: n=17,523; NILM: n=15,164; ASC-US+: n=2359)				<.001
Primary	8439 (48.16)	7350 (48.47)	1089 (46.16)	
Middle	4896 (27.94)	4126 (27.21)	770 (32.64)	
High	2398 (13.68)	2071 (13.66)	327 (13.86)	
Graduate	1790 (10.22)	1617 (10.66)	173 (7.34)	
Region (overall: n=17,523; NILM: n=15,164; ASC-US+: n=2359)				<.001
Urban	6059 (34.58)	5624 (37.09)	435 (18.44)	
Rural	11,464 (65.42)	9540 (62.91)	1924 (81.56)	
Family history of cancer (overall: n=17,523; NILM: n=15,164; ASC-US+: n=2359)				.27
No	17,170 (97.99)	14,866 (98.03)	2304 (97.67)	
Yes	353 (2.01)	298 (1.97)	55 (2.33)	
Menopause (overall: n=17,523; NILM: n=15,164; ASC-US+: n=2359)				.43
No	9031 (51.54)	7797 (51.42)	1234 (52.31)	
Yes	8492 (48.46)	7367 (48.58)	1125 (47.69)	
Gravidity (overall: n=17,523; NILM: n=15,164; ASC-US+: n=2359)				<.001
0	547 (3.12)	497 (3.28)	50 (2.12)	
1-2	8498 (48.5)	7448 (49.12)	1050 (44.51)	
≥3	8478 (48.38)	7219 (47.6)	1259 (53.37)	
Parity (overall: n=17,523; NILM: n=15,164; ASC-US+: n=2359)				.005
0	602 (3.43)	543 (3.58)	59 (2.5)	
1-2	14,402 (82.19)	12,474 (82.26)	1928 (81.73)	
≥3	2519 (14.38)	2147 (14.16)	372 (15.77)	
Cervical screening[c] (overall: n=17,514; NILM: n=15,157; ASC-US+: n=2357)				.002
Never	13,621 (77.77)	11,726 (77.36)	1895 (80.4)	
Within 3 years	3047 (17.4)	2674 (17.64)	373 (15.83)	
>3 years ago	846 (4.83)	757 (4.99)	89 (3.78)	
History of other cancers (overall: n=17,523; NILM: n=15,164; ASC-US+: n=2359)				.45
No	16,954 (96.75)	14,665 (96.71)	2289 (97.03)	
Yes	569 (3.25)	499 (3.29)	70 (2.97)	
Postcoital bleeding (overall: n=17,523; NILM: n=15,164; ASC-US+: n=2359)				.32
No	17,369 (99.12)	15,026 (99.09)	2343 (99.32)	

Characteristics	Overall, n (%)	Groups		*P* value
		NILM[a], n (%)	ASC-US+[b], n (%)	
Yes	154 (0.88)	138 (0.91)	16 (0.68)	
Abnormal leukorrhea (overall: n=17,523; NILM: n=15,164; ASC-US+: n=2359)				.92
No	16,471 (94)	14,252 (93.99)	2219 (94.07)	
Yes	1052 (6)	912 (6.01)	140 (5.93)	
PE[e]: cervix abnormality[c] (overall: n=17,413; NILM: n=15,067; ASC-US+: n=2346)				<.001
Normal	10,715 (61.53)	9389 (62.31)	1326 (56.52)	
Abnormal	6698 (38.47)	5678 (37.69)	1020 (43.48)	
Trichomonas infection[c] (overall: n=16,926; NILM: n=14,629; ASC-US+: n=2297)				.96
No	16,411 (96.96)	14,183 (96.95)	2228 (97)	
Yes	515 (3.04)	446 (3.05)	69 (3)	
Candida infection[c] (overall: n=16,926; NILM: n=14,629; ASC-US+: n=2297)				.97
No	16,159 (95.47)	13,967 (95.47)	2192 (95.43)	
Yes	767 (4.53)	662 (4.53)	105 (4.57)	
Gardnerella infection[c] (overall: n=16,926; NILM: n=14,629; ASC-US+: n=2297)				.54
No	16,858 (99.6)	14,568 (99.58)	2290 (99.7)	
Yes	68 (0.4)	61 (0.42)	7 (0.3)	

[a]NILM: negative for intraepithelial lesion or malignancy.

[b]ASC-US+: ThinPrep cytologic test result worse than atypical squamous cells of undetermined significance.

[c]The sum does not equal the total number because of the existence of missing values.

[d]BMI categories: underweight (<18.5), normal (18.5-25), and overweight (≥25).

[e]PE: pelvic examination.

The Risk Factors for Cytological Abnormalities

Table 2 shows the results of the univariate and multivariate logistic regression, which assessed the risk factors of ASC-US+ for participants with non-16/18 hrHPV. A higher incidence of ASC-US+ was observed in women who attended middle or high school (OR 1.31, 95% CI 1.17-1.45; *P*<.001, and OR 1.32, 95% CI 1.14-1.53; *P*<.001, respectively) and those living in rural areas (OR 2.58, 95% CI 2.29-2.90; *P*<.001). The likelihood of ASC-US+ increased with gravidity ≥3 (OR 2.77, 95% CI 1.19-6.45; *P*=.02) and cervix abnormalities detected in PE (OR 1.22, 95% CI 1.11-1.34; *P*<.001). The risk of ASC-US+ was lower in the women who had cervical screening >3 years ago (OR 0.79, 95% CI 0.62-1.00; *P*=.048) than in those with no previous screening. When stratified by rural or urban areas, the results showed that middle or high school education (OR 1.34, 95% CI 1.19-1.50; *P*<.001, and OR 1.42, 95% CI 1.20-1.68; *P*<.001, respectively) and gravidity ≥3 (OR 3.48, 95% CI 1.12-10.82; *P*=.03) were associated with significantly increased risk in women living in rural areas. Cervix abnormalities detected in PE was associated with an increased risk for ASC-US+ in both rural (OR 1.21, 95% CI 1.09-1.34; *P*<.001) and urban (OR 1.28, 95% CI 1.04-1.58; *P*=.02) areas (Figure 2).

Table 2. Risk factors of ASC-US+[a] for participants with non-16/18 high-risk human papillomavirus infection explored by univariate and multivariate logistic regression.

Characteristics	Univariate logistic		Multivariate logistic			
	OR[b] (95% CI)	*P* value	Full model[c]		Simplified model[d]	
			OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years; ref[e]: <40)						
40-50	0.98 (0.86-1.12)	.79	0.99 (0.86-1.14)	.87	N/A[f]	N/A
50-60	0.95 (0.83-1.08)	.44	0.96 (0.80-1.14)	.61	N/A	N/A
≥60	0.99 (0.84-1.17)	.92	1.02 (0.82-1.28)	.83	N/A	N/A
BMI[g] (ref: normal)						
Underweight	1.12 (0.90-1.39)	.30	1.15 (0.93-1.44)	.21	N/A	N/A
Overweight	0.97 (0.87-1.07)	.54	0.94 (0.85-1.05)	.28	N/A	N/A
Education (ref: primary)						
Middle	1.26 (1.14-1.39)	<.001	1.31 (1.17-1.45)	<.001	1.30 (1.18-1.44)	<.001
High	1.07 (0.93-1.22)	.35	1.32 (1.14-1.53)	<.001	1.35 (1.18-1.56)	<.001
Graduate	0.72 (0.61-0.86)	<.001	1.05 (0.87-1.27)	.61	1.09 (0.91-1.31)	.35
Region (ref: urban)						
Rural	2.61 (2.34-2.91)	<.001	2.58 (2.29-2.90)	<.001	2.60 (2.32-2.91)	<.001
Family history of cancer	1.19 (0.89-1.59)	.24	1.04 (0.76-1.43)	.80	N/A	N/A
Menopause	0.96 (0.88-1.05)	.42	0.97 (0.84-1.11)	.65	N/A	N/A
Gravidity (ref: 0)						
1-2	1.40 (1.04-1.89)	.03	2.28 (0.98-5.29)	.06	2.17 (0.99-4.78)	.05
≥3	1.73 (1.29-2.33)	<.001	2.77 (1.19-6.45)	.02	2.67 (1.21-5.88)	.02
Parity (ref: 0)						
1-2	1.42 (1.08-1.87)	.01	0.61 (0.28-1.32)	.21	0.56 (0.27-1.16)	.12
≥3	1.59 (1.19-2.13)	.002	0.61 (0.28-1.35)	.23	0.56 (0.27-1.18)	.13
Screening (ref: never)						
Within 3 years	0.86 (0.77-0.97)	.02	0.94 (0.83-1.07)	.36	0.94 (0.83-1.07)	.34
>3 years ago	0.73 (0.58-0.91)	.006	0.79 (0.62-1.00)	.048	0.81 (0.64-1.01)	.07
History of other cancers	0.90 (0.70-1.16)	.410	0.97 (0.74-1.28)	.85	N/A	N/A
Postcoital bleeding	0.74 (0.44-1.25)	.264	0.71 (0.41-1.23)	.22	N/A	N/A
Abnormal leukorrhea	0.99 (0.82-1.18)	.88	0.89 (0.74-1.08)	.26	N/A	N/A
PE[h]: cervix abnormality	1.27 (1.16-1.39)	<.001	1.22 (1.11-1.34)	<.001	1.23 (1.13-1.35)	<.001
Trichomonas infection	0.98 (0.76-1.27)	.91	0.85 (0.65-1.10)	.22	N/A	N/A
Candida infection	1.01 (0.82-1.25)	.92	0.91 (0.73-1.13)	.40	N/A	N/A
Gardnerella infection	0.73 (0.33-1.60)	.43	0.69 (0.31-1.52)	.36	N/A	N/A

[a]ASC-US+: ThinPrep cytologic test result worse than atypical squamous cells of undetermined significance.

[b]OR: odds ratio.

[c]Full model: including all variables.

[d]Simplified model: including the variables with *P*<.10 in the univariate logistic regression.

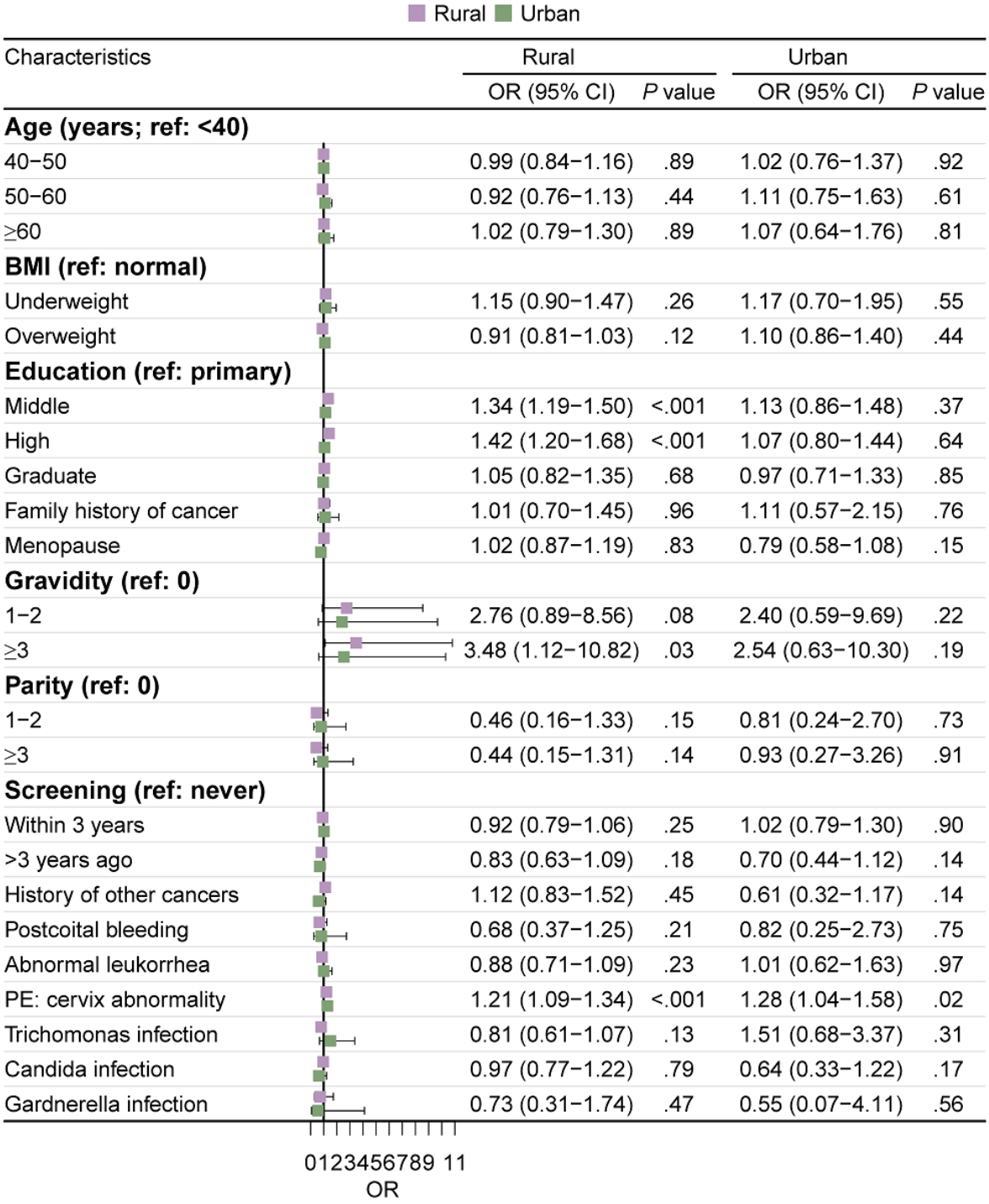
[e]ref: reference.

[f]N/A: not applicable.

[g]BMI categories: underweight (<18.5), normal (18.5-25), and overweight (≥25).

[h]PE: pelvic examination.

Figure 2. Multivariate logistic regression analysis stratified on area to explore risk factors for cytological abnormalities among individuals infected with non-16/18 high-risk human papillomavirus. OR: odds ratio; PE: pelvis examination.



Characteristics	Rural		Urban	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years; ref: <40)				
40−50	0.99 (0.84−1.16)	.89	1.02 (0.76−1.37)	.92
50−60	0.92 (0.76−1.13)	.44	1.11 (0.75−1.63)	.61
≥60	1.02 (0.79−1.30)	.89	1.07 (0.64−1.76)	.81
BMI (ref: normal)				
Underweight	1.15 (0.90−1.47)	.26	1.17 (0.70−1.95)	.55
Overweight	0.91 (0.81−1.03)	.12	1.10 (0.86−1.40)	.44
Education (ref: primary)				
Middle	1.34 (1.19−1.50)	<.001	1.13 (0.86−1.48)	.37
High	1.42 (1.20−1.68)	<.001	1.07 (0.80−1.44)	.64
Graduate	1.05 (0.82−1.35)	.68	0.97 (0.71−1.33)	.85
Family history of cancer	1.01 (0.70−1.45)	.96	1.11 (0.57−2.15)	.76
Menopause	1.02 (0.87−1.19)	.83	0.79 (0.58−1.08)	.15
Gravidity (ref: 0)				
1−2	2.76 (0.89−8.56)	.08	2.40 (0.59−9.69)	.22
≥3	3.48 (1.12−10.82)	.03	2.54 (0.63−10.30)	.19
Parity (ref: 0)				
1−2	0.46 (0.16−1.33)	.15	0.81 (0.24−2.70)	.73
≥3	0.44 (0.15−1.31)	.14	0.93 (0.27−3.26)	.91
Screening (ref: never)				
Within 3 years	0.92 (0.79−1.06)	.25	1.02 (0.79−1.30)	.90
>3 years ago	0.83 (0.63−1.09)	.18	0.70 (0.44−1.12)	.14
History of other cancers	1.12 (0.83−1.52)	.45	0.61 (0.32−1.17)	.14
Postcoital bleeding	0.68 (0.37−1.25)	.21	0.82 (0.25−2.73)	.75
Abnormal leukorrhea	0.88 (0.71−1.09)	.23	1.01 (0.62−1.63)	.97
PE: cervix abnormality	1.21 (1.09−1.34)	<.001	1.28 (1.04−1.58)	.02
Trichomonas infection	0.81 (0.61−1.07)	.13	1.51 (0.68−3.37)	.31
Candida infection	0.97 (0.77−1.22)	.79	0.64 (0.33−1.22)	.17
Gardnerella infection	0.73 (0.31−1.74)	.47	0.55 (0.07−4.11)	.56

The Risk Factors for High-Grade Cytological Abnormalities

Table 3 shows the risk factors for high-grade cytological abnormalities. Education, region, cervical screening, and cervix abnormalities detected in PE were included in the multivariate analysis as their *P* values were <.10 in the univariate analysis. Among these factors, significant differences were observed with middle school education (OR 1.45, 95% CI 1.07-1.98; *P*=.02), rural region (OR 1.52, 95% CI 1.10-2.10; *P*=.01), and cervix abnormality (OR 1.72, 95% CI 1.30-2.26; *P*<.001).

Table 3. Risk factors of high-grade cytological abnormalities for participants with non-16/18 high-risk human papillomavirus infection explored by univariate and multivariate logistic regression.

Characteristics	Univariate logistic		Multivariate logistic			
	OR[a] (95% CI)	*P* value	Full model[b]		Simplified model[c]	
			OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years; ref[d]: <40)						
40-50	0.98 (0.86-1.12)	.79	0.99 (0.86-1.14)	.87	N/A[e]	N/A
50-60	0.95 (0.83-1.08)	.44	0.96 (0.80-1.14)	.61	N/A	N/A
≥60	0.99 (0.84-1.17)	.92	1.02 (0.82-1.28)	.83	N/A	N/A
BMI[f] (ref: normal)						
Underweight	1.12 (0.90-1.39)	.30	1.15 (0.93-1.44)	.21	N/A	N/A
Overweight	0.97 (0.87-1.07)	.54	0.94 (0.85-1.05)	.28	N/A	N/A
Education (ref: primary)						
Middle	1.26 (1.14-1.39)	<.001	1.31 (1.17-1.45)	<.001	1.30 (1.18-1.44)	<.001
High	1.07 (0.93-1.22)	.35	1.32 (1.14-1.53)	<.001	1.35 (1.18-1.56)	<.001
Graduate	0.72 (0.61-0.86)	<.001	1.05 (0.87-1.27)	.61	1.09 (0.91-1.31)	.35
Region (ref: urban)						
Rural	2.61 (2.34-2.91)	<.001	2.58 (2.29-2.90)	<.001	2.60 (2.32-2.91)	<.001
Family history of cancer	1.19 (0.89-1.59)	.24	1.04 (0.76-1.43)	.80	N/A	N/A
Menopause	0.96 (0.88-1.05)	.42	0.97 (0.84-1.11)	.65	N/A	N/A
Gravidity (ref: 0)						
1-2	1.40 (1.04-1.89)	.03	2.28 (0.98-5.29)	.06	2.17 (0.99-4.78)	.05
≥3	1.73 (1.29-2.33)	<.001	2.77 (1.19-6.45)	.02	2.67 (1.21-5.88)	.02
Parity (ref: 0)						
1-2	1.42 (1.08-1.87)	.01	0.61 (0.28-1.32)	.21	0.56 (0.27-1.16)	.12
≥3	1.59 (1.19-2.13)	.002	0.61 (0.28-1.35)	.23	0.56 (0.27-1.18)	.13
Screening (ref: never)						
Within 3 years	0.86 (0.77-0.97)	.02	0.94 (0.83-1.07)	.36	0.94 (0.83-1.07)	.34
>3 years ago	0.73 (0.58-0.91)	.006	0.79 (0.62-1.00)	.048	0.81 (0.64-1.01)	.07
History of other cancers	0.90 (0.70-1.16)	.410	0.97 (0.74-1.28)	.85	N/A	N/A
Postcoital bleeding	0.74 (0.44-1.25)	.264	0.71 (0.41-1.23)	.22	N/A	N/A
Abnormal leukorrhea	0.99 (0.82-1.18)	.88	0.89 (0.74-1.08)	.26	N/A	N/A
PE[g]: cervix abnormality	1.27 (1.16-1.39)	<.001	1.22 (1.11-1.34)	<.001	1.23 (1.13-1.35)	<.001
Trichomonas infection	0.98 (0.76-1.27)	.91	0.85 (0.65-1.10)	.22	N/A	N/A
Candida infection	1.01 (0.82-1.25)	.92	0.91 (0.73-1.13)	.40	N/A	N/A
Gardnerella infection	0.73 (0.33-1.60)	.43	0.69 (0.31-1.52)	.36	N/A	N/A

[a]OR: odds ratio.

[b]Full model: including all variables.

[c]Simplified model: including the variables with *P*<.10 in the univariate logistic regression.

[d]ref: reference.

[e]N/A: not applicable.

[f]BMI categories: underweight (<18.5), normal (18.5-25), and overweight (≥ 25).

[g]PE: pelvic examination.

Discussion

Principal Findings

Middle or high school education, living in rural areas, gravidity ≥3, and cervix abnormalities detected in PE were the risk factors for ASC-US+ in this study. In addition, receiving cervical screening >3 years ago was negatively associated with the prevalence of ASC-US+ among women with non-16/18 hrHPV infections. Our findings may have important implications for the prevention and control of cervical cancer in non-16/18 hrHPV–positive individuals. High-risk groups identified by their risk factors should be carefully diagnosed and treated according to medical advice to prevent adverse outcomes.

We observed that age had no effect in this study. Considering the large sample size of this study (n=17,523) and broad age range (from 30 to 78 years old), we believe that the result of null effect of age on cytological abnormalities in women infected with non-16/18 hrHPV is reliable. Some previous studies also did not find an effect of age on cytological abnormalities in all women or those infected with HPV [24,25,28]. However, in some studies, the risk of cytological abnormalities increased significantly with age [22,26]. This inconsistency may be due to differences in race, social environment, behavior, and habits in different areas.

Education was an important risk factor for cytological abnormalities. Women with middle and high school education were more likely to have cytological abnormalities than those with primary school education. Previous studies have also shown that women with middle and high school education are at a higher risk for cervical cancer [4,27,35]. The reason may be that women with primary school education tend to marry earlier and have more stable sexual partners. Previous studies have reported that both women and their husbands' lifetime number of sexual partners were significantly positively correlated with cervical cancer risk [36].

Women in rural areas had a higher probability of cytological abnormalities. Poor sanitation, insufficient knowledge about cervical cancer, and poor awareness of prevention in rural areas [37] could increase vulnerability to cervical cancer. In addition, women in rural areas have a lower frequency of gynecologic examination and cervical cancer screening than those in urban areas [38], resulting in an inability to detect abnormalities and receive timely treatment. Therefore, efforts should be intensified in rural areas to popularize cervical cancer prevention knowledge and reduce the incidence of cervical cancer. Furthermore, risk factors for cytological abnormalities differ in rural and urban areas. Among rural women, middle or high school education and gravidity ≥3 were associated with an increased risk of cytological abnormalities, whereas such results were not observed in urban women. This finding means that narrowing and eventually addressing the socioeconomic gap is imperative for cervical cancer prevention.

The prevalence of cytological abnormalities significantly increased when gravidity was ≥3, which may be related to hormonal changes during pregnancy [39]. Female sex hormones (estrogen and progesterone) may affect immune function [40]. Unstable sex hormone levels reduce immunity in women, thus lowering the resistance to hrHPV, weakening the ability to clear hrHPV, and resulting in an increased probability of cytological abnormalities. Women with high gravidity who are infected with hrHPV are recommended to consult their physician for further diagnosis in a timely manner. In addition to complying with the cervical cancer screening guidelines [41], it is recommended that women who are infected with non-16/18 hrHPV undergo HPV examination and cytology test again 1 year later, even if their TCT results were NILM.

Women with cervix abnormalities in PE are more likely to have cytological abnormalities. Previous studies have shown that the appearance of the cervix is correlated with the incidence of cervical cancer [42]. In the United Kingdom, both clinical practice guidelines on the diagnosis of cancer [43] and the National Institute for Health and Care Excellence guidelines [44] recommend visualizing the cervix to facilitate timely diagnosis of women with cervical cancer. Although no such guidelines exist in the United States, the American College of Obstetricians and Gynecologists Committee on Gynecologic Practice suggests a similar approach [45]. Therefore, PE is recommended to be added to the physical examination in women to detect the abnormal appearance of the cervix and facilitate early treatment, thereby lowering the incidence of cervical cancer.

Some cohort studies have shown that cervical cytology screening can reduce the incidence of cervical cancer by detecting precancerous lesions and early-stage cancer [18,46]. We found that cervical screening performed >3 years ago was a protective factor against cytological abnormalities. However, such protective effects were not observed when screening was performed within 3 years. Women with cytological abnormalities are particularly recommended to undergo regular follow-up cytological screening to monitor the progression or regression of cervical abnormalities. Women who screened for cervical cancer within 3 years were more likely to have previous cervical abnormalities than those screened >3 years ago. Further, women who were screened for cervical cancer >3 years ago were likely to have normal results on their last cervical cancer examination, indicating a low risk of current cytological abnormalities. Undoubtedly, well-organized screening programs have been documented to reduce the incidence and mortality of cervical cancer [17,47,48]. Women are advised to adhere to the Cervical Cancer Screening Program, which is expected to expand worldwide. It is recommended that women with non-16/18 hrHPV–positive status undergo regular cervical cancer screenings regardless of disease status and follow up with doctors if abnormalities are detected upon screening.

Comparison With Prior Work

To the best of our knowledge, this is the first study investigating cytological abnormalities in women infected with non-16/18 hrHPV. A few previous studies have explored the influencing factors of cytological abnormalities in all women; however, they did not focus on this overlooked subpopulation of those infected with non-16/18 hrHPV. Compared with previous studies, one of the strengths of this study is the large sample size of 17,523 individuals collected from multiple centers, which

guarantees high statistical power and good precision of the estimates. In addition, we considered other potential influencing factors, including demographic characteristics, menstruation and fertility, PE results, and vaginal microenvironment infection.

Limitations

Our study has some limitations. First, this was a cross-sectional study without detailed information from previous screening results, and all subjects were infected with non-16/18 hrHPV detected by the current screening. The HPV genotype was not considered in this study because the selected Cobas HPV test could not detect specific types of non-16/18 hrHPV. This information on specific HPV genotypes and the persistence of infection may have an impact on abnormalities according to previous research [27,49]. Second, this study included only Chinese women; the risk factors for cytological abnormalities may differ according to ethnicity, social environment, and behavioral habits. Therefore, caution should be exercised when extrapolating the conclusions to other populations. Third, personal behaviors, such as cigarette smoking and long-term oral contraceptive use, which have been proven to be cofactors in cervical cancer [50], were not controlled in our study. As a result, the relationship between these factors and cytological abnormalities could not be investigated. Finally, reporting and recall biases may exist because of the use of a self-reported questionnaire.

Conclusion

This large-scale, cross-sectional study assessed the prevalence and risk factors of cytological abnormalities in 17,523 Chinese women infected with non-16/18 hrHPV. Middle or high school education, living in rural areas, gravidity ≥3, and cervix abnormalities detected in PE were found to be risk factors for cytological abnormalities, whereas receiving cervical screening >3 years ago was associated with a reduced prevalence of cytological abnormalities. In addition, middle school education, living in rural regions, and cervix abnormality were risk factors for high-grade cytological abnormalities. More attention should be paid to improving diagnostic, management, and vaccination strategies among individuals with non-16/18 hrHPV infections.

Acknowledgments

The authors are grateful to the clinical staff of the hospitals who participated in the cervical cancer screening project in Xiangyang City, China, from January 2017 to February 2018. This work was supported by the National Natural Science Foundation of China (82002192, 82003552, and 81972449) and the Guangdong Basic and Applied Basic Research Foundation (2020A1515011161 and 2021A1515220065).

Data Availability

The data are not publicly available because of privacy and ethical restrictions. The data supporting the findings of this study are available upon request from the corresponding author.

Authors' Contributions

HX, C-QO, and JY contributed equally to the correspondence work. HX, C-QO, and JY initiated the study. HX and MY collected the data. TX cleaned the data and performed the statistical analysis. TX, JY, and C-QO drafted the manuscript. CW, TY, XX, LS, and HX revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire in both English and Chinese.
[PDF File (Adobe PDF File), 74 KB - publichealth_v8i12e38628_app1.pdf]

References

1. 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 May;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Qiu H, Cao S, Xu R. Cancer incidence, mortality, and burden in China: a time-trend analysis and comparison with the United States and United Kingdom based on the global epidemiological data released in 2020. Cancer Commun (Lond) 2021 Oct;41(10):1037-1048 [FREE Full text] [doi: 10.1002/cac2.12197] [Medline: 34288593]
3. Cao W, Chen H, Yu Y, Li N, Chen W. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. Chin Med J (Engl) 2021 Mar 17;134(7):783-791 [FREE Full text] [doi: 10.1097/CM9.0000000000001474] [Medline: 33734139]
4. Wu P, Xiong H, Yang M, Li L, Wu P, Lazare C, et al. Co-infections of HPV16/18 with other high-risk HPV types and the risk of cervical carcinogenesis: a large population-based study. Gynecol Oncol 2019 Dec;155(3):436-443. [doi: 10.1016/j.ygyno.2019.10.003] [Medline: 31604662]

5. Walboomers JMM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999 Sep;189(1):12-19. [doi: 10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F] [Medline: 10451482]
6. Jun S, Park ES, Kim J, Kang J, Lee JJ, Bae Y, et al. Comparison of the Cobas 4800 HPV and HPV 9G DNA chip tests for detection of high-risk human papillomavirus in cervical specimens of women with consecutive positive HPV tests But negative Pap smears. PLoS One 2015 Oct 15;10(10):e0140336 [FREE Full text] [doi: 10.1371/journal.pone.0140336] [Medline: 26469982]
7. Castellsagué X. Natural history and epidemiology of HPV infection and cervical cancer. Gynecol Oncol 2008 Sep;110(3 Suppl 2):S4-S7. [doi: 10.1016/j.ygyno.2008.07.045] [Medline: 18760711]
8. Robadi IA, Pharaon M, Ducatman BS. The importance of high-risk human papillomavirus types other than 16 and 18 in cervical neoplasia. Arch Pathol Lab Med 2018 Jun;142(6):693-695 [FREE Full text] [doi: 10.5858/arpa.2017-0563-RA] [Medline: 29848032]
9. Xu HH, Lin A, Chen YH, Dong SS, Shi WW, Yu JZ, et al. Prevalence characteristics of cervical human papillomavirus (HPV) genotypes in the Taizhou area, China: a cross-sectional study of 37 967 women from the general population. BMJ Open 2017 Jun 09;7(6):e014135 [FREE Full text] [doi: 10.1136/bmjopen-2016-014135] [Medline: 28601819]
10. Mammas IN, Vageli D, Spandidos DA. Geographic variations of human papilloma virus infection and their possible impact on the effectiveness of the vaccination programme. Oncol Rep 2008 Jul;20(1):141-145. [Medline: 18575729]
11. Bao H, Jin C, Wang S, Song Y, Xu Z, Yan X, et al. Prevalence of cervicovaginal human papillomavirus infection and genotypes in the pre-vaccine era in China: a nationwide population-based study. J Infect 2021 Apr;82(4):75-83. [doi: 10.1016/j.jinf.2021.02.017] [Medline: 33610682]
12. Lyons YA, Kamat AA, Zhou H, Mody DR, Schwartz MR, Hobday C, et al. Non-16/18 high-risk HPV infection predicts disease persistence and progression in women with an initial interpretation of LSIL. Cancer Cytopathol 2015 Jul;123(7):435-442 [FREE Full text] [doi: 10.1002/cncy.21549] [Medline: 25903015]
13. Dong L, Hu S, Zhang Q, Feng R, Zhang L, Zhao X, et al. Risk prediction of cervical cancer and precancers by type-specific human papillomavirus: evidence from a population-based cohort study in China. Cancer Prev Res (Phila) 2017 Dec;10(12):745-751. [doi: 10.1158/1940-6207.CAPR-17-0088] [Medline: 28916509]
14. Mendoza RP, Haidary T, Gabutan E, Zhou YY, Bukhari Z, Connelly C, et al. Mixed and nonvaccine high risk HPV types are associated with higher mortality in Black women with cervical cancer. Sci Rep 2021 Jul 07;11(1):14064 [FREE Full text] [doi: 10.1038/s41598-021-93485-1] [Medline: 34234252]
15. Schiffman M, Vaughan LM, Raine-Bennett TR, Castle PE, Katki HA, Gage JC, et al. A study of HPV typing for the management of HPV-positive ASC-US cervical cytologic results. Gynecol Oncol 2015 Sep;138(3):573-578 [FREE Full text] [doi: 10.1016/j.ygyno.2015.06.040] [Medline: 26148763]
16. Wang J, Andrae B, Sundström K, Ström P, Ploner A, Elfström KM, et al. Risk of invasive cervical cancer after atypical glandular cells in cervical screening: nationwide cohort study. BMJ 2016 Feb 11;352:i276 [FREE Full text] [doi: 10.1136/bmj.i276] [Medline: 26869597]
17. Fontham ETH, Wolf AMD, Church TR, Etzioni R, Flowers CR, Herzig A, et al. Cervical cancer screening for individuals at average risk: 2020 guideline update from the American Cancer Society. CA Cancer J Clin 2020 Sep;70(5):321-346 [FREE Full text] [doi: 10.3322/caac.21628] [Medline: 32729638]
18. Pry JM, Manasyan A, Kapambwe S, Taghavi K, Duran-Frigola M, Mwanahamuntu M, et al. Cervical cancer screening outcomes in Zambia, 2010-19: a cohort study. Lancet Glob Health 2021 Jun;9(6):e832-e840 [FREE Full text] [doi: 10.1016/S2214-109X(21)00062-0] [Medline: 34019837]
19. Marcus JZ, Cason P, Downs LS, Einstein MH, Flowers L. The ASCCP Cervical Cancer Screening Task Force endorsement and opinion on the American Cancer Society updated cervical cancer screening guidelines. J Low Genit Tract Dis 2021 Jul 01;25(3):187-191. [doi: 10.1097/LGT.0000000000000614] [Medline: 34138787]
20. Moy LM, Zhao F, Li L, Ma J, Zhang Q, Chen F, et al. Human papillomavirus testing and cervical cytology in primary screening for cervical cancer among women in rural China: comparison of sensitivity, specificity, and frequency of referral. Int J Cancer 2010 Aug 01;127(3):646-656 [FREE Full text] [doi: 10.1002/ijc.25071] [Medline: 19960441]
21. Terasawa T, Hosono S, Sasaki S, Hoshi K, Hamashima Y, Katayama T, et al. Comparative accuracy of cervical cancer screening strategies in healthy asymptomatic women: a systematic review and network meta-analysis. Sci Rep 2022 Jan 07;12(1):94 [FREE Full text] [doi: 10.1038/s41598-021-04201-y] [Medline: 34997127]
22. Zhong G, Wang Y, Xie Q, Lin R, Yao T. HPV-specific risk assessment of cervical cytological abnormalities. BMC Cancer 2021 Aug 24;21(1):949 [FREE Full text] [doi: 10.1186/s12885-021-08703-w] [Medline: 34429079]
23. Thomas JO, Ojemakinde KO, Ajayi IO, Omigbodun AO, Fawole OI, Oladepo O. Population-based prevalence of abnormal cervical cytology findings and local risk factors in Ibadan, Nigeria: implications for cervical cancer control programs and human papilloma virus immunization. Acta Cytol 2012;56(3):251-258. [doi: 10.1159/000337444] [Medline: 22555526]
24. Al Sekri E, Al Musalhi A, Al Abri K, Al Kindi H, Al Salmani AA. Prevalence of cytological abnormalities in Papanicolaou smears and risk factors for cervical cancer among women in Muscat, Oman. Sultan Qaboos Univ Med J 2021 Nov 25;21(4):598-603 [FREE Full text] [doi: 10.18295/squmj.4.2021.041] [Medline: 34888080]

25. Thorsteinsson K, Storgaard M, Katzenstein TL, Ladelund S, Rønsholt FF, Johansen IS, et al. Prevalence and distribution of cervical high-risk human papillomavirus and cytological abnormalities in women living with HIV in Denmark - the SHADE. BMC Cancer 2016 Nov 08;16(1):866 [FREE Full text] [doi: 10.1186/s12885-016-2881-1] [Medline: 27821088]
26. Chaung KV, Zheng Y, Martella AT, Stoecker JB, Cote DR, Augustine JJ, et al. Risk factors for abnormal cervical cytology in women undergoing kidney transplant evaluation. Exp Clin Transplant 2019 Feb;17(1):31-36 [FREE Full text] [doi: 10.6002/ect.2017.0064] [Medline: 29251578]
27. Jing L, Zhong X, Huang W, Liu Y, Wang M, Miao Z, et al. HPV genotypes and associated cervical cytological abnormalities in women from the Pearl River Delta region of Guangdong province, China: a cross-sectional study. BMC Infect Dis 2014 Jul 12;14:388 [FREE Full text] [doi: 10.1186/1471-2334-14-388] [Medline: 25016305]
28. Rizzuto I, Nalam M, Jiang J, Linder A, Rufford B. Risk factors for HPV persistence and cytology anomalies at follow-up after treatment for cervical dysplasia. Int J Gynaecol Obstet 2018 May 21;141(2):240-244. [doi: 10.1002/ijgo.12431] [Medline: 29250779]
29. Liang Y, Chen M, Qin L, Wan B, Wang H. A meta-analysis of the relationship between vaginal microecology, human papillomavirus infection and cervical intraepithelial neoplasia. Infect Agent Cancer 2019 Oct 26;14(1):29 [FREE Full text] [doi: 10.1186/s13027-019-0243-8] [Medline: 31673281]
30. Yang M, Li L, Jiang C, Qin X, Zhou M, Mao X, et al. Co-infection with trichomonas vaginalis increases the risk of cervical intraepithelial neoplasia grade 2-3 among HPV16 positive female: a large population-based study. BMC Infect Dis 2020 Sep 01;20(1):642 [FREE Full text] [doi: 10.1186/s12879-020-05349-0] [Medline: 32873233]
31. Lim AW, Ramirez AJ, Hamilton W, Sasieni P, Patnick J, Forbes LJ. Delays in diagnosis of young females with symptomatic cervical cancer in England: an interview-based study. Br J Gen Pract 2014 Oct;64(627):e602-e610 [FREE Full text] [doi: 10.3399/bjgp14X681757] [Medline: 25267045]
32. Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M, Forum Group Members, Bethesda 2001 Workshop. The 2001 Bethesda System: terminology for reporting results of cervical cytology. JAMA 2002 Apr 24;287(16):2114-2119. [doi: 10.1001/jama.287.16.2114] [Medline: 11966386]
33. Liu C, Li L, Song K, Zhan Z, Yao Y, Gong H, et al. A nomogram for predicting mortality in patients with COVID-19 and solid tumors: a multicenter retrospective cohort study. J Immunother Cancer 2020 Sep;8(2):e1314 [FREE Full text] [doi: 10.1136/jitc-2020-001314] [Medline: 32895296]
34. Macios A, Didkowska J, Wojciechowska U, Komerska K, Glińska P, Kamiński MF, et al. Risk factors of cervical cancer after a negative cytological diagnosis in Polish cervical cancer screening programme. Cancer Med 2021 May;10(10):3449-3460 [FREE Full text] [doi: 10.1002/cam4.3857] [Medline: 33934537]
35. Vesco KK, Whitlock EP, Eder M, Burda BU, Senger CA, Lutz K. Risk factors and other epidemiologic considerations for cervical cancer screening: a narrative review for the U.S. Preventive Services Task Force. Ann Intern Med 2011 Nov 15;155(10):698-705, W216 [FREE Full text] [doi: 10.7326/0003-4819-155-10-201111150-00377] [Medline: 22006929]
36. Chelimo C, Wouldes TA, Cameron LD, Elwood JM. Risk factors for and prevention of human papillomaviruses (HPV), genital warts and cervical cancer. J Infect 2013 Mar;66(3):207-217. [doi: 10.1016/j.jinf.2012.10.024] [Medline: 23103285]
37. Song B, Ding C, Chen W, Sun H, Zhang M, Chen W. Incidence and mortality of cervical cancer in China, 2013. Chin J Cancer Res 2017 Dec;29(6):471-476 [FREE Full text] [doi: 10.21147/j.issn.1000-9604.2017.06.01] [Medline: 29353969]
38. Wen X, Wen D, Yang Y, Chen Y, Akazawa K, Liu Y, et al. Urban-rural disparity in cervical cancer in China and feasible interventions for tackling the rural excess. Medicine (Baltimore) 2019 Jan;98(1):e13907 [FREE Full text] [doi: 10.1097/MD.0000000000013907] [Medline: 30608415]
39. Song L, Lyu Y, Ding L, Li X, Gao W, Wang M, et al. revalence and genotype distribution of high-risk human papillomavirus infection in women with abnormal cervical cytology: a population-based study in Shanxi province, China. Cancer Manag Res 2020 Jun 23;12:12583-12591 [FREE Full text] [doi: 10.2147/CMAR.S269050] [Medline: 33324103]
40. Kaushic C, Roth KL, Anipindi V, Xiu F. Increased prevalence of sexually transmitted viral infections in women: the role of female sex hormones in regulating susceptibility and immune responses. J Reprod Immunol 2011 Mar;88(2):204-209. [doi: 10.1016/j.jri.2010.12.004] [Medline: 21296427]
41. Huh WK, Ault KA, Chelmow D, Davey DD, Goulart RA, Garcia FAR, et al. Use of primary high-risk human papillomavirus testing for cervical cancer screening: interim clinical guidance. Gynecol Oncol 2015 Feb;136(2):178-182. [doi: 10.1016/j.ygyno.2014.12.022] [Medline: 25579107]
42. Bates CK, Carroll N, Potter J. The challenging pelvic examination. J Gen Intern Med 2011 Jun;26(6):651-657 [FREE Full text] [doi: 10.1007/s11606-010-1610-8] [Medline: 21225474]
43. Clinical practice guidance for the assessment of young women aged 20-24 with abnormal vaginal bleeding. Department of Health. 2010 Mar. URL: https://webarchive.nationalarchives.gov.uk/ukgwa/20130123201324/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_113478 [accessed 2022-08-14]
44. Suspected cancer: recognition and referral. National Institute for Health and Care Excellence (NICE). 2021 Dec 15. URL: https://www.nice.org.uk/guidance/ng12 [accessed 2022-12-01]
45. American College of Obstetricians and Gynecologists. ACOG committee opinion no. 557: management of acute abnormal uterine bleeding in nonpregnant reproductive-aged women. Obstet Gynecol 2013 Apr;121(4):891-896. [doi: 10.1097/01.AOG.0000428646.67925.9a] [Medline: 23635706]

46. Bucchi D, Chiavarini M, Bianconi F, Galeotti M, Gili A, Stracci F. Immigration, screening, and cervical cancer incidence: an application of Age-Period-Cohort analysis. Eur J Cancer Prev 2019 Nov;28(6):529-536 [FREE Full text] [doi: 10.1097/CEJ.0000000000000495] [Medline: 30431461]
47. Yabroff KR, Lawrence WF, King JC, Mangan P, Washington KS, Yi B, et al. Geographic disparities in cervical cancer mortality: what are the roles of risk factor prevalence, screening, and use of recommended treatment? J Rural Health 2005 Mar;21(2):149-157. [doi: 10.1111/j.1748-0361.2005.tb00075.x] [Medline: 15859052]
48. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin 2021 Jan 12;71(1):7-33 [FREE Full text] [doi: 10.3322/caac.21654] [Medline: 33433946]
49. Rezhake R, Chen F, Hu S, Zhao X, Zhang X, Cao J, et al. Triage options to manage high-risk human papillomavirus-positive women: a population-based cross-sectional study from rural China. Int J Cancer 2020 Oct 15;147(8):2053-2064 [FREE Full text] [doi: 10.1002/ijc.33001] [Medline: 32249409]
50. Gierisch JM, Coeytaux RR, Urrutia RP, Havrilesky LJ, Moorman PG, Lowery WJ, et al. Oral contraceptive use and risk of breast, cervical, colorectal, and endometrial cancers: a systematic review. Cancer Epidemiol Biomarkers Prev 2013 Nov;22(11):1931-1943. [doi: 10.1158/1055-9965.EPI-13-0298] [Medline: 24014598]

Abbreviations

AGC: atypical glandular cells
ASC-H: atypical squamous cells not possible excluding high-grade squamous intraepithelial lesion
ASC-US: atypical squamous cells of undetermined significance
ASC-US+: ThinPrep cytologic test result worse than atypical squamous cells of undetermined significance
GVIF: generalized variance inflation factor
HPV: human papillomavirus
hrHPV: high-risk human papillomavirus
HSIL: high-grade squamous intraepithelial lesion
LSIL: low-grade squamous intraepithelial lesion
NILM: negative for intraepithelial lesion or malignancy
OR: odds ratio
PE: pelvic examination
TCT: ThinPrep cytologic test

Edited by A Mavragani, G Eysenbach; submitted 10.04.22; peer-reviewed by M Li, C Liu; comments to author 25.07.22; revised version received 14.08.22; accepted 11.10.22; published 08.12.22.

Please cite as:
Xiao T, Ou CQ, Yang J, Wang C, Yang M, Yu T, Shen L, Xu X, Xing H
The Risk Factors for Cervical Cytological Abnormalities Among Women Infected With Non-16/18 High-Risk Human Papillomavirus: Cross-sectional Study
JMIR Public Health Surveill 2022;8(12):e38628
URL: https://publichealth.jmir.org/2022/12/e38628
doi:10.2196/38628
PMID:36480259

Original Paper

Incidence of and Experiences with Abortion Attempts in Soweto, South Africa: Respondent-Driven Sampling Study

Caitlin Gerdts[1], MHS, PhD; Ruvani T Jayaweera[1], MPH, PhD; Relebohile Motana[2], BA; Tshegofatso Bessenaar[2], BA, MA; Paul Wesson[3], BA, PhD

[1]Ibis Reproductive Health, Oakland, CA, United States

[2]Ibis Reproductive Health, Johanesburg, South Africa

[3]Department of Epidemiology and Biostatistics, School of Medicine, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Caitlin Gerdts, MHS, PhD
Ibis Reproductive Health
1736 Franklin St
Oakland, CA, 94612
United States
Phone: 1 510 822 2686
Email: cgerdts@ibisreproductivehealth.org

Abstract

Background: Estimation of abortion incidence, particularly in settings where most abortions occur outside of health facility settings, is critical for understanding information gaps and service delivery needs in different settings. However, the existing methods for measuring out-of-facility abortion incidence are plagued with methodological challenges. Respondent-driven sampling (RDS) may offer a methodological improvement in the estimation of abortion incidence.

Objective: In this study, we tested the feasibility of using RDS to recruit participants into a study about abortion and estimated the proportion of people who ever attempted abortion as well as 1-year and 5-year incidence of abortion (both in-facility and out-of-facility settings) among women of reproductive age in Soweto, South Africa.

Methods: Participants were eligible if they identified as a woman; were aged between 15 and 49 years; spoke English, Tswana, isiZulu, Sotho, or Xhosa; and lived in Soweto. Working with community partners, we identified 11 seeds who were provided with coupons to refer eligible peers to the study. Upon arrival at the study site, the recruits completed an interviewer-administered questionnaire that solicited information about demographic characteristics, social network composition, health behaviors, sexual history, pregnancy history, and experience with abortion; recruits also received 3 recruitment coupons. Recruitment was tracked using coupon numbering. We used the RDS-II estimator to estimate the population proportions of demographic characteristics and our primary outcome, the proportion of people who ever attempted abortion.

Results: Between April 4, 2018, and December 17, 2018, 849 eligible participants were recruited into the study. The estimated proportion of people who ever attempted abortion was 12.1% (95% CI 9.7%-14.4%). A total of 7.1% (95% CI 5.4%-8.9%) reported a facility-based abortion, and 4.4% (95% CI 3.0%-5.8%) reported an out-of-facility abortion.

Conclusions: The estimated proportion of people who ever attempted abortion of 12% (102/849) in our study likely represents a substantial underestimation of the actual proportion of abortion attempts among this study population—representing a failure of the RDS method to generate more reliable estimates of abortion incidence in our study. We caution against the use of RDS to measure the incidence of abortion because of persistent concerns with underreporting but consider potential alternative applications of RDS with respect to the study of abortion.

(JMIR Public Health Surveill 2022;8(12):e38045) doi:10.2196/38045

KEYWORDS

induced abortion; respondent-driven sampling; self-managed abortion; abortion incidence

Introduction

In contexts where abortion is legally restricted or where other barriers exist, abortion commonly occurs without the involvement of the formal health care sector [1,2] using a variety of methods ranging from safe, World Health Organization–recommended medications [3] to the ingestion of harmful substances. Forthwith, we will refer to all such abortions as *out-of-facility* abortions. The most recent global estimates suggest that approximately 45% of abortions worldwide from 2010 to 2014 took place outside a health care facility, whereas in specific settings, out-of-facility abortions comprised 70% to 80% of all abortions [4]. Researchers have studied out-of-facility abortion for decades [5]; however, the existing data sources on out-of-facility abortions often suffer from selection bias, misclassification, and underreporting and have led to documented underestimates of abortion incidence [6] and unreliable data on the characteristics of abortion seekers and outcomes of abortion in such contexts [7,8].

Multiple innovations in the estimation of out-of-facility abortion incidence have been tested in recent years, none of which have emerged as a reliable gold standard [9-12]. As out-of-facility abortion becomes an increasingly common and supported model for abortion around the globe, there is a pressing need for new and innovative research methods that can more accurately measure the prevalence, incidence, and characteristics of out-of-facility abortions.

Respondent-driven sampling (RDS), a sampling methodology that relies upon social networks to identify populations engaging in stigmatized, illicit, or otherwise hidden behaviors, may offer a previously untested alternative to measuring out-of-facility abortion. Studies that use RDS begin with a small nonrandom sample of point people (ie, seeds) within social networks engaging in hidden or stigmatized behaviors, who are interviewed and provided with referral coupons to recruit others within the same social network (ie, the target population). RDS has been used to estimate the prevalence of sensitive and illegal behaviors among hidden populations such as people who inject drugs, sex workers, and men who have sex with men; and relies upon social networks to identify populations for whom no valid sampling frame exists [13-19]. To account for potential selection bias because of peer-to-peer recruitment, RDS inference methods inversely weight participants according to their social network size. Inference from the RDS data additionally requires several assumptions around the recruitment process. These assumptions include the following: all relationships between recruiters and their recruits are reciprocal, the composition of the final sample is independent of the composition of the initial seeds, sampling mimics sampling with replacement, participants can accurately estimate their degree, and recruiters randomly recruit from within their social network [20]. As RDS studies are typically conducted among populations with no valid sampling frame, an empirical assessment of whether RDS yields a representative sample is impossible in most contexts. Studies that have been able to assess these assumptions or compare RDS estimates with population estimates have found that although RDS generally yields a representative sample, RDS estimators often fail to reduce bias when it does exist, and recruitment assumptions are often not met [20-22]. However, RDS does allows for the recruitment of individuals who would not likely be identified or reached via traditional sampling methods.

RDS has never before been implemented to study abortion, and although RDS has most commonly been used to measure outcomes among a stigmatized population, this study is, to our knowledge, the first example of using RDS to measure abortion (a stigmatized outcome) among a general population. We hypothesized that RDS could be well suited to the measurement of out-of-facility abortion for a range of reasons. First, population-representative surveys, such as household surveys, may systematically exclude young women, women living in informal settlements, or female refugees. Furthermore, traditional direct survey techniques often result in participants underreporting their abortion experiences [1,2]. RDS has the potential to reach a broader population than the existing methods for abortion measurement, and the process of being recruited into the study by someone known to the participants may generate trust between the participant-recruiter and the researcher and encourage the disclosure of sensitive experiences.

In the Republic of South Africa, the Choice on Termination of Pregnancy Act, passed in 1996, allows for the legal termination of pregnancy on request up to 12-week gestation; under socioeconomic, incest, rape, and medical grounds from 12 to 20 weeks; and to save a pregnant person's life after 20 weeks. Abortion services are provided free of charge in the public sector. However, barriers to abortion access in South Africa remain: a shortage of trained and willing providers [23] and a lack of dedicated facilities in which to perform abortions [24] can result in waiting lists that cause delays for abortion seekers, often beyond the legal gestational limit [25,26]. The most recent global estimates of abortion incidence, from 2015 to 2019, suggest an annual average of 30 abortions per 1000 women of reproductive age [27], and no reliable estimates exist for the proportion of abortions that occur within or outside facility settings in South Africa. Although out-of-facility abortions are widely known to occur in South Africa [24,28,29], their prevalence, safety, and effectiveness remain unknown. Although some data exist on people's experiences with out-of-facility abortion in South Africa [30,31], reliable information about the prevalence of and people's experiences with abortions that occur outside of the formal health system is needed both to inform improvements in abortion services, as well as to inform the development of resources about abortion that meet the needs and experiences of people in South Africa

In this study, we tested the feasibility of using RDS to sample participants and estimate the proportion of people who have ever attempted abortion—both those that occurred in-facility settings and those that occurred outside-of-facility settings—among the women of reproductive age in Soweto, South Africa. To assess feasibility, we considered (1) our ability to reach the proposed sample size, (2) whether RDS inference methods generated a sample similar to the source population, and (3) whether abortion was underreported.

Methods

Recruitment and Procedures

This study was conducted in Soweto, South Africa, from April 2018 to December 2018. Soweto is a large township within the city of Johannesburg with a total population of 1.3 million [32]. We used RDS, a well-established sampling method for populations for which there is no sampling frame, to calculate the proportion of people who have ever attempted abortion as well as the 1-year and 5-year cumulative incidence of abortion among women aged 15 to 49 years in Soweto.

With the assistance of well-known community-based organizations that provided a range of services, including, but not limited to, reproductive health services in Soweto, we recruited 11 women to serve as our initial seeds for RDS recruitment. These women were of various ages, income levels, and sexual and reproductive health experiences, including those with prior abortion. In accordance with the RDS methodology, seeds were members of the target population and purposively selected by our research team to initiate recruitment chains. We selected 2 study sites based on recommendations from our community-based organization partners about accessibility within the community and considerations of confidentiality—specifically, locations where people commonly gather or seek a range of services not specific to reproductive health. Seeds presented at one of the 2 study sites and completed an interviewer-administered questionnaire on their experiences with abortion. After completing the questionnaire, seeds were given 3 coupons to refer eligible peers to the study. Recruitment coupons contained information about the eligibility criteria, instructions on how to schedule an interview, and information about the study incentives. Potential participants contacted the study phone number via SMS text message or call and answered a short screening questionnaire to assess their eligibility. Participants were eligible if they identified as a woman; were aged between 15 and 49 years; spoke English, Tswana, isiZulu, Sotho, or Xhosa; and lived in Soweto. It is important to acknowledge that people of all genders have and need for abortions, and not all of them identify as women; some identify as men or another gender, and some people who identify as women do not have the capacity to carry a pregnancy. In the context of this study, we recruited people who identified as "women" and have referred to the study population accordingly throughout this paper.Eligible participants were invited to schedule an in-person interview at one of the 2 possible sites. The participants aged <18 years arrived at the interview with a signed parental consent form. Upon arrival at the study site, eligibility was confirmed, and consent was obtained. The consented participants completed an interviewer-administered questionnaire—administered by trained members of the study team who were South African women of reproductive age and spoke the above languages—and received 3 recruitment coupons. Participants received a participation incentive of 75 South African rand (approximately US $4) and a recruitment incentive of 50 South African rand (approximately US $2.50) for each eligible participant they successfully recruited. Participants returned to the study site to collect their recruitment incentive and completed an additional survey on their experiences participating in and recruiting for the study. Recruitment was tracked using coupon numbering.

Ethics Approval

Ethics approval for this study was obtained from the Human Sciences Research Council Research Ethics Committee in South Africa (REC 10/18/11/15: Experiences of women who self-induce abortion in Soweto, South Africa). The amount of compensation for participation in the study was arrived at in extensive consultation with the Human Sciences Research Council. No identifying information was collected from the participants. Study-related documents, including coupons, did not disclose abortion incidence as the primary aim of the study.

Instruments and Measurement

The main instrument in our study was a quantitative survey with questions on demographic characteristics, social network composition, health behaviors, sexual history, pregnancy history, and experience with abortion. The follow-up instrument contained questions on the recruitment process, including questions on refusals. Categories for out-of-facility providers and methods were informed by existing literature [28,30,33] as well as findings from formative research that comprised in-depth interviews conducted with 19 women from Soweto who had attempted to terminate a pregnancy outside the formal health setting [34]. In addition, once draft instruments were developed, we conducted cognitive interviews with 5 participants from the formative research phase (all of whom had consented to be recontacted) to ensure that the instruments were understandable and that the answer choices were appropriate. Minor refinements to the terminology and answer choices were made following the cognitive interview phase.

The primary outcome of interest for this study was the proportion of people who ever attempted abortion (facility-based or out-of-facility abortion), measured as the weighted proportion of women in the study who reported attempting at least 1 abortion in their lifetime. In addition, we calculated the 1-year and 5-year incidence of abortion attempts. We defined *out-of-facility* abortion as any abortion attempt, successful or unsuccessful, that did not take place under the supervision of (nor with a prescription from) a physician, nurse, or other advanced practice clinician at a government-run or privately operated health care facility. We defined facility-based abortion as any abortion attempt, successful or unsuccessful, that took place under the supervision of a physician, nurse, or other advanced practice clinician at a government-run or privately operated health care facility. The key sociodemographic variables used to compare the representativeness of our sample with the source population (women of reproductive age in Soweto) were age, educational attainment, employment status, and home language. Consistent with the RDS literature [21,35], we assessed network size using the following question: "How many women of reproductive age who live in Soweto do you know, who also know you, that you have seen in the past week?"

We additionally collected data on network characteristics such as recruiter-recruit relationships and recruitment experiences to assess whether several RDS recruitment assumptions were

met in this study; full findings from this methodological assessment are published elsewhere [36].

Statistical Analysis

Using the method proposed by Salganik [37] for calculating the desired sample size for a sample proportion, we arrived at a minimum sample size of 834 participants, which enabled us to detect a proportion of people who attempted abortion of 50% (417/834; maximally conservative estimate), with an SE <0.03 and assuming a design effect of 3. Data management was conducted using R (version 4.0.2; R Foundation for Statistical Computing) [38] and Stata (version 14; StataCorp LLC) [39]. We used RDS Analyst [40] to examine recruitment patterns, equilibrium (the point in recruitment at which the sample proportions of sociodemographic characteristics stabilize), homophily, waves of recruitment, and mean network size and compute weighted estimates of our primary outcomes. We used the RDS-II estimator to estimate the population proportions of demographic characteristics and our primary outcomes [41]. The RDS-II estimator reweights the sample population to account for *homophily*, the tendency of participants to recruit other participants who share similar characteristics [14]. Participants are weighted by the inverse of their degree (social network size); for example, participants with a degree of 10 would be given a weight of 1/10. We used imputed *visibility* for our measure of degree (effective network size), which incorporated self-reported social network size, the number of successful recruits, and the time to recruit to estimate each participant's inclusion probability. Visibility was imputed using the *impute.visibility_mle* function in RDS Analyst [40]. We calculated 95% CIs using 1000 bootstrap replications.

Results

Recruitment

Between April 4, 2018, and December 17, 2018, 849 eligible participants were recruited into the study. Recruitment occurred over 36 weeks, and the longest recruitment chain lasted 17 waves, with a mean of 6.6 recruitment waves for active seeds. A total of 2 seeds did not recruit any participants, and 56.5% (480/849) of the sample originated from 1 seed. Approximately one-third (n=837, 36.8%) of the 2277 distributed coupons were returned. Recruitment patterns based on lifetime experience of abortion are shown in Figure 1. A methodological assessment of RDS assumptions and recruitment dynamics has been previously published [36], and the key findings are summarized below. There was strong homophily (chi-square test for independence, $P<.05$) for age, educational attainment, employment status, and lifetime experiences with abortion, suggesting a strong tendency to recruit individuals with similar characteristics to theirs as compared with random recruitment. Sample proportions for age, home language, educational attainment, and socioeconomic indicators stabilized (reached equilibrium) by approximately 300 to 500 participants, well before our estimated sample size.

Figure 1. Recruitment tree from a respondent-driven sampling study of women aged 15 to 49 years in Soweto, South Africa (N=849). Each node represents a participant connected to their recruits and recruiters. Nodes in blue indicate a participant who reported any lifetime experience of abortion.

Study Population

The unweighted sample proportions for the selected demographic characteristics, along with weighted population proportions, are reported in Table 1. Table 1 presents the population estimates of the selected demographic characteristics based on publicly available data. The unweighted median age was 27 (IQR 22-36) years. Approximately one-fifth (20.2%, 95% CI 17.4%-23.1%) of the target population were currently in school, and most had at least some secondary education (52.5%, 95% CI 49.1%-56.0%) or completed secondary education (39.6%, 95% CI 36.2%-43.1%). Most were unemployed (83.1%, 95% CI 80.5%-85.8%). Although a statistical comparison of the RDS-II sociodemographic estimates to the estimated source population estimates is not possible, the sample characteristics are largely similar to the source population for all variables, except for employment status.

Table 1. Sociodemographic characteristics of women participating in a respondent-driven sampling survey, Soweto, South Africa, 2018 (N=849).

Sociodemographic characteristics	Sample proportion (unweighted)		RDS-II[a] weighted proportion		Estimated proportion in source population[b] (%)
	Participants, n (%)	95% CI	Participants (%)	95% CI	
Age (years)					
15-19	133 (15.7)	13.4-18.3	16.1	14.1-18.0	13.7
20-24	184 (21.7)	19-24.6	22.2	19.5-24.8	18.8
25-29	166 (19.6)	17-22.4	19.2	17.6-20.8	19.5
30-34	127 (15)	12.7-17.5	14.9	12.0-17.7	15.4
35-39	109 (12.8)	10.7-15.3	12.6	9.8-15.5	12.6
40-44	78 (9.2)	7.4-11.3	9	6.6-11.4	10.6
45-49	52 (6.1)	4.7-8.0	6.1	3.6-8.7	9.6
Educational attainment					
Currently in school	169 (19.9)	17.4-22.8	20.2	17.4-23.1	__[c]
No schooling	1 (0.1)	0.0-0.8	0.1	0.1-0.2	6.8
Some primary	9 (1.1)	0.6-2.0	1.1	0.9-1.3	6.3
Completed primary	30 (3.5)	2.5-5.0	4	3.7-4.4	6.3
Some secondary	450 (53.1)	49.8-56.5	52.5	49.1-56.0	76.1
Completed secondary	335 (39.6)	36.3-42.9	39.6	36.2-43.1	76.1
Higher education	22 (2.6)	1.7-3.9	2.6	1.0-4.3	10.8
Relationship status					
Never married	622 (73.8)	70.7-76.6	74.1	70.8-77.3	45.6
Married (traditional or civil)	55 (6.5)	5.0-8.4	6.2	5.4-7.1	32.8
Living together	152 (18)	15.6-20.8	18	14.8-21.3	10.6
Divorced	9 (1.1)	0.6-2.0	1	0.7-1.4	3.1
Separated	5 (0.6)	0.2-1.4	0.6	0.4-0.9	0.9
Employment status					
Employed	138 (16.3)	13.9-18.9	16.9	14.2-19.6	70.6
Unemployed or student	710 (83.7)	81.1-86.1	83.1	80.5-85.8	29.4
Housing type					
Formal housing	676 (79.7)	76.9-82.3	79.3	76.4-82.2	81.3
Traditional housing	2 (0.2)	0.1-0.9	0.2	0.0-1.0	0.1
Informal housing	160 (18.9)	16.4-21.6	19.4	19.4-19.4	18
Other	10 (1.2)	0.6-2.2	1.1	0.0-4.0	0.6
Water source					
Piped water inside the dwelling	427 (50.4)	47-53.7	51.1	47.6-54.6	60.3
Piped water inside the yard	402 (47.4)	44.1-50.8	46.9	43.4-50.4	31.8
Piped water from access point outside	19 (2.2)	1.4-3.5	2	0.9-3.0	7.1
Toilet facilities					
Flush toilet (connected to sewage)	820 (96.7)	95.3-97.7	97	95.8-98.2	88.6
Flush toilet (with septic tank)	8 (0.9)	0.5-1.9	0.9	0.3-1.6	1.6
Pit toilet with ventilation	14 (1.7)	1.0-2.8	1.4	0.5-2.3	3.4
Pit toilet without ventilation	6 (0.7)	0.3-1.6	0.7	0.1-1.3	1.7
Resources					

Sociodemographic characteristics	Sample proportion (unweighted)		RDS-II[a] weighted proportion		Estimated proportion in source population[b] (%)
	Participants, n (%)	95% CI	Participants (%)	95% CI	
Connected to electricity	838 (98.8)	97.8-99.4	99	98.9-99.1	87.5
Owns a television	787 (92.8)	90.8-94.4	93.3	91.5-95.2	86.5
Has a radio	622 (73.3)	70.2-76.2	73	70.0-76.0	70.1
Has a landline	80 (9.4)	7.6-11.6	9.6	7.5-11.8	—
Has a cellphone	841 (99.2)	98.3-99.6	99.1	99.1-99.2	95.5
Home language					
Afrikaans	2 (0.2)	0.1-0.9	0.2	0.0-3.2	1.3
English	4 (0.5)	0.2-1.3	0.5	0.3-0.7	2.3
IsiXhosa	66 (7.8)	6.2-9.8	7.7	4.4-11.1	8.7
IsiZulu	329 (38.9)	35.7-42.3	39.9	39.6-40.1	37.1
Sepedi	20 (2.4)	1.5-3.6	2.4	1.1-3.7	5.1
Sesotho	216 (25.6)	22.7-28.6	24.5	21.8-27.3	15.5
Setswana	59 (7)	5.4-8.9	7.2	4.8-9.6	12.9
Tshivenda	35 (4.1)	3.0-5.7	4	2.0-6.0	4.5
Xitsonga	100 (11.8)	9.8-14.2	12.1	11.8-12.5	8.9
Shona	13 (1.5)	0.9-2.6	1.3	1.0-1.6	0
Other	1 (0.1)	0-0.8	0.1	0.0-2.0	3.7

[a]RDS-II: respondent-driven sampling II.

[b]Source population proportion estimates are from the 2016 South Africa Community Survey, localized to Johannesburg for age, educational attainment, housing type, water source, toilet type, and electricity access. Data on relationship status and other resources were obtained from the 2016 South Africa Community Survey, localized to Gauteng Province. Data on employment status are from the Labor Force Survey and from the Quarterly Labour Force Survey published by Statistics South Africa; data represent unemployment rates among women in South Africa from July 2018 to September 2018.

[c]Census data not available for comparison.

Abortion Attempts

The RDS-II estimates of the proportion of people who ever attempted abortion was 12.1% (95% CI 9.7%-14.4%; Table 2). A total of 7.1% (95% CI 5.4%-8.9%) reported a facility-based abortion, and 4.4% (95% CI 3.0%-5.8%) reported an out-of-facility abortion. The true design effect for the main outcome, any abortion attempt, was 1.14. Most participants (RDS-II estimate: 61.8%, 95% CI 58.4%-65.2%) reported that their best friend had an abortion (not displayed in the tables).

Because of likely underreporting, we present the unweighted proportions for various abortion experiences. A total of 106 out of 849 (12.5%) participants reported at least one abortion attempt at any point in their lifetime, and 9 (n=106, 8.5%) of them did not provide any subsequent information about their experiences. Among the remaining 97 (91.5%) participants who reported an abortion attempt and answered additional questions related to their abortion experience, 85 (88%) attempted an abortion once, 8 (8%) reported 2 abortion attempts, and 4 (4%) reported ≥3 abortion attempts. When asked about their most recent abortion attempt, 60 (62%) participants reported that they went to a health care facility. Among those who went to a health care facility, 9 (15%) attempted to do something on their own to end their pregnancy before seeking facility-based health care, most commonly taking a laxative. At the health care facility, 28 out of 60 (47%) participants reported taking medications for abortion, 10 (17%) participants reported a surgical procedure, and 2 (3%) participants did not know what method was used to end their pregnancy. Of the participants who received medications (n=28), 2 (7%) did not have a complete abortion and continued with their pregnancy. Of the participants who went to a health care facility (n=60), 17 (28%) did not ultimately end up having an abortion because they decided that they wanted to continue with their pregnancy, they were counseled to continue with the pregnancy, or their gestational age was beyond the legal limit.

Of the 97 participants, 37 (38%) of the participants who did not report going to a health care facility for their most recent abortion attempt reported using methods such as laxatives, aspirin, strong tea or coffee, pesticides, bleach, or combinations of the above. Of these 37 participants, 22 (59%) reported successfully terminating their pregnancy.

Table 2. Proportion of ever attempting abortion, 1-year incidence of abortion, and 5-year incidence of abortion in a respondent-driven sampling survey, Soweto, South Africa, 2018 (N=849).

Lifetime experience of abortion	Unweighted estimate			RDS-II[a] estimate		Design effect
	Participants, N[b]	Participants, n (%)	95% CI	Participants (%)	95% CI	
Abortion attempt						
People who attempted to have an abortion	849	106 (12.49)	10.4-14.9	12.1	9.7-14.4	1.14
People who attempted to have an in-facility abortion[c]	841	63 (7.49)	5.9-9.5	7.1	5.4-8.9	1.05
People who attempted to have an out-of-facility abortion[c]	841	37 (4.3)	3.2-6.0	4.4	3.0-5.8	1.03
Incidence of abortion[d,e]						
1-year incidence of abortion attempts (2017)	840	9 (10.71)	3.74-17.70	9.14	1.90-16.40	1.18
5-year incidence of abortion attempts (2013-2017)	840	42 (50)	33.82-66.18	47.3	35.64-58.96	0.54

[a]RDS-II: respondent-driven sampling II.

[b]The total sample of 849 participants denotes those who attempted to have an abortion. The location of abortion attempt is missing for 8 participants; therefore, the N value is smaller for the location rows. The data are missing for the year of abortion for 9 participants; therefore, the N value for incidence is 840.

[c]Data on the type of abortion (in vs out of facility) are missing for 8 participants.

[d]Data on the year of abortion are missing for 9 participants.

[e]Incidence of abortion per 1000 women.

Discussion

Principal Findings

In our study, we explored the feasibility of applying the RDS methodology to estimate the proportion of women of reproductive age who have ever attempted abortion in Soweto, South Africa. The estimated proportion of ever attempting an abortion of 12% and 1-year incidence of 9.1 abortion attempts per 1000 women of reproductive age in our study likely represents a substantial underestimation of the actual abortion experiences in this study population [27]. Although no directly comparable measures exist, recently published, country-specific estimates of abortion incidence report an annual estimated 30 abortions occur per 1000 women of reproductive age in South Africa, representing a figure nearly 3 times the magnitude of the comparable estimate from our study [42]. We posit that this underestimation represents a failure of the RDS method to generate more reliable estimates of the abortion incidence in our study.

We previously published a methodological assessment of whether several RDS assumptions were met in these data [36]. In that paper, we found that although the approximation of sampling with replacement was met, the participants did not consistently report the same degree, nor did they randomly recruit from within their social network. It is likely that the failure to meet the assumptions yielded a sample with different employment characteristics than those of the target population, which was not resolved by standard RDS methods. However, without gold standard abortion estimates for the target population by sociodemographic characteristics, it is challenging to assess the impact of failing to meet these assumptions on inference for abortion. Although it is plausible that some of the underestimation of abortion may have been due to the overrepresentation of unemployed participants in the sample, it is more likely to have been due to underreporting.

Strengths and Limitations

This study highlights the limitations of RDS in measuring abortion. Although the social networking literature is lacking on the subject of abortion, public health evidence suggests that those who have abortions outside the formal health sector communicate with members of their social network to obtain information about self-managed or community-based abortions [43-45]. We hypothesized that RDS could offer a previously untested alternative method to more accurately measure abortion incidence by similarly relying on peer recruitment to help reduce the underreporting of abortion. However, it is possible that because we recruited from a general population of women of reproductive age, peer recruitment operated in the opposite direction in our study; if participants were recruited into the study by members of their social network who they would not want to know about their prior abortions, they might have been less likely to report their abortions to the researchers conducting the study. In this context, it is notable that most participants in our study reported that their best friend had had an abortion—potentially indicating a willingness, as seen in other studies, to discuss the abortion experiences of others but not themselves. In addition, as we lack representative sociodemographic data on reproductive-aged women localized to Soweto, we were unable to directly validate whether RDS sampling generated a sample with demographic characteristics similar to those of the overall population of women of

reproductive age living in Soweto. However, based on the best available estimates (population-based data from Gauteng Province and the city of Johannesburg, where Soweto is located), our sample differs substantially from the source population, particularly with respect to employment, even after adjusting using RDS inference procedures.

However, the limitations of our study should be considered in the context of the strengths. We successfully recruited a large sample of women of reproductive age, demonstrating that RDS can be used to recruit a sample of participants who are willing to participate in a study about reproductive health, answer questions about abortion, and participate in peer recruitment. We hope that the lessons learned from our study will be instructive to future researchers exploring the use of novel sampling approaches for measuring abortion.

Despite the failure of RDS to generate more reliable estimates of abortion incidence, it may be a method best suited for sampling when selecting on stigmatizing characteristics (injection drug use, men having sex with men, sex work, and now, abortion), as it has most commonly been applied. For example, RDS could be used to sample a population with out-of-facility abortion experience and estimate the proportion of that population that has experienced one or more outcome of interest (eg, using medication abortion, seeking health care in the formal health sector, or experiencing complications). Other population size estimation methods could be deployed to arrive at estimates of prevalence in an RDS study that is specific to abortion experiences.

It is conceivable that RDS could reduce underreporting of abortion if it were deployed to estimate abortion incidence among a highly socially networked population (potentially in humanitarian settings, among sex workers, etc). However, it is also possible that asking questions about stigmatizing experiences in any general population–based survey will be subject to underreporting—especially when interviews are administered face-to-face. Using tools such as Audio Computer Assisted Self-Interviewing and other technologies has been shown to reduce underreporting in studies of some stigmatized behaviors and could prove useful in the context of abortion as well [46,47].

Conclusions

Accurate estimates of abortion incidence within and outside formal health settings are vital for developing targeted and effective programs, policies, and interventions to increase the access to safe abortions. In certain highly networked populations, RDS may prove to be a useful tool in the toolkit of abortion researchers, but to ensure that people seeking abortion have the information and support they need, regardless of where or how their abortion takes place, more work is needed to develop and validate tools that more accurately measure not only the incidence of abortion but also the experiences, quality, and outcomes of abortions.

Acknowledgments

The authors would like to thank Lifeline Soweto for their assistance in data collection. They would also like to thank Isabel Muñoz, Willi McFarland, and Heidi Moseson for their valuable input on study design and analysis as well as the attendees of the International Union for the Scientific Study of Population Seminar on Incidence and Safety of Abortion: New Evidence and Improvements in Measurement (Watamu, Kenya, December 2018).

Conflicts of Interest

None declared.

References

1. Moseson H, Herold S, Filippa S, Barr-Walker J, Baum SE, Gerdts C. Self-managed abortion: a systematic scoping review. Best Pract Res Clin Obstet Gynaecol 2020 Feb;63:87-110. [doi: 10.1016/j.bpobgyn.2019.08.002] [Medline: 31859163]
2. Sorhaindo A, Sedgh G. Scoping review of research on self-managed medication abortion in low-income and middle-income countries. BMJ Glob Health 2021 May;6(5):e004763 [FREE Full text] [doi: 10.1136/bmjgh-2020-004763] [Medline: 33986002]
3. Medical management of abortion. World Health Organization. 2018. URL: https://apps.who.int/iris/handle/10665/278968 [accessed 2022-09-07]
4. Ganatra B, Gerdts C, Rossier C, Johnson Jr BR, Tunçalp Ö, Assifi A, et al. Global, regional, and subregional classification of abortions by safety, 2010-14: estimates from a Bayesian hierarchical model. Lancet 2017 Nov 25;390(10110):2372-2381 [FREE Full text] [doi: 10.1016/S0140-6736(17)31794-4] [Medline: 28964589]
5. Adler AJ, Filippi V, Thomas SL, Ronsmans C. Quantifying the global burden of morbidity due to unsafe abortion: magnitude in hospital-based studies and methodological issues. Int J Gynaecol Obstet 2012 Sep;118 Suppl 2:S65-S77. [doi: 10.1016/S0020-7292(12)60003-4] [Medline: 22920625]
6. Rossier C. Estimating induced abortion rates: a review. Stud Fam Plann 2003 Jun;34(2):87-102. [doi: 10.1111/j.1728-4465.2003.00087.x] [Medline: 12889341]
7. Scott RH, Filippi V, Moore AM, Acharya R, Bankole A, Calvert C, et al. Setting the research agenda for induced abortion in Africa and Asia. Int J Gynaecol Obstet 2018 Aug;142(2):241-247. [doi: 10.1002/ijgo.12525] [Medline: 29745418]

8. Gray N, Azzopardi P, Kennedy E, Willersdorf E, Creati M. Improving adolescent reproductive health in Asia and the Pacific: do we have the data? A review of DHS and MICS surveys in nine countries. Asia Pac J Public Health 2013 Mar;25(2):134-144. [doi: 10.1177/1010539511417423] [Medline: 21807622]
9. Sully E, Giorgio M, Anjur-Dietrich S. Estimating abortion incidence using the network scale-up method. Dem Res 2020 Dec 16;43:1651-1684. [doi: 10.4054/DemRes.2020.43.56]
10. Stillman M, Leong E, Utomo B, Dadun D, Aryanty RI, Sedgh G, et al. An application of the confidante method to estimate induced abortion incidence in Java, Indonesia. Int Perspect Sex Reprod Health 2020 Sep 29;46:199-210. [doi: 10.1363/46e0120] [Medline: 33001827]
11. Keogh SC, Otupiri E, Chiu DW, Polis CB, Hussain R, Bell SO, et al. Estimating the incidence of abortion: a comparison of five approaches in Ghana. BMJ Glob Health 2020 Apr 02;5(4):e002129. [doi: 10.1136/bmjgh-2019-002129]
12. Moseson H, Treleaven E, Gerdts C, Diamond-Smith N. The list experiment for measuring abortion: what we know and what we need. Stud Fam Plann 2017 Dec;48(4):397-405. [doi: 10.1111/sifp.12042] [Medline: 29148056]
13. Wang J, Carlson RG, Falck RS, Siegal HA, Rahman A, Li L. Respondent-driven sampling to recruit MDMA users: a methodological assessment. Drug Alcohol Depend 2005 May 09;78(2):147-157. [doi: 10.1016/j.drugalcdep.2004.10.011] [Medline: 15845318]
14. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. Soc Probl 1997 May;44(2):174-199. [doi: 10.2307/3096941]
15. Heckathorn DD. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. Soc Probl 2002 Feb;49(1):11-34. [doi: 10.1525/sp.2002.49.1.11]
16. Salganik MJ, Heckathorn DD. 5. Sampling and estimation in hidden populations using respondent-driven sampling. Sociol Methodol 2004 Dec;34(1):193-240. [doi: 10.1111/j.0081-1750.2004.00152.x]
17. Decker MR, Marshall BD, Emerson M, Kalamar A, Covarrubias L, Astone N, et al. Respondent-driven sampling for an adolescent health study in vulnerable urban settings: a multi-country study. J Adolesc Health 2014 Dec;55(6 Suppl):S6-12 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.07.021] [Medline: 25454005]
18. Reisner SL, Mimiaga MJ, Johnson CV, Bland S, Case P, Safren SA, et al. What makes a respondent-driven sampling "seed" productive? Example of finding at-risk Massachusetts men who have sex with men. J Urban Health 2010 May;87(3):467-479 [FREE Full text] [doi: 10.1007/s11524-010-9439-3] [Medline: 20354911]
19. Ludlam AH, Saxton PJ, Dickson NP, Adams J. Respondent-driven sampling among gay and bisexual men: experiences from a New Zealand pilot study. BMC Res Notes 2015 Oct 09;8:549 [FREE Full text] [doi: 10.1186/s13104-015-1449-5] [Medline: 26453538]
20. Gile KJ, Johnston LG, Salganik MJ. Diagnostics for respondent-driven sampling. J R Stat Soc Ser A Stat Soc 2015 Jan;178(1):241-269 [FREE Full text] [doi: 10.1111/rssa.12059] [Medline: 27226702]
21. McCreesh N, Frost SD, Seeley J, Katongole J, Tarsh MN, Ndunguse R, et al. Evaluation of respondent-driven sampling. Epidemiology 2012 Jan;23(1):138-147 [FREE Full text] [doi: 10.1097/EDE.0b013e31823ac17c] [Medline: 22157309]
22. Li J, Valente TW, Shin HS, Weeks M, Zelenev A, Moothi G, et al. Overlooked threats to respondent driven sampling estimators: peer recruitment reality, degree measures, and random selection assumption. AIDS Behav 2018 Jul;22(7):2340-2359 [FREE Full text] [doi: 10.1007/s10461-017-1827-1] [Medline: 28660381]
23. Favier M, Greenberg JM, Stevens M. Safe abortion in South Africa: "We have wonderful laws but we don't have people to implement those laws". Int J Gynaecol Obstet 2018 Oct;143 Suppl 4:38-44. [doi: 10.1002/ijgo.12676] [Medline: 30374986]
24. South Africa: Barriers to safe and legal abortion in South Africa. Amnesty International. 2017 Feb 1. URL: https://www.amnesty.org/en/documents/afr53/5423/2017/en/ [accessed 2022-09-07]
25. Jewkes R, Gumede T, Westaway M, Dickson K, Brown H, Rees H. Why are women still aborting outside designated facilities in metropolitan South Africa? BJOG 2005 Sep;112(9):1236-1242. [doi: 10.1111/j.1471-0528.2005.00697.x] [Medline: 16101602]
26. Harries J, Orner P, Gabriel M, Mitchell E. Delays in seeking an abortion until the second trimester: a qualitative study in South Africa. Reprod Health 2007 Sep 20;4:7 [FREE Full text] [doi: 10.1186/1742-4755-4-7] [Medline: 17883835]
27. Bearak JM, Popinchalk A, Beavin C, Ganatra B, Moller A, Tunçalp Ö, et al. Country-specific estimates of unintended pregnancy and abortion incidence: a global comparative analysis of levels in 2015-2019. BMJ Glob Health 2022 Mar;7(3):e007151 [FREE Full text] [doi: 10.1136/bmjgh-2021-007151] [Medline: 35332057]
28. Harries J, Gerdts C, Momberg M, Greene Foster D. An exploratory study of what happens to women who are denied abortions in Cape Town, South Africa. Reprod Health 2015 Mar 21;12:21 [FREE Full text] [doi: 10.1186/s12978-015-0014-y] [Medline: 25884479]
29. Gerdts C, DePiñeres T, Hajri S, Harries J, Hossain A, Puri M, et al. Denial of abortion in legal settings. J Fam Plann Reprod Health Care 2015 Jul;41(3):161-163 [FREE Full text] [doi: 10.1136/jfprhc-2014-100999] [Medline: 25511805]
30. Gerdts C, Raifman S, Daskilewicz K, Momberg M, Roberts S, Harries J. Women's experiences seeking informal sector abortion services in Cape Town, South Africa: a descriptive study. BMC Womens Health 2017 Oct 02;17(1):95 [FREE Full text] [doi: 10.1186/s12905-017-0443-6] [Medline: 28969631]

31. Harries J, Daskilewicz K, Bessenaar T, Gerdts C. Understanding abortion seeking care outside of formal health care settings in Cape Town, South Africa: a qualitative study. Reprod Health 2021 Sep 23;18(1):190 [FREE Full text] [doi: 10.1186/s12978-021-01243-3] [Medline: 34556120]
32. 2018 Statistics South Africa. Statistics South Africa, Republic of South Africa. 2022 Sep 7. URL: https://www.statssa.gov.za/?m=2018 [accessed 2022-09-07]
33. Constant D, Grossman D, Lince N, Harries J. Self-induction of abortion among women accessing second-trimester abortion services in the public sector, Western Cape Province, South Africa: an exploratory study. S Afr Med J 2014 Apr;104(4):302-305. [doi: 10.7196/samj.7408] [Medline: 25118559]
34. Ramirez AM, Jayaweera RT, Mtimkulu M, Bessenaar T, Chollampat N, Gerdts C. Understanding abortion seeking and care outside of legal public sector services in South Africa: a qualitative study. Contraception 2017 Oct 1;96(4):277-278. [doi: 10.1016/j.contraception.2017.07.061]
35. McCormick TH, Salganik MJ, Zheng T. How many people do you know?: Efficiently estimating personal network size. J Am Stat Assoc 2010 Mar 01;105(489):59-70 [FREE Full text] [doi: 10.1198/jasa.2009.ap08518] [Medline: 23729943]
36. Jayaweera R, Gerdts C, Wesson P, Motana R, Bessenaar T. Respondent-driven sampling to assess cumulative lifetime incidence of abortion in Soweto, South Africa: a methodological assessment. Am J Epidemiol 2022. [doi: 10.2196/preprints.38045]
37. Salganik MJ. Variance estimation, design effects, and sample size calculations for respondent-driven sampling. J Urban Health 2006 Nov;83(6 Suppl):i98-112 [FREE Full text] [doi: 10.1007/s11524-006-9106-x] [Medline: 16937083]
38. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2022 Sep 7. URL: https://www.r-project.org/ [accessed 2022-09-07]
39. Stata Statistical Software: Release 15. StataCorp. College Station, TX, USA: StataCorp LLC; 2017. URL: https://www.stata.com/ [accessed 2022-09-01]
40. Handcock M, Fellows I, Gile K. RDS Analyst: Software for the Analysis of Respondent-Driven Sampling Data. RDS Analyst. 2016. URL: http://wiki.stat.ucla.edu/hpmrg/index.php?title=RDS_Analyst_Install&action=edit [accessed 2022-09-01]
41. Volz E, Heckathorn DD. Probability based estimation theory for respondent driven sampling. J Off Stat 2008 Mar;24(1):79-97.
42. Bearak J, Popinchalk A, Ganatra B, Moller AB, Tunçalp Ö, Beavin C, et al. Unintended pregnancy and abortion by income, region, and the legal status of abortion: estimates from a comprehensive model for 1990-2019. Lancet Glob Health 2020 Sep;8(9):e1152-e1161 [FREE Full text] [doi: 10.1016/S2214-109X(20)30315-6] [Medline: 32710833]
43. Rossier C. Abortion: an open secret? Abortion and social network involvement in Burkina Faso. Reprod Health Matters 2007 Nov;15(30):230-238. [doi: 10.1016/S0968-8080(07)30313-3] [Medline: 17938091]
44. Baum SE, Ramirez AM, Larrea S, Filippa S, Egwuatu I, Wydrzynska J, et al. "It's not a seven-headed beast": abortion experience among women that received support from helplines for medication abortion in restrictive settings. Health Care Women Int 2020 Oct;41(10):1128-1146. [doi: 10.1080/07399332.2020.1823981] [Medline: 33156737]
45. Zurbriggen R, Keefe-Oates B, Gerdts C. Accompaniment of second-trimester abortions: the model of the feminist Socorrista network of Argentina. Contraception 2018 Feb;97(2):108-115 [FREE Full text] [doi: 10.1016/j.contraception.2017.07.170] [Medline: 28801052]
46. Lindberg L, Scott RH. Effect of ACASI on reporting of abortion and other pregnancy outcomes in the US national survey of family growth. Stud Fam Plann 2018 Sep;49(3):259-278 [FREE Full text] [doi: 10.1111/sifp.12068] [Medline: 30040126]
47. Pershad J, Mugerwa KY, Filippi V, Mehrtash H, Adu-Bonsaffoh K, Bello FA, et al. Prevalence and determinants of self-reported anxiety and stress among women with abortion-related complications admitted to health facilities in Eastern and Southern Africa: a cross-sectional survey. Int J Gynaecol Obstet 2022 Jan;156 Suppl 1:53-62. [doi: 10.1002/ijgo.14042] [Medline: 35014698]

Abbreviations

RDS: respondent-driven sampling

Edited by Y Khader; submitted 16.03.22; peer-reviewed by S Bell, A M; comments to author 02.05.22; revised version received 08.09.22; accepted 10.10.22; published 08.12.22.

Please cite as:
Gerdts C, Jayaweera RT, Motana R, Bessenaar T, Wesson P
Incidence of and Experiences with Abortion Attempts in Soweto, South Africa: Respondent-Driven Sampling Study
JMIR Public Health Surveill 2022;8(12):e38045
URL: https://publichealth.jmir.org/2022/12/e38045
doi:10.2196/38045
PMID:36480253

Original Paper

Comprehensive Comparisons of Family Health Between Families With One Immigrant Parent and Native Families in Taiwan: Nationwide Population-Based Cohort Study

Yi-Lung Chen[1,2*], PhD; Hsing-Ying Ho[1,2*], MSc

[1]Department of Healthcare Administration, Asia University, Taichung, Taiwan

[2]Department of Psychology, Asia University, Taichung, Taiwan

[*]all authors contributed equally

Corresponding Author:
Yi-Lung Chen, PhD
Department of Healthcare Administration
Asia University
500 Lioufeng Rd, Wufeng
Taichung, 41354
Taiwan
Phone: 886 4 23323456
Email: elong@asia.edu.tw

Abstract

Background: Mothers and children in families with one immigrant parent have been reported to be healthier than those in native families; however, the health of the fathers in these families has rarely been discussed in literature.

Objective: We aimed to comprehensively compare the health of all the family members between families with one immigrant parent (native fathers, immigrant mothers, and their children) and native families (native fathers, native mothers, and their children).

Methods: We conducted a cohort study by using the Taiwan Maternal and Child Health Database to recruit live-born children and their parents from 2004 to 2016. Overall, we identified 90,670 fathers, 91,270 mothers, and 132,457 children in families with one immigrant parent and 1,666,775 fathers, 1,734,104 mothers, and 2,637,191 children in native families and followed up with them from 2004 to 2017. The outcomes comprised common physical and mental disorders, catastrophic illnesses, mortality, and child adversities and accidents. The covariates comprised the child's year of birth, parental age, low-income status, and physical or mental disorder status. Logistic regression was performed to compare the risks of the outcomes between families with one immigrant parent and native families.

Results: The parents in families with one immigrant parent were more likely to be of low-income status and were older than the parents in native families. After adjusting for the covariates, fathers in families with one immigrant parent were found to have higher risks of physical and mental disorders, catastrophic illness, and mortality than fathers in native families. Conversely, mothers in families with one immigrant parent had lower risks of physical and mental disorders, catastrophic illness, and mortality than mothers in native families. Finally, the children in families with one immigrant parent generally had better physical and mental health but higher risks for leukemia, liver diseases, autism spectrum disorder, and road traffic accidents than children in native families.

Conclusions: The health status of the members of families with one immigrant parent was nonhomogeneous, and the poorer general health of fathers in such families suggests health inequalities in families with one immigrant parent.

(JMIR Public Health Surveill 2022;8(12):e33624) doi:10.2196/33624

KEYWORDS

international marriage immigrant family; family health; physical health; mental health; mortality; reduced inequalities; good health and well-being

Introduction

Currently, international migration is a common globalization phenomenon, and migrants compose an important part of the population in many countries. The number of international migrants reached 272 million worldwide in 2019 [1]. Marriage is one of the main causes of migration, especially in Asian countries [1] such as Japan, South Korea [2], Singapore [3], and Taiwan [4]. Individuals often migrate to wealthier neighboring countries to improve their living conditions [1]. In Taiwan, 95.6% of the naturalization applications from 1993 to 2019 were due to international marriage [5], which is defined as a marriage formed between 2 individuals from different countries of origin, resulting in 565,000 married migrants with a temporary or permanent residence permit as of 2019 [6]. Most married immigrants in Taiwan are females, and most of them migrated from East or Southeast Asian countries [4]. International marriage contributed to 9.1% of the overall marriages in Taiwan from 1998 to 2020 [4]. The newborns of married immigrants comprised 8.6% of the overall newborns in Taiwan from 1998 to 2020 [7]. These findings indicate that families with one immigrant parent have become an important part of the population in Taiwan, and this study focused on families with one immigrant parent (families consisting of a native father, an immigrant mother, and a child) and native families (families consisting of a native father, a native mother, and a child).

The literature is not in agreement regarding the differences in terms of general health between married female immigrants and married native women, and 2 different theories have been proposed to explain these inconsistent findings [8,9]. The first theory is the healthy immigrant effect, which asserts that immigrants generally have better health because healthier people are more likely to immigrate to seek a better life, and medical examinations required by immigration authorities in host countries may also prevent less healthy individuals from immigrating [8]. The healthy immigrant effect has been supported among married female immigrants in Taiwan [10]. Furthermore, although married female immigrants reported higher acculturative stress and lower spousal support, they reported fewer depressive symptoms than native women [11]. Another study indicated that married female immigrants had a better quality of life, fewer stressful life events, and a lower prevalence of major depressive disorder than married native women [12]. Moreover, fewer married female immigrants reported prenatal and postpartum depression and physical disorders than married native women [13]. The other theory is the salmon bias effect, which proposes that sick or older immigrants return to their countries of origin [14]. Thus, immigrants may not be truly healthy, and the disease and mortality of immigrants may be underestimated and that is why they return to their countries of origin. Some studies have partially supported the salmon bias effect in married female immigrants [10,14-16]. Specifically, married female immigrants were reported to be more likely to experience physical and mental disorders than native populations in Asian countries [10,15]. Furthermore, depression was one of the main concerns among mental disorders; married female immigrants reported having a higher prevalence of depression during the antenatal (31.8% vs 18.6%, respectively) [16] and postpartum period (41.1% vs 8.4%, respectively) [17,18] than native women. The higher risk of anxiety was another concern [19]. With regard to physical disorders, the risk of viral hepatitis in married female immigrants was higher than that in native women [20]. However, unlike married female immigrants, their native male spouses receive much less attention, and the difference in the general health between the native male spouses of native women and the native male spouses of married female immigrants remains unclear.

One evident feature of the native male spouses of married female immigrants is the low socioeconomic status (SES) [21,22], including older age [17,23], low education levels [24,25], employment in unskilled labor positions [26], and low income [25], thus hindering native men from getting married to native women and having possibly poorer health. Marriage migration in Taiwan originated in the 1960s, when retired veterans had difficulties getting married, which resulted in the development of international marriage brokerage agencies. Furthermore, by 1990, both the out-migration of industries and the import of foreign labor had a great impact on the employment of men in unskilled labor positions in Taiwan. In addition, the education levels of Taiwanese women increased in the 1990s. These situations made spouse selection difficult for men with low education levels and partially increased the average age of marriage for Taiwanese individuals [27]. As a result, these Taiwanese men with low SES tended to utilize international marriage brokerage agencies. In addition, the low SES of these native male spouses may be related to negative impacts on their health. Although several types of health status have been reported in native male spouses, including mental or physical disabilities [25,28], chronic diseases, serious illnesses [25], and general health issues [25,29], significant limitations existed in these studies. First, the studies were small-scale cross-sectional studies with sample sizes ranging from 140 to 1827 participants. Second, these studies only reported the prevalence of diseases, and a comparison with native male spouses who married native women is lacking. Third, because these studies were based on informant reports and not self-reports, the reliability and validity of the studies may be limited. Therefore, more comprehensive and large-scale studies with direct information from native male spouses of female immigrants are warranted for a better understanding of their health.

The mental health of children of married female immigrants has been reported to be generally worser than the mental health of children of native women, while comparison studies on physical health are relatively limited. Specifically, for mental health, more externalizing (eg, delinquent behavior) [19,23,30] and internalizing (eg, anxiety, depression) behavioral problems [19,30] were observed in the children of married female immigrants, although the results varied in school and family settings [30]. Furthermore, the depression levels of the children of married female immigrants were more likely to be affected by family factors [31]. In terms of physical health, newborns of married female immigrants had a lower risk of neonatal mortality than newborns of native women after adjusting for demographic confounders [32].

There are debatable findings on child adversity between children of native families and those of immigrant families. Some studies have shown that children of immigrant families experience a higher rate of maltreatment and domestic violence [33] and road traffic accidents [34], possibly as a result of low SES [33]. However, mental and physical disorders and child adversity have not been examined in Taiwan, and a comprehensive comparison of mental and physical disorders, adversity, and accidents between families with one immigrant parent and native families is lacking.

The aim of this study was to examine the family health between families with one immigrant parent and native families, and this analysis included the health of all the family members, that is, mothers (married female immigrants or native women), native fathers, and children. To address the aforementioned research gaps such as the small-scale cross-sectional study designs and comparisons of specific diseases, this study used a nationwide population-based cohort database to comprehensively compare the health of the family members between families with one immigrant parent and native families. We compared the risks of common physical and mental disorders, catastrophic illnesses, and mortality between all members of families with one immigrant parent and native families and the risks of domestic violence, maltreatment, sexual assault, and road traffic accidents between children of families with one immigrant parent and those of native families.

Methods

Population

The population data were derived from the Taiwan National Health Insurance Research Database, a medical claims database that includes data on all the medical visits for ambulatory care, emergency care, and hospitalization, which is compulsory social insurance for citizens, immigrants, foreign workers, and foreign students. Up to 99.9% of Taiwan's population is enrolled in this database [35]. We used the Maternal and Child Health Database from the Taiwan National Health Insurance Research Database to extract complete information of live-born children regarding gestational age at birth and the identities of their parents. The Maternal and Child Health Database includes 99.78% of all births nationwide from 2004 to 2016 in Taiwan [36], which was followed up to 2017.

Ethics Approval

This study was approved by the Research Ethics Committee of the China Medical University and Hospital (approval: CMUH108-REC1-142).

Measures

Exposure

The exposure in this study was international married immigrant status. We used the record in the Taiwan Birth Certificate Registration to identify the nationality of the participants, and because the Taiwan Birth Certificate Registration data set from the Taiwan National Health Insurance Research Database contains only data of the mothers, a linkage to the Taiwan Maternal and Child Health Database was made to obtain the complete data of the children and their fathers. The inclusion criteria in this study were as follows: (1) live-born children included in the Taiwan Birth Certificate Registration and (2) no missing data of the children and their fathers and mothers. One exclusion criterion is that we excluded married female immigrants from South Korea, Japan, and western countries because the development index in these countries was close to or higher than that in Taiwan, and we restricted families with one immigrant parent to those in East or Southeast Asian countries, which were defined as China, Vietnam, Indonesia, Thailand, the Philippines, Malaysia, Myanmar, and Cambodia. The exposure group included family members from families with one immigrant parent (native fathers, immigrant mothers, and their children), whereas the nonexposure group included native families (native fathers, native mothers, and their children). A comparison was made separately for individual family members between families with one immigrant parent and native families.

Outcome

The outcomes in this study comprised common physical and mental disorders, Charlson comorbidity index, catastrophic illnesses, mortality, adversities, and accidents. The Charlson comorbidity index was originally designed as a measure to examine the risk of 1-year mortality from comorbid diseases in a longitudinal study of general hospital patients by taking the seriousness of comorbid diseases into account and weighting them to calculate a comorbidity score [37]. If participants have 1 of the 19 diseases, they receive a corresponding weight score. The assigned weights for diseases were as follows: 1 for myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, connective tissue disease, ulcer disease, mild liver disease, and diabetes; 2 for diabetes with end-organ damage, hemiplegia, moderate or severe renal disease, any tumor, leukemia, and lymphoma; 3 for moderate or severe liver disease; and 6 for metastatic solid tumor and AIDS. The potential range of the Charlson comorbidity index is from 0 to 37, with a higher score indicating worse comorbidity. It was subsequently adapted for use as an index of general health [38]. The Charlson comorbidity index and common mental disorders were used to represent physical health and mental health in this study and were determined based on the International Classification of Diseases, Ninth Revision and Tenth Revision codes in the Taiwan National Health Insurance Research Database. We excluded some uncommon mental disorders specific to children and adults because some disorders are differentially prevalent in different age groups. For example, regarding physical health, dementia is included in the Charlson comorbidity index, but dementia is not diagnosed in children, whereas for mental health, conduct disorder or oppositional defiant disorder (CD/ODD) and tic disorders are childhood mental disorders, and when individuals with conduct disorder reach adulthood, their symptoms may be exhibited as antisocial personality disorder [39]. Antisocial personality disorder is usually underdiagnosed and undertreated, and symptoms of tic disorders are usually relieved in adulthood. Specifically, for fathers and mothers, we used all disorders in the Charlson comorbidity index [40] and common mental disorders in

adulthood (ie, autism spectrum disorder [ASD], attention deficit and hyperactivity disorder [ADHD], anxiety disorders, major depressive disorder, bipolar disorder, and schizophrenia); for children, we excluded myocardial infarction and dementia when using Charlson comorbidity index, and we included 2 additional childhood disorders (ie, tic disorder and CD/ODD). Participants were considered to have diseases if they received at least one inpatient diagnosis or more than 2 outpatient diagnoses from 2004 to 2017. Catastrophic illness was determined from the Registry for Catastrophic Illness Patients database, and the categories of diseases for catastrophic illnesses is listed in Table S1 of Multimedia Appendix 1. Mortality was determined by the Cause of Death Data. Child adversities and accidents were defined as the experience of domestic violence, maltreatment, sexual assault, and road traffic accidents and were extracted from the Family Violence Data and the Reported Data of Protection of Child and Youths, the Reported Data of Sexual Assault, and the Traffic Accident Data. Detailed information on these data sets is available in Multimedia Appendix 1.

Covariates

For the analysis of the physical and mental health of the parents, we controlled for age, low-income status, and geographical location (ie, northern, central, southern, and eastern Taiwan). The definition of the geographical location was based on the National Development Council [41]. For the analysis of physical and mental health in children, to control for hereditary factors, we examined the children's physical and mental health and further controlled for their parents' physical health (Charlson comorbidity index) and mental health (schizophrenia, bipolar disorder, ASD, and ADHD), as well as age, sex and low-income status.

Statistical Analysis

SAS version 9.4 (SAS Institute Inc) was used for data management and analysis. Descriptive statistics were applied to indicate the frequency with percentage for categorical variables (ie, sex, low-income status, physical and mental disorders, mortality, and child adversities and accidents) and the mean with SD for continuous variables (ie, age and Charlson comorbidity index score). We used logistic regression for binary outcome variables (ie, physical and mental disorders) and linear regression for continuous outcome variables (ie, Charlson comorbidity index score) to compare the sociodemographic variables and risks of family health among fathers, mothers, and children in families with one immigrant parent and those in native families. For parents, we first performed an unadjusted analysis to compare the sociodemographic variables between families with one immigrant parent and native families and reported crude odds ratios or regression coefficients and 95% CIs. Furthermore, we performed an unadjusted analysis to examine family health while controlling for the sociodemographic variables.

Since the mothers in families with one immigrant parent did not have data on the health care utilization before their immigration unlike the mothers in native families, we performed a sensitivity analysis to restrict a similar period of health care utilization (the start date of health care utilization was the delivery date of their first child) as the native women. Furthermore, we compared the health between the fathers and mothers stratified by families with one immigrant parent and native families by using a similar analysis, and a moderation analysis was performed to examine whether the health between fathers and mothers differed between families with one immigrant parent and native families.

A similar analytical strategy was performed for between the children of families with one immigrant parent and those of native families. Furthermore, in the adjusted model, we first included parental age as a covariate, and we further controlled for the physical (ie, Charlson comorbidity index score) and mental health (ie, ASD, ADHD, and schizophrenia for mental disorders) of the parents separately when we examined the physical and mental health between children of families with one immigrant parent and those of native families. Finally, we further examined the risks of adversity and accidents between them.

Results

Table 1 summarizes the sociodemographic data and the general health of the parents in families with one immigrant parent and those of parents in native families. We included 90,670 fathers, 91,270 mothers, and 132,457 children from families with one immigrant parent and 1,666,775 fathers, 1,734,104 mothers, and 2,637,191 children from native families over a period of 12 years. The fathers (age, 44.8 years vs 40.2 years, respectively) and mothers (age, 35.2 years vs 34.4 years, respectively) in families with one immigrant parent were older than the fathers and mothers in native families. Families with one immigrant parent were more likely to have low-income status than native families. After adjusting for age and low-income status, fathers in families with one immigrant parent had worser physical health than fathers in native families (indicated by the Charlson comorbidity index score with a regression coefficient of 0.05), especially with regard to cardiovascular diseases, cerebrovascular diseases, dementia, diabetes, renal diseases, tumors, and AIDS, with a range of adjusted odds ratios (aORs) from 1.13 to 1.45. Moreover, the risk of catastrophic illness in the fathers of families with one immigrant parent was higher than that in the fathers of native families. The fathers in families with one immigrant parent, in addition to poor physical health, had comparatively poor mental health, specifically with regard to ASD, major depressive disorder, bipolar disorder, and schizophrenia, with a range of aORs from 1.13 to 3.12. Further, the mortality rate in the fathers of families with one immigrant parent was higher than that in the fathers of native families (aOR 1.30, 95% CI 1.22-1.38).

Table 1. Sociodemographic variables and physical and mental health status of fathers and mothers in families with one immigrant parent and of those in native families.

Variable	Father			Mother		
	Families with one immigrant parent (n=90,670)	Native families (n=1,666,775)	Odds ratio[a] (95% CI)	Families with one immigrant parent (n=91,270)	Native families (n=1,734,104)	Odds ratio[a] (95% CI)
Sociodemographics[b]						
Age (years), mean (SD)	44.8 (7.4)	40.2 (0.8)	4.6 (4.5 to 4.6)	35.2 (5.1)	34.4 (6.0)	0.8 (0.7 to 0.9)
Low income, n (%)	9053 (9.98)	74,186 (4.45)	2.38 (2.33 to 2.44)	9108 (9.98)	86,163 (4.97)	2.12 (2. 07 to 2.17)
Geographical location in Taiwan, n (%)						
Northern	44,492 (49.07)	805,225 (48.31)	1.00	45,889 (50.28)	816,943 (47.11)	1.00
Central	19,313 (21.30)	418,668 (25.12)	0.83 (0.82 to 0.85)	20,598 (22.57)	444,424 (25.63)	0.83 (0.81 to 0.84)
Southern	23,683 (26.12)	404,455 (24.27)	1.06 (1.04 to 1.08)	22,748 (24.92)	428,076 (24.69)	0.94 (0.93 to 0.96)
Eastern	3183 (3.51)	38,427 (2.30)	1.50 (1.44 to 1.56)	2034 (2.23)	44,661 (2.58)	0.81 (0.77 to 0.85)
Physical disorders[c]						
Charlson comorbidity index, mean (SD)	1.1 (2.0)	0.8 (1.6)	0.05 (0.04 to 0.06)	0.2 (0.7)	0.6 (1.2)	–0.2 (–0.2 to –0.3)
Myocardial infarction, n (%)	828 (0.91)	6889 (0.41)	1.17 (1.04 to 1.27)	6 (0.01)	573 (0.03)	0.16 (0.06 to 0.42)
Congestive heart failure, n (%)	488 (0.54)	3201 (0.19)	1.45 (1.20 to 1.59)	55 (0.06)	2334 (0.13)	0.48 (0.34 to 0.62)
Peripheral vascular disease, n (%)	875 (0.97)	8291 (0.50)	1.28 (1.18 to 1.36)	148 (0.16)	6280 (0.36)	0.52 (0.42 to 0.59)
Cerebrovascular disease, n (%)	2889 (3.19)	23,924 (1.44)	1.27 (1.21 to 1.31)	246 (0.27)	12,261 (0.71)	0.41 (0.35 to 0.48)
Dementia, n (%)	1383 (1.53)	14,266 (0.86)	1.40 (1.31 to 1.51)	1010 (1.11)	34,136 (1.97)	0.57 (0.50 to 0.62)
Chronic pulmonary disease, n (%)	1807 (1.99)	13,902 (0.83)	0.97 (0.94 to 1.02)	121 (0.13)	6548 (0.38)	0.38 (0.36 to 0.40)
Connective tissue disease, n (%)	10,145 (11.19)	167,583 (10.05)	0.85 (0.80 to 0.97)	3782 (4.14)	188,174 (10.85)	0.34 (0.29 to 0.40)
Ulcer disease, n (%)	1045 (1.15)	17,395 (1.04)	0.95 (0.91 to 0.97)	651 (0.71)	39,232 (2.26)	0.56 (0.53 to 0.59)
Mild liver disease, n (%)	17,135 (18.90)	272,341 (16.34)	0.92 (0.89 to 0.97)	6865 (7.52)	241,908 (13.95)	0.49 (0.43 to 0.55)
Diabetes, n (%)	9905 (10.92)	87,297 (5.24)	1.36 (1.27 to 1.41)	834 (0.91)	43,403 (2.50)	0.42 (0.39 to 0.51)
Diabetes with end-organ damage, n (%)	2952 (3.26)	23,862 (1.43)	1.32 (1.25 to 1.39)	101 (0.11)	8777 (0.51)	0.24 (0.21 to 0.28)
Hemiplegia, n (%)	12,422 (13.70)	192,840 (11.57)	1.25 (1.16 to 1.32)	1725 (1.89)	80,778 (4.66)	0.42 (0.32 to 0.52)
Moderate or severe renal disease, n (%)	7182 (7.92)	121,505 (7.29)	1.20 (1.16 to 1.24)	1620 (1.77)	63,672 (3.67)	0.32 (0.30 to 0.34)
Any tumor, n (%)	3519 (3.88)	35,537 (2.13)	1.13 (1.06 to 1.24)	263 (0.29)	17,287 (1)	0.15 (0.13 to 0.17)
Leukemia, n (%)	1598 (1.76)	13,880 (0.83)	1.05 (0.80 to 1.31)	117 (0.13)	20,526 (1.18)	0.49 (0.30 to 0.70)
Lymphoma, n (%)	114 (0.13)	1455 (0.09)	0.92 (0.76 to 1.09)	30 (0.03)	1397 (0.08)	0.31 (0.20 to 0.45)
Moderate or severe liver disease, n (%)	133 (0.15)	1946 (0.12)	0.89 (0.85 to 0.92)	27 (0.03)	1831 (0.11)	0.59 (0.52 to 0.66)
Metastatic solid tumor, n (%)	743 (0.82)	6129 (0.37)	1.13 (1.02 to 1.25)	101 (0.11)	5596 (0.32)	0.45 (0.36 to 0.53)
AIDS, n (%)	70 (0.08)	648 (0.04)	1.38 (1.06 to 1.79)	22 (0.02)	476 (0.03)	0.70 (0.44 to 1.11)
Catastrophic illness, n (%)	5262 (5.80)	46,715 (2.80)	1.43 (1.40 to 1.45)	775 (0.85)	50,324 (2.90)	0.32 (0.30 to 0.34)
Mental disorders, n (%)[c]						
Autism spectrum disorder	10 (0.01)	97 (0.01)	2.73 (1.41 to 5.31)	3 (0.003)	91 (0.005)	0.18 (0.02 to 1.80)
Attention-deficit/hyperactivity disorder	57 (0.06)	2010 (0.12)	1.20 (0.90 to 1.56)	8 (0.01)	1564 (0.09)	0.08 (0.04 to 0.16)

Variable	Father			Mother		
	Families with one immigrant parent (n=90,670)	Native families (n=1,666,775)	Odds ratio[a] (95% CI)	Families with one immigrant parent (n=91,270)	Native families (n=1,734,104)	Odds ratio[a] (95% CI)
Anxiety disorders	2027 (2.24)	33,331 (2)	0.87 (0.84 to 0.92)	880 (0.96)	44,898 (2.59)	0.38 (0.36 to 0.41)
Major depressive disorder	4345 (4.79)	59,754 (3.59)	1.13 (1.09 to 1.19)	1627 (1.78)	104,100 (6)	0.30 (0.28 to 0.32)
Bipolar disorder	918 (1.01)	10,496 (0.63)	1.42 (1.33 to 1.53)	236 (0.26)	16,740 (0.97)	0.26 (0.23 to 0.30)
Schizophrenia	1142 (1.26)	5403 (0.32)	3.12 (2.93 to 3.34)	182 (0.20)	6176 (0.36)	0.50 (0.44 to 0.55)
Mortality, n (%)[b]	2608 (2.88)	18,907 (1.13)	1.30 (1.22 to 1.38)	33 (0.04)	6613 (0.38)	0.10 (0.06 to 0.15)

[a]The values in this column could be odds ratio or the regression coefficient.

[b]Crude analysis was conducted without any adjustment.

[c]Analysis was adjusted for age, geographical location, and low-income status.

Conversely, mothers in families with one immigrant parent had better physical health (indicated by the Charlson comorbidity index score with an adjusted regression coefficient of –0.2), with lower risks in most physical disorders and catastrophic illnesses than mothers in native families (aOR range 0.15-0.57). Moreover, mothers in families with one immigrant parent had better mental health than mothers in native families, with a range of aORs from 0.08 to 0.50. Further, the mortality rate in mothers of families with one immigrant parent was lower than that in mothers of native families (aOR 0.10, 95% CI 0.06-0.15). We further restricted the time period of health care utilization after the delivery of the first child to make the time period in mothers of families with one immigrant parent and those of native families comparable, thereby resulting in a similar pattern of general health between the 2 groups (Table S2 of Multimedia Appendix 1).

We observed that the fathers had statistically poorer health with regard to most physical and mental disorders than the mothers, regardless of whether they belonged to native families or families with one immigrant parent. Furthermore, based on the moderation analysis, we found that such discrepancies in physical and mental disorders between fathers and mothers were more statistically profound in families with one immigrant parent, except for ASD (Table S3 of Multimedia Appendix 1).

Table 2 summarizes the sociodemographic data and the general health between the children of families with one immigrant parent and those of native families. The children of families with one immigrant parent had slightly better physical health than the children of native families (indicated by the Charlson comorbidity index score with an adjusted regression coefficient of –0.01), with a lower risk of cerebrovascular disease, chronic pulmonary disease, and connective tissue disease (aOR range 0.74-0.91) but a higher risk of leukemia (aOR 1.31) and liver diseases (aOR 1.24). A similar pattern was also found for mental health: children of families with one immigrant parent had comparatively lower risks of ADHD, CD/ODD, and anxiety disorders (aOR range 0.62-0.88). However, a higher risk of ASD was observed in the children of families with one immigrant parent than in the children of native families (aOR 1.13). In addition to general health, the adversities and accidents experienced by the children are summarized in Table 3. A higher risk of road traffic accidents was observed among children of families with one immigrant parent than among children of native families (aOR 1.11, 95% CI 1.07-1.16).

Table 2. Sociodemographic variables and physical and mental health of children of families with one immigrant parent and of those of native families.

Variable	Children of families with one immigrant parent (n=132,457)	Children of native families (n=2,637,191)	Odds ratio[a] (95% CI)
Sociodemographics[b]			
Sex (boys), n (%)	68,304 (51.57)	1,364,032 (51.72)	0.99 (0.98 to 1.00)
Age (years), mean (SD)	6.6 (3.8)	6.7 (4.1)	0.1 (0.1 to 0.1)
Low income, n (%)	14,152 (10.68)	147,543 (5.59)	2.02 (1.98 to 2.06)
Geographical location in Taiwan, n (%)			
Northern	61,836 (46.68)	1,221,120 (46.30)	1.00
Central	35,066 (26.47)	691,353 (26.22)	1.00 (0.99 to 1.02)
Southern	32,918 (24.85)	666,032 (25.26)	0.98 (0.96 to 0.99)
Eastern	2637 (1.99)	58,681 (2.23)	0.88 (0.85 to 0.92)
Physical disorders[c]			
Charlson comorbidity index, mean (SD)	0.2 (0.5)	0.3 (0.5)	–0.01 (–0.01 to –0.01)
Congestive heart failure, n (%)	159 (0.12)	3229 (0.12)	0.96 (0.81 to 1.18)
Peripheral vascular disease, n (%)	25 (0.02)	363 (0.01)	1.32 (0.84 to 2.07)
Cerebrovascular disease, n (%)	141 (0.11)	3302 (0.13)	0.82 (0.68 to 0.99)
Chronic pulmonary disease, n (%)	70 (0.05)	1572 (0.06)	0.91 (0.88 to 0.95)
Connective tissue disease, n (%)	25,652 (19.37)	593,158 (22.49)	0.74 (0.56 to 0.98)
Ulcer disease, n (%)	61 (0.05)	2026 (0.08)	1.09 (0.93 to 1.26)
Mild liver disease, n (%)	229 (0.17)	4605 (0.17)	1.23 (0.98 to 1.55)
Diabetes, n (%)	63 (0.05)	1863 (0.07)	0.80 (0.61 to 1.05)
Diabetes with end-organ damage, n (%)	5 (0.004)	205 (0.01)	0.66 (0.26 to 1.64)
Hemiplegia, n (%)	96 (0.07)	1842 (0.07)	0.83 (0.63 to 1.09)
Moderate or severe renal disease, n (%)	352 (0.27)	6073 (0.23)	0.96 (0.84 to 1.10)
Any tumor, n (%)	233 (0.18)	5373 (0.20)	1.11 (0.85 to 1.40)
Leukemia, n (%)	78 (0.06)	1499 (0.06)	1.31 (1.02 to 1.73)
Lymphoma, n (%)	76 (0.06)	1220 (0.05)	1.33 (0.87 to 1.96)
Moderate or severe liver disease, n (%)	32 (0.02)	485 (0.02)	1.24 (1.11 to 1.39)
Metastatic solid tumor, n (%)	27 (0.02)	453 (0.02)	1.33 (0.87 to 2.02)
AIDS, n (%)	3 (0.002)	51 (0.002)	0.74 (0.08 to 5.81)
Catastrophic illness	1690 (1.28)	36,746 (1.39)	0.96 (0.89 to 1.01)
Mental disorders, n (%)[d]			
Autism spectrum disorder	1063 (0.80)	20,251 (0.77)	1.13 (1.03 to 1.20)
Attention-deficit/hyperactivity disorder	3665 (2.77)	90,698 (3.44)	0.88 (0.84 to 0.91)
Tic disorder	440 (0.33)	11,201 (0.43)	0.89 (0.80 to 1.05)
Conduct disorder/oppositional defiant disorder	219 (0.33)	6268 (0.24)	0.77 (0.68 to 0.89)
Anxiety disorders	26 (0.17)	929 (0.04)	0.62 (0.40 to 0.96)
Major depressive disorder	24 (0.02)	833 (0.03)	0.70 (0.46 to 1.09)
Bipolar disorder	5 (0.004)	317 (0.01)	0.44 (0.19 to 1.05)
Schizophrenia	3 (0.002)	59 (0.002)	0.03 (0.72 to 2.79)
Mortality, n (%)[b]	684 (0.52)	14,389 (0.55)	1.06 (0.97 to 1.20)

[a]The values in this column could be odds ratio or the regression coefficient.

[b]Crude analysis was conducted without any adjustment.

[c]Analysis was adjusted for child year of birth, geographical location, sex, parental age, low-income status, and physical disorder status (ie, Charlson comorbidity index).

[d]Analysis was adjusted for child year of birth, geographical location, sex, parental age, low-income status, and mental disorder status (ie, autism spectrum disorder, attention deficit and hyperactivity disorder, and schizophrenia).

Table 3. Child adversity and accidents reported among children of families with one immigrant parent and among those of native families.[a]

Variable	Children of families with one immigrant parent (n=132,457), n (%)	Children of native families (n=2,637,191), n (%)	Adjusted odds ratio (95% CI)
Domestic violence	117 (0.09)	1913 (0.07)	1.08 (0.90-1.31)
Maltreatment	844 (0.64)	20,044 (0.76)	0.62 (0.59-0.70)
Sexual assault survivors	213 (0.16)	3675 (0.14)	1.00 (0.87-1.19)
Road traffic accident	2690 (2.03)	48,589 (1.84)	1.11 (1.07-1.16)

[a]Analysis was adjusted for child year of birth, geographical location, sex, and low-income status.

Discussion

Principal Findings

This study provides a comprehensive view of the general health among members of families with one immigrant parent and among those of native families in Taiwan. Specifically, we found that the fathers in families with one immigrant parent had a higher mortality rate and poorer physical and mental health than the fathers in native families. Conversely, the mothers in families with one immigrant parent had lower mortality rates and better physical and mental health than the mothers in native families. Similarly, the children in families with one immigrant parent showed slightly better physical and mental health than the children in native families. As per our findings, poorer general health in fathers of families with one immigrant parent should be considered as an important public health issue.

Methodological Considerations

Some methodological considerations need to be mentioned before further discussion. First, there is some diversity in the immigrant families in Taiwan; although most immigrant families are formed through international marriage brokerage agencies in Taiwan, not all families adopt this approach. Families formed by love marriage may have better health than those formed through brokerage because strong bonds between couples have positive impacts on physical and mental health [42]. Second, the medical data of the mothers and children of families with one immigrant parent may not reflect their real health status, because the medical information of these immigrant mothers before their immigration was not collected in our national registered database. Moreover, although we controlled for the time issues by restricting the period to after the delivery of their first child to address the lack of data for married immigrants, some medical barriers such as language difficulties [43], inadequate health literacy [44], and inconvenient access to health care institutions [45] may lead to lower utilization of health care services. Furthermore, because mothers are the main caregivers of children in Taiwan, the lower utilization of health care services is also expected to extend to the children of families with one immigrant parent. Third, some immigrant mothers without legal residence permits may not be covered in the national registered birth data set. Fourth, the age ranges of our sample were limited because the national registered birth data sets were established from 2004 to 2016. Therefore, the medical data may not cover common diseases in subsequent adulthood. Finally, the medical data of individuals in Taiwan may differ from those in other countries that have a private health insurance system rather than a national health insurance system, and some information such as education level, employment status, and marital status (eg, divorce or separation) is not fully available in our databases.

With regard to the general health of the fathers in families with one immigrant parent, we found that fathers were a special subpopulation with health vulnerability; they had higher mortality and morbidity due to various physical and mental disorders, comprising cardiovascular diseases, cerebrovascular diseases, dementia, diabetes, renal diseases, tumors, AIDS, ASD, major depressive disorder, bipolar disorder, and schizophrenia, after adjusting for sociodemographic variables. Low SES may be the main reason for explaining their health vulnerability [46,47], and some studies have also reported low SES in fathers of families with one immigrant parent. Low education levels [24,25] have been reported to be associated with poor physical and mental health [48,49], which may be further mediated by inadequate health literacy and health promotion behaviors [50,51]. Moreover, most of the fathers in families with one immigrant parent were low-skilled laborers [26], and low-skilled labor has been reported to lead to a shorter life expectancy [52] and higher health risks [53] due to harmful work styles, the lack of health promotion behaviors, and unhealthy living habits (eg, no regular exercise, smoking, poor diet) [53]. Furthermore, education and occupation are correlated with each other [54]. Our findings for AIDS was in line with a study on the relationship between SES and AIDS [55], which also emphasized the neglect of AIDS prevention in the heterosexual population of Taiwan. As a result, for fathers of families with one immigrant parent, both low SES and poor health [56] made it difficult to find suitable partners among native women, and marriage brokerage agencies thus became another option to get married.

We observed that mothers in families with one immigrant parent were generally healthier than mothers in native families. The healthy immigrant effect is a possible reason to explain our

finding, where healthy people have more advantages to migrate abroad. The salmon bias effect may be less likely to contribute to explaining our findings because the salmon bias effect has been reported more profoundly in mortality cases and in older people [9]. However, we observed comprehensive differences in terms of physical and mental health and mortality in the mothers of families with one immigrant parent, and these mothers were generally middle-aged adults. Most likely, the health differences in terms of physical and mental health and mortality between the mothers in families with one immigrant parent and the mothers in native families can be explained by the healthy immigrant effect, and the health differences may be slightly overestimated because of the salmon bias effect. It is worth mentioning that some evidence was reported in line with the healthy immigrant effect in Taiwan. Married immigrants were observed to have better mental health [12], including lower risks of major depressive disorder [11] and postpartum depression [13], after adjusting for sociodemographic variables, whereas some findings were not in line with the healthy immigrant effect, and the inconsistent findings may be a result of the differences in methodology, culture, and sample characteristics. For studies with contradictory findings in Taiwan, the authors overlooked common confounding effects (eg, sociodemographic variables) between immigrant status and mental health [16,18,57]. In Korea, married immigrants were observed to have poor mental health after accounting for the sociodemographic variables [17,19]. The different findings between Taiwan and Korea may be due to the difference in the xenophobic atmosphere [58-60] or the relative health status in the native populations [61]. Furthermore, the healthy immigrant effect was found to disappear gradually after immigration [62].

We found that the children of families with one immigrant parent generally had slightly better health than the children of native families, but they had higher risks for some diseases (ie, leukemia, liver diseases, and ASD) and road traffic accidents. Parental health and the sociodemographic status may explain these risks. The better health of the mothers in families with one immigrant parent may explain the better health of the children [63]. In contrast, poor paternal SES may adversely affect the health of children [64]. Furthermore, we found that both the fathers and children of families with one immigrant parent had higher risks of ASD, suggesting the importance of genetic heritability. Finally, the children of families with one immigrant parent had higher risks of accidents, which may be explained by the usage of different types of vehicles. Motorcycles, rather than passenger cars, may be a more affordable choice for families with one immigrant parent because the cost of transportation is much lower; however, motorcycles have been reported to lead to a higher risk of traffic accidents than passenger cars [65].

Conclusions

Our study is the first national cohort study, to the best of our knowledge, to comprehensively elucidate the health status of families with one immigrant parent in Taiwan with substantial evidence, and our findings indicate that family health is nonhomogeneous within such families. We found that fathers in families with one immigrant parent generally had a poor physical and mental health but not the mothers. Moreover, the children of families with one immigrant parent generally had slightly better general health than the children of native families, but they had higher risks for some diseases (ie, leukemia, liver diseases, and ASD) and road traffic accidents. These results indicate that since there are health inequalities within the members of families with one immigrant parent, they should be provided with adequate prenatal care and parenting education.

Acknowledgments

This study was supported by grants from the Ministry of Science and Technology, Taiwan (MOST 109-2314-B-468-001-MY2) and National Science and Technology Council, Taiwan (NSTC 111-2314-B-468-003). The supporters had no role in the design, analysis, interpretation, or publication of this study.

Authors' Contributions

YLC conceptualized and led the design of this study. YLC and HYH contributed to the draft, interpretation, and statistical analysis in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and methods.
[DOCX File , 41 KB - publichealth_v8i12e33624_app1.docx]

References

1. International Organization for Migration. World Migration Report 2020. Geneva, Switzerland: International Organization for Migration; 2020.
2. Hiroo K, Chul Woo L. International marriage migrants to rural areas in South Korea and Japan: A comparative analysis. Geog Rev Japan B 2009;81(1):60-67. [doi: 10.4157/geogrevjapanb.81.60]

3. Population in brief 2020. National Population and Talent Division. 2020. URL: https://www.strategygroup.gov.sg/files/media-centre/publications/population-in-brief-2020.pdf [accessed 2022-12-08]
4. The number of marriages by type of marriage and nationality of bride and groom. Ministry of the Interior, Taiwan. 2020. URL: https://www.ris.gov.tw/app/portal/346 [accessed 2022-08-21]
5. The number of Taiwan naturalizations. Ministry of the Interior, Taiwan. URL: https://www.gender.ey.gov.tw/gecdb/Stat_Statistics_DetailData.aspx?sn=2Q9lfo%24P0c647DizBjI92A%40%40 [accessed 2022-08-21]
6. The number of foreign spouses by residence permit. Ministry of the Interior, Taiwan. URL: https://www.immigration.gov.tw/5385/7344/7350/8887/?alias=settledown [accessed 2022-08-21]
7. The number of newborns by biological mother's nationality of origin, age and education level. Ministry of the Interior, Taiwan. URL: https://data.gov.tw/dataset/32945 [accessed 2022-08-19]
8. Constant AF, García-Muñoz T, Neuman S, Neuman T. Micro and Macro Determinants of Health: Older Immigrants in Europe. SSRN Journal 2015:1-37. [doi: 10.2139/ssrn.2547855]
9. Di Napoli A, Rossi A, Alicandro G, Ventura M, Frova L, Petrelli A. Salmon bias effect as hypothesis of the lower mortality rates among immigrants in Italy. Sci Rep 2021 Apr 13;11(1):8033 [FREE Full text] [doi: 10.1038/s41598-021-87522-2] [Medline: 33850208]
10. Yu Z, Bowers B, Yeoh BSA. A Scoping Review of the Health of East and Southeast Asian Female Marriage Migrants. J Immigr Minor Health 2020 Mar;22(1):182-211. [doi: 10.1007/s10903-019-00901-w] [Medline: 31152303]
11. Huang S, Yang H. Is There a Healthy Immigrant Effect Among Women Through Transnational Marriage? Results from Immigrant Women from Southeast Asian Countries in Taiwan. J Immigr Minor Health 2018 Mar;20(1):178-187. [doi: 10.1007/s10903-016-0513-0] [Medline: 27743156]
12. Chou FH, Chen P, Liu R, Ho C, Tsai K, Ho W, et al. A comparison of quality of life and depression between female married immigrants and native married women in Taiwan. Soc Psychiatry Psychiatr Epidemiol 2010 Sep;45(9):921-930. [doi: 10.1007/s00127-009-0130-5] [Medline: 19730760]
13. Chen H, Chien L. A comparative study of domestic decision-making power and social support as predictors of postpartum depressive and physical symptoms between immigrant and native-born women. PLoS One 2020;15(4):e0231340 [FREE Full text] [doi: 10.1371/journal.pone.0231340] [Medline: 32267897]
14. Diaz E, Poblador-Pou B, Gimeno-Feliu L, Calderón-Larrañaga A, Kumar BN, Prados-Torres A. Multimorbidity and Its Patterns according to Immigrant Origin. A Nationwide Register-Based Study in Norway. PLoS One 2015;10(12):e0145233 [FREE Full text] [doi: 10.1371/journal.pone.0145233] [Medline: 26684188]
15. Ahn J, Kim T, Roh EH, Song J. Health of International Marriage Immigrant Women in South Korea: A Systematic Review. J Immigr Minor Health 2018 Jun;20(3):717-728. [doi: 10.1007/s10903-017-0604-6] [Medline: 28584961]
16. Tsao Y, Creedy D, Gamble J. A Comparison of Life Stress and Depressive Symptoms in Pregnant Taiwanese and Immigrant Women. J Nurs Res 2016 Sep;24(3):272-281. [doi: 10.1097/jnr.0000000000000137] [Medline: 26998771]
17. Choi SY, Kim EJ, Ryu E, Chang KO, Park MN. Postpartum depression and parental self-efficacy: a comparison of native Korean and Vietnamese immigrant mothers in Korea. J Transcult Nurs 2012 Apr;23(2):181-187. [doi: 10.1177/1043659611434057] [Medline: 22294339]
18. Chien L, Tai C, Yeh M. Domestic decision-making power, social support, and postpartum depression symptoms among immigrant and native women in Taiwan. Nurs Res 2012;61(2):103-110. [doi: 10.1097/NNR.0b013e31824482b6] [Medline: 22307142]
19. Lee SH, Park YC, Hwang J, Im JJ, Ahn D. Mental health of intermarried immigrant women and their children in South Korea. J Immigr Minor Health 2014 Mar;16(1):77-85. [doi: 10.1007/s10903-012-9747-7] [Medline: 23184349]
20. Liu C, Chang N, Chou P. Seroprevalence of HBV in immigrant pregnant women and coverage of HBIG vaccine for neonates born to chronically infected immigrant mothers in Hsin-Chu County, Taiwan. Vaccine 2007 Nov 01;25(44):7706-7710. [doi: 10.1016/j.vaccine.2007.07.056] [Medline: 17767981]
21. Liu Y, Chen H, Lee J, Chu K, Chien L. Factors Associated With Drinking Behavior Among Immigrant Women in Taiwan. Subst Use Misuse 2017 Apr 16;52(5):674-682. [doi: 10.1080/10826084.2016.1253745] [Medline: 28135890]
22. Lee L, Yin T, Yu S. Prenatal examination utilization and its determinants for immigrant women in Taiwan: an exploratory study. J Nurs Res 2009 Mar;17(1):73-82. [doi: 10.1097/JNR.0b013e3181999ee8] [Medline: 19352231]
23. Chang J, Huang W, Chen Y, Gau SS. The mental health of immigrant mother's offspring in Taiwan: A national epidemiological study. J Formos Med Assoc 2020 Mar;119(2):601-609 [FREE Full text] [doi: 10.1016/j.jfma.2019.08.033] [Medline: 31543300]
24. Lin L, Hung C. Vietnamese women immigrants' life adaptation, social support, and depression. J Nurs Res 2007 Dec;15(4):243-254. [doi: 10.1097/01.jnr.0000387621.95306.98] [Medline: 18080969]
25. Chen M, Tang C, Jeng H, Chiu A. The maternal and child healthcare needs of new immigrants in Taipei. J Nurs Res 2008 Dec;16(4):307-320. [doi: 10.1097/01.jnr.0000387318.50880.b5] [Medline: 19061177]
26. Huang YC, Mathers NJ. Postnatal depression and the experience of South Asian marriage migrant women in Taiwan: survey and semi-structured interview study. Int J Nurs Stud 2008 Jun;45(6):924-931. [doi: 10.1016/j.ijnurstu.2007.02.006] [Medline: 17418193]

27. Wang H. Immigration trends and policy changes in Taiwan. Asian and Pacific Migration Journal. URL: https://twsouthernsoc.nsysu.edu.tw/wp-content/uploads/2021/03/Immigration-Trends-and-Policy-Changes-in-Taiwan-.pdf [accessed 2022-11-30]
28. Tsai S, Sun W. Evaluating the satisfaction of immigrant women from a rural community regarding family functioning and health-related quality of life. Women Health 2013;53(2):135-153. [doi: 10.1080/03630242.2013.767302] [Medline: 23517512]
29. Hsiao C, Chien M, Wu H, Chiao C. Correlates of health-promoting lifestyle behaviors among Vietnamese female immigrants in Taiwan. Women Health 2017 Mar;57(3):342-357. [doi: 10.1080/03630242.2016.1160966] [Medline: 26959383]
30. Yang H, Kuo Y, Wang L, Yang C. Culture, parenting, and child behavioral problems: a comparative study of cross-cultural immigrant families and native-born families in Taiwan. Transcult Psychiatry 2014 Aug;51(4):526-544. [doi: 10.1177/1363461514532306] [Medline: 24803539]
31. Lin F, Tung H, Hsieh Y, Lin J. Interactive influences of family and school ecologies on the depression status among children in marital immigrant families. Res Dev Disabil 2011;32(6):2027-2035. [doi: 10.1016/j.ridd.2011.08.009] [Medline: 21985985]
32. Hsieh W, Hsieh C, Jeng S, Liao H, Su Y, Lin S, et al. Favorable neonatal outcomes among immigrants in Taiwan: evidence of healthy immigrant mother effect. J Womens Health (Larchmt) 2011 Jul;20(7):1083-1090. [doi: 10.1089/jwh.2011.2809] [Medline: 21668384]
33. Euser EM, van Ijzendoorn MH, Prinzie P, Bakermans-Kranenburg MJ. Elevated child maltreatment rates in immigrant families and the role of socioeconomic differences. Child Maltreat 2011 Feb;16(1):63-73. [doi: 10.1177/1077559510385842] [Medline: 21041234]
34. Kazi M, Ferdous M, Rumana N, Vaska M, Turin T. Injury among the immigrant population in Canada: exploring the research landscape through a systematic scoping review. Int Health 2019 May 01;11(3):203-214. [doi: 10.1093/inthealth/ihy086] [Medline: 30452624]
35. Hsieh C, Su C, Shao S, Sung S, Lin S, Yang Kao Y, et al. Taiwan's National Health Insurance Research Database: past and future. CLEP 2019 May;11:349-358. [doi: 10.2147/clep.s196293]
36. Li CY, Chen LH, Chiou MJ, Liang FW, Lu TH. Set-up and future applications of the Taiwan Maternal and Child Health Database (TMCHD). Taiwan Journal of Public Health 2016;35(2):209-220. [doi: 10.6288/TJPH201635104053]
37. Charlson ME, Pompei P, Ales KL, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
38. Ventimiglia E, Capogrosso P, Boeri L, Serino A, Colicchia M, Ippolito S, et al. Infertility as a proxy of general male health: results of a cross-sectional survey. Fertil Steril 2015 Jul;104(1):48-55 [FREE Full text] [doi: 10.1016/j.fertnstert.2015.04.020] [Medline: 26006735]
39. American PA. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). Washington, DC: American Psychiatric Assoc Pub; 2013.
40. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sørensen HT. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. BMC Med Res Methodol 2011 May 28;11:83 [FREE Full text] [doi: 10.1186/1471-2288-11-83] [Medline: 21619668]
41. Urban and regional development statistics of Republic of China (Taiwan). National Development Council. URL: https://www.ndc.gov.tw/Content_List.aspx?n=AE22CFA2B47614B7 [accessed 2022-08-18]
42. Bhattacharyya KK. The sacred relationship between marriage, spirituality and healthy aging in Hinduism. Journal of Religion, Spirituality & Aging 2019 Sep 24;32(2):135-148. [doi: 10.1080/15528030.2019.1670771]
43. Bricard D, Jusot F, Trannoy A, Tubeuf S. Inequality of opportunities in health and death: an investigation from birth to middle age in Great Britain. Int J Epidemiol 2020 Oct 01;49(5):1739-1748 [FREE Full text] [doi: 10.1093/ije/dyaa130] [Medline: 33011793]
44. Lee F, Wang H, Tsai H, Lin M. Factors associated with receiving Pap tests among married immigrant women of Vietnamese origin in southern Taiwan. Women Health 2016;56(3):243-256. [doi: 10.1080/03630242.2015.1088113] [Medline: 26362802]
45. Liang Y, Chang H, Lin Y, Lin L, Chen W. Factors affecting adequate prenatal care and the prenatal care visits of immigrant women to Taiwan. J Immigr Minor Health 2014 Mar;16(1):44-52. [doi: 10.1007/s10903-012-9734-z] [Medline: 23065308]
46. Wilkinson R, Marmot M. Social Determinants of Health: The Solid Facts. Geneva, Switzerland: World Health Organization; 2003.
47. Chan CQH, Lee KH, Low LL. A systematic review of health status, health seeking behaviour and healthcare utilisation of low socioeconomic status populations in urban Singapore. Int J Equity Health 2018 Apr 02;17(1):39 [FREE Full text] [doi: 10.1186/s12939-018-0751-y] [Medline: 29609592]
48. Veenstra G, Vanzella-Yang A. Does household income mediate the association between education and health in Canada? Scand J Public Health 2021 Dec;49(8):857-864. [doi: 10.1177/1403494820917534] [Medline: 32400282]
49. Milner A, Blakely T, Disney G, Kavanagh A, LaMontagne A, Aitken Z. Do employment factors reduce the effect of low education on mental health? A causal mediation analysis using a national panel study. Int J Epidemiol 2018 Oct 01;47(5):1423-1431. [doi: 10.1093/ije/dyy128] [Medline: 29992247]

50. Yamashita T, Bardo AR, Cummins PA, Millar RJ, Sahoo S, Liu D. The Roles of Education, Literacy, and Numeracy in Need for Health Information during the Second Half of Adulthood: A Moderated Mediation Analysis. J Health Commun 2019;24(3):271-283. [doi: 10.1080/10810730.2019.1601303] [Medline: 30982431]
51. Park CL, Cho D, Moore PJ. How does education lead to healthier behaviours? Testing the mediational roles of perceived control, health literacy and social support. Psychol Health 2018 Nov;33(11):1416-1429. [doi: 10.1080/08870446.2018.1510932] [Medline: 30450977]
52. Brønnum-Hansen H, Foverskov E, Andersen I. Occupational inequality in health expectancy in Denmark. Scand J Public Health 2020 May;48(3):338-345. [doi: 10.1177/1403494819882138] [Medline: 31763956]
53. Väisänen D, Kallings LV, Andersson G, Wallin P, Hemmingsson E, Ekblom-Bak E. Lifestyle-associated health risk indicators across a wide range of occupational groups: a cross-sectional analysis in 72,855 workers. BMC Public Health 2020 Nov 04;20(1):1656 [FREE Full text] [doi: 10.1186/s12889-020-09755-6] [Medline: 33148214]
54. Andersen R, van de Werfhorst HG. Education and occupational status in 14 countries: the role of educational institutions and labour market coordination. Br J Sociol 2010 Jun;61(2):336-355. [doi: 10.1111/j.1468-4446.2010.01315.x] [Medline: 20579057]
55. Lee C, Jen I, Lan Y, Yen Y, Chuang P, Chen M, et al. AIDS incidence trends at presentation and during follow-up among HIV-at-risk populations: a 15-year nationwide cohort study in Taiwan. BMC Public Health 2018 May 03;18(1):589 [FREE Full text] [doi: 10.1186/s12889-018-5500-z] [Medline: 29720151]
56. Breslau J, Miller E, Jin R, Sampson N, Alonso J, Andrade L, et al. A multinational study of mental disorders, marriage, and divorce. Acta Psychiatr Scand 2011 Dec;124(6):474-486 [FREE Full text] [doi: 10.1111/j.1600-0447.2011.01712.x] [Medline: 21534936]
57. Tsao Y, Creedy D, Gamble J. An Exploration of Parenting Stress in Immigrant and Taiwanese Mothers. J Perinat Neonatal Nurs 2015;29(4):287-295. [doi: 10.1097/JPN.0000000000000123] [Medline: 26262945]
58. Kim Y, Son I, Wie D, Muntaner C, Kim H, Kim S. Don't ask for fair treatment? A gender analysis of ethnic discrimination, response to discrimination, and self-rated health among marriage migrants in South Korea. Int J Equity Health 2016 Jul 19;15(1):112 [FREE Full text] [doi: 10.1186/s12939-016-0396-7] [Medline: 27430432]
59. Ariely G. Global identification, xenophobia and globalisation: A cross-national exploration. Int J Psychol 2017 Dec;52 Suppl 1:87-96. [doi: 10.1002/ijop.12364] [Medline: 27425583]
60. Straiton ML, Aambø AK, Johansen R. Perceived discrimination, health and mental health among immigrants in Norway: the role of moderating factors. BMC Public Health 2019 Mar 20;19(1):325 [FREE Full text] [doi: 10.1186/s12889-019-6649-9] [Medline: 30894173]
61. Noh J, Kim J, Yang Y, Park J, Cheon J, Kwon YD. Body mass index and self-rated health in East Asian countries: Comparison among South Korea, China, Japan, and Taiwan. PLoS One 2017;12(8):e0183881 [FREE Full text] [doi: 10.1371/journal.pone.0183881] [Medline: 28846742]
62. Bas-Sarmiento P, Saucedo-Moreno MJ, Fernández-Gutiérrez M, Poza-Méndez M. Mental Health in Immigrants Versus Native Population: A Systematic Review of the Literature. Arch Psychiatr Nurs 2017 Mar;31(1):111-121. [doi: 10.1016/j.apnu.2016.07.014] [Medline: 28104048]
63. Waters E, Doyle J, Wolfe R, Wright M, Wake M, Salmon L. Influence of parental gender and self-reported health and illness on parent-reported child health. Pediatrics 2000 Dec;106(6):1422-1428. [doi: 10.1542/peds.106.6.1422] [Medline: 11099598]
64. Mock-Muñoz de Luna CJ, Vitus K, Torslev MK, Krasnik A, Jervelund SS. Ethnic inequalities in child and adolescent health in the Scandinavian welfare states: The role of parental socioeconomic status - a systematic review. Scand J Public Health 2019 Nov;47(7):679-689. [doi: 10.1177/1403494818779853] [Medline: 29956595]
65. Feng C. Road accidents profile in Taiwan. IATSS Research 2007;31(1):138-139. [doi: 10.1016/s0386-1112(14)60192-8]

Abbreviations

ADHD: attention deficit and hyperactivity disorder
aOR: adjusted odds ratio
ASD: autism spectrum disorder
CD/ODD: conduct disorder or oppositional defiant disorder
SES: socioeconomic status

Edited by H Bradley; submitted 16.09.21; peer-reviewed by J Sonney, SH Jang; comments to author 25.01.22; revised version received 19.02.22; accepted 11.10.22; published 19.12.22.

Please cite as:
Chen YL, Ho HY
Comprehensive Comparisons of Family Health Between Families With One Immigrant Parent and Native Families in Taiwan: Nationwide Population-Based Cohort Study
JMIR Public Health Surveill 2022;8(12):e33624
URL: https://publichealth.jmir.org/2022/12/e33624
doi:10.2196/33624
PMID:36534459

Original Paper

The Global, Regional, and National Burdens of Cervical Cancer Attributable to Smoking From 1990 to 2019: Population-Based Study

Ruixia Yuan[1,2], PhD; Fang Ren[3], PhD; Yingying Xie[4], MPH; Kaixiang Li[1,2], MCS; Zhuang Tong[1,2], MCS

[1]Clinical Big Data Center, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

[2]Henan Academy of Medical Big Data, Zhengzhou, China

[3]Department of Gynecology, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

[4]Department of Scientific Research, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Corresponding Author:
Zhuang Tong, MCS
Clinical Big Data Center
the First Affiliated Hospital of Zhengzhou University
No 1 East Jianshe Road, Erqi District
Zhengzhou, 450052
China
Phone: 86 15638523311
Fax: 86 0371 66862923
Email: ttongzhuang@126.com

Abstract

Background: Cervical cancer is the fourth most common cause of cancer death in women worldwide. Smoking is one of the risk factors for cervical cancer. Understanding the global distribution of the disease burden of cervical cancer attributable to smoking and related changes is of clear significance for the prevention and control of cervical cancer in key populations and for tobacco control. As far as we know, research on the burden of cervical cancer attributable to smoking is lacking.

Objective: We estimated the disease burden and mortality of cervical cancer attributable to smoking and related trends over time at the global, regional, and national levels.

Methods: Data were obtained from the Global Burden of Disease study website. Age-standardized rates were used to facilitate comparisons of mortality and disability-adjusted life years (DALYs) at different levels. The estimated annual percentage change (EAPC) was used to assess trends in the age-standardized mortality rate (ASMR) and the age-standardized DALY rate (ASDR). A Pearson correlation analysis was used to evaluate correlations between the sociodemographic index and the age-standardized rates.

Results: In 2019, there were 30,136.65 (95% uncertainty interval [UI]: 14,945.09-49,639.87) cervical cancer–related deaths and 893,735.25 (95% UI 469,201.51-1,440,050.85) cervical cancer–related DALYs attributable to smoking. From 1990 to 2019, the global burden of cervical cancer attributable to smoking showed a decreasing trend around the world; the EAPCs for ASMR and ASDR were –2.11 (95% CI –2.16 to –2.06) and –2.22 (95% CI –2.26 to –2.18), respectively. In terms of age characteristics, in 2019, an upward trend was observed for age in the mortality of cervical cancer attributable to smoking. Analysis of the trend in DALYs with age revealed an initially increasing and then decreasing trend. From 1990 to 2019, the burden of disease in different age groups showed a downward trend. Among 204 countries, 180 countries showed downward trends, 10 countries showed upward trends, and the burden was stable in 14 countries. The Pearson correlation analysis revealed a significant negative correlation between sociodemographic index and the age-standardized rates of cervical cancer attributable to smoking (ρ=–0.228, P<.001 for ASMR and ρ=–0.223, P<.001 for ASDR).

Conclusions: An increase over time in the absolute number of cervical cancer deaths and DALYs attributable to smoking and a decrease over time in the ASMR and ASDR for cervical cancer attributable to smoking were observed in the overall population, and differences in these variables were also observed between countries and regions. More attention should be paid to cervical cancer prevention and screening in women who smoke, especially in low- and middle-income countries.

(JMIR Public Health Surveill 2022;8(12):e40657) doi:10.2196/40657

KEYWORDS

global burden of disease; cervical cancer; smoking women; time trends

Introduction

Cervical cancer is the fourth most common cause of cancer death in women worldwide [1]. It is estimated that in 2020, more than 600,000 new cases of cervical cancer and as many as 342,000 cervical cancer–related deaths occurred around the world [2]. Although rare in the population younger than 15 years, cervical cancer occurs in women of all ages, with the highest incidence observed in women aged 40 to 60 years [3]. It thus poses a serious threat to women's health. Persistent infection with human papillomavirus (HPV), especially types 16 and 18, is the main cause of cervical cancer [4,5]. Socioeconomic conditions [6], smoking, exposure to secondhand smoke, factors related to sexual behavior and childbirth, inflammation of the female reproductive system, and a family history of malignant tumors have also been identified as risk factors for cervical cancer [7-9].

Among the above risk factors, we focus here on the influence of smoking. First, smoking is a known independent risk factor for cervical cancer. A meta-analysis of the results of 9 studies with a low risk of bias [10] showed a significant correlation between smoking and cervical cancer (odds ratio [OR] 3.05, 95% CI 1.73-5.38). Second, smoking increases the rate of persistent high-risk HPV infection [11]. Third, smoking has been reported to reduce survival among patients with cervical cancer [12]. In 2019, the age-standardized prevalence of smoking among women aged 15 years or older was 6.62%, and smoking was identified as the cause of 1.51 million deaths, or 5.84% of all deaths, in this population [13].

It is particularly important to pay attention to key groups of patients with cervical cancer [14]. To the best of our knowledge, studies on the disease burden of cervical cancer attributable to smoking, especially the spatiotemporal trends, are still lacking. An understanding of the temporal and spatial changes in this burden at the global level would significantly enhance the development of active prevention and control measures tailored to local conditions. In this study, we used data from the Global Burden of Disease (GBD) study to estimate the global burden of cervical cancer attributable to smoking from 1990 to 2019 and describe changes in this burden over time. Our study aimed to answer the following questions: (1) How much of the disease burden of cervical cancer is attributable to smoking in the female population, and how is this burden changing? (2) Are the distribution and trends of the disease burden of cervical cancer attributable to smoking different in different countries and regions of the world, and are these differences related to the level of socioeconomic development? and (3) What are the age characteristics of the disease burden of cervical cancer attributable to smoking?

Methods

Study Data

The data used in this study were obtained from the 2019 GBD study, which provides a comprehensive and systematic assessment of the global burdens of 369 diseases, injuries, and impairments and 87 risk factors in 204 countries and territories during the 1990 to 2019 period. For this study, we extracted data on the burden of cervical cancer attributable to smoking, including mortality (calculated as the number of cervical cancer deaths × 100,000 / female population) [15] and disability-adjusted life years (DALYs; the sum of years of life lost due to premature mortality and years lived with disability) [16], at the global, regional, and national levels.

In the current study, spatial division of the sample was achieved using 3 GBD division methods. The first method is based on the sociodemographic index (SDI), a comprehensive index used to measure the level of social development in a geographic area. The included countries and territories were divided by SDI into 5 superregions: low SDI, low-middle SDI, middle SDI, high-middle SDI, and high SDI. The second method involves dividing the world into 21 geographic regions according to epidemiological similarity and geographical proximity (eg, East Asia, Australasia, and Central Europe). The third method involves simple division by country and territory (for a total of 204 countries). Furthermore, the sample population of women aged 30 to 79 years was divided into 5-year age groups (eg, 30-34 years), and those aged 80 years and older were combined into one group, to yield 11 age groups.

Definition of Smoking Exposure

According to the GBD risk factor collaborators [17], smoking exposure is defined as the current or previous use of any smoking tobacco product other than electronic cigarettes or vaporizers. For current smokers, exposure was estimated using 2 continuous measures: the number of cigarettes per smoker per day and the number of pack-years. For former smokers, exposure was estimated using the number of years since smoking cessation.

Statistical Analysis

In the GBD study, population data from the report *World Population Prospects, 2012 Revision* was used for standardization [18]. Age-standardized rates (ASRs) were used to eliminate the influence of different population groups and differences in the age structure of populations over time, thus ensuring the comparability of statistical indicators. We used absolute numbers and ASRs (per 100,000 people) with 95% uncertainty intervals (UIs) to describe the burden and mortality of cervical cancer attributable to smoking. We further used the age-standardized mortality rate (ASMR) and age-standardized DALY rate (ASDR) to compare the burden of cervical cancer attributable to smoking over time and between regions.

The estimated annual percentage change (EAPC) was used to assess trends in the ASMR and ASDR. The natural logarithm

of the ASR was fitted to the following regression line model: ln (ASR) = α + βx + , where x is the calendar year. EAPCs and 95% CIs were derived from the following regression model: y = 100 × (exp (β) − 1), where y is the EAPC [19]. An ASR was considered to increase during the observation period when the lower limit of the 95% CI of the EAPC was greater than zero, and to decrease when the upper limit of the 95% CI was less than zero; otherwise, the ASR was considered stable. A Pearson correlation analysis was used to evaluate the correlations between SDI and ASRs.

The methods used to estimate disease burdens from indicators in the GBD have been described previously [16,20]. All statistical analyses were performed using R (version 4.1.3; R Foundation for Statistical Computing). A 2-sided *P* value <.05 was considered to indicate a statistically significant difference.

Ethical Considerations

This study was approved by the Ethics Committee of Scientific Research and Clinical Trials of the First Affiliated Hospital of Zhengzhou University (NO. 2020-KY-167).

Results

Global Burden of Cervical Cancer Attributable to Smoking in 2019 and Temporal Trends From 1990 to 2019

Detailed results on the global burden of mortality and disability-adjusted life-years due to cervical cancer attributable to smoking in 1990 and 2019 and temporal trends from 1990 to 2019 are provided in Table 1, Table 2, and Table S1 in Multimedia Appendix 1. In 2019, the numbers of female deaths and DALYs attributable to cervical cancer worldwide were approximately 280,479.04 and 8,955,012.78, respectively. Of these, 30,136.65 (95% UI 14,945.09-49,639.87) cervical cancer–related deaths and 893,735.25 (95% UI 469,201.51-1,440,050.85) cervical cancer–related DALYs were attributable to smoking.

Overall, a decreasing trend was observed in the global burden of cervical cancer attributable to smoking over the studied period. The ASMR decreased from 1.28 per 100,000 (95% UI 0.65-2.06) in 1990 to 0.69 per 100,000 (95% UI 0.35-1.14) in 2019, a decrease of 85.5%, with an EAPC of –2.11 (95% CI –2.16 to –2.06) during this period. The ASDR also decreased from 39.31 per 100,000 (95% UI 21.03-62.13) in 1990 to 20.75 per 100,000 (95% UI 10.85-33.51) in 2019, with an EAPC of –2.22 (95% CI –2.26 to –2.18).

Table 1. The global burden of death due to cervical cancer attributable to smoking in 1990 and 2019 and temporal trends from 1990 to 2019.

	1990		2019		Estimated annual percentage change from 1990 to 2019 (95% CI)
	Deaths, n (95% UI[a])	Age-standardized mortality rate, n × 10^{-5} (95% UI)	Deaths, n (95% UI)	Age-standardized mortality rate, n × 10^{-5} (95% UI)	
Global burden	27,421.87 (13,984.40 to 6721.02)	1.28 (0.65 to 2.06)	30,136.65 (14,945.09 to 49,639.87)	0.69 (0.35 to 1.14)	–2.11 (–2.16 to –2.06)
Burden by sociodemographic index					
Low	1830.31 (847.48 to 3193.99)	1.43 (0.67 to 2.47)	2665.07 (1196.19 to 4655.28)	0.95 (0.44 to 1.64)	–1.52 (–1.59 to –1.44)
Low-middle	3986.82 (2018.69 to 6721.02)	1.28 (0.64 to 2.13)	5040.14 (2422.69 to 8955.89)	0.69 (0.34 to 1.25)	–2.25 (–2.32 to –2.08)
Middle	6434.95 (3243.32 to 10,665.85)	1.19 (0.58 to 1.98)	7557.27 (3565.50 to 13,116.97)	0.57 (0.27 to 1.00)	–2.69 (–2.81 to –2.57)
High-middle	7475.99 (3799.58 to 12,052.88)	1.27 (0.65 to 2.04)	8453.26 (4173.18 to 13,819.28)	0.79 (0.40 to 1.29)	–1.52 (–1.61 to –1.43)
High	7669.50 (3471.33 to 12,062.72)	1.43 (0.68 to 2.22)	6390.02 (2793.35 to 10,482.46)	0.75 (0.34 to 1.19)	–2.17 (–2.28 to –2.05)

[a]UI: uncertainty interval.

Table 2. The global burden of disability-adjusted life-years due to cervical cancer attributable to smoking in 1990 and 2019 and temporal trends from 1990 to 2019.

	1990		2019		Estimated annual percentage change from 1990 to 2019 (95% CI)
	DALYs[a], n (95% UI[b])	Age-standardized DALY rate, n × 10^{-5} (95% UI)	DALYs, n (95% UI)	Age-standardized DALY rate, n × 10^{-5} (95% UI)	
Global burden	863,494.73 (464,325.44 to 1,365,563.86)	39.31 (21.03 to 62.13)	893,735.25 (469,201.51 to 1,440,050.85)	20.75 (10.85 to 33.51)	–2.22 (–2.26 to –2.18)
Burden by sociodemographic index					
Low	61,306.72 (26,248.83 to 107,671.43)	43.51 (19.69 to 76.28)	85,793.88 (34,342.57 to 153,086.77)	27.53 (11.86 to 48.17)	–1.68 (–1.76 to –1.69)
Low-middle	127,334.6 (64,663.76 to 208,130.21)	37.08 (18.61 to 61.43)	150,971.38 (71,542.69 to 275,063.88)	19.69 (9.37 to 35.44)	–2.32 (–2.44 to –2.19)
Middle	197,494.21 (103,798.33 to 328,143.85)	33.63 (17.30 to 55.60)	216,363.48 (107,107.20 to 372,743.62)	15.65 (7.73 to 26.95)	–2.75 (–2.85 to –2.66)
High-middle	242,456.35 (131,242.40 to 377,394.25)	41.44 (22.49 to 64.40)	260,369.48 (137,295.16 to 409,749.35)	25.66 (13.93 to 40.26)	–1.58 (–1.66 to –1.49)
High	234,111.80 (117,412.53 to 360,250.81)	46.78 (24.10 to 70.67)	179,289.26 (85,852.04 to 285,069.42)	23.88 (12.15 to 36.84)	–2.27 (–2.38 to –2.16)

[a]DALY: disability-adjusted life-year.

[b]UI: uncertainty interval.

Burden by SDI Superregion

Analysis of the 5 superregions divided by SDI revealed that the absolute number of cervical cancer deaths attributable to smoking ranged from 2665.07 (95% UI 1196.19-4655.28) in the low-SDI region to 8453.26 (95% UI 4173.18-13,819.28) in the high-middle–SDI region. The lowest and highest absolute numbers of DALYs were 85,793.88 (95% UI 34,342.57-153,086.77) in the low-SDI region and 260,369.48 (95% UI 137,295.16-409,749.35) in the high-middle–SDI region, respectively.

The highest ASMR and ASDR values were 0.95 per 100,000 (95% UI 0.44-1.64) and 27.53 per 100,000 (95% UI 11.86-48.17), respectively, in the low-SDI region, and the lowest ASMR and ASDR values were 0.6 per 100,000 (95% UI 0.35-1.14) and 15.65 per 100,000 (95% UI 7.73-26.95), respectively, in the middle-SDI region.

Downward trends in the ASMR and ASDR were observed in all 5 SDI superregions from 1990 to 2019. The largest declines were observed in the middle-SDI region, with EAPCs of –2.69 (95% CI –2.81 to –2.57) and –2.75 (95% CI –2.85 to –2.66) for the ASMR and ASDR, respectively. The smallest declines in the ASMR and ASDR were observed in the low-SDI and high-middle–SDI regions, respectively, with EAPCs of –1.51 (95% CI –1.59 to –1.44) and –1.58 (95% CI –1.66 to –1.49).

Burden by Geographic Region

Among the 21 geographic regions, the largest absolute numbers of cervical cancer–related deaths and DALYs attributable to smoking in 2019 were observed in East Asia, and the smallest absolute numbers were observed in Oceania. In the same year, the highest ASMR was 2.84 per 100,000 (95% CI 1.33-5.01), and the highest ASDR was 95.03 per 100,000 (95% CI 52.82-142.33), observed in Oceania and southern Latin America, respectively, whereas the lowest ASMR was 0.23 per 100,000 (95% CI 0.09-0.43), and the lowest ASDR was 6.71 per 100,000 (95% CI 3.05-11.69), observed in North Africa and the Middle East, respectively.

Downward trends in the cervical cancer burden attributable to smoking were observed in most geographic regions from 1990 to 2019. The largest decrease was observed in Central Latin America, with an EAPC of –4.52 (95% CI –4.73 to –4.30) for the ASMR, and in tropical Latin America, with an EAPC of –4.38 (95% CI –4.58 to –4.17) for the ASDR. In contrast, an upward trend was observed in Eastern Europe, with EAPCs of 0.76 (95% CI 0.45-1.07) and 1.2 (95% CI 0.85-1.55) for the ASMR and ASDR, respectively.

Burden by Country

In 2019, the highest ASMR of cervical cancer attributable to smoking across all 204 included countries and territories was 483.2 times the lowest ASMR; the highest value of 24.16 per 100,000 (95% CI 13.27-38.83) was observed in Kiribati, and the lowest value of 0.05 per 100,000 (95% CI 0.02-0.11) was observed in Egypt. Kiribati and Egypt also had the highest and lowest ASDRs, at 734.33 per 100,000 (95% CI 385.97-1,184.35) and 1.17 per 100,000 (95% CI 0.39-2.57), respectively. Figures 1 and 2 show maps depicting the burden and mortality of cervical cancer attributable to smoking in 2019 in the 204 included countries, with darker and lighter colors indicating more or less severe conditions, respectively. Briefly, high burdens of cervical cancer attributable to smoking were observed in countries in South Asia, southern Africa, southern South America, and Greenland in North America, whereas low burdens

were observed in countries in Europe, North America, and Oceania.

A downward trend in the ASMR of cervical cancer attributable to smoking was observed in most countries over time (ie, the EAPC was less than zero and the upper limit of the 95% CI was less than zero), with the strongest trends observed in Mexico, Thailand, Singapore, the Maldives, and Denmark. In contrast, upward trends in the ASMR (ie, the EAPC was greater than zero and the lower limit of the 95% CI was greater than zero) were observed in 10 countries over time (listed in descending order according to the EAPC value): Lesotho, the Russian Federation, Bulgaria, Afghanistan, Albania, Uzbekistan, Italy, Kyrgyzstan, Bosnia and Herzegovina, and Guinea-Bissau. In addition, 16 countries showed stable ASMR over time (the 95% CI included zero). Similarly, downward trends in the ASDR were observed in most countries over time. However, upward trends were observed in Lesotho, the Russian Federation, Bulgaria, Afghanistan, Albania, Uzbekistan, Italy, Kyrgyzstan, Bosnia and Herzegovina, and the Democratic People's Republic of Korea, while ASDR was stable in 14 countries. Figures 3 and 4 depict the degree of change in the burden of cervical cancer attributable to smoking from 1990 to 2019. In the figures, darker colors indicate a stronger downward trend and uncolored areas on the map indicate stability.

Figure 1. The global distribution of ASMRs of cervical cancer attributable to smoking in 2019. ASMR: age-standardized mortality rate.

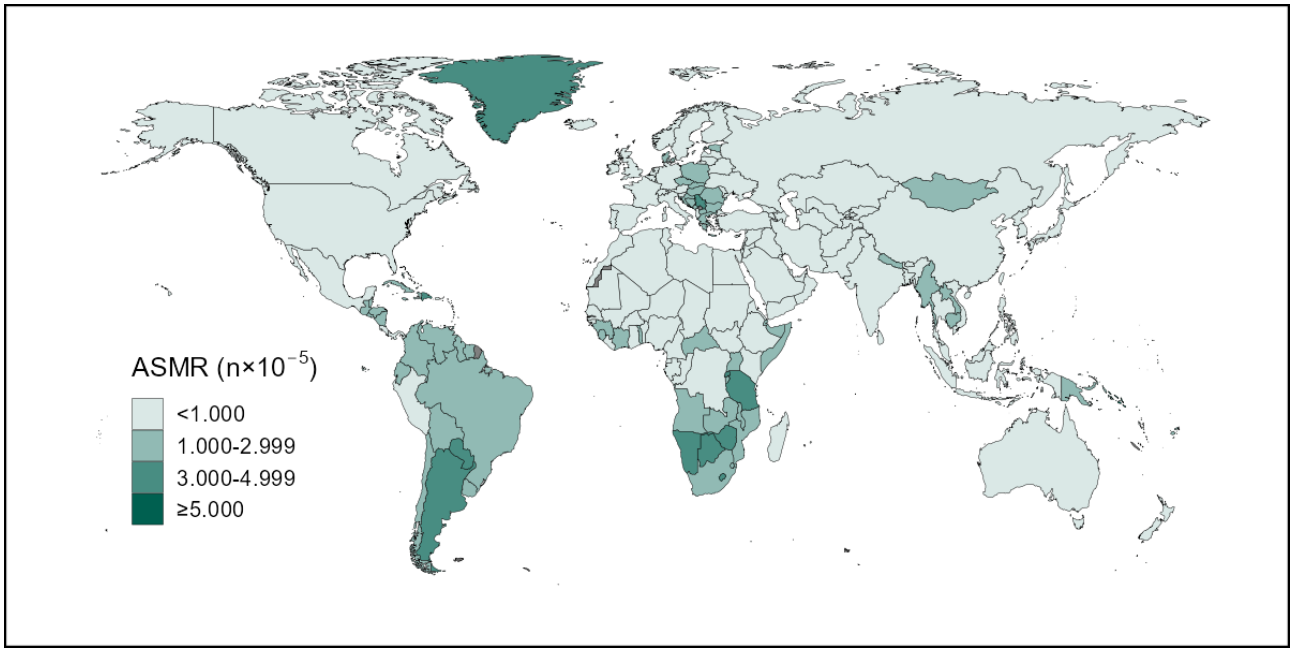


Figure 2. The global distribution of ASDRs of cervical cancer attributable to smoking in 2019. ASDR: age-standardized disability-adjusted life-year rate.

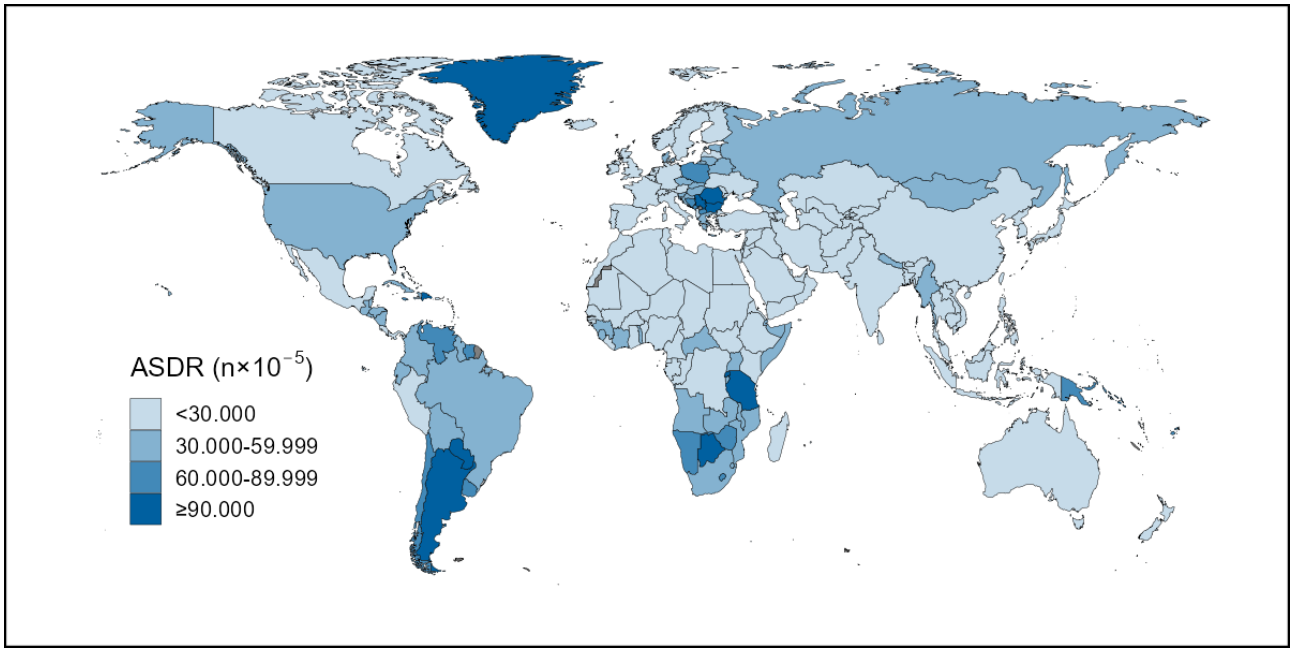


Figure 3. EAPCs in the ASMR of cervical cancer attributable to smoking by country from 1990 to 2019. EAPC: estimated annual percentage change; ASMR: age-standardized mortality rate.

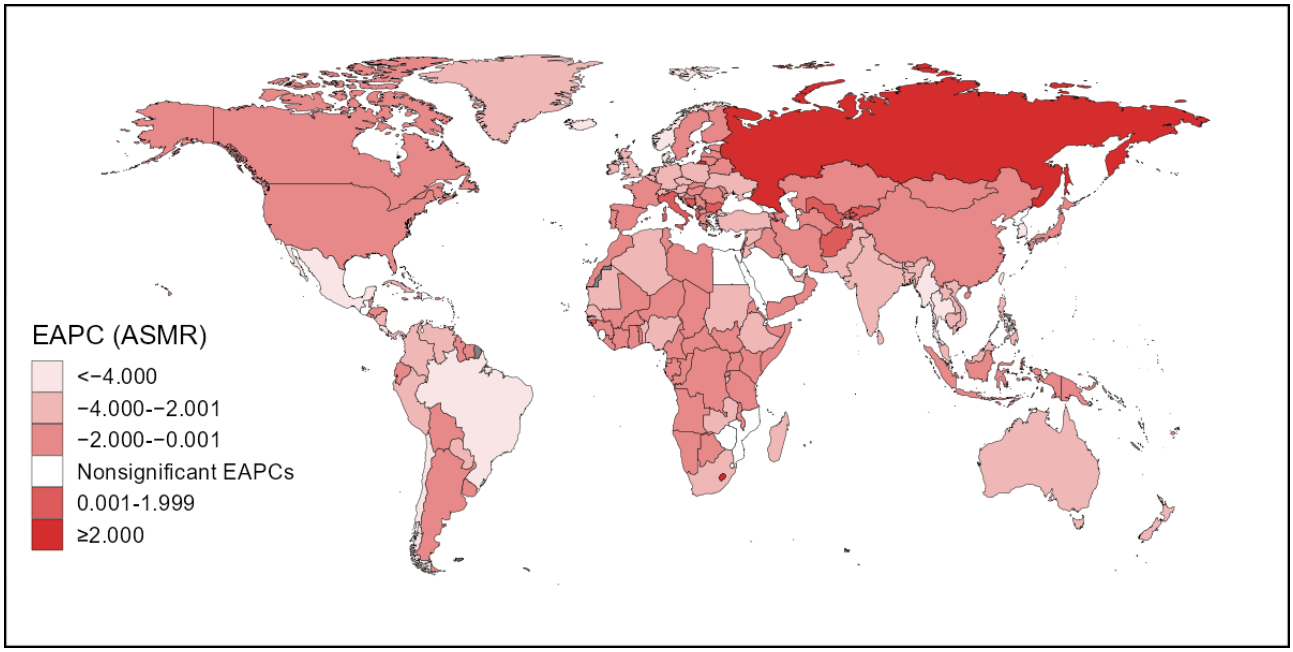
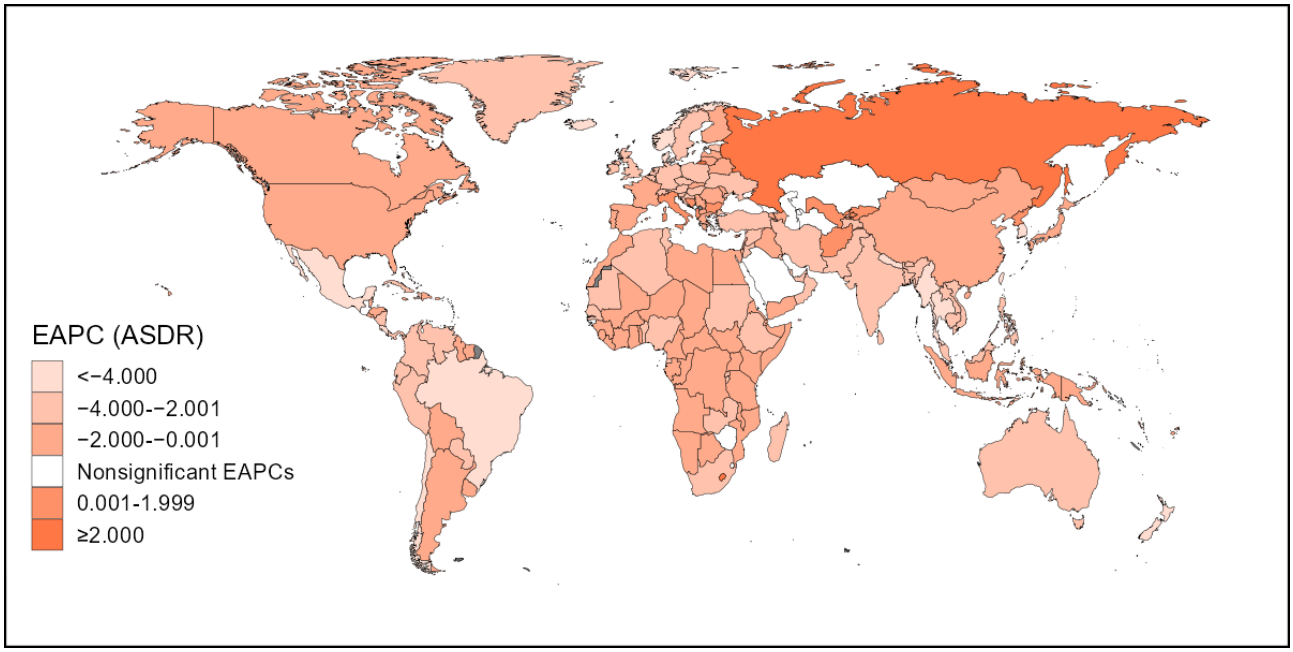


Figure 4. EAPCs in the ASDR of cervical cancer attributable to smoking by country from 1990 to 2019. EAPC: estimated annual percentage change; ASDR: age-standardized disability-adjusted life-year rate.



Characteristics of and Temporal Changes in the Cervical Cancer Burden Attributable to Smoking by Age

In 2019, an upward trend was observed in the mortality of cervical cancer attributable to smoking with age. The highest mortality rate was observed in the population older than 80 years, except for a low value in the group aged 75 to 79 years (Multimedia Appendix 2, Figure S1). Analysis of the trend in DALYs with age revealed an initially increasing and then decreasing trend, with the highest DALYs observed in the group aged 55 to 59 years (Multimedia Appendix 3, Figure S2).

Figures S3 and S4 in Multimedia Appendices 4 and 5 present the EAPCs of the ASRs of cervical cancer attributable to smoking for all age groups over the 1990 to 2019 period. Decreasing trends were observed in the ASRs of mortality and DALYs. The most obviously decreasing trend was observed in the group aged 35 to 39 years, with EAPCs of –3.19 (95% CI –3.36 to –3.03) and –3.18 (95% CI –3.35 to –3.01) for ASMR and ASDR, respectively.

Correlations Between SDI and EAPCs of the ASRs

Figures S5 and S6 in Multimedia Appendices 6 and 7 present scatterplots of correlations between SDI and EAPCs of the ASDR and ASMR in 2019. A Pearson correlation analysis

revealed a significant negative correlation between SDI and the ASMR and ASDR of cervical cancer attributable to smoking (ρ=–0.228, *P*<.001 for ASMR and ρ=–0.223, *P*<.001 for ASDR).

Discussion

Principal Findings

Our study estimated time trends in the burden of cervical cancer attributable to smoking from 1990 to 2019 at the global, regional, and national levels. In terms of spatial distribution, the worldwide distribution of this burden in 2019 was quite different in each country and region. In terms of time trends, 10 countries showed increasing trends from 1990 to 2019 and 14 countries were stable. These countries were distributed in Europe, Central Asia, and South Africa; other countries and regions showed decreasing trends.

The burden of cervical cancer related to smoking in the whole population showed an increase in the absolute number of cases and a decrease over time for the ASMR and ASDR, as well as differences in different countries and regions. This may be related, first, to the distribution characteristics of female smokers and female smoking rates around the world. Global smoking prevalence among women aged 15 years or older declined by 37.7% in 2019 compared with 1990, but population growth led to an increase in the total number of female smokers. There were 146 million current female smokers aged 30 years or older in 2019. At the same time, the prevalence of smoking among women aged 15 years or older varied greatly in different countries and regions, with the age-standardized prevalence of smoking tobacco use ranging from 0.696% in Eritrea to 42.3% in Greenland [13]. Differences in the prevalence of smoking among women around the world inevitably led to differences in the burden of disease attributable to smoking, including cervical cancer.

Second, the implementation of tobacco control measures may also have affected the spatial distribution characteristics and temporal trends of the burden of cervical cancer attributable to smoking. In 2005, the World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC) entered into force [21], and based on a consensus on the importance of tobacco control, many countries and regions around the world ratified the treaty [22]. In 2017, Gravely et al [23] used WHO data from 126 countries to examine the association between the highest levels of implementation of key demand-reduction measures of the WHO FCTC from 2007 to 2014 and smoking prevalence from 2005 to 2015. Gravely concluded that the decline in smoking prevalence was closely related to the implementation of key tobacco-control measures over the treaty's first decade and suggested that tobacco-related morbidity and mortality will continue to decline in the future. However, it cannot be ignored that there are huge gaps in the implementation of tobacco control in different countries and regions [24]. Moreover, there is a gender blind spot in tobacco control action, and the problems and challenges of female smoking are not fully recognized, especially in low- and middle-income countries [25].

Tobacco control is one of the most important cancer prevention behaviors. However, for smokers, a focus on cancer screening has the greatest potential benefit because it allows the timely detection of disease and early intervention [26]. Although smokers are at higher risk for some types of cancer, Byrne et al [27] showed that smokers are less accepting of cancer-related screening than nonsmokers. Therefore, in order to continue to reduce the burden of cervical cancer attributable to smoking, it is necessary, first, to continue to implement tobacco control measures while paying full attention to gender, so as to reduce women's tobacco use and improve women's status, and second, to further improve cervical cancer screening methods, so as to increase the screening rate for female smokers.

According to our study, the distribution of and trends in the burden of cervical cancer attributable to smoking also differed in different age groups. Mortality due to cervical cancer related to smoking increases with age, which may be related to longer exposure to smoking [28]. DALYs due to cervical cancer attributable to smoking were highest in the group aged 55 to 64 years, which had the largest absolute number of patients among age groups. Compared with patients in older age groups, the decreasing trend in the burden of cervical cancer attributable to smoking has been more pronounced in younger age groups over the last 3 decades. This may be a benefit of the promotion of cervical cancer vaccines, the continuous improvement of screening measures, and the implementation of tobacco control measures.

The results of this study show that the level of socioeconomic development is correlated with the burden of cervical cancer attributable to smoking and trends in this burden. First, countries or regions with different economic development levels have different levels of burden of cervical cancer attributable to smoking. In 2019, this burden was mainly concentrated in low- and middle-SDI countries in southern Africa, South America, and Asia. This is consistent with the global distribution of the overall burden of cervical cancer; 86% of the global burden of cervical cancer is in Africa, Latin America and the Caribbean, and Asia [29]. Therefore, it is necessary to pay special attention to the fact that although the female smoking rate is relatively low in low- and middle-income countries and regions, the burden of cervical cancer cannot be ignored. Secondly, SDI has a negative correlation with trends in this burden in the past 3 decades, which may be related to the more obvious effect of some protective factors in countries or regions with higher economic development levels. These protective factors include more effective tobacco control [30], better screening methods for cervical cancer in women, more widespread screening coverage, and more secure medical resources [31].

Study Limitations

This study has some limitations. First, the availability of certain data on the disease burden of cervical cancer attributable to smoking in low-income countries is poor, while the GBD study relied heavily on existing epidemiological studies to estimate global disease prevalence. However, all national data calculation standards were consistent, so data quality was guaranteed. Second, the GBD study did not present data on the disease burden of cervical cancer from second-hand smoke (ie,

household or occupational exposure), which is an important mode of tobacco exposure in women. If this risk factor were taken into account, the disease burden of cervical cancer due to tobacco would be higher than the current estimate, and the trend might also be different. Third, we only estimated the burden of cervical cancer attributable to smoking, but the combined effects of smoking and other risk factors may increase or complicate the burden of cervical cancer.

Conclusion

The distribution of the disease burden of cervical cancer attributable to smoking and trends in this burden differ in different countries and regions of the world. More attention should be paid to cervical cancer prevention and screening in women who smoke, especially in low- and middle-income countries.

Acknowledgments

This work was supported by grants from the Medical Science and Technology Research Plan Joint Construction Project of Henan Province (LHGJ20190223).

Data Availability

The datasets generated and analyzed during this study are available in the Global Burden of Disease (GBD) study 2019 data repository (https://ghdx.healthdata.org/gbd-results-tool).

Authors' Contributions

All authors contributed to the study conception and design. Data curation was performed by FR and KL. RY and YX rechecked the data. Statistical analysis and data visualization were performed by RY and ZT. The original draft of the manuscript was written by RY. Review and editing of the manuscript were performed by FR and ZT.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The global burden and mortality of cervical cancer attributable to smoking in 1990 and 2019 and temporal trends from 1990 to 2019 (Geographic Regions).
[DOCX File , 28 KB - publichealth_v8i12e40657_app1.docx]

Multimedia Appendix 2
Age-specific rates of global cervical cancer deaths attributable to smoking in 2019.
[PNG File , 147 KB - publichealth_v8i12e40657_app2.png]

Multimedia Appendix 3
Age-specific rates of global cervical cancer DALYs attributable to smoking in 2019. DALY: disability-adjusted life-year.
[PNG File , 158 KB - publichealth_v8i12e40657_app3.png]

Multimedia Appendix 4
EAPCs in ASMR of cervical cancer attributable to smoking by age group from 1990 to 2019. EAPC: estimated annual percentage change; ASMR: age-standardized mortality rate.
[PNG File , 180 KB - publichealth_v8i12e40657_app4.png]

Multimedia Appendix 5
EAPCs in ASDR of cervical cancer attributable to smoking by age group from 1990 to 2019. EAPC: estimated annual percentage change; ASDR: age-standardized disability-adjusted life-year rate.
[PNG File , 178 KB - publichealth_v8i12e40657_app5.png]

Multimedia Appendix 6
The correlations between the SDI index and EAPCs of the ASMR. ASMR: age-standardized mortality rate; EAPC: estimated annual percentage change.
[PNG File , 308 KB - publichealth_v8i12e40657_app6.png]

Multimedia Appendix 7

The correlations between the SDI index and EAPCs of the ASDR. ASDR: age-standardized disability-adjusted life-year rate; EAPC: estimated annual percentage change.
[PNG File , 309 KB - publichealth_v8i12e40657_app7.png]

References

1. Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet 2019 Jan 12;393(10167):169-182. [doi: 10.1016/S0140-6736(18)32470-X] [Medline: 30638582]
2. 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 May;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
3. Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. Lancet Glob Health 2020 Feb;8(2):e191-e203 [FREE Full text] [doi: 10.1016/S2214-109X(19)30482-6] [Medline: 31812369]
4. Crosbie EJ, Einstein MH, Franceschi S, Kitchener HC. Human papillomavirus and cervical cancer. Lancet 2013 Sep 07;382(9895):889-899. [doi: 10.1016/S0140-6736(13)60022-7] [Medline: 23618600]
5. de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, Retrospective International Survey and HPV Time Trends Study Group. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol 2010 Nov;11(11):1048-1056. [doi: 10.1016/S1470-2045(10)70230-8] [Medline: 20952254]
6. Safaeian F, Ghaemimood S, El-Khatib Z, Enayati S, Mirkazemi R, Reeder B. Burden of cervical cancer in the eastern Mediterranean region during the years 2000 and 2017: retrospective data analysis of the global burden of disease study. JMIR Public Health Surveill 2021 May 12;7(5):e22160 [FREE Full text] [doi: 10.2196/22160] [Medline: 33978592]
7. Fontham ETH, Wolf AMD, Church TR, Etzioni R, Flowers CR, Herzig A, et al. Cervical cancer screening for individuals at average risk: 2020 guideline update from the American Cancer Society. CA Cancer J Clin 2020 Sep;70(5):321-346 [FREE Full text] [doi: 10.3322/caac.21628] [Medline: 32729638]
8. Kashyap N, Krishnan N, Kaur S, Ghai S. Risk factors of cervical cancer: a case-control study. Asia Pac J Oncol Nurs 2019 Jul;6(3):308-314 [FREE Full text] [doi: 10.4103/apjon.apjon_73_18]
9. Zhao M, Gu R, Ding S, Luo L, Jia Y, Gao C, et al. Risk factors of cervical cancer among ethnic minorities in Yunnan Province, China: a case-control study. Eur J Cancer Prev 2022 May 01;31(3):287-292 [FREE Full text] [doi: 10.1097/CEJ.0000000000000704] [Medline: 34267110]
10. Nagelhout G, Ebisch RM, Van Der Hel O, Meerkerk G, Magnée T, De Bruijn T, et al. Is smoking an independent risk factor for developing cervical intra-epithelial neoplasia and cervical cancer? A systematic review and meta-analysis. Expert Rev Anticancer Ther 2021 Jul;21(7):781-794. [doi: 10.1080/14737140.2021.1888719] [Medline: 33663309]
11. Kaderli R, Schnüriger B, Brügger LE. The impact of smoking on HPV infection and the development of anogenital warts. Int J Colorectal Dis 2014 Aug;29(8):899-908. [doi: 10.1007/s00384-014-1922-y] [Medline: 24935346]
12. Coker AL, DeSimone CP, Eggleston KS, Hopenhayn C, Nee J, Tucker T. Smoking and survival among Kentucky women diagnosed with invasive cervical cancer: 1995-2005. Gynecol Oncol 2009 Feb;112(2):365-369 [FREE Full text] [doi: 10.1016/j.ygyno.2008.10.013] [Medline: 19036421]
13. GBD 2019 Tobacco Collaborators. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet 2021 Jun 19;397(10292):2337-2360 [FREE Full text] [doi: 10.1016/S0140-6736(21)01169-7] [Medline: 34051883]
14. Soto M, Martinez-Gutierrez J, Momany M, Capurro D, Ciampi Spode F, Cea E, et al. Preferences of underserved Chilean women on a mobile technology intervention for cervical cancer screening: qualitative study. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e196 [FREE Full text] [doi: 10.2196/mhealth.9494] [Medline: 30459141]
15. Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. Lancet 1997 May 17;349(9063):1436-1442. [doi: 10.1016/S0140-6736(96)07495-8] [Medline: 9164317]
16. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
17. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1923-1994 [FREE Full text] [doi: 10.1016/S0140-6736(18)32225-6] [Medline: 30496105]
18. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015 Jan 10;385(9963):117-171 [FREE Full text] [doi: 10.1016/S0140-6736(14)61682-2] [Medline: 25530442]
19. Hankey B, Ries L, Kosary C, Feuer E, Merrill R, Clegg L, et al. Partitioning linear trends in age-adjusted rates. Cancer Causes Control 2000 Jan;11(1):31-35. [doi: 10.1023/a:1008953201688] [Medline: 10680727]

20. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1223-1249 [FREE Full text] [doi: 10.1016/S0140-6736(20)30752-2] [Medline: 33069327]
21. Roemer R, Taylor A, Lariviere J. Origins of the WHO Framework Convention on Tobacco Control. Am J Public Health 2005 Jun;95(6):936-938. [doi: 10.2105/AJPH.2003.025908] [Medline: 15914812]
22. Chung-Hall J, Craig L, Gravely S, Sansone N, Fong GT. Impact of the WHO FCTC over the first decade: a global evidence review prepared for the Impact Assessment Expert Group. Tob Control 2019 Jun 07;28(Suppl 2):s119-s128 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054389] [Medline: 29880598]
23. Gravely S, Giovino GA, Craig L, Commar A, D'Espaignet ET, Schotte K, et al. Implementation of key demand-reduction measures of the WHO Framework Convention on Tobacco Control and change in smoking prevalence in 126 countries: an association study. Lancet Public Health 2017 Apr;2(4):e166-e174 [FREE Full text] [doi: 10.1016/S2468-2667(17)30045-2] [Medline: 29253448]
24. Puska P, Daube M, WHO FCTC Impact Assessment Expert Group. Impact assessment of the WHO Framework Convention on Tobacco Control: introduction, general findings and discussion. Tob Control 2019 Jun;28(Suppl 2):s81-s83 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054429] [Medline: 30181384]
25. Amos A, Greaves L, Nichter M, Bloch M. Women and tobacco: a call for including gender in tobacco control research, policy and practice. Tob Control 2012 Mar;21(2):236-243. [doi: 10.1136/tobaccocontrol-2011-050280] [Medline: 22166266]
26. Zoorob R, Mejia M, Matas J, Wang H, Salemi J, Levine R. Smoking history and adherence to cancer-related recommendations in a primary care setting. Transl Behav Med 2022 Feb 16;12(2):362-367. [doi: 10.1093/tbm/ibab156] [Medline: 34918162]
27. Byrne MM, Davila EP, Zhao W, Parker D, Hooper MW, Caban-Martinez A, et al. Cancer screening behaviors among smokers and non-smokers. Cancer Epidemiol 2010 Oct;34(5):611-617. [doi: 10.1016/j.canep.2010.06.017] [Medline: 20655820]
28. Sugawara Y, Tsuji I, Mizoue T, Inoue M, Sawada N, Matsuo K, Research Group for the Development and Evaluation of Cancer Prevention Strategies in Japan. Cigarette smoking and cervical cancer risk: an evaluation based on a systematic review and meta-analysis among Japanese women. Jpn J Clin Oncol 2019 Jan 01;49(1):77-86. [doi: 10.1093/jjco/hyy158] [Medline: 30407555]
29. Vaccarella S, Laversanne M, Ferlay J, Bray F. Cervical cancer in Africa, Latin America and the Caribbean and Asia: Regional inequalities and changing trends. Int J Cancer 2017 Nov 15;141(10):1997-2001 [FREE Full text] [doi: 10.1002/ijc.30901] [Medline: 28734013]
30. Smith CE, Hill SE, Amos A. Impact of population tobacco control interventions on socioeconomic inequalities in smoking: a systematic review and appraisal of future research directions. Tob Control 2020 Sep 29;30(e2):e87-e95 [FREE Full text] [doi: 10.1136/tobaccocontrol-2020-055874] [Medline: 32994297]
31. Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to the Human Development Index (2008-2030): a population-based study. Lancet Oncol 2012 Aug;13(8):790-801. [doi: 10.1016/S1470-2045(12)70211-5] [Medline: 22658655]

Abbreviations

ASDR: age-standardized disability-adjusted life-year rate
ASMR: age-standardized mortality rate
ASR: age-standardized rate
DALY: disability-adjusted life-year
EAPC: estimated annual percentage change
FCTC: Framework Convention on Tobacco Control
GBD: Global Burden of Disease
SDI: sociodemographic index
UI: uncertainty interval
WHO: World Health Organization

Edited by G Eysenbach; submitted 01.07.22; peer-reviewed by J Kim, Y Zhou; comments to author 21.10.22; revised version received 31.10.22; accepted 30.11.22; published 23.12.22.

Please cite as:
Yuan R, Ren F, Xie Y, Li K, Tong Z
The Global, Regional, and National Burdens of Cervical Cancer Attributable to Smoking From 1990 to 2019: Population-Based Study
JMIR Public Health Surveill 2022;8(12):e40657
URL: https://publichealth.jmir.org/2022/12/e40657
doi:10.2196/40657
PMID:36563035

Original Paper

Drug Abuse Ontology to Harness Web-Based Data for Substance Use Epidemiology Research: Ontology Development Study

Usha Lokala[1], MSci; Francois Lamy[2], PhD; Raminta Daniulaityte[3], PhD; Manas Gaur[1], PhD; Amelie Gyrard[4], PhD; Krishnaprasad Thirunarayan[5], PhD; Ugur Kursuncu[1], PhD; Amit Sheth[1], PhD

[1]AI Institute, University of South Carolina, Columbia, SC, United States

[2]Department of Society and Health, Mahidol University, Salaya, Thailand

[3]College of Health Solutions, Arizona State Univeristy, Phoenix, AZ, United States

[4]Department of IoT and AI, Trialog Information Technology & Services, Ile-de-France, France

[5]Department of Computer Science & Engineering, Wright State University, Dayton, OH, United States

Corresponding Author:
Usha Lokala, MSci
AI Institute, University of South Carolina
1112 Greene St
Columbia, SC, 29208
United States
Phone: 1 803 777 9707
Email: nlokala@email.sc.edu

Abstract

Background: Web-based resources and social media platforms play an increasingly important role in health-related knowledge and experience sharing. There is a growing interest in the use of these novel data sources for epidemiological surveillance of substance use behaviors and trends.

Objective: The key aims were to describe the development and application of the drug abuse ontology (DAO) as a framework for analyzing web-based and social media data to inform public health and substance use research in the following areas: determining user knowledge, attitudes, and behaviors related to nonmedical use of buprenorphine and illicitly manufactured opioids through the analysis of web forum data Prescription Drug Abuse Online Surveillance; analyzing patterns and trends of cannabis product use in the context of evolving cannabis legalization policies in the United States through analysis of Twitter and web forum data (eDrugTrends); assessing trends in the availability of novel synthetic opioids through the analysis of cryptomarket data (eDarkTrends); and analyzing COVID-19 pandemic trends in social media data related to 13 states in the United States as per Mental Health America reports.

Methods: The domain and scope of the DAO were defined using competency questions from popular ontology methodology (101 ontology development). The 101 method includes determining the domain and scope of ontology, reusing existing knowledge, enumerating important terms in ontology, defining the classes, their properties and creating instances of the classes. The quality of the ontology was evaluated using a set of tools and best practices recognized by the semantic web community and the artificial intelligence community that engage in natural language processing.

Results: The current version of the DAO comprises 315 classes, 31 relationships, and 814 instances among the classes. The ontology is flexible and can easily accommodate new concepts. The integration of the ontology with machine learning algorithms dramatically decreased the false alarm rate by adding external knowledge to the machine learning process. The ontology is recurrently updated to capture evolving concepts in different contexts and applied to analyze data related to social media and dark web marketplaces.

Conclusions: The DAO provides a powerful framework and a useful resource that can be expanded and adapted to a wide range of substance use and mental health domains to help advance big data analytics of web-based data for substance use epidemiology research.

(JMIR Public Health Surveill 2022;8(12):e24938) doi:10.2196/24938

KEYWORDS

ontology; knowledge graph; semantic web; illicit drugs; cryptomarket; social media

Introduction

Background

Illicit drug use is a complex social phenomenon generating a variety of public health issues that affect individuals and their communities. In its 2020 report, the United Nations Office on Drugs and Crime estimated that 5.4% of the world population used illicit drugs in 2018 while 0.7% of the whole population is affected by substance use disorder [1]. Individuals affected by substance use disorder are at risk of experiencing a variety of adverse psychiatric and physical health effects such as unintentional overdoses or disease infections (eg, HIV and hepatitis C). Individual drug use also potentially impacts the well-being of others, affecting local communities and neighborhoods [2], which in turn creates the contextual conditions and social determinants linked to individual drug use initiation [3]. Although cannabis remains by far the most consumed illicit drug with more potent forms potentially linked to adverse consequences [4], opioid and amphetamine-type drugs remain more frequently associated with psychiatric and physical harms [5].

Although illicit substance use represents an endemic phenomenon affecting modern societies, recent years have seen radical and rapid changes in terms of the variety of substances available, the growing role played by the internet, and the decriminalization or legalization of several illicit substances in an increasing number of countries. For example, the European Monitoring Centre for Drugs and Drug Addiction has identified and listed approximately 400 novel psychoactive substances since 2015 [6], while cryptomarkets located on the dark net have become increasingly important platforms for the distribution of novel psychoactive substances and other illicit or prescription drugs [7,8]. These changes call for more timely methods of data collection, allowing the monitoring of both demand and supply sides. In this ever-changing environment, user-generated content on illicit drug use shared on social media represents a rich source of unsolicited and unfiltered self-disclosures of attitudes and practices related to substance use [9]. Furthermore, web-based sources of distribution can be harnessed to provide updates on the illicit drug supply trade and new trends [10].

These unfiltered web-based communications and advertisements offer a rich source of data sensitive to changing and emerging drug use trends, and can be used to complement and enhance existing epidemiological surveillance systems.

Semantic web-based approaches play a key role in enhancing and improving big data analytics for such complex domains as substance use. The semantic web is an extension of the web in which a set of design principles and technologies have been created to capture the meaning of information [11]. An ontology is defined as a specification of shared concepts and relationships among them, consisting of a schema and a knowledge base of instances [12].

Ontologies also play key roles in the development of (1) semantic web applications, (2) semantic annotation of data, and (3) tools for querying and reasoning [13]. However, to apply semantic web tools effectively, there is a need for a domain-specific ontology to represent the main entities of value described in the social media posts and their relationships [14].

There has been a broad range of research developing ontologies for social media data. For instance, the work proposed by Kim et al [15] aimed to develop an ontology dedicated to obesity for investigating obesity-related social media posts and detecting sentiments, emotions, and opinions posted on specific social media. Their ontology was evaluated by mapping concepts from ontology with similar terms found in tweets related to obesity, and is only limited to 8 superclasses related to broader perspectives of any biomedical ontology. This study is limited to social media posts for improving upon the ontology, and the keywords are vastly distributed among the top 2 obesity types (abdomen and thigh) and top 3 management types (diet, exercise, and drug therapy) and are only limited to the general population in social media.

There are fewer ontologies related to the domain of mental health. For example, Jung et al. [16] proposed to design an ontology using an entity-attribute-value triplet data model dedicated to adolescent depression in order to analyze related social media. This ontology was developed using clinical guidelines and unstructured social media posts with 777 terms divided into *risk factors*, *signs and symptoms*, *screening*, *diagnosis*, *treatment*, *and prevention*. This work is mainly limited to the extraction of data solely from adolescent depression-related social media posts.

Several prior ontologies were developed for the analysis of the prescription drug domain. For example, the prescription drugs ontology [17] aims at improving the semantics of drug prescriptions and prospectively enabling the interoperability of prescription data by reusing classes and object properties from the information artifact ontology [18], the ontology for biomedical investigations [19], the ontology for general medical science [20], the ontology for medically related social entities [21], and the drug ontology [22]. However, these ontologies focus on medical uses of prescribed drugs and do not include concepts or slang terms related to the use of illicit drugs and addiction.

As the opioid crisis has deepened in recent years, efforts to analyze the opioid research on social media and make policy decisions have intensified. In a recent study, a specific knowledge graph called Opioid Drug Knowledge Graph (ODKG) [23] was developed to capture opioid-related drugs and related entities in eHealth records. As the drug abuse ontology (DAO) also contains information about opioid-related drugs, we compared the ODKG and DAO in terms of their coverage of relevant entities in opioid-related social media corpus (Twitter) and observed that the DAO outperformed the ODKG by order of magnitude. As the DAO was designed to also cover slang terms that are common in social media, it performed well by retrieving 7 million more tweets than the ODKG (2 million) from a resource of 1.2 billion crawled tweets during the COVID-19 pandemic [24].

The key aims of this paper were to describe the process of development, evaluation, and application of the DAO to facilitate and enhance social media and web-based analytics for

substance use epidemiology research. This paper describes the process of DAO development in the context of 4 research projects out of which 3 are National Institutes of Health (NIH)–sponsored studies that aimed to harness web-based and social media data for substance use epidemiology research: (1) Prescription Drug Abuse Online Surveillance (PREDOSE) project that aimed to characterize user knowledge, attitudes, and behaviors related to nonmedical use of buprenorphine and other illicitly manufactured opioids through the analysis of web forum data [25-27]; (2) eDrugTrends project that focused on patterns and trends of cannabis product use in the context of evolving cannabis legalization policies in the United States through the analysis of Twitter and web forum data [28-32]; (3) eDarkTrends project that aimed to identify availability trends of novel synthetic opioids through the analysis of crypto market data [33-35]; and (4) COVID-19 pandemic trends in social media data related to 13 states in the United States and its mental health impact.

The terminology related to machine learning (ML), natural language processing (NLP), and ontology design used in this paper is organized alphabetically in Textbox 1.

Textbox 1. Descriptions of machine learning (ML), natural language processing (NLP), and ontology terms used in this paper.

Terminology and description

- 101 ontology [36]: the 101 ontology is a guideline to create an ontology and offers step by step process. It leverages the authors' experiences developing and maintaining ontologies in several ontology environments like Protégé.
- Bootstrap and bagged random Forest with contextual features (BRF-CF): Random Forest is one of the most popular ML algorithms. It is a type of ensemble ML algorithm called bootstrap or bagging.
- Class, data property, individual count: these terms are used as the signatures for the imports closure of the active ontology. In other words, the number of distinct classes, object properties, data properties, and individuals are mentioned in the ontology. The numbers here include built-in entities, such as owl: Thing if they are explicitly mentioned in the ontology.
- Community Ontology Repository [37]: this is the repository of ontologies hosted by Earth Science Information Partner's members that would let users try out semantic technologies, understand their benefits, and explore possible applications that used semantic resources.
- Depression and drug abuse BERT: BERT is a bidirectional encoder representations from transformers and is a transformer-based ML technique for NLP. We fine-tune BERT models on corpora that are representative of depression and drug abuse.
- DBpedia [38]: DBpedia is a crowd-sourced community effort to extract structured content from the information created in various Wikipedia projects.
- Diagnostic and Statistical Manual for Mental Disorders (DSM)-5: It is the taxonomic and diagnostic manual developed and published by the American Psychiatric Association. It is an authoritative guide for mental health care professionals in the diagnosis of mental disorders.
- Entity, concept: the entity is referred to as an encompassing concept for classes, individuals, and properties. Concept and class are simply synonyms.
- F1 score: It is the weighted average of precision and recall. This score takes both false positives and false negatives into account. F1 is usually more useful than accuracy score.
- False positive, true positive: a false alarm is also known as a false positive. A false positive is a result that indicates a given condition exists when it does not. For example, the model indicates that cannabis can cause pain when it does not cause pain. A true positive is an outcome where the model correctly predicts the positive class. Similarly, a true negative is an outcome where the model correctly predicts the negative class. A false positive is an outcome where the model incorrectly predicts the positive class.
- Horizontal linguistic features, vertical linguistic features, fine-grained features: while training an ML model, we organized our feature set into 3 broad groups: horizontal linguistic features, vertical linguistic features, and fine-grained features. Contextual Features (or embedding of a social media post) with Modulations (CFwM) and without Modulations (CFw/oM) are 2 additional feature set created using Word2Vec.
- Ontology metrics [39]: the metrics list the numbers for structures and representation of ontology in Protégé as it is the most widely used tool to create an ontology. Axioms associate class and properties and are a combination of logical and nonlogical attributes. The number of distinct classes, object properties, data properties, and individuals reported is focused on the evaluation of the structure of DAO.
- Oops (ontology pitfall scanner), vapor, triple checker [40]: these are Semantic Web (SemWeb) validation or documentation tools that help to improve ontologies. Oops detect common pitfalls in ontology automatically and provide recommendations to fix them.
- Owl file: the W3C web Ontology Language is a SemWeb language designed to represent rich and complex knowledge about things, groups of things, and relations between things.
- PerfectO methodology [40]: PerfectO references, classifies, and provides tools to encourage SemWeb best practices to achieve semantic interoperability by focusing on ontology improvement.
- Precision, recall: precision is the proportion of times that when you predict it is positive and it actually turns out to be positive, whereas recall is like accuracy over just the positives—it is the proportion of times you labeled positive correctly over the number of times it was actually positive.
- Protégé: protégé is a free, open-source ontology editor and framework for building intelligent systems.
- SEDO [41]: It stands for Semantic Encoding and Decoding Optimization. It is a procedure to modulate the word embedding (vectors) of a word. SEDO modulates the embeddings of each word in the Reddit content of the user based on the proximity of the word to the Diagnostic and Statistical Manual for Mental Disorders-5th edition category.
- Vanilla BERT: Vanilla BERT is a variation of the attention-based BERT model and provides a pretrained starting point layer for neural networks.
- WebVOWL [42]: It is a web application for the interactive visualization of ontologies which is one of the ontology visual representations.

Evolution of the DAO

As social media and other web resources play an increasingly important role in health-related knowledge and experience sharing [43], there is a need for an ontology explicitly dedicated to the domain of substance use research. The DAO was developed to formalize concepts, entities, and relationships relevant to the domains of addictions and mental health to harness its use on social media data. Our approach, built on the integration of semantic web technologies, enhances traditional ML and NLP techniques for automatic extraction and representation of relevant data and facilitates analysis and interpretation related to the specific goals of each study.

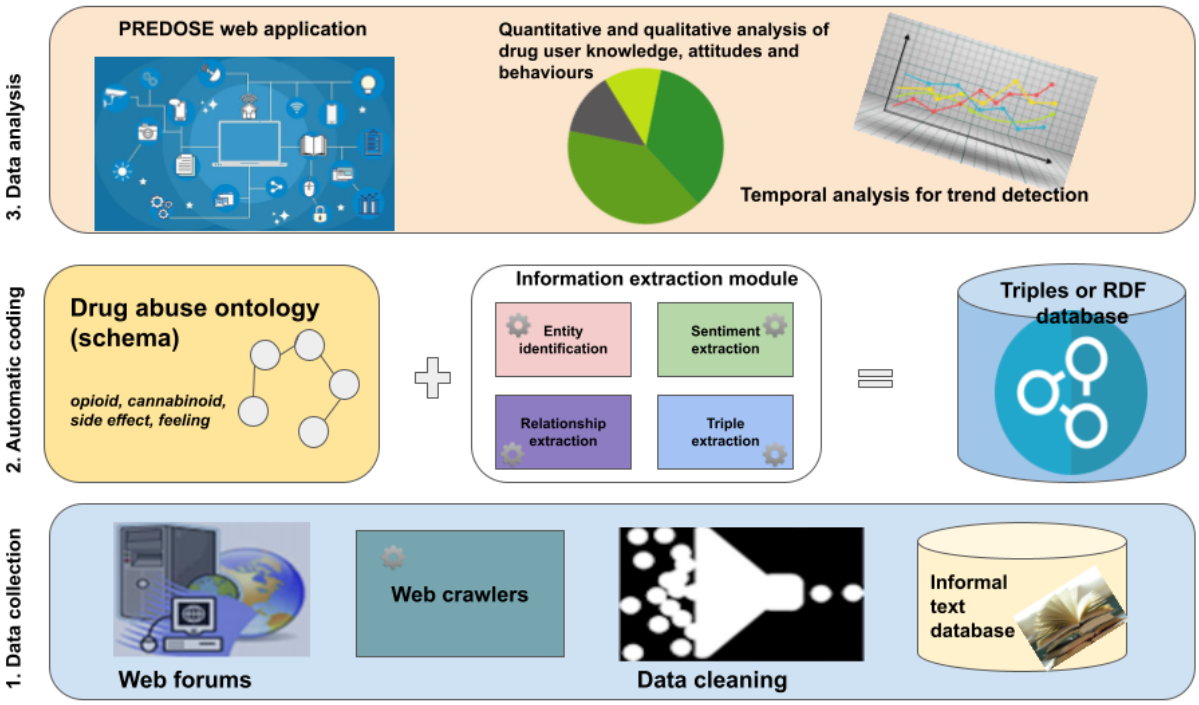
Prescription Drug Abuse Online Surveillance

This study focuses on web forum data related to the nonmedical use of buprenorphine [26,27] approved in late 2002 by the United States Food and Drug Administration for the treatment

of opioid addiction. Use of buprenorphine was defined as nonprescribed when used without medical supervision. Although there is always a level of uncertainty in disambiguating prescribed versus nonprescribed use in web-based discussions, some of the questions and practices shared by individuals provided indicators about nonprescribed use (eg, saying that Suboxone was obtained from a friend; that *bupe* was snorted; or that it was cut up and used in smaller amounts). Buprenorphine (Suboxone, Subutex, etc) is the only controlled substance that may be prescribed for the treatment of opioid addiction by a licensed physician in an office-based setting. The overall purpose of PREDOSE was to study user-generated web forum discussions about the illicit use of Suboxone (buprenorphine or naloxone), Subutex (buprenorphine), and other buprenorphine products by applying novel information processing techniques to facilitate qualitative and quantitative analysis [26]. Along with Twitter and Reddit, we also used 3 web forums that provided venues for people to freely share drug use experiences and post questions, comments, and opinions about different drugs. One of these web forums used in our research was Bluelight [44] (please note that in compliance with Institutional Review Board guidelines at Wright State University, the names of the other 2 forums have not been disclosed in this paper). Our team has developed a research collaboration with the Bluelight team and was able to obtain deidentified data updates directly from Bluelight. Data from these forums were collected using custom-built web crawlers. We chose to study buprenorphine because there was at that time (2011-2012) a growing body of evidence that buprenorphine was used and that there was relatively little knowledge about the patterns and trends of its nonmedical use in the United States. As buprenorphine use is linked to a broader domain of illicit opioid use and addiction, the initial versions of the DAO included detailed representation of the opioid class drugs, including slang and brand name terminology. The DAO developed for the PREDOSE project also included other classes of drugs, such as cannabis and stimulant-type drugs, because polysubstance use is common among illicit opioid users. Figure 1 [26] demonstrates the use of the DAO ontology within our PREDOSE architecture, which comprises three main modules:

1. Data collection module that collected approximately 1 million posts (1,066,502) from 35,974 users.
2. Automatic coding module that semantically annotated the posts using the DAO ontology.
3. Data analysis and interpretation module to visualize the keywords (eg, loperamide and buprenorphine) found within posts and referenced within the DAO ontology.

Figure 1. Use of the drug abuse ontology within Prescription Drug Abuse Online Surveillance (PREDOSE). RDF: Resource Description Framework.



eDrugTrends

This is our second project that received funding from NIH and National Institute on Drug Abuse (NIDA) in 2014 [45]. This study focused on social media data related to cannabis and synthetic cannabinoid use in the context of evolving cannabis legalization policies in the United States. The aim of this study was to develop eDrugTrends, a comprehensive software platform for semiautomated processing and visualization of thematic, sentiment, spatiotemporal, and social network dimensions of social media data (Twitter and web forums) on cannabis and synthetic cannabinoid use. The study also aimed to (1) identify and compare trends in knowledge, attitudes, and behaviors related to cannabis and synthetic cannabinoid use across United States regions with different cannabis legalization policies using Twitter and web forum data and (2) analyze social network characteristics and identify key influencers in cannabis and synthetic cannabinoid–related discussions on Twitter. For addressing these aims of the eDrugTrends platform, the DAO was expanded further to include a more comprehensive

representation of emerging cannabis products, synthetic cannabinoid products, health-related consequences, and mental health conditions.

eDarkTrends

This is the third project using the DAO. This study was funded through the NIH and NIDA time-sensitive mechanism [46], which started in 2017. The eDarkTrends project was orientated toward novel synthetic opioids, such as illicitly manufactured fentanyl that have emerged over the past few years and were and still are significant contributors to the increase in unintentional opioid-related overdose mortality in the United States [35,47,48]. However, epidemiological surveillance on cryptomarket data was limited at the time (2017). The study's overall goal was to harness cryptomarket data to conduct surveillance of illicit fentanyl, fentanyl analog, and other novel synthetic opioid availability trends over time and identify new substances as they emerge in the Darknet environment. Ultimately, eDarkTrends aimed at providing a powerful tool for epidemiological surveillance, enhancing the capacities of early warning systems to capture changes in the fentanyl and other illicit synthetic opioid supply and availability. For addressing the specific needs of the project, the DAO was further expanded to include a comprehensive and detailed representation of novel illicit synthetic opioid domains (eg, carfentanil, furanyl fentanyl, U-47700, and MT-45).

COVID-19 Pandemic

In addition, we applied the DAO on COVID-19 social media data analysis to analyze the social media data related to the pandemic. The intent is that the COVID-19 pandemic has alleviated community-wide depression and has led to increased drug use [49]. The impact of the COVID-19 pandemic on mental health was investigated in recent studies [50-52]. For this, we proposed a novel framework for assessing the spatiotemporal-thematic progression of depression, drug use, and informativeness of the underlying news content across different states in the United States [53]. The DAO is used along with the Medical Subject Headings terms hierarchy in the Unified Medical Language System, the Diagnostic and Statistical Manual for Mental Disorders-5th edition (DSM-5) lexicon [41], which are collectively referred to as the Mental Health and Drug Abuse Knowledge base (MHDA-Kb) to spot additional entities.

Methods

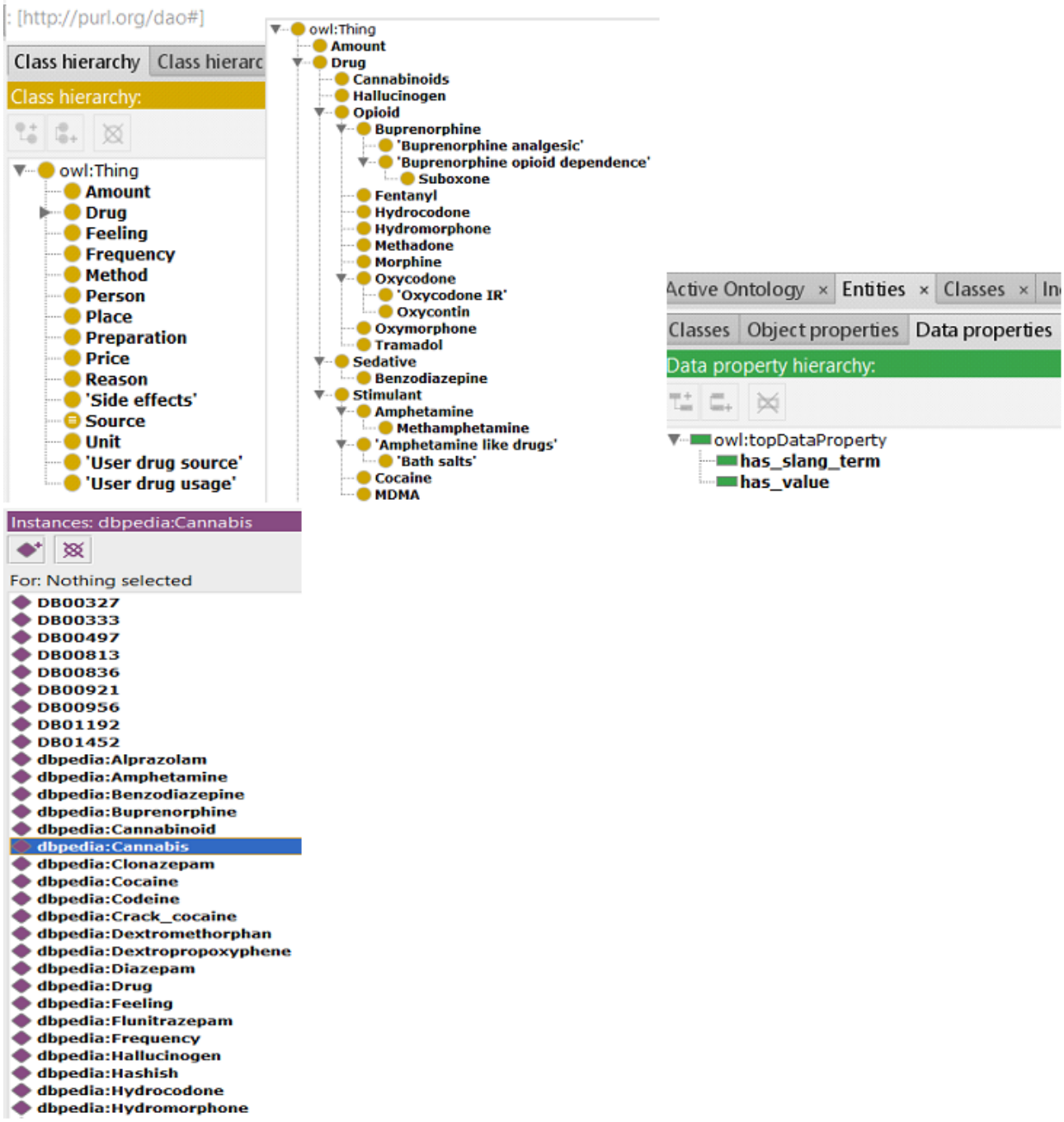
Overview

The ontology was manually developed by the domain expert coauthors (FL and RD), who used a range of sources, including (1) key epidemiological data sources and reports accessible through the NIDA [54], Drug Enforcement Agency [55], European Monitoring Centre for Drugs Addiction [56], and RxNorm [57]; (2) prior peer-reviewed publications related to illicitly manufactured opioids, cannabis, and other drugs [58-61]; and (3) ongoing manual assessment and examination of web-based social media sources related to selected substances [25,27,62]. Sources of types 1 and 2 provided primary concepts while sources of type 3 were important in identifying alternative concepts, including synonyms and street names. To develop the DAO, we followed the well-known 101 ontology development methodology [63]. The 101 method includes (1) determining the domain and scope of ontology, (2) reusing existing knowledge, (3) enumerating important terms in ontology, and (4) defining the classes and their properties and creating instances of the classes.

Design

Figure 2 provides an overview of the DAO ontology. Protégé [64], a popular ontology editor, was used to build the ontology as a tree of subclasses. The ontology was designed as a catalog of concepts related to substance use. Hence, classes of psychoactive substances (eg, cannabinoids and opioids) were created and populated with subtypes of substances (eg, morphine and fentanyl). Each substance was defined by its name and, when applicable, information regarding its pharmaceutical or brand name (*has_brand_name*), slang or street name (*has_street_name*), and chemical designation (*has_chemical_formula*) were added. This latter information was collected through different sources: pharmaceutical or brand names were based on existing medical or pharmacological dictionaries, slang or street names were based on the domain knowledge of the second and third authors (RD and FL), and chemical designations mostly concerned synthetic cannabinoid receptor agonists and were based on academic literature as well as on seizure data (eg, the National Forensic Laboratory Information System and Europol). The DAO was also enhanced with concepts and slang terms related to those concepts regarding unit (eg, *caps, ml, and bottle)*, purity, and form of preparation (*eg, crush and eyeball)* to enable the identification and analysis of triple in text content [65]. For example, one instance of the drug *Morphine* is *Poppy_Tea*, which has the slang terms *Pod* and *Poppy_Pods* used on social media.

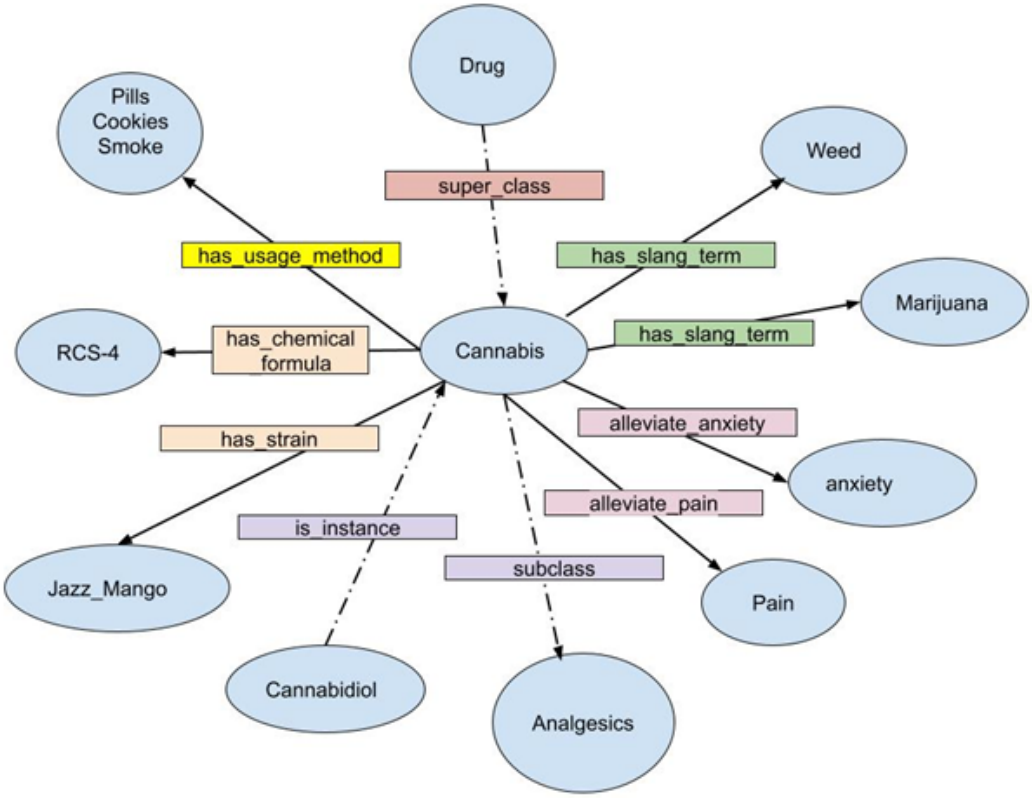
Figure 2. Drug abuse ontology in Protégé (concepts, object properties, data properties, and instances).



Instantiation

This is defined as creating instances of classes in a hierarchy. The instance of a class has its own class and fills a value. The instance has its own properties. For example, *Fentanyl* belongs to the class *Opioid* and has its own properties such as *has_brand_name*, *has_synonym*, *has_slang_term*, etc. The DAO ontology reuses instances from the DBpedia data set [66] (eg, buprenorphine). Figure 3 is the WebVOWL (web application for the interactive visualization of ontologies) representation of the DAO focused on the entity Cannabis derived from the visual data web [67]. Figure 2 shows the tree of drug names implemented as a web ontology format (owl) file within the DAO ontology. In Figure 2, entities, object properties, instances, and data properties are represented in yellow, green, and purple tags, respectively, which clearly depict the nature of classes, instances, hierarchies, and relationships for each entity.

Figure 3. Web-based visualization of OWL ontologies (WebVOWL) representation of the drug abuse ontology, focused on the cannabis concept. RCS-4: 1-pentyl-3-(4-methoxybenzoyl)indole.



Ethics Approval

This research is done in compliance with institutional review board guidelines at Wright State University. The names of the selected websites have not been disclosed in this manuscript. Our project involves analysis of Twitter data that is publicly available and that has been anonymized. It does not involve any direct interaction with any individuals or their personally identifiable data. Furthermore, our data set does not include any interaction with human participants. Our data set does not contain any images as per our data use safety agreement. Thus, this study was reviewed by the Wright State University Institutional Review Board and received an exemption determination.

Results

Evaluation

The DAO ontology was evaluated following the semantic web best practices recognized by the International Semantic Web Conference Resource Track guidelines [68], which provide the following criteria: (1) impact, (2) reusability, (3) design and technical quality, and (4) availability. We have also followed the PerfectO methodology [40], which synthesizes a set of additional best practices and eases their achievements [69]. We have discussed the results of applying the following criteria to our DAO:

1. Impact and reusability: the DAO has been exploited in 4 scenarios, as mentioned earlier. Automatic documentation can be provided using the Live OWL documentation environment [70], and the DAO documentation is available in Community ontology repository [71].
2. Design, technical quality, and availability: the design of the ontology is available on the web as a graph visualization using web-based visualization of ontologies (WebVOWL) [72,73]. We improved the ontology using Oops (Ontology Pitfall Scanner) tools that automatically detect common pitfalls and provide recommendations to fix them. Oops loaded with the DAO can be tested on the web [71,74]. The Linked data validator, Vapour tool integrated with the DAO [75] was used to check dereferencing uniform resource identifier and content negotiation. Finally, Resource description framework Triple-Checker checks whether the existing ontologies have been correctly used within our DAO [76].
3. Ontology metrics: the DAO was also evaluated, as shown in Table 1, with respect to several ontology metrics [77]. The metrics list the numbers for the structures and representation of ontology in Protégé, as it is the most widely used tool to create ontology [78]. Axioms associate class and properties and are a combination of logical and nonlogical axioms [79]. The number of distinct classes, object properties, data properties, and individuals reported in Table 1 are focused on the evaluation of the structure of the DAO.

Table 1. Drug abuse ontology metrics: the ontology metrics view displays entity and axiom count for the axioms in the active ontology [39].

Metric	Count, n	Description
Ontology metrics		
Axioms	4876	Combined logical and nonlogical axiom count
Logical axiom count	3478	The number of logical axioms
Declaration axioms count	1185	The number of declaration axioms
Class count	316	The number of distinct classes, object properties, data properties and individuals that are mentioned in the ontology
Object property count	12	The number of distinct classes, object properties, data properties and individuals that are mentioned in the ontology
Data property count	13	The number of distinct classes, object properties, data properties and individuals that are mentioned in the ontology
Individual count	845	The number of distinct classes, object properties, data properties and individuals that are mentioned in the ontology
Class axiom		
SubClassOf	313	The number of SubClassOf axioms in the ontology. A subclass axiom states that a class is a subclass of another class
Individual axioms		
Data property assertion	2317	A data property assertion states that the individual is connected by the data property expression to the literal.
ClassAssertion	830	A class assertion states that the individual is an instance of the class expression.
AnnotationAssertion	213	An annotation assertion states that the annotation subject is an anonymous individual with the annotation property and value.

The subsequent sections demonstrate the results with the DAO in different platforms and the evolution of the DAO with each use case.

The DAO Within PREDOSE

Figure 4 [26,80] describes how the texts are automatically annotated using the DAO. In the text shown in Figure 4, we identify drug entities, dosage, time interval, route of administering the drug, etc. In the DAO, buprenorphine is defined as the subclass of *Subutex* and *Suboxone*. It has the slang terms *Bupe* and *Bupey*. The term *Bupe* identified in the text would not have been possible without defining it as a slang term in the DAO. The DAO is capable of mapping units (eg, mg→MILLIGRAM) and slang terms (eg, bupe—buprenorphine) based on a lexical lookup in the ontology. Similarly, other concepts, such as the route of administration *injected*, are also identified in the text. In NLP-related tasks, such as lexical, semantic, and syntactic analysis of textual data, adding ontology works as an external source of knowledge in identifying triples and entities in data. Conceptualizing the domain in data acts as a prior requirement for processing further information (lexicon and rule-based grammar) about it [81] (Figure 5 [80]). When evaluating 601 web forum posts with the DAO, we achieved 84.9% precision and 72.5% recall in information extraction tasks. In particular, out of 3639 annotations, 2640 were predicted correct (true positives), whereas 683 slang terms are incorrect (false positives). As far as the recall is concerned, only 999 out of 3639 annotations are missed (false negatives) [26]. For triple extraction with the DAO, we achieved 33% precision across 197 evaluated triple patterns (66 were correct and 131 were incorrect). For relation extraction with the DAO, we achieved 36% precision across 183 phrases (66 were correct and 117 were incorrect). Another finding (Figure 6 [25]) is that our analysis of web forums with the DAO revealed that loperamide was widely used as a treatment for withdrawal symptoms related to opioid addiction, where buprenorphine and methadone are commonly prescribed. A total of 3 toxicology studies following this work led to a Food and Drug Administration warning in 2016 [25,82]. A video demo [83] on the PREDOSE platform is available on the web. The PREDOSE platform indicates a need for additional enhancements in information extraction and automated data coding techniques.

Figure 4. Automatic annotation of texts with the drug abuse ontology (DAO) [80].

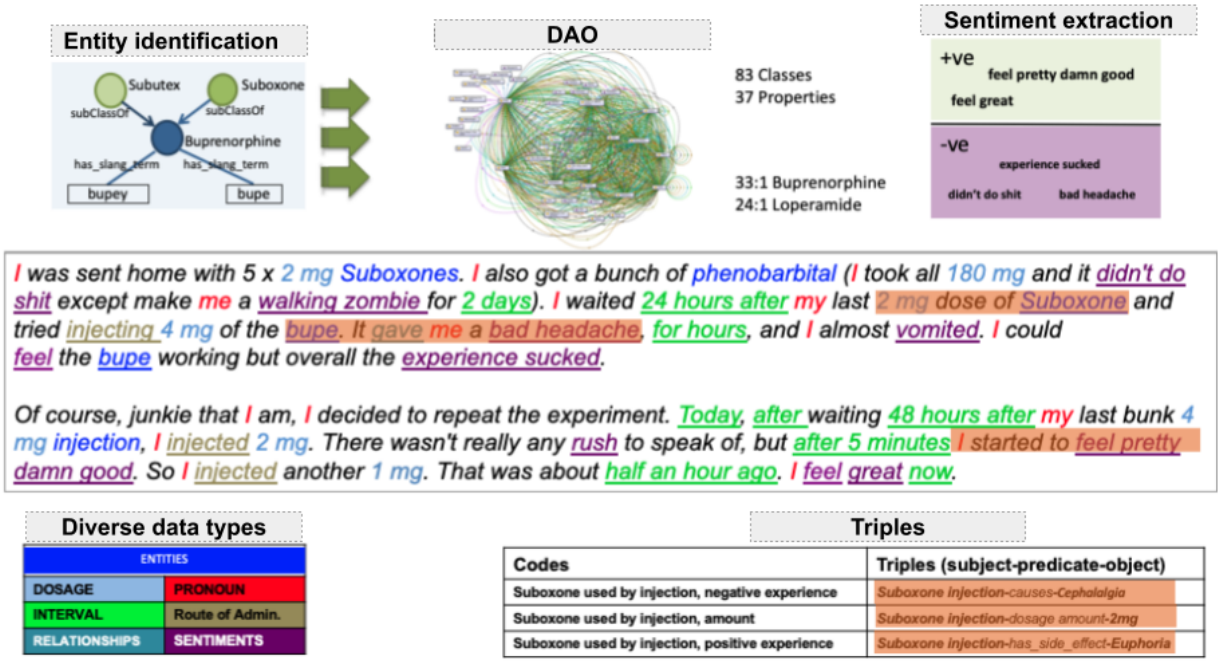


Figure 5. Benefits of ontologies with lexicons and rule-based grammar [80].

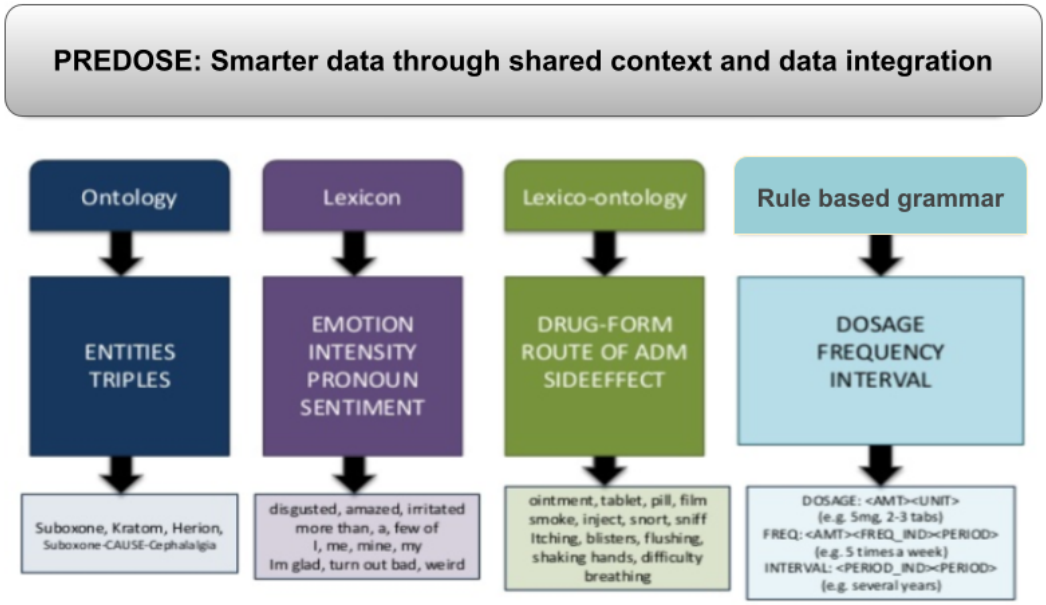
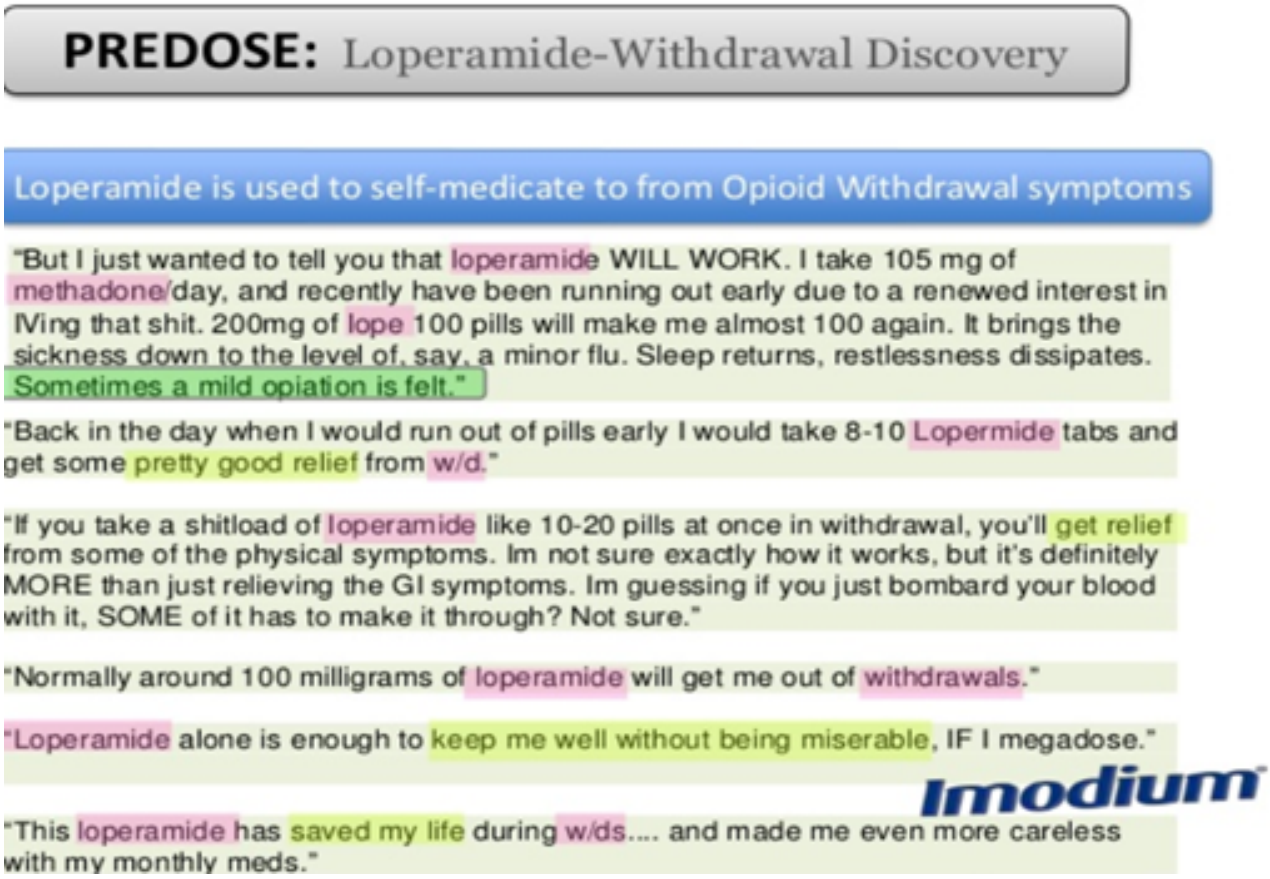


Figure 6. Loperamide discovery and its use in self-medication for opioid withdrawal.



eDrugTrends (Monitoring Drug Trends on Social Media)

The eDrugTrends project aimed to analyze trends in knowledge, attitudes, and behaviors related to the use of cannabis and synthetic cannabinoids on web forums and Twitter [26,28-31]. Figure 7 [79] shows the application of the DAO ontology within the eDrugTrends architecture, which includes 4 stages: (1) data collection, (2) data processing, (3) data access tools for exploration and visualization, and (4) quantitative and qualitative analyses and interpretation. From the social science or substance use epidemiology perspective, the data processing and information extraction stages correspond with the coding task that prepares raw data for further analysis and interpretation. During data processing, the DAO came into the picture by playing an important role in identifying entities in the data that are exact names or synonyms or slang terms or street names of a drug. We generated embedding vectors using the DAO for domain-specific word embedding models and built an ML model to classify users by their types (individual, agency, and retailer) on Twitter by classifying their marijuana-related conversations [28]. We achieved this using multimodal embeddings extracted from people, content, and network views, achieving an 8% improvement over the empirical baseline [28]. We evaluated our approach using the average F1-score for each user type individual (P), informed agency (I), and retailer (R). The F1 scores for the individual classes P, I, and R were 95%, 42%, and 73%, respectively. The descriptive statistics of the training set at the Twitter user account level used for this study, which involved semantic filtering [84] using the DAO, are shown in Table 2.

Figure 7. Architecture of the eDrugTrends project.

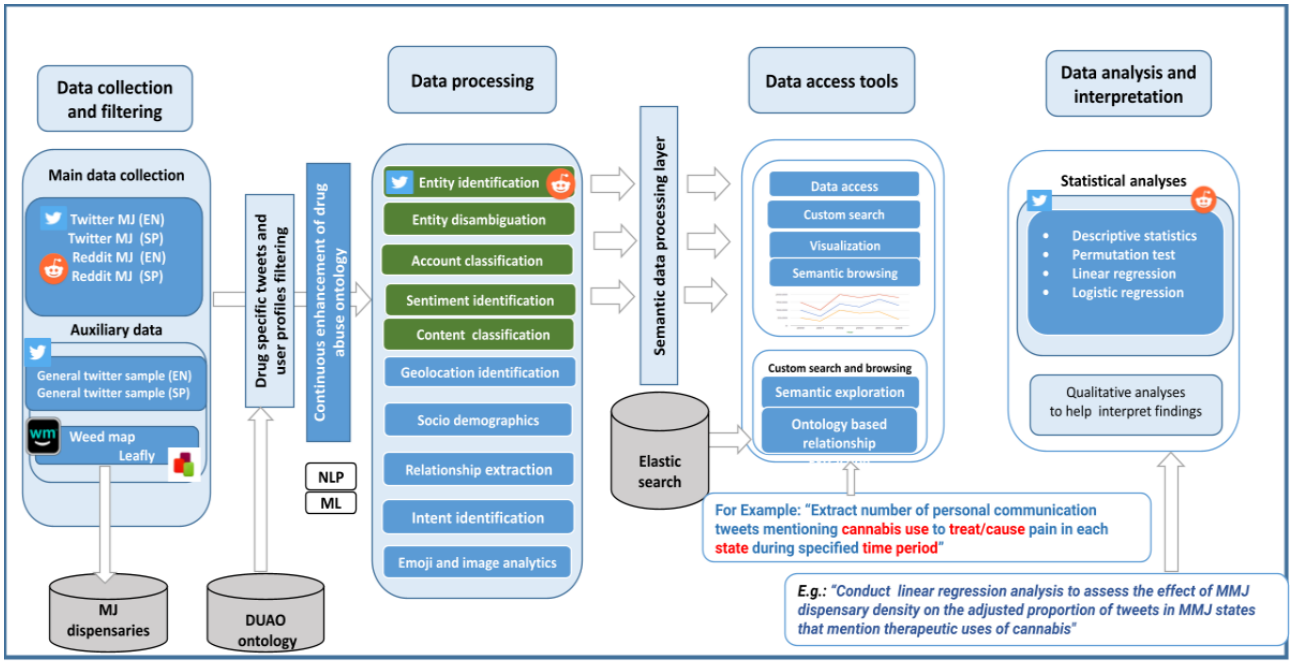


Table 2. Descriptive information of user accounts on Twitter extracted using the drug abuse ontology [28].

Features	Personal accounts	Retail accounts	Informed agency	Total
Number of tweets	9836	1928	338	12,102
Number of profile pictures	4394	476	111	4981
Number of users with description	3884	461	108	4453
Number of retweets	955	24	964	1943
Number of mentions	94	6	307	407

Enhancing the DAO With DSM-5

The motive for enhancing the DAO with DSM-5 is to provide actionable information to clinicians about the mental health of a patient in diagnostic terms for web-based interventions. We chose Reddit data for this study as the concepts, instances, and relations associated with drugs are semantically connected to mental health communications on social media, especially on Reddit. In our Reddit corpus, the drug use–related categories form a substantial portion (48%; corpus size is 2.5 million posts from 15 mental health subreddits by 268,104 users) of the data set in size. However, the DAO still lacked concepts directly related to mental health diagnostic disorders as defined in DSM-5 that are present in the International Classification of Diseases 10th edition [85], Systematized Nomenclature of Medicine-Clinical Terms [86], and DataMed [87]. In a recent study [41] on matching mental conditions of user posts on Reddit to DSM-5 diagnostic disorders, we enhanced the DAO with knowledge derived from DSM-5, which includes 20 chapters (Table 3), consistent with International Classification of Diseases 10th edition and NIH's research domain criteria [88] for mental health. The enhanced DAO includes representations of mental health disorders and related symptoms that were developed following the DSM-5 classification [89]. For example, references for *Cannabis Use Disorder* include terms such as *addicted to cannabis*, *addicted to Marijuana*, and *Jazz_mango addict*. References to the feeling of *anxiety* or *anxious* include such terms as *antsy, worried*, and *agitated*. These lay terms were added to the DAO manually using synonym dictionaries and by manually examining Reddit conversations related to depression, anxiety, and other mental health conditions.

Table 3. Demonstration of improvement in the number of DSM-5[a] category–related concepts being captured before and after including the DAO[b] [41].

DSM-5 category	DSM-5–related concepts captured without the DAO, n	DSM-5–related concepts captured with the DAO, n
Dissociative disorder	20	20
Anxiety disorder	40	87
Substance use and addictive disorder	39	123
Schizophrenia spectrum	77	77
Sleep-wake disorder	14	19
Paraphilic disorders	14	14
Gender dysphoria	15	15
Neurodevelopmental disorders	25	53
Sexual dysfunctions	23	23
Personality disorders	76	98
Trauma and stressor related disorder	25	28
Disruptive, impulse, control, and conduct disorder	34	34
Psychotic disorders	85	87
Bipolar and related disorders	75	84
Elimination disorders	18	18
Depressive disorders	71	107
Obsessive-compulsive related disorder	43	60
Feeding and eating disorders	32	39
Neurocognitive disorders	80	80
Suicidal behavior or ideation	34	47

[a]DSM-5: Diagnostic and Statistical Manual for Mental Disorders-5th edition.

[b]DAO: drug abuse ontology.

The DAO, curated and enhanced by DSM-5 concepts, was used in a weakly supervised setting to label Reddit posts with DSM-5 categories. In a comparative analysis with the state-of-the-art research by Park and Conway [90], Saravia et al [91], and Gkotsis et al [92], we observed that expansion of the DAO with DSM-5 helped improve the accuracy of our entity identification tools (reduced false positives by 92%). These results are shown in Figure 8. We further assessed the meaningfulness of the prediction through a reliability assessment with a domain expert, which gave an agreement score of 84%. In addition, the incorporation of slang terms from the DAO to match and process the informal social media data improved both coverage and recall (Table 4). Thus, we demonstrated that semantic weighting of contextual features from the content using the DAO and DSM-5 knowledge could significantly improve the robustness of the artificial intelligence system. As web-based content is mapped to a clinically acceptable vocabulary, the system brings in explainability. Furthermore, Table 3 shows the improvement in the number of concepts extracted from the DAO being captured in our Reddit Corpus that relate to DSM-5, 20 chapters before and after adding slang terms.

Figure 8. Results illustrating that domain-specific knowledge bases lower false alarm rates in identifying Diagnostic and Statistical Manual for Mental Disorders-5th edition (DSM-5) categories to tag posts in mental health subreddits. DAO: drug abuse ontology.

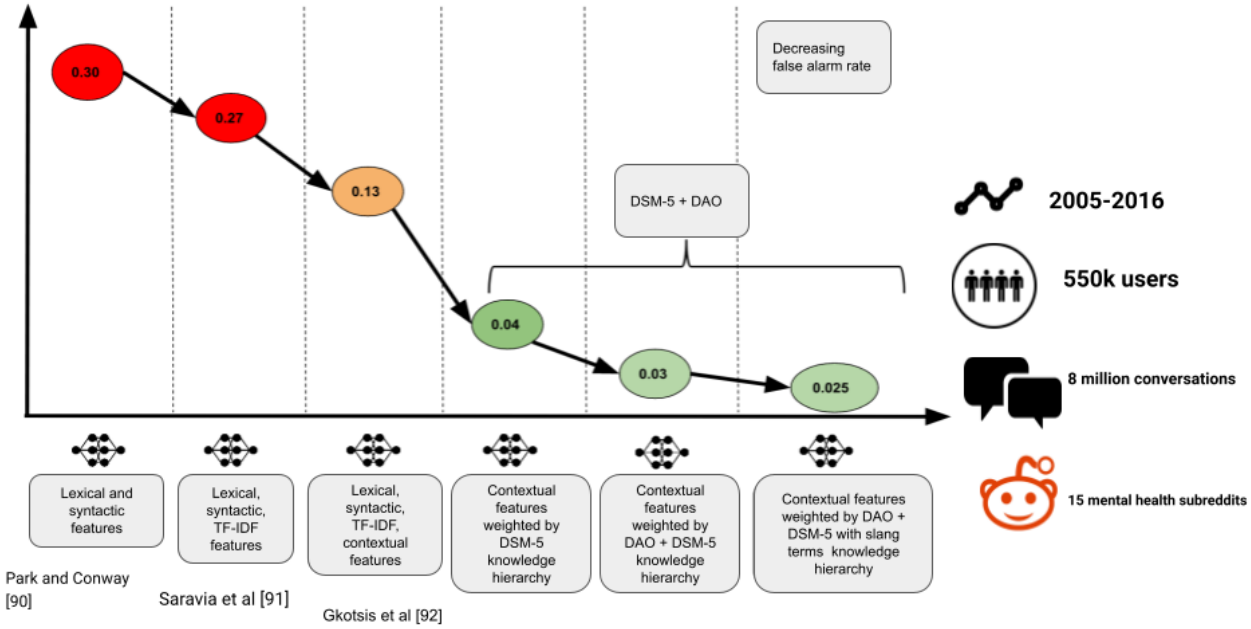


Table 4. Ablation study on contextual features and their modulation using SEDO[a] weights generated from either DSM-5[b] or its enrichment using the DAO[c] and slang terms[d].

Method (with horizontal linguistic features, vertical linguistic features, and fine-grained features)	Precision	Recall	F1-score
BRF[e] with CF[f]	0.60	0.54	0.57
BRF-CF (SEDO weights generated from DSM-5 lexicon without the DAO)	0.87	0.77	0.82
BRF-CF (SEDO weights generated from DSM-5 lexicon with the DAO without slang terms)	0.87	0.80	0.83
BRF-CF (SEDO weights generated from DSM-5 lexicon without the DAO with slang terms)	0.85	0.82	0.83
BRF-CF (SEDO weights generated from DSM-5 lexicon with the DAO with slang terms)	0.88	0.83	0.85

[a]SEDO: Semantic Encoding and Decoding Optimization.

[b]DSM-5: Diagnostic and Statistical Manual for Mental Disorders-5th edition.

[c]DAO: drug abuse ontology.

[d]This table demonstrates the improvement of models with the enhanced DAO.

[e]BRF: balanced random forest.

[f]CF: contextual features.

The base model for the ablation study is a balanced random forest with horizontal linguistic features (number of definite articles, words per post, first-person pronouns, pronouns, and subordinate conjunctions), vertical linguistic features (number of part-of-speech tags, similarity between the posts, intrasubreddit similarity, and intersubreddit similarity), and fine-grained features (sentiment, emotion, and readability scores).

eDarkTrends (Monitoring Drug Trends on Cryptomarkets)

The DAO also plays an essential role in identifying relevant entities and analyzing data from the Darknet cryptomarkets (eg, Agora, Dream Market, and Empire Market) to quantify and assess the availability of fentanyl, fentanyl analogs, and other novel synthetic opioids on the cryptomarkets [25,26]. The snapshot of the Darknet Marketplace is shown in Figure 9 [33].

The terms and slang terms associated with instances populating the DAO opioid subclass, as well as the dosage (eg, gram, mL, and ounce) and form (eg, tablet and powder) classes were compiled as regular expressions and used as expression patterns in the dedicated named entity recognition (NER) algorithm specifically designed for Darknet data [35]. The DAO was inductively augmented with abbreviations and terms specific to the cryptomarket environment (eg, fuff for fluoro-furanyl fentanyl or FE for finalize early) to ensure that only relevant data on novel synthetic opioids were collected. The NER allows capturing the types and quantities of novel synthetic opioids advertised on crypto markets; for example, the NER would provide the following information about the advertisement *FENTANYL TRANSDERMAL PATCHES 100 mcg per h* as class: fentanyl-type; name: fentanyl; dosage: 0.0001 g per h; form: transdermal. The results regarding the average numbers of fentanyl, fentanyl analogs, and other nonpharmaceutical

synthetic opioids advertised on cryptomarkets identified are shown in Table 5. The crawls considered to obtain these results were the dark web posts collected from the Agora and Dream markets in the years 2015 and 2018 [35]. We also classified vendors on Darknet markets (Dream, Tochka, and Wall Street are the marketplaces used for this study) using the DAO. The summary of our findings related to unique vendors, substance, location, vendor descriptions, and the number of withdrawal transactions is shown in Table 6.

Figure 9. Screenshot of the Darknet marketplace.

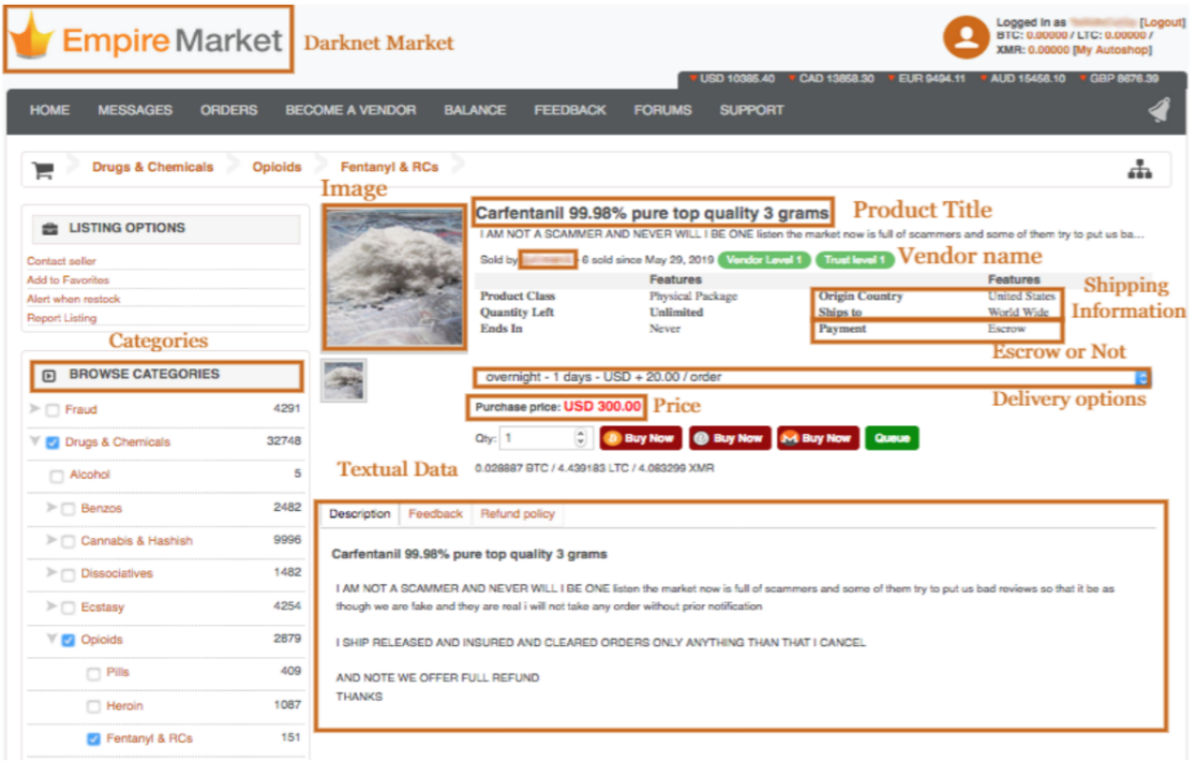


Table 5. Average number of fentanyl, fentanyl analogs, and other nonpharmaceutical synthetic opioids advertised on cryptomarkets extracted with the drug abuse ontology [34].

Types of substances	Average number of advertisements per day, by month (number of crawls)				
	Agora			Dream Market	
	March 2015	April 2015	May 2015	March 2018	April 2018
Fentanyl[a]	130	174	139	207	216
Fentanyl analogs					
Acetyl fentanyl	44	39	41	3	1
Butyr fentanyl	12	10	17	6	7
Carfentanil	0	0	0	12	5
Furanyl fentanyl	0	0	1	31	39
Methoxy Acetyl fentanyl	0	0	0	14	14
4-fluroIsoButyr fentanyl	0	0	0	19	16
3-methoxyMethyl fentanyl	0	0	0	2	2
Total, fentanyl analogs	56	49	59	87	84
Other NP[b] synthetic opioids					
U-47,700	5	4	5	0	3
W-18	5	4	5	0	0
MT-45	9	8	9	0	0
AH-7921	0	0	1	0	0
U-48,800	0	0	0	1	7
U-49,900	0	0	0	0	1
U-4TDP	0	0	0	0	4
U-50,488	0	0	0	8	4
MPF-47700	0	0	0	0	5
Total, other NP synth opioids	19	16	20	9	24
Other opioids[c]	827	1061	1152	3211	3137
Total (any opioids)	1033	1300	1370	3512	3460

[a]Includes mentions of fentanyl, China white heroine, synthetic heroine, and mentions of pharmaceutical fentanyl such as Duragesic, fentanyl patches, and fentanyl transdermal system.

[b]NP: nonpharmaceutical.

[c]Includes mentions of heroin, opium, morphine and other types of pharmaceutical opioids (eg, hydrocodone, oxycodone, and hydromorphone) excluding pharmaceutical fentanyl.

Table 6. Summary of data set extracted from Darknet markets using the drug abuse ontology [33].

Marketplace	Withdrawal number of transactions	Bitcoin	US $ equivalent	Unique number of vendors	Unique number of substances	Unique number of locations	Unique number of descriptions
Dream	261	99.1503695	197,589.12	1448	852	356	16,800
Tochka	2990	0.70483642	5072.33	408	313	44	1829
Wall Street	7755	2.572515	18,729.40	466	290	29	1723

COVID-19 Scenario

We performed a spatiotemporal analysis of the psychological impact of the novel COVID-19 using approximately 1.2 billion tweets from January 1 to April 10, 2020 [93,94]. The concepts related to addiction and mental health in the COVID-19–related data were semiautomatically recognized using the entities and slang terms mentioned in the DAO. Approximately 90 related concepts and 140 slang terms were used to extract tweets mentioning illicit drug use, alcoholism, and pharmacological drug misuse. Furthermore, suicide risk factors such as insomnia and depression were observed in the tweets extracted using the DAO. Similarly, we studied the negative media exposure from

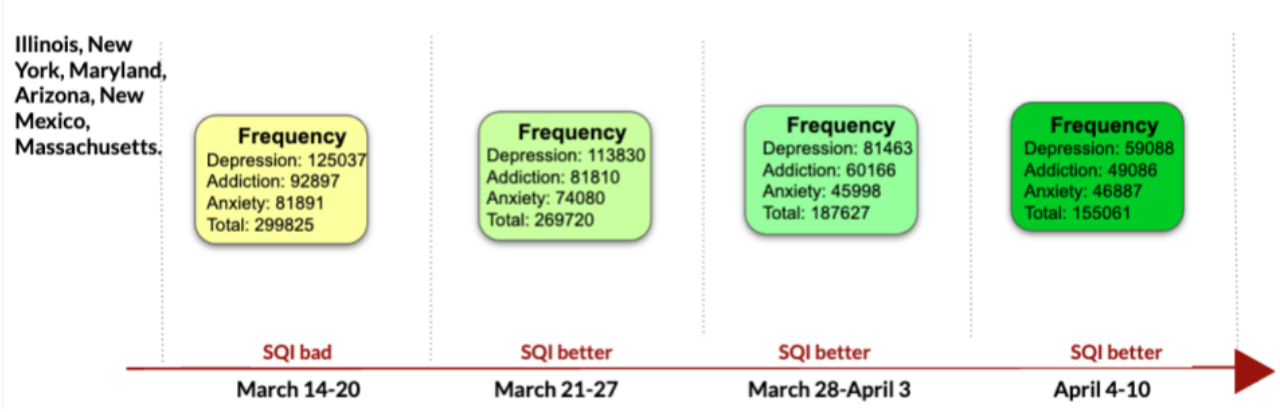
approximately 700,000 news articles published during the COVID-19 pandemic by fine-tuning the bidirectional encoder representations from transformers (BERT) model with the DAO [53]. The 3 months (January, February, and March) in the year 2020 were considered for our earlier study, as this period had a huge COVID-19 spread as per the Mental Health America report [95]. We used 10 of the 13 states recognized as high-spread areas in this report. The 3 states that are not included in Table 7 are Washington, Wyoming, and Idaho. These 3 states were not included, as the related data were not present in our data set cohort. In this work, we reported the state-wise labels (ie, depressive, drug abusive, and informative) for each month using deep learning models vanilla BERT, depression BERT, and drug use BERT, as shown in Table 7. The definitions of these deep learning models are described in Textbox 1. This study is followed by analyzing the Social Quality Index, which aggregates mental health components (depression and anxiety), addiction, and substance use disorders, considering tweets in the period March to April 2020. The Social Quality Index and tweets for states Illinois, New York, Maryland, Arizona, New Mexico, and Massachusetts are shown in Figure 10 [94].

Table 7. Evaluation of BERT[a] models for Mental Health America states over 3 months (January, February, and March 2020) [53,94].

Mental Health America states with depression and drug use	vanillaBERT (2020; months)	Druguse-BERT (2020; months)	Depression BERT (2020; months)
Tennessee	February and March	February and March	February and March
Alabama	February	February and March	February
Oklahoma	March	February and March	February and March
Kansas	February	January and February	January and February
Montana	March	February	February and March
South Carolina	March	March	February and March
Alaska	February and March	January, February, and March	February and March
Utah	March	March	March
Oregon	None	February	None
Nevada	February	February	February

[a]BERT: bidirectional encoder representations from transformers.

Figure 10. Social quality index (SQI) pattern of improvement in conditions as the decline in the number of tweets on depression, addiction, and anxiety.



Discussion

Strengths and Limitations

The DAO is an ongoing project that can be continuously improved and expanded to handle additional topic areas and emerging substance use issues and trends. DAO development requires intensive, hands-on involvement of experts in the field of substance use research (domain experts). We acknowledge a limitation to our approach in that our DAO development team did not include persons with lived experiences of substance use disorders. In the future, it would be important to also involve individuals who use drugs to help develop and refine DAO sections and terms. The DAO can provide a tool and a framework for interdisciplinary collaborative teams to carry this work forward. The DAO ontology has been proven effective in several scenarios, as demonstrated in *Evaluation* section (Section 3). Table 8 summarizes the evolution and improvement of the ontology use according to the needs of the projects. The public health findings described in this document of associated projects, with a focus on person, place, and time, are referenced in Table 8.

Table 8. Summary of the drug abuse ontology implemented in projects.

Domain	Related publications	Manuscript section	Data type	Findings reference
Buprenorphine, loperamide, other opioids	Cameron et al [26], Daniulaityte et al [25,82]	PREDOSE[a] [26]	Web forum data	Figures 4 and 5
User types in marijuana-related posts on social media	Kursuncu et al [28], Lamy et al [31]	eDrugTrends [28-31,96]	Twitter data, web forums, and Bluelight	Table 2
Depression DSM-5	Gaur et al [41]	eDrugTrends [45]	Web forums, Reddit, and Twitter	Tables 3 and 4
Fentanyl, fentanyl analogs, Clustering of dark web vendors	Usha et al [35], Kumar et al [33], Lamy et al [34]	eDarkTrends [46]	Social media and cryptomarket	Tables 5 and 6
COVID-19	Gaur et al [53,88]	COVID-19: public health study [97]	Social media	Figure 10; Table 7

[a]PREDOSE: Prescription Drug Abuse Online Surveillance.

Principal Findings and Conclusions

In this study, we developed and evaluated the DAO as a framework for identifying concepts, entities, and relationships of interest in social media posts. The DAO developed in this study comprises 315 classes, 31 relationships, and 814 instances with 2 to 4 levels deeper. Our ontology was designed to study social media data, dark web data, and web forums. The DAO is primarily used for knowledge extraction and is broadly applicable to these platforms.

The superclasses of our ontology integrate all concepts regarding health conditions, individual-related, network-related, and society (public policies), sources (dealers, internet, medical, self-produced), spatiotemporal, and substance-related classes. The integrated ontology developed in this study is suitable for analyzing social media posts and dark web posts to understand network-related characteristics, location and time issues, identifying new trends, synonyms, slang items, and new drugs.

Our ontology incorporates terminology not only extracted from DSM-5 but also various terms and slang used on social media and other web posts. The terminology with all the medical terms, synonyms, and slang terms representing all the substances enabled a rich collection of terms in social media and dark web data. Our ontology also helps in topic discovery and entity extraction from social media and dark web data. In addition, we used ontology to extract information in the description of each product in dark web marketplaces to identify substances that are being sold that are not known, such as synthetic drugs, research chemicals, synthetic cannabinoids, and synthetic heroin.

Following well-known software development methodologies (eg, agile methodology), the ontology is constantly being updated according to the needs of current addiction-based research. The DAO stands as a machine-processable resource that describes a collection of addiction domain-related objects and classes, and is growing with the needs of the new ongoing projects. For instance, the current ontology is being enriched with knowledge from the dark web. In future work, the ontology will be linked to other ontologies (eg, MEDDRA [98], a Medical Dictionary for Regulatory Activities) to design the drug abuse knowledge graph. Another research contribution would be to automatically update the DAO with new concepts and properties, inspired by the algorithm that allows users to interactively build topic-specific ontologies using suggestions retrieved from a knowledge graph [99]. Glossary of the terms used in this paper is provided in Multimedia Appendix 1.

Acknowledgments

This work was supported in part by the National Institute on Drug Abuse (NIDA) grant 5R01DA039454-02 *Trending: Social Media Analysis to Monitor Cannabis and Synthetic Cannabinoid Use*; the National Science Foundation award 1761931 *Spokes: MEDIUM: MIDWEST: Collaborative: Community-Driven Data Engineering for Substance Abuse Prevention in the Rural Midwest*; the NIDA grant 5R21DA044518-02 *eDarkTrends: Monitoring Darknet Markets to Track Illicit Synthetic Opioid Trends*; and the National Institutes of Health grant R21 DA030571-01A1 *A Study of Social Web Data on Buprenorphine Abuse using Semantic Web Technology*. Any opinions, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation, National Institutes of Health, or NIDA.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Glossary of terms used in this paper.
[DOCX File , 16 KB - publichealth_v8i12e24938_app1.docx]

References

1. UNODC World Drug Report 2020: Global drug use rising; while COVID-19 has far reaching impact on global drug markets. United Nations. 2020 Jul 25. URL: https://www.unodc.org/unodc/press/releases/2020/June/media-advisory---global-launch-of-the-2020-world-drug-report.html [accessed 2022-04-09]
2. Boardman JD, Finch BK, Ellison CG, Williams DR, Jackson JS. Neighborhood disadvantage, stress, and drug use among adults. J Health Soc Behav 2001 Jun;42(2):151-165. [Medline: 11467250]
3. Spooner C, Hetherington K. Social Determinants of Drug Use. Sydney, New South Wales: University of New South Wales; 2004.
4. Hall W, Lynskey M. Assessing the public health impacts of legalizing recreational cannabis use: the US experience. World Psychiatry 2020 Jun 11;19(2):179-186 [FREE Full text] [doi: 10.1002/wps.20735] [Medline: 32394566]
5. Ross EJ, Graham DL, Money KM, Stanwood GD. Developmental consequences of fetal exposure to drugs: what we know and what we still must learn. Neuropsychopharmacology 2015 Jan;40(1):61-87 [FREE Full text] [doi: 10.1038/npp.2014.147] [Medline: 24938210]
6. European Drug Report 2019: trends and developments. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). 2019 Jun. URL: https://www.emcdda.europa.eu/publications/edr/trends-developments/2019_en [accessed 2021-02-15]
7. Kruithof K, Aldridge J, Hétu DD, Sim M, Dujso E, Hoorens S. The role of the 'dark web' in the trade of illicit drugs. RAND Corporation. 2016. URL: https://www.rand.org/pubs/research_briefs/RB9925.html [accessed 2022-07-29]
8. Aldridge J, Décary-Hétu D. Hidden wholesale: the drug diffusing capacity of online drug cryptomarkets. Int J Drug Policy 2016 Sep;35:7-15 [FREE Full text] [doi: 10.1016/j.drugpo.2016.04.020] [Medline: 27260863]
9. Kursuncu U, Gaur M, Lokala U, Thirunarayan K, Sheth A, Arpinar I. Predictive analysis on Twitter: techniques and applications. In: Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining. Cham: Springer; 2019.
10. Soska K, Christin N. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In: Proceedings of the 24th USENIX Conference on Security Symposium. 2015 Presented at: SEC'15: Proceedings of the 24th USENIX Conference on Security Symposium; Aug 12 - 14, 2015; Washington, D.C.
11. Cardoso J, Sheth A. The semantic web and its applications. In: Semantic Web Services, Processes and Applications. Boston, MA: Springer; 2006.
12. Uschold M, Gruninger M. Ontologies and semantics for seamless connectivity. SIGMOD Rec 2004 Dec;33(4):58-64 [FREE Full text] [doi: 10.1145/1041410.1041420]
13. Grimm S, Abecker A, Völker J, Studer R. Ontologies and the semantic web. In: Handbook of Semantic Web Technologies. Berlin, Heidelberg: Springer; 2011.
14. Horrocks I. Ontologies and the semantic web. Commun ACM 2008 Dec;51(12):58-67. [doi: 10.1145/1409360.1409377]
15. Kim AR, Park H, Song T. Development and evaluation of an obesity ontology for social big data analysis. Healthc Inform Res 2017 Jul;23(3):159-168 [FREE Full text] [doi: 10.4258/hir.2017.23.3.159] [Medline: 28875050]
16. Jung H, Park H, Song T. Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals. J Med Internet Res 2017 Jul 24;19(7):e259 [FREE Full text] [doi: 10.2196/jmir.7452] [Medline: 28739560]
17. Ethier J, Barton A, Taseen R. An ontological analysis of drug prescriptions. Applied Ontol 2018 Nov 09;13(4):273-294. [doi: 10.3233/ao-180202]
18. Ceusters W. An information artifact ontology perspective on data collections and associated representational artifacts. Stud Health Technol Inform 2012;180:68-72. [Medline: 22874154]
19. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The ontology for biomedical investigations. PLoS One 2016 Apr 29;11(4):e0154556 [FREE Full text] [doi: 10.1371/journal.pone.0154556] [Medline: 27128319]
20. Scheuermann R, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. Summit Transl Bioinform 2009 Mar 01;2009:116-120 [FREE Full text] [Medline: 21347182]
21. Hicks A, Hanna J, Welch D, Brochhausen M, Hogan WR. The ontology of medically related social entities: recent developments. J Biomed Semantics 2016 Jul 12;7(1):47 [FREE Full text] [doi: 10.1186/s13326-016-0087-8] [Medline: 27406187]
22. Hanna J, Joseph E, Brochhausen M, Hogan WR. Building a drug ontology based on RxNorm and other sources. J Biomed Semantics 2013 Dec 18;4(1):44 [FREE Full text] [doi: 10.1186/2041-1480-4-44] [Medline: 24345026]
23. Kamdar M, Hamamsy T, Shelton S, Vala A, Eftimov T, Zou J. A knowledge graph-based approach for exploring the U.S. opioid epidemic. arXiv 2019 [FREE Full text]
24. Motlagh F, Shekarpour S, Sheth A, Thirunarayan K, Raymer M. Predicting public opinion on drug legalization: social media analysis and consumption trends. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2019 Presented at: ASONAM '19: International Conference on Advances in Social Networks Analysis and Mining; Aug 27 - 30, 2019; Vancouver British Columbia Canada. [doi: 10.1145/3341161.3344380]

25. Daniulaityte R, Carlson R, Falck R, Cameron D, Perera S, Chen L, et al. "I just wanted to tell you that loperamide WILL WORK": a web-based study of extra-medical use of loperamide. Drug Alcohol Depend 2013 Jun 01;130(1-3):241-244 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.11.003] [Medline: 23201175]
26. Cameron D, Smith GA, Daniulaityte R, Sheth AP, Dave D, Chen L, et al. PREDOSE: a semantic web platform for drug abuse epidemiology using social media. J Biomed Inform 2013 Dec;46(6):985-997 [FREE Full text] [doi: 10.1016/j.jbi.2013.07.007] [Medline: 23892295]
27. Daniulaityte R, Carlson R, Brigham G, Cameron D, Sheth A. "Sub is a weird drug:" a web-based study of lay attitudes about use of buprenorphine to self-treat opioid withdrawal symptoms. Am J Addict 2015 Aug 25;24(5):403-409 [FREE Full text] [doi: 10.1111/ajad.12213] [Medline: 26009867]
28. Kursuncu U, Gaur M, Lokala U, Illendula A, Thirunarayan K, Daniulaityte R. What's ur type? Contextualized classification of user types in marijuana-related communications using compositional multiview embedding. In: Proceedings of the 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI). 2018 Presented at: 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI); Dec 03-06, 2018; Santiago, Chile. [doi: 10.1109/wi.2018.00-50]
29. Daniulaityte R, Nahhas RW, Wijeratne S, Carlson RG, Lamy FR, Martins SS, et al. "Time for dabs": analyzing Twitter data on marijuana concentrates across the U.S. Drug Alcohol Depend 2015 Oct 01;155:307-311 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.07.1199] [Medline: 26338481]
30. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. "When 'Bad' is 'Good'": identifying personal communication and sentiment in drug-related Tweets. JMIR Public Health Surveill 2016 Oct 24;2(2):e162 [FREE Full text] [doi: 10.2196/publichealth.6327] [Medline: 27777215]
31. Lamy FR, Daniulaityte R, Sheth A, Nahhas RW, Martins SS, Boyer EW, et al. "Those edibles hit hard": exploration of Twitter data on cannabis edibles in the U.S. Drug Alcohol Depend 2016 Jul 01;164:64-70 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.04.029] [Medline: 27185160]
32. Daniulaityte R, Lamy FR, Smith GA, Nahhas RW, Carlson RG, Thirunarayan K, et al. "Retweet to pass the blunt": analyzing geographic and content features of cannabis-related Tweeting across the United States. J Stud Alcohol Drugs 2017 Nov;78(6):910-915 [FREE Full text] [doi: 10.15288/jsad.2017.78.910] [Medline: 29087826]
33. Kumar R, Yadav S, Daniulaityte R, Lamy F, Lokala U. eDarkFind: unsupervised multi-view learning for Sybil Account detection. In: Proceedings of The Web Conference 2020. 2020 Presented at: WWW '20: The Web Conference 2020; Apr 20 - 24, 2020; Taipei Taiwan. [doi: 10.1145/3366423.3380263]
34. Lamy F, Daniulaityte R, Barratt M, Lokala U, Sheth A, Carlson R. Listed for sale: analyzing data on fentanyl, fentanyl analogs and other novel synthetic opioids on one cryptomarket. Drug Alcohol Depend 2020 Jun 12;213:108115 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108115] [Medline: 32585419]
35. Lokala U, Lamy FR, Daniulaityte R, Sheth A, Nahhas RW, Roden JI, et al. Global trends, local harms: availability of fentanyl-type drugs on the dark web and accidental overdoses in Ohio. Comput Math Organ Theory 2019 Mar 25;25(1):48-59 [FREE Full text] [doi: 10.1007/s10588-018-09283-0] [Medline: 32577089]
36. Ontology101. Protege Wiki. URL: https://protegewiki.stanford.edu/wiki/Ontology101, [accessed 2022-04-10]
37. Welcome to ESIP's Community Ontology Repository, or COR. Community Ontology Repository. URL: http://esipfed.github.io/cor/ [accessed 2022-04-10]
38. About DBpedia. DBpedia. URL: https://wiki.dbpedia.org/about [accessed 2022-04-09]
39. Ontology metrics. Protégé 5 Documentation. URL: http://protegeproject.github.io/protege/views/ontology-metrics/ [accessed 2022-04-09]
40. Gyrard A. PerfectO: semantic web best practices. PerfectO. 2021 Oct. URL: http://perfectsemanticweb.appspot.com/ [accessed 2020-02-20]
41. Gaur M, Kursuncu U, Alambo A, Sheth A, Daniulaityte R, Thirunarayan K. "Let me tell you about your mental health!": contextualized classification of reddit posts to DSM-5 for web-based intervention. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018 Presented at: CIKM '18: The 27th ACM International Conference on Information and Knowledge Management; Oct 22 - 26, 2018; Torino Italy. [doi: 10.1145/3269206.3271732]
42. WebVOWL - Web-based visualization of ontologies. VOWL. URL: http://vowl.visualdataweb.org/webvowl.html [accessed 2022-04-10]
43. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Given LM, et al. Social media use among patients and caregivers: a scoping review. BMJ Open 2013 May 09;3(5):e002819 [FREE Full text] [doi: 10.1136/bmjopen-2013-002819] [Medline: 23667163]
44. Bluelight homepage. Bluelight. URL: https://www.bluelight.org/xf/ [accessed 2021-02-15]
45. eDrugTrends Center for Interventions, Treatment, and Addictions Research (CITAR). Wright State University. URL: https://medicine.wright.edu/citar/edrugtrends [accessed 2020-02-20]
46. eDarkTrends | Center for Interventions, Treatment, and Addictions Research (CITAR). Wright State University. URL: https://medicine.wright.edu/citar/edarktrends [accessed 2020-02-20]
47. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and opioid-involved overdose deaths — United States, 2013–2017. MMWR Morb Mortal Wkly Rep 2018 Dec 21;67(5152):1419-1427 [FREE Full text] [doi: 10.15585/mmwr.mm6751521e1]

48. Wilson N, Kariisa M, Seth P, Smith H, Davis NL. Drug and opioid-involved overdose deaths - United States, 2017-2018. MMWR Morb Mortal Wkly Rep 2020 Mar 20;69(11):290-297 [FREE Full text] [doi: 10.15585/mmwr.mm6911a4] [Medline: 32191688]
49. Panchal N, Kamal R, Orgera K, Cox C, Garfield R, Hamel L, et al. The implications of COVID-19 for mental health and substance use. Kaiser Family Foundation. 2020 Aug 21. URL: https://abtcounseling.com/wp-content/uploads/2020/09/The-Implications-of-COVID-19-for-Mental-Health-and-Substance-Use-_-KFF.pdf [accessed 2022-07-29]
50. Garfin D, Silver R, Holman E. The novel coronavirus (COVID-2019) outbreak: amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357 [FREE Full text] [doi: 10.1037/hea0000875] [Medline: 32202824]
51. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560. [doi: 10.1016/s2215-0366(20)30168-1]
52. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen Psych 2020 Mar 06;33(2):e100213. [doi: 10.1136/gpsych-2020-100213]
53. Alambo A, Gaur M, Thirunarayan K. Depressive, drug abusive, or informative: knowledge-aware study of news exposure during COVID-19 outbreak. arXiv 2020 [FREE Full text]
54. National Institute on Drug Abuse. NIDA. URL: https://www.drugabuse.gov/ [accessed 2021-02-11]
55. United States Drug Enforcement Administration homepage. United States Drug Enforcement Administration. URL: https://www.dea.gov/ [accessed 2021-02-11]
56. 2022 European Drug Report. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). URL: https://www.emcdda.europa.eu/emcdda-home-page_en [accessed 2021-02-11]
57. RxNorm. NIH U.S. National Library of Medicines. URL: https://www.nlm.nih.gov/research/umls/rxnorm/index.html [accessed 2021-02-11]
58. Daniulaityte R, Carlson RG, Kenne DR. Initiation to pharmaceutical opioids and patterns of misuse: preliminary qualitative findings obtained by the ohio substance abuse monitoring network. J Drug Issues 2016 Aug 03;36(4):787-808. [doi: 10.1177/002204260603600402]
59. Daniulaityte R, Carlson RG. "To numb out and start to feel nothing": experiences of stress among crack-cocaine using women in a midwestern city. J Drug Issues 2011 Jan 01;41(1):1-24 [FREE Full text] [doi: 10.1177/002204261104100101] [Medline: 21625334]
60. Daniulaityte R, Falck R, Carlson RG. Illicit use of buprenorphine in a community sample of young adult non-medical users of pharmaceutical opioids. Drug Alcohol Depend 2012 May 01;122(3):201-207 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.09.029] [Medline: 22036303]
61. Daniulaityte R, Carlson RG, Kenne DR. Methamphetamine use in Dayton, Ohio: preliminary findings from the Ohio Substance Abuse Monitoring Network. J Psychoactive Drugs 2007 Sep;39(3):211-221. [doi: 10.1080/02791072.2007.10400607] [Medline: 18159774]
62. Lamy FR, Daniulaityte R, Zatreh M, Nahhas RW, Sheth A, Martins SS, et al. "You got to love rosin: solventless dabs, pure, clean, natural medicine." Exploring Twitter data on emerging trends in rosin tech marijuana concentrates. Drug Alcohol Depend 2018 Feb 01;183:248-252 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.10.039] [Medline: 29306816]
63. Noy NF, McGuinness DL. Ontology Development 101: A Guide to Creating Your First Ontology. Stanford, CA: Stanford University; 2001.
64. Musen MA, Protégé Team. The protégé project: a look back and a look forward. AI Matters 2015 Jun 16;1(4):4-12 [FREE Full text] [doi: 10.1145/2757001.2757003] [Medline: 27239556]
65. Lin Y, Liu Z, Sun M, Liu Y, Zhu X. Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2015 Presented at: Twenty-Ninth AAAI Conference on Artificial Intelligence; Jan 25–30, 2015; Austin, Texas, USA URL: https://ojs.aaai.org/index.php/AAAI/article/view/9491
66. Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes PN, et al. DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia. Semant Pragmat 2015;6(2):167-195 [FREE Full text] [doi: 10.3233/SW-140134]
67. Lohmann S, Negru S, Haag F, Ertl T. Visualizing ontologies with VOWL. Semantic Web 2016 May 27;7(4):399-419 [FREE Full text] [doi: 10.3233/SW-150200]
68. Call for resources track papers. ISWC. URL: http://iswc2018.semanticweb.org/call-for-resources-track-papers/ [accessed 2020-02-20]
69. Gyrard A, Atemezing G, Serrano M. PerfectO: an online toolkit for improving quality, accessibility, and classification of domain-based ontologies. In: Semantic IoT: Theory and Applications. Cham: Springer; 2021.
70. LODE - Live OWL Documentation Environment homepage. LODE - Live OWL Documentation Environment. URL: http://www.essepuntato.it/lode [accessed 2020-02-21]
71. Lokala U. Drug abuse ontology. Earth Science Information Partners Community Ontology Repository. 2020 Jul 27. URL: http://cor.esipfed.org/ont/~ushanri/DAO [accessed 2022-02-15]

72. Lohmann S, Negru S, Haag F, Ertl T. Visualizing ontologies with VOWL. Semantic Web 2016 May 27;7(4):399-419. [doi: 10.3233/sw-150200]
73. Visual Notation for OWL Ontologies. Visual Dataweb. 2020 Jul 27. URL: http://vowl.visualdataweb.org/webvowl-old/webvowl-old.html#iri=http://cor.esipfed.org/ont/~ushanri/DAO%C2%A0 [accessed 2022-02-15]
74. Poveda-Villalón M, Gómez-Pérez A, Suárez-Figueroa MC. OOPS! (OntOlogy Pitfall Scanner!): an on-line tool for ontology evaluation. Intl J Semantic Web Inform Syst 2014;10(2) [FREE Full text] [doi: 10.4018/ijswis.2014040102]
75. Vapour report. Vapour Linked Data Validator. URL: http://linkeddata.uriburner.com:8000/vapour?uri=http://cor.esipfed.org/ont/~ushanri/DAO [accessed 2022-02-15]
76. Graphite PHP linked data library. RDF Triple-Checker. URL: http://graphite.ecs.soton.ac.uk/checker/?uri=http://cor.esipfed.org/ont/~ushanri/DAO [accessed 2022-02-15]
77. Ontology metrics. Protégé 5 Documentation. URL: http://protegeproject.github.io/protege/views/ontology-metrics/ [accessed 2021-02-15]
78. García J, García-Peñalvo FJ, Therón R. A survey on ontology metrics. In: Knowledge Management, Information Systems, E-Learning, and Sustainability Research. Berlin, Heidelberg: Springer; 2010.
79. Web ontology language (OWL) abstract syntax and semantics section 2. Abstract syntax. W3C Working Draft. URL: https://www.w3.org/TR/2002/WD-owl-semantics-20021108/syntax.html [accessed 2021-02-15]
80. Sheth A, Perera S, Wijeratne S, Thirunarayan K. Knowledge will propel machine understanding of content: extrapolating from current examples. In: Proceedings of the International Conference on Web Intelligence. 2017 Presented at: WI '17: International Conference on Web Intelligence 2017; Aug 23 - 26, 2017; Leipzig Germany URL: https://datamed.org [doi: 10.1145/3106426.3109448]
81. Estival D, Nowak C, Zschorn A. Towards ontology-based natural language processing. In: Proceeedings of the Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology. 2004 Presented at: NLPXML '04: Proceeedings of the Workshop on NLP and XML (NLPXML-2004): RDF/RDFS and OWL in Language Technology; Jun 1, 2004; Barcelona, Spain.
82. Daniulaityte R, Carlson RG, Falck RS, Cameron DH, Udayanaga S, Chen L, et al. A web-based study of self-treatment of opioid withdrawal symptoms with loperamide. Wright State University. 2012. URL: https://corescholar.libraries.wright.edu/knoesis/624/ [accessed 2022-04-08]
83. PREDOSE Demo. Youtube. 2013 Sep 16. URL: https://www.youtube.com/watch?v=gCFPzMgEPQM [accessed 2020-02-16]
84. Sheth A, Kapanipathi P. Semantic filtering for social data. IEEE Internet Comput 2016 Jul;20(4):74-78. [doi: 10.1109/MIC.2016.86]
85. International Classification of Diseases, Version 10. BioPortal. URL: https://bioportal.bioontology.org/ontologies/ICD10 [accessed 2020-02-20]
86. SNOMED CT. BioPortal. URL: https://bioportal.bioontology.org/ontologies/SNOMEDCT [accessed 2020-02-20]
87. bioCaddie Core Development Team. Home - DataMed. URL: https://datamed.org [accessed 2020-02-20]
88. Research domain criteria (RDoC). Research domain criteria (RDoC). In: National Institute of Mental Health (NIMH). URL: https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc [accessed 2022-04-10]
89. Alambo A, Gaur M, Lokala U, Kursuncu U, Thirunarayan K, Gyrard A, et al. Question answering for suicide risk assessment using reddit. In: Proceedings of the 2019 IEEE 13th International Conference on Semantic Computing (ICSC). 2019 Presented at: 2019 IEEE 13th International Conference on Semantic Computing (ICSC); Jan 30 - Feb 01, 2019; Newport Beach, CA, USA. [doi: 10.1109/ICOSC.2019.8665525]
90. Park A, Conway M. Harnessing reddit to understand the written-communication challenges experienced by individuals with mental health disorders: analysis of texts from mental health communities. J Med Internet Res 2018 Apr 10;20(4):e121 [FREE Full text] [doi: 10.2196/jmir.8219] [Medline: 29636316]
91. Saravia E, Chang C, De Lorenzo RJ, Chen Y. MIDAS: mental illness detection and analysis via social media. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2016 Presented at: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); Aug 18-21, 2016; San Francisco, CA, USA. [doi: 10.1109/ASONAM.2016.7752434]
92. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJ, Dobson RJ, et al. Characterisation of mental health conditions in social media using Informed Deep Learning. Sci Rep 2017 Mar 22;7:45141 [FREE Full text] [doi: 10.1038/srep45141] [Medline: 28327593]
93. Gaur M, Khandelwal V, Kurşuncu U, Pallagani V. Measuring spatio-temporal psychological impact of novel coronavirus through social quality index. YouTube. URL: https://youtu.be/XzYrn0PEzNk [accessed 2020-06-27]
94. Gaur M, Kursuncu U, Khandelwal V, Pallagani V, Shalin V, Sheth A. Psychidemic: measuring the spatio-temporal psychological impact of novel choronovirus with a social quality index. In: Proceedings of the Computing Research Association Annual Conference. 2020 Presented at: Computing Research Association Annual Conference; 2020; -. [doi: 10.1007/bf03256752]
95. Ranking the states. Mental Health America. URL: https://www.mhanational.org/issues/ranking-states [accessed 2021-02-11]
96. Daniulaityte R, Carlson RG, Golroo F, Wijeratne S, Boyer EW, Martins SS, et al. "Time for Dabs": analyzing Twitter data on butane hash oil use. In: 2015 Abstract Book. Fairborn, Ohio: Wright State University; 2015.

97. Covid19 - knoesis wiki. Wiki. URL: http://wiki.aiisc.ai/index.php/Covid19 [accessed 2022-04-08]
98. Mozzicato P. MedDRA. Pharmaceut Med 2009;23:65-75.
99. Böhm K, Ortiz M. A tool for building topic-specific ontologies using a knowledge graph. In: Proceedings of the 31st International Workshop on Description Logics co-located with 16th International Conference on Principles of Knowledge Representation and Reasoning (KR 2018). 2018 Presented at: 31st International Workshop on Description Logics co-located with 16th International Conference on Principles of Knowledge Representation and Reasoning (KR 2018); Oct 27-29, 2018; Tempe, Arizona, US.

Abbreviations

BERT: bidirectional encoder representations from transformers
DAO: drug abuse ontology
DSM-5: Diagnostic and Statistical Manual for Mental Disorders-5th edition
ML: machine learning
NER: named entity recognition
NIDA: National Institute on Drug Abuse
NIH: National Institutes of Health
NLP: natural language processing
ODKG: Opioid Drug Knowledge Graph
PREDOSE: Prescription Drug Abuse Online Surveillance

Edited by H Bradley; submitted 10.10.20; peer-reviewed by M Meacham, P Foulquié, N Dasgupta; comments to author 21.12.20; revised version received 15.02.21; accepted 10.05.22; published 23.12.22.

Please cite as:
Lokala U, Lamy F, Daniulaityte R, Gaur M, Gyrard A, Thirunarayan K, Kursuncu U, Sheth A
Drug Abuse Ontology to Harness Web-Based Data for Substance Use Epidemiology Research: Ontology Development Study
JMIR Public Health Surveill 2022;8(12):e24938
URL: https://publichealth.jmir.org/2022/12/e24938
doi:10.2196/24938
PMID:36563032

Original Paper

Multifactor Quality and Safety Analysis of Antimicrobial Drugs Sold by Online Pharmacies That Do Not Require a Prescription: Multiphase Observational, Content Analysis, and Product Evaluation Study

Tim Ken Mackey[1,2,3], MAS, PhD; Alan K Jarmusch[4], PhD; Qing Xu[3,5], MAS; Kunyang Sun[4], BS; Aileen Lu[4], BS; Shaden Aguirre[4], BS; Jessica Lim[4], BS; Simran Bhakta[4], BS; Pieter C Dorrestein[4,6,7], PhD

[1]Global Health Program, Department of Anthropology, University of California, San Diego, La Jolla, CA, United States

[2]Global Health Policy and Data Institute, San Diego, CA, United States

[3]S-3 Research, San Diego, CA, United States

[4]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, United States

[5]Department of Healthcare Technology and Policy, University of California, San Diego, La Jolla, CA, United States

[6]Department of Pharmacology, University of California, San Diego, La Jolla, CA, United States

[7]Department of Pediatrics, University of California, San Diego, La Jolla, CA, United States

Corresponding Author:
Tim Ken Mackey, MAS, PhD
Global Health Program
Department of Anthropology
University of California, San Diego
9500 Gilman Drive
MC: 0505
La Jolla, CA, 92093
United States
Phone: 1 9514914161
Email: tkmackey@ucsd.edu

Abstract

Background: Antimicrobial resistance is a significant global public health threat. However, the impact of sourcing potentially substandard and falsified antibiotics via the internet remains understudied, particularly in the context of access to and quality of common antibiotics. In response, this study conducted a multifactor quality and safety analysis of antibiotics sold and purchased via online pharmacies that did not require a prescription.

Objective: The aim of this paper is to identify and characterize "no prescription" online pharmacies selling 5 common antibiotics and to assess the quality characteristics of samples through controlled test buys.

Methods: We first used structured search queries associated with the international nonproprietary names of amoxicillin, azithromycin, amoxicillin and clavulanic acid, cephalexin, and ciprofloxacin to detect and characterize online pharmacies offering the sale of antibiotics without a prescription. Next, we conducted controlled test buys of antibiotics and conducted a visual inspection of packaging and contents for risk evaluation. Antibiotics were then analyzed using untargeted mass spectrometry (MS). MS data were used to determine if the claimed active pharmaceutical ingredient was present, and molecular networking was used to analyze MS data to detect drug analogs as well as possible adulterants and contaminants.

Results: A total of 109 unique websites were identified that actively advertised direct-to-consumer sale of antibiotics without a prescription. From these websites, we successfully placed 27 orders, received 11 packages, and collected 1373 antibiotic product samples. Visual inspection resulted in all product packaging consisting of pill packs or blister packs and some concerning indicators of potential poor quality, falsification, and improper dispensing. Though all samples had the presence of stated active pharmaceutical ingredient, molecular networking revealed a number of drug analogs of unknown identity, as well as known impurities and contaminants.

Conclusions: Our study used a multifactor approach, including web surveillance, test purchasing, and analytical chemistry, to assess risk factors associated with purchasing antibiotics online. Results provide evidence of possible safety risks, including substandard packaging and shipment, falsification of product information and markings, detection of undeclared chemicals, high variability of quality across samples, and payment for orders being defrauded. Beyond immediate patient safety risks, these falsified and substandard products could exacerbate the ongoing public health threat of antimicrobial resistance by circulating substandard product to patients.

(JMIR Public Health Surveill 2022;8(12):e41834) doi:10.2196/41834

KEYWORDS

online pharmacy; antimicrobial resistance; drug safety; cyberpharmacies; public health; health website; online health; web surveillance; patient safety

Introduction

The growing spread of antimicrobial resistance (AMR) is a global public health and security threat gaining increased attention from public health practitioners, clinicians, and policy makers alike [1]. The World Health Organization (WHO) estimates that US $1.2 trillion in additional health expenditure per annum is expected due to AMR by 2050 [2]. A recent systematic analysis estimated that AMR led to the direct cause of 1.27 million deaths in 2019 alone [3]. Hence, the future health, environmental, and economic costs of AMR have made it a priority global health issue that needs to be addressed [4].

Growth in AMR is driven by several factors, including misuse and overuse of prescription antibiotics that may not be rationally prescribed or subject to sufficient professional oversight, including nonprescription or over-the-counter dispensing, as well as potential illegal sourcing from nonauthorized channels [5-7]. This includes antibiotics sold via the internet, where documented direct-to-consumer sale can enable patients to choose their dosages, duration of treatment, and type of treatment, and enabling vendors to dispense the product without requiring a valid prescription [8-11]. These online pharmacies may also be conduits for patient exposure to substandard and falsified medicines, with anti-infective classes of medications widely reported as counterfeited and offered for sale online [6,12-15].

Due to the clear public health risks posed by online pharmacies engaged in questionable sourcing, previous studies have focused on conducting surveys of antibiotic purchasing behavior or examining characteristics of online sellers and the general availability of antibiotic products to better characterize the risk. For example, an early study on the topic published in 2009 identified 138 vendors selling antibiotics without a prescription and the sale of several antibiotic therapeutic classes by conducting “no prescription”–related keyword searches on Google and Yahoo search engines [16]. A more recent study published in 2017 reinforced these results by identifying 20 unique online pharmacies located in the United Kingdom offering the sale of antibiotics, with 45% not requiring a valid prescription [6]. Additionally, a 2020 study conducted a nationwide cross-section assessment of online and community pharmacies in China and found that 79% of online pharmacies did not require a valid prescription [7]. Though these studies provide important empirical evidence and reaffirm the use of the internet as an unregulated and potentially illegal point of access for antibiotics, no studies to our knowledge have used a combination of these methods to evaluate actual product safety and quality features of the products offered.

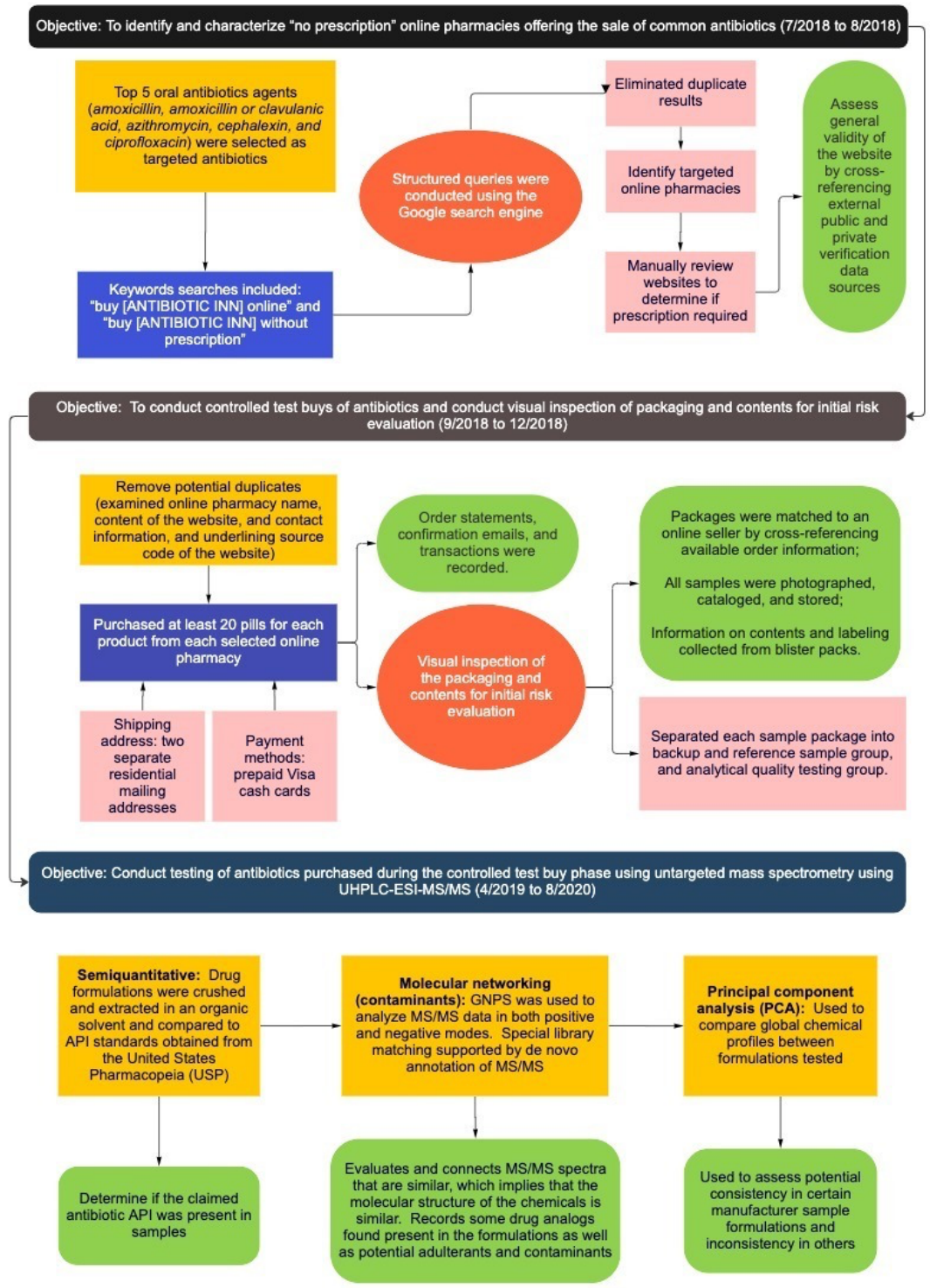
In response, this study expands on prior studies by first identifying and characterizing online pharmacies selling prescription antibiotics with a specific focus on common medications. We then conduct test purchases of antibiotics detected from “no prescription” online pharmacies and visually inspect packaging and products for possible safety concerns. Finally, the study conducts chemical analysis of antibiotic samples using mass spectrometry (MS) and molecular networking to assess quality characteristics and possible risk indicators.

Methods

Overview

This multifactor risk and quality assessment was conducted in 3 phases (Figure 1). The first phase used structured search engine queries to identify and characterize “no prescription” online pharmacies offering the sale of common antibiotics. The second phase used websites identified in the first phase to conduct controlled test buys of antibiotics and conducted visual inspection of packaging and contents for risk evaluation. The third and final phase involved testing antibiotics purchased during the controlled test buy phase using untargeted mass spectrometry through ultra–high performance liquid chromatography–electrospray ionization tandem mass spectrometry. This study did not involve research on human participants. We describe each phase in detail below.

Figure 1. Study strategy and summary of methods. API: active pharmaceutical ingredient; GNPS: global natural product social molecular networking; MS: mass spectrometry; UHPLC-ESI-MS/MS: ultra–high performance liquid chromatography–electrospray ionization tandem mass spectrometry.



Structured Web Searches

To identify websites and specific vendors advertising the sale of common antibiotics without a prescription, we selected the top 5 oral antibiotics agents prescribed to outpatient in the United States (amoxicillin, amoxicillin and clavulanic acid, azithromycin, cephalexin, and ciprofloxacin) according to data from the US Centers for Disease Control and Prevention ("targeted antibiotics") as they constitute commonly prescribed antibiotics in the jurisdiction of focus of this study [17]. Structured queries were conducted using the Google search engine with cookie files removed and the incognito privacy setting selected in Chrome browser. Keyword searches included the following: "buy [ANTIBIOTIC] online" and "buy [ANTIBIOTIC] without prescription" repeated for all 5 targeted antibiotics using a protocol similar to other published studies

(Multimedia Appendix 1) [6,18]. All URLs or hyperlinks populated in the first 10 pages of search results were included for analysis. Searches were conducted between April and May 2018.

After collating search engine results by cataloging returned URL or hyperlinks, we first eliminated all duplicate results and then conducted website content analysis to identify and classify websites that met the criteria for *online pharmacies*—websites that purport to operate as an internet pharmacy and include an e-commerce shopping cart to enable direct-to-consumer purchase (see Multimedia Appendix 1 for other website classifications assessed) [6,8,19,20]. URLs or websites that were classified as *online pharmacies* were then manually reviewed to determine if they required a valid prescription to place an order. This included assessing if the website claimed to be a "no prescription" online pharmacy and confirming whether a valid prescription was required prior to the ordering process, or if the website did not require a prescription or claimed to use a medical questionnaire in lieu of a prescription requirement.

We also assessed the general validity of the website or URL by cross-referencing external public and private verification data sources including LegitScript (a web-based service that monitors online pharmacies for compliance with applicable laws and regulations and classifies illegal and legitimate websites), the National Association of Boards of Pharmacy (NABP) Not Recommended List, and the European Union (EU) logo for online sale of medicines introduced by Directive 2011/62/EU ("EU Common Logo"). To further identify and characterize the potential location (both physical address and IP address) and owner of websites reviewed, we also cross-referenced data from the Internet Corporation for Assigned Names and Numbers, WHOIS lookup tool, to assess the registrar, registrant name, registrant country, registrant address, IP address, and IP Server for each URL. Content analysis of websites was conducted from May to June 2018.

Controlled Test Buys and Visual Inspection

After completion of website classification, online pharmacies were evaluated for selection in our control test purchase phase using a study inclusion and exclusion criteria protocol (Multimedia Appendix 1). Beginning in July 2018, we began to conduct controlled purchases of targeted antibiotics from online pharmacies that purported to sell at least 1 of the 5 targeted antibiotics. To avoid placing orders from the same owner or group of multiple online pharmacies or affiliated websites, we examined the similarity of the online pharmacy name, content of the website, contact information, and underlining source code of the website (JavaScript) to remove potential duplicates. This generated a smaller sample of websites that comprised our final set of websites for our test purchase process.

Based on the final set of online pharmacies generated, all targeted antibiotics were advertised as sold by pack (eg, 15 pills/pack, 30 pills/pack), with the number of pills in a pack and the price varied. To generate enough samples for the phase 3 analytical quality testing, we purchased at least 20 pills for each product from each selected online pharmacy and set a minimum criterion for the dose of active ingredient (Multimedia Appendix 1). Purchases began in August 2018 and were made using prepaid Visa cash cards with shipment orders set to 2 separate residential mailing addresses on the West Coast of the United States. All order statements, confirmation emails, and transactions made to Visa cash card statements were recorded.

After receiving packages, we conducted visual inspection of packaging for initial risk evaluation. Packages were matched to an online seller by cross-referencing available order information, specific targeted antibiotics purchased, and cash card transaction record with all samples photographed, cataloged, and stored in a secure location. External packaging was then physically inspected for known product falsification risk characteristics and then opened for inspection and confirmation of antibiotics purchased based on information on blister packs (Table 1). We note that all packaging, regardless of country of origin, included English language markings on blister packages and labeling. We then separated each sample package into 2 groups, with the first group kept as a backup and reference sample and the second group used for purposes of analytical quality testing in phase 3 of the study.

Table 1. Risk characteristics of packaging and medicines' visual inspection.

Risk characteristic	Description	Characteristics identified
Packaging characteristics		
Type of package	Type of packaging used in shipments	• Box • Envelop
Package damage	Inspect if there is any damage of the shipping package	• Yes, package is damaged. • No, package is in good condition.
Postal shipping provider	Identification of shipping service or carrier used by the seller	• India Post service • Express Mail Service • Prepaid Germany Postfach service • Singapore Post United States Postal Service
Shipping metadata	Return address and package tracking number	N/A[a]
Item characteristics		
Types of drugs in each package	Detail of the content of each package	• Quantity of drug in each package • Identification and cataloging of drug formulations or names • Identification of any unordered product or free samples in the package
Prescription requirement warning	If "prescription required" warning is on the package (eg, blister pack or pouch) of drug	• Yes, "prescription required" warning on the package • No, no prescription required warning
Type of package (factory packaging)	Types of medical packages for each type of drug	• Blister packs • Pouches
Factory packaging damage	Inspect to determine if any damage to drugs shipped (how many)	• Yes, (number of damaged drugs or pills) • No, no damaged drugs

[a]N/A: not applicable.

Analytical Quality Testing

Targeted antibiotics collected in the controlled test buy phase were then prepared for analytical testing and sent for analysis using untargeted MS through ultra–high performance liquid chromatography–electrospray ionization tandem mass spectrometry (Multimedia Appendix 1). Drug formulations were crushed and extracted in an organic solvent and compared to active pharmaceutical ingredient (API) standards obtained from the US Pharmacopeia. Semiquantitative results—defined as integrated MS signal over time, yielding peak areas that are reflective of amount but cannot be used to determine absolute physical quantity or concentration—were used to determine if the claimed antibiotic API was present in the samples. Molecular networking using the global natural product social molecular networking platform [21] was used to analyze MS/MS data in both positive and negative modes. Molecular networking evaluates and connects MS/MS spectra that are similar, which implies that the molecular structure of the chemicals is similar.

Data from targeted antibiotic samples were collected using an ultra-high performance liquid chromatograph (Vanquish, Thermo) coupled with an Orbitrap mass spectrometer (QExactive, Thermo). A processing method was created in Xcalibur (Thermo) to integrate the values (MS1 data used) of the APIs claimed to be in the drug formulations tested, including samples of sildenafil and tadalafil that were sent to study members unsolicited by vendors. By special library matching supported by de novo annotation of MS/MS, some drug analogs, as well as adulterants and contaminants, were found present in the formulations.

Results

Online Pharmacy Characteristics

We collected 109 unique URLs (from a total of 135 URLs), of which 98 (89.9%) were classified as an online pharmacy, and 21 (19.3%) online pharmacies registered more than one URL. The vast majority (n=85, 78.0%) of these websites were classified as either "rogue" or "unapproved" by Legitscript, and 62 (56.9%) were on the NABP Not Recommended List. Additionally, only 19 (17.4%) had the EU common logo (Table 2). Further, based on available WHOIS data, 20 (18.3%) used commercially available domain masking and enhanced privacy services to hide their location and ownership. Among websites with available location data, the top 5 registrant countries were the United States, Russia, Barbados, Canada, and the United Kingdom, though registered locations included a broader set of countries that covered North America, South America, Europe, and Asia. After applying our inclusion and exclusion criteria

for selection for test purchasing, 27 online pharmacies were selected for controlled test buys.

Table 2. Online pharmacy verification summary (N=109).

Variable	Value, n (%)
LegitScript	
Legitimate	2 (1.8)
Certified	3 (2.8)
No information	19 (17.4)
Rogue	69 (63.3)
Unapproved	16 (14.7)
EU[a] common logo	
Not verified	90 (82.6)
Verified	19 (17.4)
NABP[b]	
Verified	2 (1.8)
Information is not available	26 (23.9)
Not recommended list	62 (56.9)

[a]EU: European Union.

[b]NABP: National Association of Boards of Pharmacy.

Controlled Buy Results and Packaging Analysis

Orders were placed with 27 online pharmacies and resulted in 1373 antibiotic product samples collected. This study defines product samples as a single pill or capsule that was collected from packaging sent to the research team, with packaging primarily consisting of blister packages with different numbers of pills or capsules. During the process of ordering, we received phone calls, as well as emails from vendors to confirm order details with phone calls not taken, but study team members responded to emails from vendors by simply confirming order details. Some online pharmacies included emails requesting the customer to verify transactions with the credit card–issuing bank, including emails with suspicious verification requests (eg, “If Support Bank can contact you for verification payment, please do not tell about buying medicines from XXX Pharmacy. You can tell them that you Paid for ‘FAMILY PHOTOS CONVERTED TO CD and FLASH DRIVE, or WEBSITE DESIGN ETC”).

Among all 27 orders, only 13 (48.1%) were successfully completed through the website’s online ordering process, which resulted in a confirmed transaction. However, 2 (7.4%) of these transactions resulted in fraud (eg, fraudulent purchases for other e-commerce transactions were made on Visa cash card information provided to online pharmacies), and we did not receive products from these vendors. From the remaining successfully completed transactions, 11 packages were received, and based on shipping labeling and records, 10 (91%) were identified as shipped from India and 1 (9%) from Singapore with all but 1 (9%) package shipped in a mailing envelope (ie, 1 package was delivered in a small box). In addition to the targeted antibiotics ordered, 2 (18.2%) packages included other unsolicited prescription drug products typically used or indicated for erectile dysfunction (eg, sidenafil citrate).

Packages and samples were inspected for information on the name of the purported manufacturer, product warnings (eg, only dispensed with “Rx”), and any certification or logo indicating the authenticity of the product. Two envelopes had visible damage when received and contained damaged capsules when opened. Based on further packaging analysis, we found that among antibiotics, there was a high degree of variability in product characteristics (eg, color, shape, blister pack, etc), even among antibiotics purportedly from the same manufacturer. For example, in 2 packages received, both co-amoxicillin 625 mg samples are labeled as “Manufacturer X”; however, these 2 samples are packed and presented differently (Figure 2A). Some of the packages also included warning labels stating, “to be sold by retail on the prescription of a Registered Medical Practitioner only,” despite being purchased from sources that did not require a prescription. Finally, 2 blister packs also included a logo stating “WHO GMP certified company” (Figure 2B), though we were unable to confirm the validity of this certification or its origin (“WHO-GMP” certification could indicate certification from the Food and Drug Administration Maharashtra in India). Other products, as previously mentioned, arrived in damaged blister packs (Figure 2C).

Figure 2. Image of risk characteristics identified from drug packages. GMP: Good Manufacturing Practices; WHO: World Health Organization.

Analytical Testing of Samples

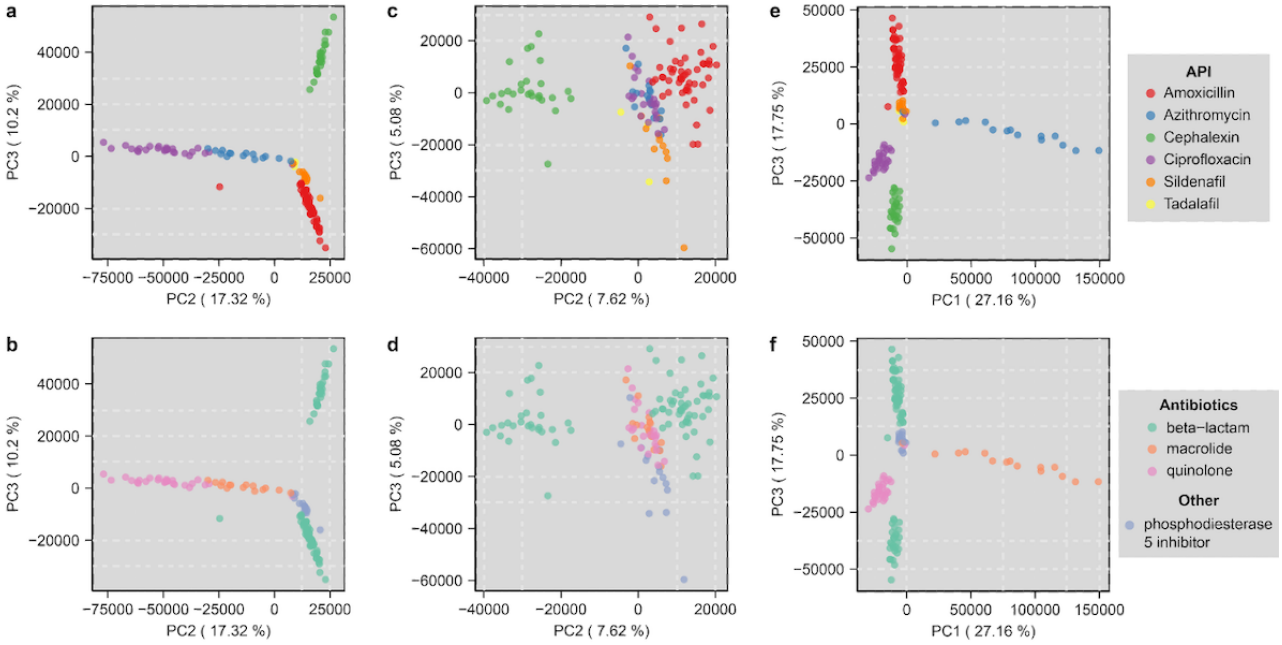
There was a total of 45 unique boxes or blister packs of drugs (from 12 manufacturers), from which 3 pills per package were analyzed through random selection, equating to 135 samples that underwent analytical testing. All targeted antibiotics purchased from internet pharmacies stated the name of the drug, stated on the label that it contained API, and, based on analytical testing, contained the stated API. However, certain samples had specific risk characteristics of interest, including chemicals that are presumed contaminants, drug-related compounds, and undeclared API.

Molecular networks via global natural product social molecular networking were used to explore differences in the untargeted MS data that might not have been listed in the formulations. Molecular networking and subsequent interpretation of potential risk factors revealed the presence of a feature annotated as octabenzone, a chemical used in sun-protection products in at least one sample from a manufacturer (sildenafil API listed) and 4 samples from another manufacturer (amoxicillin API listed; see Multimedia Appendix 1 for additional details). Dodecyl sulfate, a common surfactant, as well as tetradecyl sulfate were detected in 4 different manufacturers, which included cephalexin, ciprofloxacin, amoxicillin, and tadalafil as listed APIs, respectively. Among other chemicals observed in the molecular network, which we believe are contaminants, were flame retardants (triphenylphosphate), wetting and dispersing agents (tetramethyl-5-decyne-4,7-diol), and

plasticizers. Regarding, drug-like chemicals, we observed a few chemicals related to the claimed API (connected in the molecular network to API) such as descladinose azithromycin as well as other unknown, unannotated chemicals. Finally, the sildenafil sample also contained dapoxetine (generally used as an API for the treatment of premature ejaculation), an undeclared API.

Regarding variance across samples, untargeted MS analysis of the drug formulations using unsupervised multivariate statistics, specifically principal component analysis (PCA), was used to analyze the positive and negative mode data; in doing so, we compared the global chemical profiles between the formulations tested (Figure 3 and Multimedia Appendix 1). In the positive mode of PCA, separation of drug formulations was observed with clear grouping of samples based on the API observed in figure panels labelled principal component 2 ("PC2") compared to principal component 3 ("PC3"), indicating consistency in certain manufacturer sample formulations and inconsistency in others (Figure 3A). PCA of negative mode data resulted in clear groups of amoxicillin and cephalexin, but less clear grouping of the other drug formulations, though overall, the chemical differences between samples were lesser in the negative mode compared to the positive ionization mode. Further analysis of the individual drug package and manufacturer compared to the overall separation observed by PCA by testing replicate samples indicated variances in chemical similarity of samples (Multimedia Appendix 1). Variance may result from differences in the quantitative amount of API (not fully evaluated in this study) or differences in the excipient and other chemicals present in the formulations tested.

Figure 3. Untargeted mass spectrometric analysis of drugs formulations in positive and negative mode analyzed by principal component analysis (PCA). (a) PCA score plot of positive mode data, pareto scaled, displaying each sample as a point colored by active pharmaceutical ingredient (API) and (b) drug class; (c) PCA score plot of negative mode data, pareto scaled, colored by API and (d) drug class; (e) mid-level data fusion of positive and negative mode data displaying each sample as a point colored by API and (f) drug class.



Discussion

Principal Findings

Our study detected 109 unique websites actively advertising the sale of common antibiotics without a prescription, resulting in 27 online controlled test buy orders, 11 packages, and a total of 1373 antibiotic product samples that were evaluated using a combination of visual inspection, analytical chemistry, and molecular networking. In our sample of websites selling antibiotics, 57 masked their location or owner address with a privacy service, and of those ordered from, all the websites requested additional verification information in order to effectuate the processing of payment.

These characteristics potentially implicate risk factors for consumer and product safety associated with online drug purchasing and are generally considered characteristics of high-risk transactions that can result in identity theft or fraudulent transactions. Legitimate mail-order pharmacies do not exhibit these characteristics, and these risks are further reinforced by the lack of third-party validation for the majority of websites, as we found out after cross-referencing them with Legitscript, NABP, and MHRA verification. In fact, during the course of this study, prepaid cards used for test purchases were fraudulently charged (eg, fraudulent charges for food orders placed on prepaid cards without permission of the study team), and we were unable to recover these stolen funds.

Products shipped and received to us in the United States all originated from overseas, with almost all product samples shipped from India. As the focus of this study was from the context of a US patient or consumer, these antibiotics represent products outside of the controlled drug supply chain and have a higher risk of being adulterated, falsified, or otherwise harmful to human health. Almost all products were shipped in mailing

envelopes, which did not have adequate protection to secure the product (some products appeared to be damaged in transit), and declared goods on labels often did not include "medicines" as a description, but instead were named as other unrelated consumer goods, likely in an attempt to evade customs inspection. These characteristics are clear warning signs of unauthorized and potentially falsified medications.

Finally, upon completion of analytical testing of samples using MS, all products were determined to have stated API, though this study did not quantitatively measure the percentage of API in each sample tested. We performed unsupervised multivariate statistics (eg, PCA) on the formulations, which indicated that particular manufacturers (as indicated on the packaging) were more precise in their formulation's chemistry, whereas others displayed wider variability across replicate samples tested, possibly reflecting poor or inconsistent manufacturing practices or quality. The use of molecular networking also identified other impurities present in samples that evidence a further potential for adulteration, which may introduce unique patient safety harms.

Prior studies examining the quality of prescription drugs generally have findings along the following lines: they primarily focus on field-based sampling using different prevalence surveys and analytical techniques to test samples of antibiotics sold in physical establishments, primarily in low- to middle-income countries; they use packaging analysis or analytical techniques to test nonantibiotic drugs or dietary supplements (eg, erectile dysfunction drugs, growth hormone, diabetes drugs, stimulants, dietary supplements, etc) purchased from the internet; or they simply characterize different online sellers of antibiotics but do not purchase or test the products [13,22-31]. Our study builds on these prior studies by conducting a multifactor quality and safety analysis to generate new data points regarding the potential health and safety risks associated with the online sale of common antibiotics from no prescription providers that should be further confirmed through additional sampling and product testing. Importantly, many of the potential safety concerns identified are important in the context of broader public health and regulatory challenges aimed at addressing the global trade in substandard and falsified medicines, drug importation policy, ensuring supply chain integrity, and modernizing postmarket surveillance and pharmacovigilance approaches [32].

Limitations

This study also has certain limitations given the methodology used. First, web-based search queries for marketing and the availability of drugs have certain limitations. Primarily, searches in this study were conducted at a limited time range, with an assessment of online availability that was also limited to the study period described. However, websites are created, modified, and taken down dynamically on the internet, hence limiting the generalizability of our results. Additionally, we cannot be certain if the stated manufacturer on the label or blister pack of samples was in fact the manufacturer of that product, as we did not contact manufacturers to confirm the medication lot number or authenticity. We also could not fully ascertain if visual inspection of packaging or sample quality was degraded, damaged, or underwent other spoilage due to shipping or storage issues prior to the product being received by the study team and independent of the online pharmacy, which could have impacted quality testing and external validity of results.

Finally, though this study focused on risks associated with importation of products from "no prescription" online pharmacies, the specific risk characteristics associated with antimicrobial drugs identified in this study may also be associated with different manufacturing standards or be indicative of a failure to manufacture drugs according to US standards from sources that originate outside of the United States. Future studies should incorporate additional control or comparison groups of online pharmacies (eg, foreign online pharmacies that require a valid prescription) to better understand factors associated with the identified risk characteristics.

Conclusions

The use of infoveillance approaches, such as using structured web-based search queries, connecting results to "secret shopper" and online test buys, and evaluating sellers and products for risk characteristics, has the potential to address other online health challenges that may implicate illegal actors, such as the illegal sale of other prescription drugs, controlled substances, and illicit drugs, and even COVID-19–fraudulent products [11,33-35]. Critical to these efforts will be consensus building and the development of internationally agreed-upon standards and methodologies for initial risk evaluation when purchasing drugs online [36]. Comprehensive and specific risk assessments tailored for online sellers and even different drug classes can be developed based upon existing instruments, such as the European Directorate for the Quality of Medicines and the Asia Pacific Economic Cooperation Supply Chain Security toolkit for internet sales; weighted criteria for counterfeiting risk assessment, as suggested by Vida et al [36], can also be beneficial. The results from this study can also form the basis for future risk-based multimodal surveillance approaches and product quality assessment methodologies that can be scaled to larger data collection to establish more generalizable findings. For example, future studies should consider testing for the WHO Model Essential Medicine List "WATCH" and "RESERVE" antibiotics given their potential for higher microbial resistance and public health impact, should they be counterfeited.

Acknowledgments

The authors would like to thank the US Pharmacopeia (USP) Academic Connection program for providing access to the USP National Formulary (USP-NF) and USP for Reference Standards for the drugs tested. The authors also thank the Alliance for Safe Online Pharmacies (ASOP) foundation for its financial support through a research grant award for this project.

Data Availability

Study data are available upon reasonable request to the authors. Mass spectrometry data are available on the internet [37].

Conflicts of Interest

TKM and QX are employees of the startup company S-3 Research LLC. S-3 Research is funded and currently supported by the National Institutes of Health–National Institute on Drug Abuse through a Small Business Innovation and Research contract for opioid-related social media research and technology commercialization. PCD is an advisor to Cybele and cofounder of Ometa and Enveda with prior approval by University of California, San Diego. TKM is Editor-in-Chief of JMIR Infodemiology. AKJ, KS, AL, SA, JL, and SB declare no conflicts of interest.

Multimedia Appendix 1
Supplementary file.
[PDF File (Adobe PDF File), 6227 KB - publichealth_v8i12e41834_app1.pdf]

References

1. Jee Y, Carlson J, Rafai E, Musonda K, Huong TTG, Daza P, et al. Antimicrobial resistance: a threat to global health. Lancet Infect Dis 2018 Sep;18(9):939-940. [doi: 10.1016/S1473-3099(18)30471-7] [Medline: 30152350]
2. Antimicrobial resistance. World Health Organization. URL: https://www.who.int/health-topics/antimicrobial-resistance [accessed 2022-12-09]
3. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 2022 Feb 12;399(10325):629-655 [FREE Full text] [doi: 10.1016/S0140-6736(21)02724-0] [Medline: 35065702]
4. Prestinaci F, Pezzotti P, Pantosti A. Antimicrobial resistance: a global multifaceted phenomenon. Pathog Glob Health 2015;109(7):309-318 [FREE Full text] [doi: 10.1179/2047773215Y.0000000030] [Medline: 26343252]
5. Kuang L, Liu Y, Wei W, Song X, Li X, Liu Q, et al. Non-prescription sale of antibiotics and service quality in community pharmacies in Guangzhou, China: A simulated client method. PLoS One 2020;15(12):e0243555 [FREE Full text] [doi: 10.1371/journal.pone.0243555] [Medline: 33301450]
6. Boyd SE, Moore LSP, Gilchrist M, Costelloe C, Castro-Sánchez E, Franklin BD, et al. Obtaining antibiotics online from within the UK: a cross-sectional study. J Antimicrob Chemother 2017 May 01;72(5):1521-1528 [FREE Full text] [doi: 10.1093/jac/dkx003] [Medline: 28333179]
7. Gong Y, Jiang N, Chen Z, Wang J, Zhang J, Feng J, et al. Over-the-counter antibiotic sales in community and online pharmacies, China. Bull World Health Organ 2020 Jul 01;98(7):449-457 [FREE Full text] [doi: 10.2471/BLT.19.242370] [Medline: 32742030]
8. Orizio G, Merla A, Schulz PJ, Gelatti U. Quality of online pharmacies and websites selling prescription drugs: a systematic review. J Med Internet Res 2011 Sep 30;13(3):e74 [FREE Full text] [doi: 10.2196/jmir.1795] [Medline: 21965220]
9. Ivanitskaya L, Brookins-Fisher J, O Boyle I, Vibbert D, Erofeev D, Fulton L. Dirt cheap and without prescription: how susceptible are young US consumers to purchasing drugs from rogue internet pharmacies? J Med Internet Res 2010 Apr 26;12(2):e11 [FREE Full text] [doi: 10.2196/jmir.1520] [Medline: 20439253]
10. Fittler A, Bősze G, Botz L. Evaluating aspects of online medication safety in long-term follow-up of 136 Internet pharmacies: illegal rogue online pharmacies flourish and are long-lived. J Med Internet Res 2013 Sep 10;15(9):e199 [FREE Full text] [doi: 10.2196/jmir.2606] [Medline: 24021777]
11. Ozawa S, Billings J, Sun Y, Yu S, Penley B. COVID-19 treatments sold online without prescription requirements in the United States: Cross-sectional study evaluating availability, safety and marketing of medications. J Med Internet Res 2022 Feb 16;24(2):e27704 [FREE Full text] [doi: 10.2196/27704] [Medline: 34662286]
12. Ozawa S, Evans DR, Bessias S, Haynie DG, Yemeke TT, Laing SK, et al. Prevalence and Estimated Economic Burden of Substandard and Falsified Medicines in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. JAMA Netw Open 2018 Aug 03;1(4):e181662 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1662] [Medline: 30646106]
13. Ahmed J, Modica de Mohac L, Mackey TK, Raimi-Abraham BT. A critical review on the availability of substandard and falsified medicines online: Incidence, challenges and perspectives. J Med Access 2022 Feb 15;6:23992026221074548 [FREE Full text] [doi: 10.1177/23992026221074548] [Medline: 36204527]
14. Mackey TK, Liang BA, York P, Kubic T. Counterfeit drug penetration into global legitimate medicine supply chains: a global assessment. Am J Trop Med Hyg 2015 Jun;92(6 Suppl):59-67 [FREE Full text] [doi: 10.4269/ajtmh.14-0389] [Medline: 25897059]
15. Sun N, Gong Y, Liu J, Wu J, An R, Dong Y, et al. Prevalence of antibiotic purchase online and associated factors among Chinese residents: A nationwide community survey of 2019. Front Pharmacol 2021;12:761086 [FREE Full text] [doi: 10.3389/fphar.2021.761086] [Medline: 34803704]

16. Mainous AG, Everett CJ, Post RE, Diaz VA, Hueston WJ. Availability of antibiotics for purchase without a prescription on the internet. Ann Fam Med 2009;7(5):431-435 [FREE Full text] [doi: 10.1370/afm.999] [Medline: 19752471]
17. Outpatient antibiotic prescriptions — United States, 2014. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/antibiotic-use/community/pdfs/annual-reportsummary_2014.pdf [accessed 2022-12-09]
18. Mackey TK, Nayyar G. Digital danger: a review of the global public health, patient safety and cybersecurity threats posed by illicit online pharmacies. Br Med Bull 2016 Jun;118(1):110-126 [FREE Full text] [doi: 10.1093/bmb/ldw016] [Medline: 27151957]
19. Chavez J, Shah NA, Ruoss S, Cuomo RE, Ward SR, Mackey TK. Online marketing practices of regenerative medicine clinics in US-Mexico border region: a web surveillance study. Stem Cell Res Ther 2021 Mar 18;12(1):189 [FREE Full text] [doi: 10.1186/s13287-021-02254-4] [Medline: 33736697]
20. Mackey TK, Aung P, Liang BA. Illicit Internet availability of drugs subject to recall and patient safety consequences. Int J Clin Pharm 2015 Dec;37(6):1076-1085. [doi: 10.1007/s11096-015-0154-8] [Medline: 26148859]
21. GNPS. URL: https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp [accessed 2022-12-08]
22. Zabala GA, Bellingham K, Vidhamaly V, Boupha P, Boutsamay K, Newton PN, et al. Substandard and falsified antibiotics: neglected drivers of antimicrobial resistance? BMJ Glob Health 2022 Aug;7(8):e008587 [FREE Full text] [doi: 10.1136/bmjgh-2022-008587] [Medline: 35981806]
23. Veronin M. Packaging and labeling of pharmaceutical products obtained from the internet. J Med Internet Res 2011 Feb 15;13(1):e22 [FREE Full text] [doi: 10.2196/jmir.1441] [Medline: 21324833]
24. Gelatti U, Pedrazzani R, Marcantoni C, Mascaretti S, Repice C, Filippucci L, et al. 'You've got m@il: fluoxetine coming soon!': accessibility and quality of a prescription drug sold on the web. Int J Drug Policy 2013 Sep;24(5):392-401. [doi: 10.1016/j.drugpo.2013.01.006] [Medline: 23433969]
25. Vida RG, Fittler A, Mikulka I, Ábrahám E, Sándor V, Kilár F, et al. Availability and quality of illegitimate somatropin products obtained from the Internet. Int J Clin Pharm 2017 Feb;39(1):78-87. [doi: 10.1007/s11096-016-0398-y] [Medline: 27888454]
26. van der Gouwe D, Brunt TM, van Laar M, van der Pol P. Purity, adulteration and price of drugs bought on-line versus off-line in the Netherlands. Addiction 2017 Apr;112(4):640-648. [doi: 10.1111/add.13720] [Medline: 27936283]
27. Rahman MS, Yoshida N, Sugiura S, Tsuboi H, Keila T, Kiet HB, et al. Quality of omeprazole purchased via the internet and personally imported into Japan: comparison with products sampled in other Asian countries. Trop Med Int Health 2018 Mar;23(3):263-269 [FREE Full text] [doi: 10.1111/tmi.13028] [Medline: 29314458]
28. Zhu S, Yoshida N, Kimura K, Matsushita R, Tsuboi H. Falsified vardenafil tablets available online. J Pharm Biomed Anal 2020 Jan 05;177:112872. [doi: 10.1016/j.jpba.2019.112872] [Medline: 31525574]
29. Sanada T, Yoshida N, Matsushita R, Kimura K, Tsuboi H. Falsified tadalafil tablets distributed in Japan via the internet. Forensic Sci Int 2020 Feb;307:110143 [FREE Full text] [doi: 10.1016/j.forsciint.2020.110143] [Medline: 31931435]
30. Zhu S, Yoshida N, Tsuboi H, Matsushita R, Kimura K. Quality and Authenticity of Metformin Tablets Circulating on Japanese Websites. Ther Innov Regul Sci 2021 Jul;55(4):656-666 [FREE Full text] [doi: 10.1007/s43441-021-00262-3] [Medline: 33595786]
31. Long CS, Kumaran H, Goh KW, Bakrin FS, Ming LC, Rehman IU, et al. Online Pharmacies Selling Prescription Drugs: Systematic Review. Pharmacy (Basel) 2022 Apr 01;10(2):42 [FREE Full text] [doi: 10.3390/pharmacy10020042] [Medline: 35448701]
32. Fadlallah R, El-Jardali F, Annan F, Azzam H, Akl EA. Strategies and systems-level interventions to combat or prevent drug counterfeiting: A systematic review of evidence beyond effectiveness. Pharmaceut Med 2016;30(5):263-276 [FREE Full text] [doi: 10.1007/s40290-016-0156-4] [Medline: 27752221]
33. Mackey TK, Kalyanam J, Katsuki T, Lanckriet G. Twitter-based detection of illegal online sale of prescription opioid. Am J Public Health 2017 Dec;107(12):1910-1915. [doi: 10.2105/AJPH.2017.303994] [Medline: 29048960]
34. Kalyanam J, Mackey TK. A review of digital surveillance methods and approaches to combat prescription drug abuse. Curr Addict Rep 2017 Sep 18;4(4):397-409. [doi: 10.1007/s40429-017-0169-4]
35. Mackey TK, Li J, Purushothaman V, Nali M, Shah N, Bardier C, et al. Big data, natural language processing, and deep learning to detect and characterize illicit COVID-19 product sales: Infoveillance study on Twitter and Instagram. JMIR Public Health Surveill 2020 Aug 25;6(3):e20794 [FREE Full text] [doi: 10.2196/20794] [Medline: 32750006]
36. Vida RG, Merczel S, Jáhn E, Fittler A. Developing a framework regarding a complex risk based methodology in the evaluation of hazards associated with medicinal products sourced via the internet. Saudi Pharm J 2020 Dec;28(12):1733-1742 [FREE Full text] [doi: 10.1016/j.jsps.2020.10.018] [Medline: 33424264]
37. MassIVE MSV000083605. Center for Computational Mass Spectrometry. URL: https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=340238f1e5924651851a16a04fc685ae [accessed 2022-12-08]

Abbreviations

AMR: antimicrobial resistance
API: active pharmaceutical ingredient

EU: European Union
MS: mass spectrometry
NABP: National Association of Boards of Pharmacy
PCA: principal component analysis
WHO: World Health Organization

Edited by A Mavragani; submitted 11.08.22; peer-reviewed by A Fittler, K Adapa, J Ma; comments to author 02.09.22; revised version received 26.10.22; accepted 02.11.22; published 23.12.22.

Please cite as:
Mackey TK, Jarmusch AK, Xu Q, Sun K, Lu A, Aguirre S, Lim J, Bhakta S, Dorrestein PC
Multifactor Quality and Safety Analysis of Antimicrobial Drugs Sold by Online Pharmacies That Do Not Require a Prescription: Multiphase Observational, Content Analysis, and Product Evaluation Study
JMIR Public Health Surveill 2022;8(12):e41834
URL: https://publichealth.jmir.org/2022/12/e41834
doi:10.2196/41834
PMID:36563038

Original Paper

The Prognostic and Predictive Effects of Human Papillomavirus Status in Hypopharyngeal Carcinoma: Population-Based Study

Shi-Ping Yang[1*], MD; Xiang-Ying Lin[1*], MD; Min Hu[2*], MD; Cheng-Fu Cai[3], MD

[1]Department of Radiation Oncology, Hainan General Hospital (Hainan Affiliated Hospital of Hainan Medical University), Haikou, China

[2]School of Medicine, Xiamen University, Xiamen, China

[3]Department of Otorhinolaryngology Head and Neck Surgery, Zhongshan Hospital, School of Medicine, Xiamen University, Xiamen, China

[*]these authors contributed equally

Corresponding Author:
Cheng-Fu Cai, MD
Department of Otorhinolaryngology Head and Neck Surgery
Zhongshan Hospital, School of Medicine
Xiamen University
201-209, South Hubin Road
Xiamen, 361004
China
Phone: 86 5922292201
Email: caichengfuxmu@126.com

Abstract

Background: The role of the Human Papillomavirus (HPV) status in patients with hypopharyngeal squamous cell carcinoma (HSCC) remains controversial.

Objective: Our aim was to determine the prognostic and predictive effects of HPV status in patients with locally advanced HSCC (stage III-IVB) receiving primary radiotherapy.

Methods: Patients diagnosed with stage III-IVB HSCC between 2010 and 2016 were identified. HPV status, demographics, clinicopathological characteristics, treatment, and survival data were captured. Kaplan-Meier analysis, multivariable Cox regression analysis, and propensity score matching analysis were performed.

Results: We identified 531 patients in this study and 142 (26.7%) patients with HPV-positive diseases. No significant differences were observed between those with HPV-negative and HPV-positive diseases with regard to demographics, clinicopathological characteristics, and chemotherapy use. HPV-positive HSCC had better head and neck cancer-specific survival (HNCSS; P=.001) and overall survival (OS; P<.001) compared to those with HPV-negative tumors. Similar results were found using the multivariable Cox regression analysis. Sensitivity analyses showed that the receipt of chemotherapy was associated with significantly improving HNCSS (P<.001) and OS (P<.001) compared to not receiving chemotherapy in HPV-negative HSCC, whereas comparable HNCSS (P=.59) and OS (P=.12) were found between both treatment arms in HPV-positive HSCC. Similar results were found after propensity score matching.

Conclusions: Approximately one-quarter of HSCC may be HPV-related, and HPV-positive HSCC is associated with improved survival outcomes. Furthermore, additional chemotherapy appears to be not related to a survival benefit in patients with HPV-positive tumors who received primary radiotherapy.

(JMIR Public Health Surveill 2022;8(12):e40185) doi:10.2196/40185

KEYWORDS

hypopharyngeal carcinoma; human papillomavirus; HPV; chemotherapy; radiotherapy; prognosis; cancer; carcinoma

Introduction

Head and neck squamous cell carcinoma (HNSCC) accounts for more than 90% of all head and neck malignancies [1]. However, hypopharyngeal squamous cell carcinoma (HSCC) is relatively rare overall, accounting for approximately 3% of all HNSCCs [2]. Symptoms and signs of HSCC tend to be minimized by patients until severe distress occurs or an obvious mass is found in the neck. Because of this, most patients are diagnosed with locally advanced HNSCC (stage III or IV) [3],

with a 5-year overall survival (OS) of 30%, which has the worst prognosis compared to other sites of HNSCC [2].

Human papillomavirus (HPV) is an important factor in the carcinogenesis of HNSCC, especially in oropharyngeal squamous cell carcinoma (OSCC) [4]. The existing evidence has shown a better prognosis in patients with HPV-positive OSCC. Moreover, HPV status has also impacted the treatment decision-making of OSCC [5,6]. Several studies with a small sample size suggested that the HPV infection rate in HSCC was relatively low (1.6%-8.5%) [7-9]. However, two recent population-based studies indicated that 17.7%-23.9% of patients with HSCC had HPV-positive diseases [10,11]. This raises our question of whether HPV status has a prognostic effect on HSCC. Currently, contradictory results were observed in several previous studies regarding the prognosis of HPV status in HSCC, with some studies indicating HPV-related HSCC with significantly improved survival outcomes, whereas others have found similar survival rates between HPV-negative and HPV-positive diseases [7,10-15]. In addition, there are limited studies assessing the role of HPV status in patients who received primary radiotherapy or chemoradiotherapy regarding the HSCC [12]. In light of this, we conducted this analysis from the Surveillance, Epidemiology, and End Results (SEER) database to investigate the prognostic and predictive effect of HPV status in stage III-IVB HSCC receiving primary radiotherapy.

Methods

Patients Selection Criteria

The data for this study were captured from the SEER database [16], in which HPV status was categorized as either HPV-negative, HPV-positive, or unknown status. HPV status was determined by p16-immunohistochemistry, in situ hybridization, or polymerase chain reaction methods of pathologic specimens from either the primary hypopharyngeal tumors or metastatic cervical lymph nodes. The HPV data set was reviewed by the SEER data quality team to ensure the accuracy of HPV testing status [17].

We queried the SEER public database from 2010 to 2016 for patients diagnosed with stage III-IVB HSCC who received primary radiotherapy with or without chemotherapy. We did not include patients diagnosed before 2010 because HPV status was only added as a SEER variable in 2010. Those with unknown HPV status were excluded. In addition, we excluded patients who were treated with non-beam irradiation, including radioactive implants and radioisotopes.

Ethical Considerations

This study was approved by the ethics committee of the Hainan General Hospital (No. ZDYF2022SHFZ130).

Data Collection

The following demographics, clinicopathological characteristics, or treatment data were identified from each patient's medical record: age, gender, race, grade, tumor location, American Joint Committee on Cancer (AJCC) staging, HPV status, chemotherapy use, insurance status, and marital status. AJCC 7th staging system was used to determine the stage of patients. The primary outcome endpoints were head and neck cancer-specific survival (HNCSS) and OS. HNCSS was defined as the time from the initial diagnosis of HCSS till death due to head and neck cancer. OS was defined as the time from the initial diagnosis of HCSS till death from all causes.

Statistical Analyses

The difference in patients' characteristics and treatment data were compared using the chi-square test or Fisher exact test. HNCSS and OS curves were estimated using the Kaplan-Meier methods and compared by the log-rank test. A 1:1 propensity score matching (PSM) was conducted to balance the potential confounders. Multivariable Cox regression models were used to investigate whether HPV-positive HSCC was related to better HNCSS and OS. Sensitivity analyses were used to investigate the effect of chemotherapy on survival according to HPV status. Data analyses were conducted using SPSS statistical software (version 22.0; IBM Corp). $P<.05$ was considered to be statistically significant.

Results

Baseline Characteristics

We identified 531 patients in this study (Figure 1), including 389 (73.3%) patients with HPV-negative diseases and 142 (26.7%) patients who had HPV-positive HSCC. Table 1 lists the baseline characteristics of the study cohort. A total of 445 (83.8%) patients were male; 404 (76.1%) patients had stage IVA-IVB disease; and 466 (87.8%) patients received chemotherapy. In patients with tumor location available (n=381), 74% (n=282) of the tumor was located in pyriform sinus. No significant difference was found between HPV-negative and HPV-positive diseases with regard to age, gender, race, AJCC staging, tumor grade, tumor location, chemotherapy use, insurance status, and marital status. Patients of younger age ($P=.002$) and Non-Hispanic White patients ($P=.02$) were more likely to receive chemotherapy (Table S1 in Multimedia Appendix 1).

Figure 1. Patient selection procedure in the Surveillance, Epidemiology, and End Results (SEER) database regarding human papillomavirus (HPV) status in head and neck cancers.

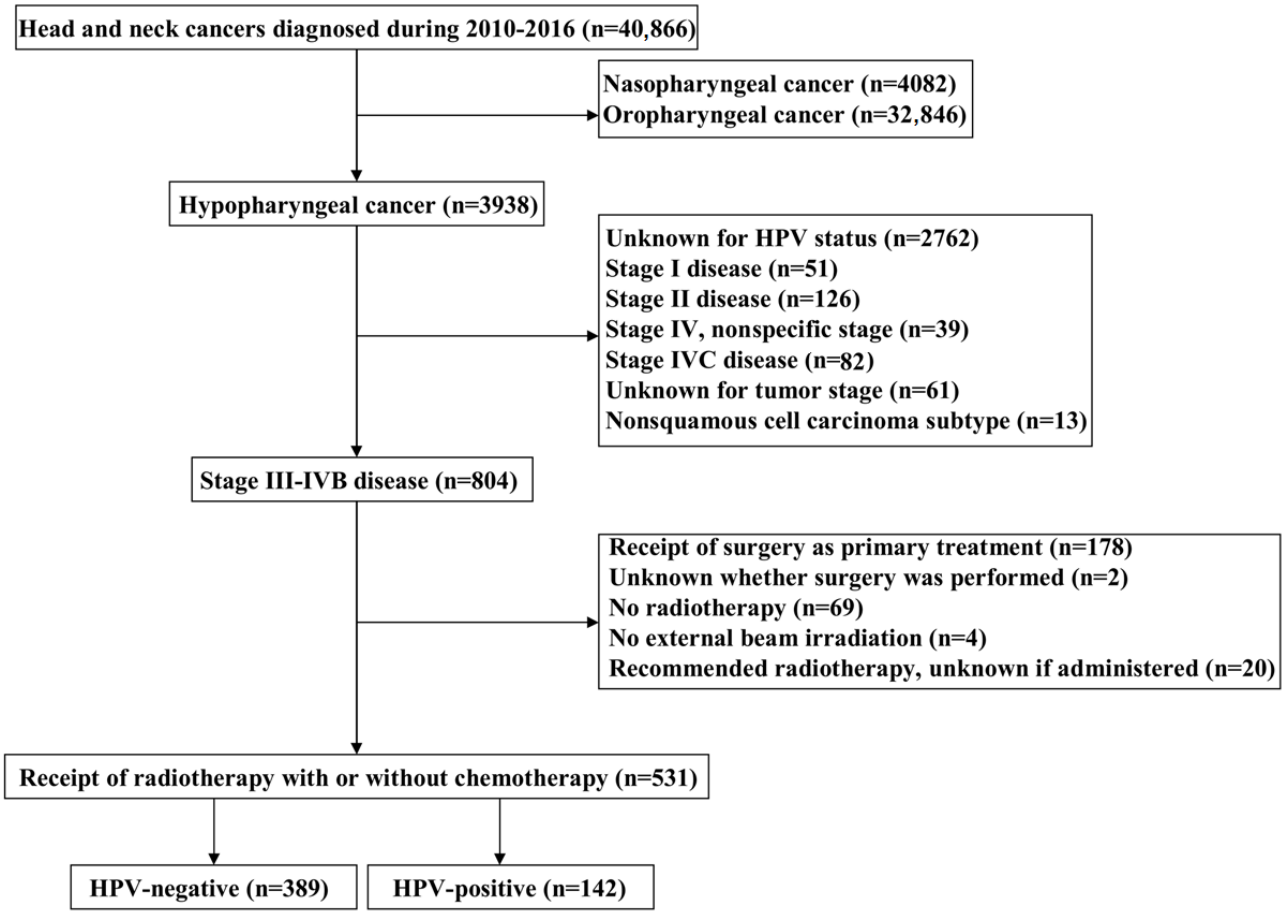


Table 1. Patients' baseline characteristics according to human papillomavirus (HPV) status.

Variables	Patients, n	HPV-negative (n=389), n (%)	HPV-positive (n=142), n (%)	*P* value
Age (years), n (%)				.73
<50	40	28 (7.2)	12 (8.5)	
50-64	242	181 (46.5)	61 (43)	
>64	249	180 (46.3)	69 (48.6)	
Gender, n (%)				.79
Male	445	325 (83.5)	120 (84.5)	
Female	86	64 (16.5)	22 (15.5)	
Race, n (%)				.06
Non-Hispanic White	379	267 (68.6)	112 (78.9)	
Non-Hispanic Black	65	54 (13.9)	11 (7.7)	
Hispanic (all)	47	34 (8.7)	13 (9.2)	
Other	40	34 (8.7)	6 (4.2)	
Grade, n (%)				.19
Well differentiated	20	17 (5.7)	3 (2.9)	
Moderately differentiated	201	155 (52)	46 (45.1)	
Poorly differentiated or undifferentiated	179	126 (42.3)	53 (52)	
Unknown	131	__[a]	—	
Tumor location, n (%)				.18
Pyriform sinus	282	205 (71.7)	77 (81.1)	
Aryepiglottic fold	40	30 (10.5)	10 (10.5)	
Postcricoid region	12	11 (3.8)	1 (1.1)	
Posterior wall	47	40 (14)	7 (7.4)	
Unknown	150	—	—	
AJCC[b] stage, n (%)				.85
III	127	95 (24.4)	32 (22.5)	
IVA	335	245 (63)	90 (63.4)	
IVB	69	49 (12.6)	20 (14.1)	
Chemotherapy, n (%)				.91
No	65	48 (12.3)	17 (12)	
Yes	466	341 (87.7)	125 (88)	
Insurance status, n (%)				.05
Insured	22	20 (5.2)	2 (1.4)	
Uninsured	503	363 (94.8)	140 (98.6)	
Unknown	6	—	—	
Marital status, n (%)				.32
Married	253	180 (48.8)	73 (54.1)	
Divorced	86	68 (18.4)	18 (13.3)	
Single	128	91 (24.7)	37 (27.4)	
Widowed	37	30 (8.1)	7 (5.2)	
Unknown	27	—	—	

[a]Not available.

[b]AJCC: American Joint Committee on Cancer.

Survival Outcomes and Prognostic Analyses Stratified by HPV Status

With a median follow-up of 16 (range 0-82) months, a total of 205 deaths were observed, including 152 patients who died of head and neck cancer-related disease. Using Kaplan-Meier survival estimates, HPV-positive patients had better survival outcomes compared to HPV-negative patients. The 3-year HNCSS in HPV-negative and HPV-positive patients was 55.3% and 74.9% (P=.001), respectively (Figure 2A). The 3-year OS in HPV-negative and HPV-positive patients was 44.5% and 70% (P<.001), respectively (Figure 2B).

Table 2 lists the results of multivariate Cox regression analyses. The results indicated that patients with positive HPV had significantly better HNCSS (hazard ratio [HR]: 0.460; P<.001) and OS (HR: 0.422; P<.001) compared to patients with negative HPV. In addition, patients who received chemotherapy had better HNCSS (HR: 0.405; P<.001) and OS (HR: 0.405; P<.001) than those without chemotherapy. Moreover, age, race, tumor location, AJCC stage, and marital status were also risk factors independently associated with HNCSS or OS.

Figure 2. Survival curve in hypopharyngeal squamous cell carcinoma according to human papillomavirus (HPV) status.

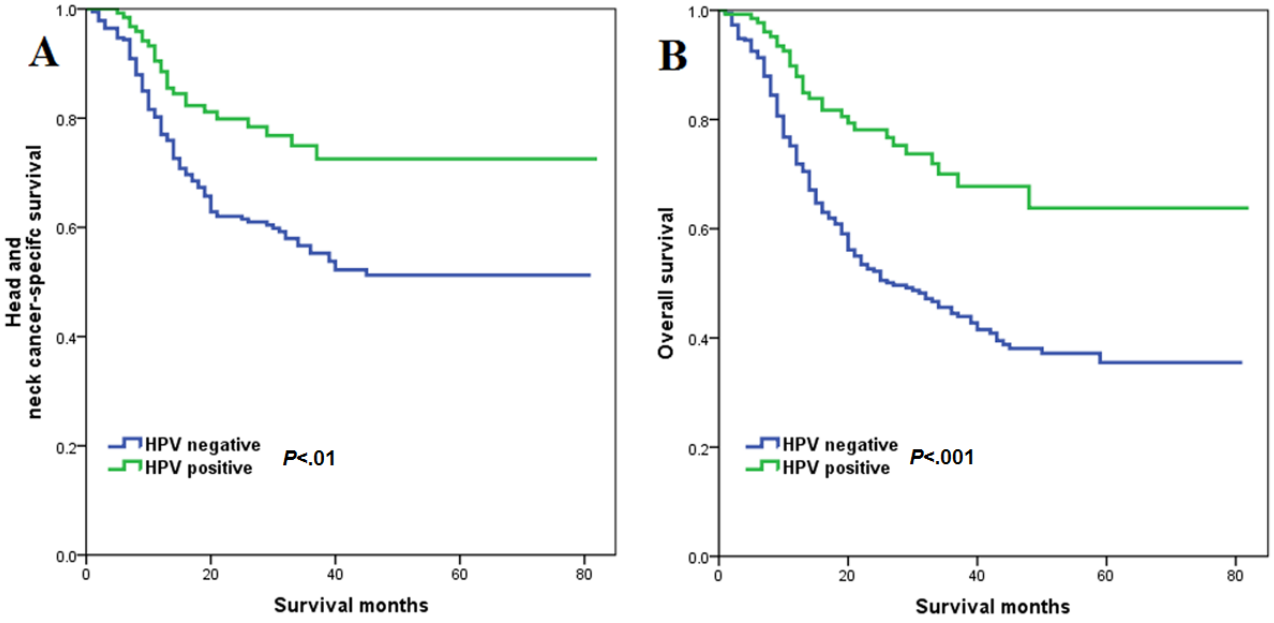


Table 2. Multivariable Cox regression analyses for factors affecting survival outcomes.

Variables	head and neck cancer-specific survival			overall survival		
	HR[a]	95% CI	*P* value	HR	95% CI	*P* value
Age (years)						
<50	1	__[b]	—	1	—	—
50-64	1.528	0.751-3.110	.24	1.363	0.752-2.468	.31
>64	2.240	1.104-4.547	.03	2.025	1.119-3.664	.02
Gender						
Male	1	—	—	1	—	—
Female	1.128	0.720-1.768	.60	1.118	0.759-1.647	.57
Race						
Non-Hispanic White	1	—	—	1	—	—
Non-Hispanic Black	1.419	0.869-2.317	.16	1.841	1.253-2.704	.002
Hispanic (all)	1.116	0.618-2.017	.72	1.034	0.620-1.725	.90
Other	1.374	0.776-2.433	.28	1.203	0.725-1.999	.48
Grade						
Well differentiated	1	—	—	1	—	—
Moderately differentiated	0.712	0.338-1.503	.37	0.604	0.316-1.152	.13
Poorly differentiated or undifferentiated	0.684	0.324-1.448	.32	0.613	0.321-1.170	.14
Tumor location						
Pyriform sinus	1	—	—	1	—	—
Aryepiglottic fold	1.484	0.789-2.790	.22	1.355	0.782-2.347	.28
Postcricoid region	2.413	1.028-5.662	.04	1.746	0.769-3.967	.18
Posterior wall	1.366	0.797-2.340	.26	1.291	0.811-2.056	.28
AJCC[c] stage						
III	1	—	—	1	—	—
IVA	1.676	1.085-2.587	.02	1.445	1.011-2.065	.04
IVB	2.513	1.408-4.484	.002	2.252	1.382-3.669	.001
Chemotherapy						
No	1	—	—	1	—	—
Yes	0.405	0.259-0.633	<.001	0.405	0.274-0.598	<.001
Insurance status						
Insured	1	—	—	1	—	—
Uninsured	0.721	0.355-1.464	.37	0.788	0.416-1.489	.46
Marital status						
Married	1	—	—	1	—	—
Divorced	1.016	0.621-1.663	.95	1.087	0.724-1.631	.69
Single	1.728	1.143-2.612	.009	1.323	0.904-1.937	.15
Widowed	1.658	0.910-3.020	.10	1.719	1.032-2.864	.04
HPV[d] status						
HPV-negative	1	—	—	1	—	—
HPV-positive	0.460	0.299-0.709	<.001	0.422	0.285-0.625	<.001

[a]HR: hazard ratio.

[b]Not applicable.

[c]AJCC: American Joint Committee on Cancer.

[d]HPV: human papillomavirus.

Effect of Chemotherapy After Stratification of HPV Status

Sensitivity analyses were used to investigate the effect of chemotherapy according to HPV status. After adjustment of age, gender, race, grade, tumor location, AJCC stage, insurance status, and marital status, the results of multivariate Cox regression analyses indicated that receipt of chemotherapy was related to better HNCSS (HR: 0.350; *P*<.001) and OS (HR: 0.342; *P*<.001) compared to not receiving chemotherapy in patients with HPV-negative HSCC, while similar HNCSS (chemotherapy vs no chemotherapy: HR 0.581; *P*=.45) and OS (chemotherapy vs no chemotherapy: HR 0.340; *P*=.07) were observed between both treatment arms in HPV-positive HSCC (Table 3). There were 31 (Table S2 in Multimedia Appendix 1) and 17 (Table S3 in Multimedia Appendix 1) pairs of patients matched using PSM in HPV-negative and HPV-positive groups, respectively. The survival curves by chemotherapy receipt according to HPV status are listed in Figure 3. Similar results were found after PSM (Table 3 and Figure 4).

Table 3. Multivariable Cox regression analyses to determine the effect of chemotherapy on survival outcomes according to HPV status.

Variables	Head and neck cancer-specific survival			Overall survival		
	HR[a]	95% CI	*P* value	HR	95% CI	*P* value
HPV[b]- negative (before PSM[c])						
Chemotherapy						
No	1	—[d]	—	1	—	—
Yes	0.350	0.216-0.567	<.001	0.342	0.224-0.523	<.001
HPV - positive (before PSM)						
Chemotherapy						
No	1	—	—	1	—	—
Yes	0.581	0.143-2.354	.45	0.340	0.106-1.087	.07
HPV -negative (after PSM)						
Chemotherapy						
No	1	—	—	1	—	—
Yes	0.445	0.204-0.971	.04	0.395	0.194-0.803	.01
HPV - positive (after PSM)						
Chemotherapy						
No	1	—	—	—	—	—
Yes	0.048	0.002-1.307	.07	—	—	.29

[a]HR: hazard ratio.

[b]HPV: human papillomavirus.

[c]PSM: propensity score matching.

[d]Not applicable.

Figure 3. The effect of additional chemotherapy—according to human papillomavirus (HPV) status—before propensity score matching on head and neck cancer-specific survival (A: HPV-negative; C: HPV-positive) and overall survival (B: HPV-negative; D: HPV-positive) among patients with hypopharyngeal squamous cell carcinoma.

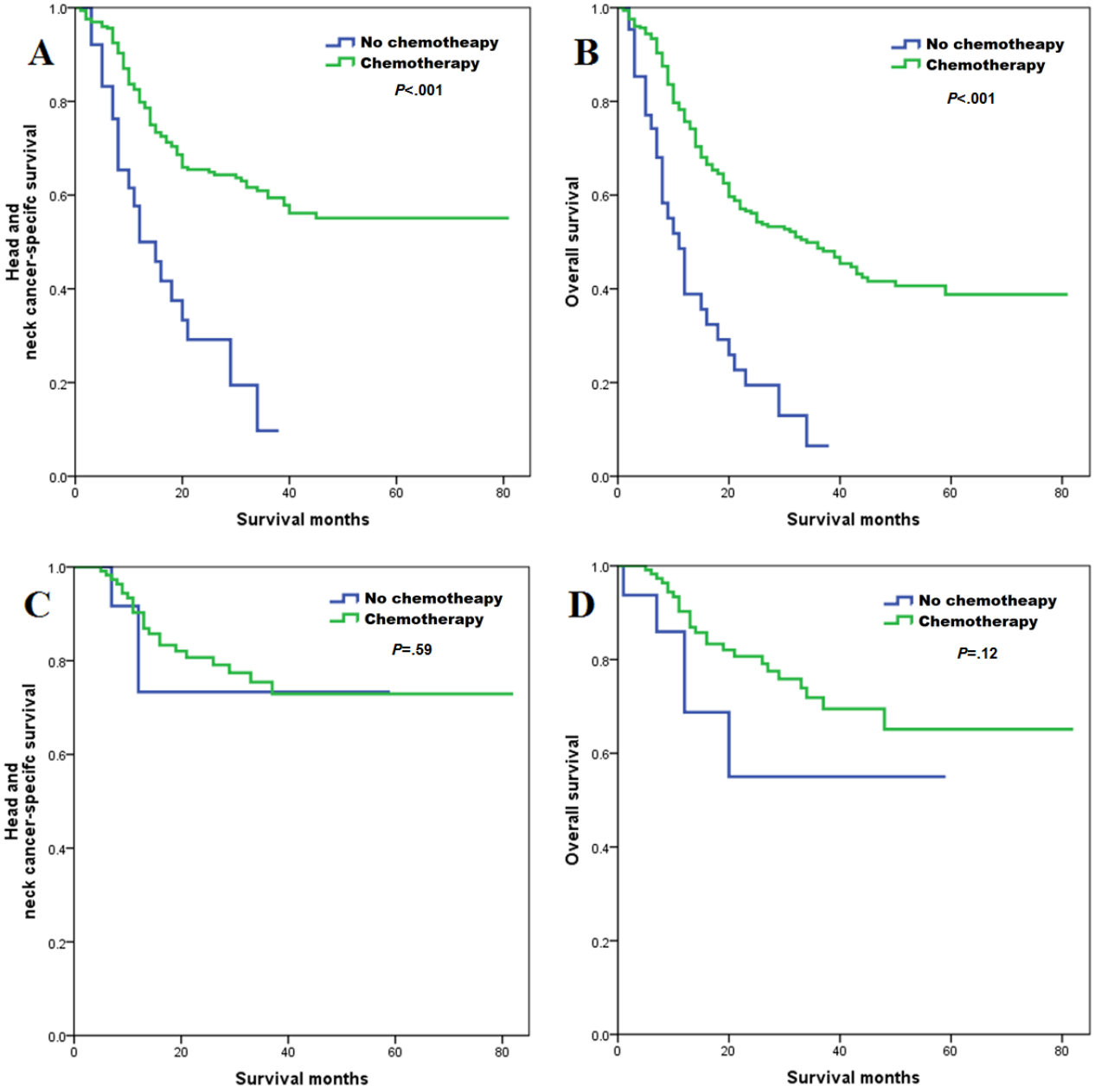
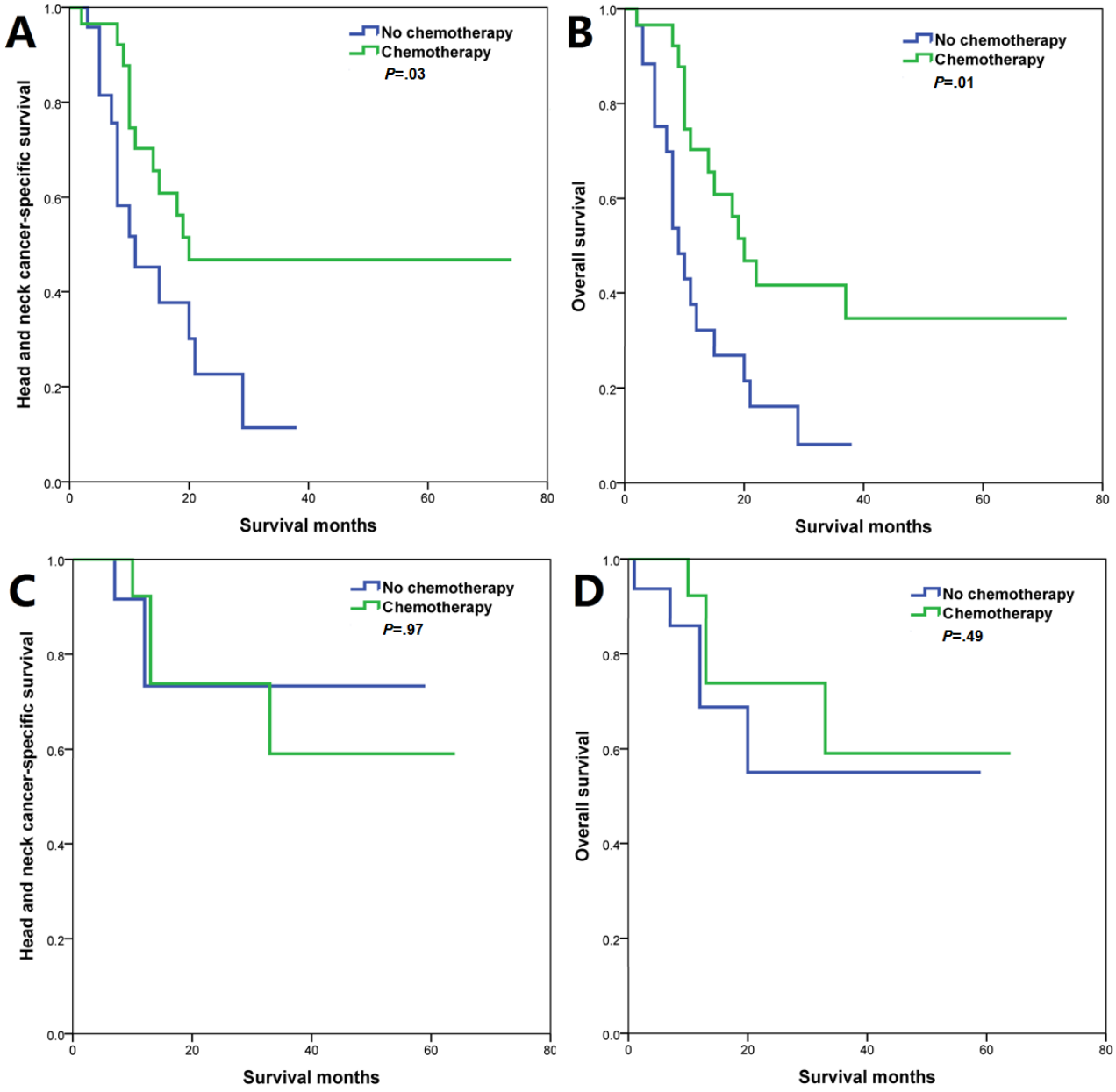


Figure 4. The effect of additional chemotherapy—according to human papillomavirus (HPV) status—after propensity score matching on head and neck cancer-specific survival (A: HPV-negative; C: HPV-positive) and overall survival (B: HPV-negative; D: HPV-positive) among patients with hypopharyngeal squamous cell carcinoma.



Discussion

Principal Findings

In this study, we first investigated the prognostic and predictive effects of HPV status of locally advanced HSCC receiving definitive radiotherapy. Our results showed that 26.7% of patients with locally advanced HSCC had HPV-positive disease. Moreover, patients with positive HPV had a better prognosis than those with negative HPV. The secondary objective of this study was to investigate whether the HPV status could predict the effect of chemotherapy on survival in patients with HSCC receiving radiotherapy. The sensitivity analyses showed that the addition of chemotherapy only improved survival outcomes in HPV-negative HSCC, but not in HPV-positive HSCC.

The etiological relation with cancers developing in the nonoropharynx parts versus the oropharynx remains unestablished. The incidence of HPV infection created a significant difference regarding tumor sites and race. In a large cohort study [10] from the National Cancer Data Base including 24740 patients with HNSCC, the percentages of HPV-positive disease by tumor location were 17.7% for hypopharynx, 11% for larynx, 10.6% for oral cavity, and 62.9% for oropharynx. A study from Japan [18] included 493 patients with HNSCC, in whom the prevalence of HPV in oropharyngeal, oral, nasopharyngeal, hypopharyngeal, and laryngeal carcinomas was 34.4%, 0%, 12%, 3.5%, and 3.9%, respectively. Another study from Thailand [8] showed that the prevalence of HPV in OSCC was only 6%, and no HPV infection was found in laryngeal and hypopharyngeal cancers. However, a case-control study in the Southern Chinese population showed that 29.4%

of oropharyngeal cancers, 16.1% of laryngeal cancers, 14.3% of hypopharyngeal cancers, and 2.2% of oral cavity cancers were HPV DNA positive [19]. In our study, the incidence of HPV-related disease in HSCC was higher than that in the above studies (26.7% vs 3.3%-17.7%), which might be due to the fact that we only included patients in stage III-IVB receiving primary radiotherapy or chemoradiotherapy. Similar results were found in OSCC, which also showed a higher percentage of HPV-related patients receiving primary radiotherapy or chemoradiotherapy [20]. In addition, the patient selection, the geographical distribution of patients, and the HPV testing methods also played a role in this variability. Moreover, the probability of tobacco use in different cohorts may also lead to a discrepancy in HPV infection rates [9].

Patients with HPV-positive OSCC were more likely to be male, younger, in an early tumor stage, and in advanced nodal stage [21]. In a recent SEER study including stage I-IV HSCC, they found a higher proportion of HPV-positive patients who were White or Hispanic [22]. Another SEER study by Abdel-Rahman [11] indicated that HPV-related HSCC was more likely to involve younger people and higher tumor grades. The Danish Head and Neck Cancer Group trials also found no significant difference regarding age and AJCC stage in laryngeal and hypopharyngeal cancers between p16-negative and p16-positive disease, while the p16-positive disease was more likely to present in female patients (29% vs 17%; *P*=.02) [13]. In our study, we also showed that a higher proportion of HPV-related tumors were found in Non-Hispanic White patients; otherwise, no significant differences were found in the demographics, clinicopathological characteristics, or chemotherapy receipt between HPV-negative and HPV-positive diseases.

The role of the HPV status in HSCC remains controversial. Hughes et al [12] included 94 patients with laryngeal or hypopharyngeal cancers (13% of patients were HPV-related), and HPV did not appear to significantly impact survival or disease control in patients in stage III-IV receiving primary chemoradiotherapy. In addition, a study from Karolinska Institute [7] included 82 patients with HSCC and found that being HPV DNA positive (n=7) was associated with better OS but not disease-specific survival compared to those being HPV DNA negative, while a similar prognosis was found between p16-negative and p16-positive diseases. Several studies including the Danish Head and Neck Cancer Group trials also showed similar outcomes between p16-negative and p16-positive diseases, suggesting that the prognostic effect may be limited to OSCC only [13-15]. However, a small portion of patients with HPV DNA-positive or p16-positive disease in the above studies limited the study to be applied to the general population. Two larger cohort studies from the National Cancer Data Base (n=1085) and SEER (n=1157) included patients with HSCC, and they found that those with HPV-positive HSCC had better OS and cancer-specific survival compared to those with HPV-negative HSCC [10,11]. To our knowledge, our study was the largest cohort study to investigate the role of HPV status in patients with HSCC receiving primary radiotherapy or chemoradiotherapy. Within this cohort, we suggest that HPV status may be an additional factor for risk stratification of HSCC, and the future revision of the AJCC staging should consider the HPV status.

HPV-positive OSCC is a distinct pathological entity and may deserve a more personalized therapeutic strategy to decrease the severe early and late toxicities, including de-escalation of radiation doses, less toxic chemotherapy treatment, and removal of chemotherapy [23-25]. In HPV-positive HNSCC, the data also suggested that intensive chemoradiotherapy approaches did not improve clinical outcomes compared to radiotherapy alone in the definitive radiotherapy setting and postoperative radiotherapy setting [26-29]. In National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology, the optimal treatment options for HSCC are induction chemotherapy, surgery, concurrent chemoradiotherapy, or clinical trials [30]. Radiotherapy or chemoradiotherapy is not recommended as primary treatment. However, in clinical practice, concurrent chemoradiotherapy remains the main treatment strategy for organ preservation among patients with HSCC. Two previous studies including patients from the SEER and National Cancer Database showed that 81.3% and 72.2% of patients have received radiotherapy or chemoradiotherapy, respectively [31,32]. In this study, we demonstrated a survival benefit of additional chemotherapy in HPV-negative HSCC, whereas radiotherapy outcome did not differ by the receipt of chemotherapy in HPV-positive HSCC. One possible explanation for this finding is that HPV-positive HSCC may be cured by primary radiotherapy. The results of HPV status have been widely used for prognostic assessment and treatment decision-making for OSCC. Our results indicated that the prognosis and treatment response for HSCC could also be individualized according to HPV status.

The reasons for the better prognosis and limited effect from chemotherapy in patients with HPV-related HSCC who received primary radiotherapy remain unsolved. The high response of HPV-associated cancer cells to radiotherapy may be related to cell cycle dysregulation, repopulation signaling, and impaired DNA repair capacity of the tumor cells [33-36]. Furthermore, the proximity of the HNSCC to lymphoid tissues may also contribute to the high radiosensitivity of HPV-related tumors, and the interaction between the virus antigens and the immune system may contribute to enhancing the radiosensitivity [13]. Whether HPV-related HSCC also possesses such enhanced radiosensitivity remains to be clarified. In addition, HPV-negative tumors often carry frequent TP53 mutations, resulting in significant radioresistance [37]. Moreover, those with HPV-positive HSCC had a better prognosis due to higher immune activity and overexpression of immune proteins compared to those with HPV-negative HSCC, which was similar to the results for OSCC [38-40].

Several limitations should be acknowledged in this study. First, the findings of our study should be viewed with caution because this is a retrospective observational study from a population-based cohort. Second, the rationale for treatment decisions among each patient group cannot be ascertained from the SEER database. Third, the HPV testing results may be heterogeneous with respect to technique, and the results of HPV testing were not centrally reviewed. However, a previous study showed high concordance among the

p16-immunohistochemistry, in situ hybridization, or polymerase chain reaction methods for detecting HPV status [41]. Fourth, several confounding factors were not measured in the SEER database, including the chemotherapy regimen, the target volume of radiotherapy, the sequence of chemotherapy and radiotherapy, details regarding patient performance status, and tobacco or alcohol exposure. Moreover, the patterns of locoregional and distant metastasis after primary radiotherapy or chemoradiotherapy were also not routinely captured in the SEER database. Finally, approximately 15% of patients should have undergone salvage surgery for recurrence after radiotherapy or chemoradiotherapy in a previous study [42]. However, SEER does not record the information regarding salvage surgery, thus it is unable to evaluate the impact of salvage surgery on the prognosis and its distribution in both groups (HPV-positive and HPV-negative). Despite the above limitations, we believe the results from our population-based study are provocative enough to warrant further investigation of the prognostic and predictive effects of the HPV status in HSCC.

Conclusions

In conclusion, our study suggests that approximately one-quarter of HSCC may be HPV-related, and HPV-positive HSCC is associated with improved survival outcomes. Furthermore, additional chemotherapy appears to be not related to a survival benefit in HPV-positive tumors receiving primary radiotherapy. More studies are required to better understand the prognostic and predictive effects of the HPV status in HSCC.

Acknowledgments

The authors acknowledge the efforts of the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER database.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[PDF File (Adobe PDF File), 58 KB - publichealth_v8i12e40185_app1.pdf]

References

1. Concu R, Cordeiro MNDS. Cetuximab and the head and neck squamous cell cancer. Curr Top Med Chem 2018 May 14;18(3):192-198. [doi: 10.2174/1568026618666180112162412] [Medline: 29332581]
2. Garneau JC, Bakst RL, Miles BA. Hypopharyngeal cancer: a state of the art review. Oral Oncol 2018 Nov;86:244-250. [doi: 10.1016/j.oraloncology.2018.09.025] [Medline: 30409307]
3. Takes RP, Strojan P, Silver CE, Bradley PJ, Haigentz MJ, Wolf GT, International HeadNeck Scientific Group. Current trends in initial management of hypopharyngeal cancer: the declining use of open surgery. Head Neck 2012 Feb 10;34(2):270-281 [FREE Full text] [doi: 10.1002/hed.21613] [Medline: 22228621]
4. Marur S, Forastiere AA. Head and neck squamous cell carcinoma: update on epidemiology, diagnosis, and treatment. Mayo Clin Proc 2016 Mar;91(3):386-396. [doi: 10.1016/j.mayocp.2015.12.017] [Medline: 26944243]
5. Wittekindt C, Klussmann J. Tumor staging and HPV-related oropharyngeal cancer. Recent Results Cancer Res 2017:206-133. [doi: 10.1007/978-3-319-43580-0_9] [Medline: 27699534]
6. Wagner S, Reder H, Sharma SJ, Würdemann N, Wittekindt C, Klußmann JP. [HPV-associated oropharyngeal cancer-incidence, trends, diagnosis, and treatment]. Urologe A 2018 Dec 12;57(12):1457-1463. [doi: 10.1007/s00120-018-0810-4] [Medline: 30421134]
7. Dalianis T, Grün N, Koch J, Vlastos A, Tertipis N, Nordfors C, et al. Human papillomavirus DNA and p16(INK4a) expression in hypopharyngeal cancer and in relation to clinical outcome, in Stockholm, Sweden. Oral Oncol 2015 Sep;51(9):857-861. [doi: 10.1016/j.oraloncology.2015.06.002] [Medline: 26120094]
8. Pongsapich W, Eakkasem N, Siritantikorn S, Pithuksurachai P, Bongsabhikul K, Chongkolwatana C. Prevalence of HPV infection in hypopharyngeal and laryngeal squamous cell carcinoma at Thailand's largest tertiary referral center. Infect Agent Cancer 2017 Nov 2;12(1):58 [FREE Full text] [doi: 10.1186/s13027-017-0167-0] [Medline: 29118826]
9. Rodrigo JP, Hermsen MA, Fresno MF, Brakenhoff RH, García-Velasco F, Snijders PJ, et al. Prevalence of human papillomavirus in laryngeal and hypopharyngeal squamous cell carcinomas in northern Spain. Cancer Epidemiol 2015 Feb;39(1):37-41. [doi: 10.1016/j.canep.2014.11.003] [Medline: 25468644]
10. Tian S, Switchenko JM, Jhaveri J, Cassidy RJ, Ferris MJ, Press RH, et al. Survival outcomes by high-risk human papillomavirus status in nonoropharyngeal head and neck squamous cell carcinomas: A propensity-scored analysis of the National Cancer Data Base. Cancer 2019 Aug 15;125(16):2782-2793 [FREE Full text] [doi: 10.1002/cncr.32115] [Medline: 31012957]
11. Abdel-Rahman O. Prognostic value of HPV status among patients with hypopharyngeal carcinoma: a population-based study. Clin Transl Oncol 2020 Sep 14;22(9):1645-1650. [doi: 10.1007/s12094-020-02289-8] [Medline: 31939101]

12. Hughes RT, Beuerlein WJ, O'Neill SS, Porosnicu M, Lycan TW, Waltonen JD, et al. Human papillomavirus-associated squamous cell carcinoma of the larynx or hypopharynx: Clinical outcomes and implications for laryngeal preservation. Oral Oncol 2019 Nov;98:20-27 [FREE Full text] [doi: 10.1016/j.oraloncology.2019.09.008] [Medline: 31536842]
13. Lassen P, Primdahl H, Johansen J, Kristensen CA, Andersen E, Andersen LJ, Danish HeadNeck Cancer Group (DAHANCA). Impact of HPV-associated p16-expression on radiotherapy outcome in advanced oropharynx and non-oropharynx cancer. Radiother Oncol 2014 Dec;113(3):310-316 [FREE Full text] [doi: 10.1016/j.radonc.2014.11.032] [Medline: 25544647]
14. Meshman J, Wang P, Chin R, John MS, Abemayor E, Bhuta S, et al. Prognostic significance of p16 in squamous cell carcinoma of the larynx and hypopharynx. Am J Otolaryngol 2017 Jan;38(1):31-37. [doi: 10.1016/j.amjoto.2016.09.007] [Medline: 27751621]
15. Ernoux-Neufcoeur P, Arafa M, Decaestecker C, Duray A, Remmelink M, Leroy X, et al. Combined analysis of HPV DNA, p16, p21 and p53 to predict prognosis in patients with stage IV hypopharyngeal carcinoma. J Cancer Res Clin Oncol 2011 Jan 8;137(1):173-181 [FREE Full text] [doi: 10.1007/s00432-010-0871-2] [Medline: 20376678]
16. SEER*Stat Database: Incidence - SEER 18 Regs Custom Data Head and Neck (select schemas with HPV recode and additional treatment fields), Nov 2018 Sub (2010-2016) - Linked To County Attributes - Total U.S., 1969-2017 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2019, based on the November 2018 submission. Surveillance, Epidemiology, and End Results (SEER) Program. 1969. URL: http://www.seer.cancer.gov [accessed 2022-05-05]
17. Fullerton ZH, Butler SS, Mahal BA, Muralidhar V, Schoenfeld JD, Tishler RB, et al. Short-term mortality risks among patients with oropharynx cancer by human papillomavirus status. Cancer 2020 Apr 01;126(7):1424-1433 [FREE Full text] [doi: 10.1002/cncr.32652] [Medline: 31930488]
18. Maruyama H, Yasui T, Ishikawa-Fujiwara T, Morii E, Yamamoto Y, Yoshii T, et al. Human papillomavirus and p53 mutations in head and neck squamous cell carcinoma among Japanese population. Cancer Sci 2014 Apr;105(4):409-417 [FREE Full text] [doi: 10.1111/cas.12369] [Medline: 24521534]
19. Chor JS, Vlantis AC, Chow T, Fung S, Ng FY, Lau C, et al. The role of human papillomavirus in head and neck squamous cell carcinoma: a case control study on a southern Chinese population. J Med Virol 2016 May 03;88(5):877-887. [doi: 10.1002/jmv.24405] [Medline: 26467027]
20. Zhan KY, Puram SV, Li MM, Silverman DA, Agrawal AA, Ozer E, et al. National treatment trends in human papillomavirus-positive oropharyngeal squamous cell carcinoma. Cancer 2020 Mar 15;126(6):1295-1305 [FREE Full text] [doi: 10.1002/cncr.32654] [Medline: 31825543]
21. Tanaka TI, Alawi F. Human papillomavirus and oropharyngeal cancer. Dent Clin North Am 2018 Jan;62(1):111-120. [doi: 10.1016/j.cden.2017.08.008] [Medline: 29126488]
22. Marshall DC, Kao DD, Bakst R, Sharma S, Ferrandino R, Rosenzweig K, et al. Prognostic role of human papilloma virus status in hypopharyngeal squamous cell carcinoma. Laryngoscope Investig Otolaryngol 2020 Oct 11;5(5):860-867 [FREE Full text] [doi: 10.1002/lio2.443] [Medline: 33134533]
23. De Virgilio A, Costantino A, Mercante G, Petruzzi G, Sebastiani D, Franzese C, et al. Present and future of de-intensification strategies in the treatment of oropharyngeal carcinoma. Curr Oncol Rep 2020 Jul 09;22(9):91. [doi: 10.1007/s11912-020-00948-1] [Medline: 32643045]
24. Chen AM, Felix C, Wang P, Hsu S, Basehart V, Garst J, et al. Reduced-dose radiotherapy for human papillomavirus-associated squamous-cell carcinoma of the oropharynx: a single-arm, phase 2 study. Lancet Oncol 2017 Jun;18(6):803-811 [FREE Full text] [doi: 10.1016/S1470-2045(17)30246-2] [Medline: 28434660]
25. Gabani P, Lin AJ, Barnes J, Oppelt P, Adkins DR, Rich JT, et al. Radiation therapy dose de-escalation compared to standard dose radiation therapy in definitive treatment of HPV-positive oropharyngeal squamous cell carcinoma. Radiother Oncol 2019 May;134:81-88. [doi: 10.1016/j.radonc.2019.01.016] [Medline: 31005228]
26. O'Sullivan B, Huang SH, Perez-Ordonez B, Massey C, Siu LL, Weinreb I, et al. Outcomes of HPV-related oropharyngeal cancer patients treated by radiotherapy alone using altered fractionation. Radiother Oncol 2012 Apr;103(1):49-56. [doi: 10.1016/j.radonc.2012.02.009] [Medline: 22410204]
27. Chen AM, Zahra T, Daly ME, Farwell DG, Luu Q, Gandour-Edwards R, et al. Definitive radiation therapy without chemotherapy for human papillomavirus-positive head and neck cancer. Head Neck 2013 Nov 17;35(11):1652-1656 [FREE Full text] [doi: 10.1002/hed.23209] [Medline: 23335285]
28. Yom S, Torres-Saavedra P, Caudell J, Waldron J, Gillison M, Truong M, et al. NRG-HN002: A Randomized Phase II Trial for Patients With p16-Positive, Non-Smoking-Associated, Locoregionally Advanced Oropharyngeal Cancer. IJROBP 2019 Nov;105(3):684-685 [FREE Full text] [doi: 10.1016/j.ijrobp.2019.08.038]
29. Skillington SA, Kallogjeri D, Lewis JS, Piccirillo JF. The role of adjuvant chemotherapy in surgically managed, p16-positive oropharyngeal squamous cell carcinoma. JAMA Otolaryngol Head Neck Surg 2017 Mar 01;143(3):253-259 [FREE Full text] [doi: 10.1001/jamaoto.2016.3353] [Medline: 27918781]
30. Clinical practice guidelines in oncology: head and neck cancers. National Comprehensive Cancer Network. 2021. URL: https://www.nccn.org/professionals/physician_gls/pdf/head-and-neck.pdf [accessed 2021-09-21]

31. Hochfelder CG, McGinn AP, Mehta V, Castellucci E, Kabarriti R, Ow TJ. Treatment sequence and survival in locoregionally advanced hypopharyngeal cancer: a surveillance, epidemiology, and end results-based study. Laryngoscope 2020 Nov 10;130(11):2611-2621 [FREE Full text] [doi: 10.1002/lary.28452] [Medline: 31821572]
32. Bates JE, Morris CG, Hitchcock KE, Dziegielewski PT, Mendenhall WM, Amdur RJ. Locally advanced hypopharyngeal and laryngeal cancer: influence of HPV status. Radiother Oncol 2019 Nov;140:6-9. [doi: 10.1016/j.radonc.2019.05.013] [Medline: 31174105]
33. Braakhuis BJM, Senft A, de Bree R, de Vries J, Ylstra B, Cloos J, et al. Expression profiling and prediction of distant metastases in head and neck squamous cell carcinoma. J Clin Pathol 2006 Dec 27;59(12):1254-1260 [FREE Full text] [doi: 10.1136/jcp.2005.035451] [Medline: 16679350]
34. Sørensen BS, Busk M, Olthof N, Speel E, Horsman MR, Alsner J, et al. Radiosensitivity and effect of hypoxia in HPV positive head and neck cancer cells. Radiother Oncol 2013 Sep;108(3):500-505. [doi: 10.1016/j.radonc.2013.06.011] [Medline: 23953409]
35. Rieckmann T, Tribius S, Grob TJ, Meyer F, Busch C, Petersen C, et al. HNSCC cell lines positive for HPV and p16 possess higher cellular radiosensitivity due to an impaired DSB repair capacity. Radiother Oncol 2013 May;107(2):242-246. [doi: 10.1016/j.radonc.2013.03.013] [Medline: 23602369]
36. Dok R, Kalev P, Van Limbergen EJ, Asbagh LA, Vázquez I, Hauben E, et al. p16INK4a impairs homologous recombination–mediated DNA repair in human papillomavirus–positive head and neck tumors. Cancer Res 2014 Jan 28;74(6):1739-1751. [doi: 10.1158/0008-5472.can-13-2479]
37. Perrone F, Suardi S, Pastore E, Casieri P, Orsenigo M, Caramuta S, et al. Molecular and cytogenetic subgroups of oropharyngeal squamous cell carcinoma. Clin Cancer Res 2006 Nov 15;12(22):6643-6651 [FREE Full text] [doi: 10.1158/1078-0432.CCR-06-1759] [Medline: 17121883]
38. Landin D, Ährlund-Richter A, Mirzaie L, Mints M, Näsman A, Kolev A, et al. Immune related proteins and tumor infiltrating CD8+ lymphocytes in hypopharyngeal cancer in relation to human papillomavirus (HPV) and clinical outcome. Head Neck 2020 Nov;42(11):3206-3217. [doi: 10.1002/hed.26364] [Medline: 32613643]
39. Huang SH, Waldron JN, Milosevic M, Shen X, Ringash J, Su J, et al. Prognostic value of pretreatment circulating neutrophils, monocytes, and lymphocytes in oropharyngeal cancer stratified by human papillomavirus status. Cancer 2015 Feb 15;121(4):545-555 [FREE Full text] [doi: 10.1002/cncr.29100] [Medline: 25336438]
40. Ward MJ, Thirdborough SM, Mellows T, Riley C, Harris S, Suchak K, et al. Tumour-infiltrating lymphocytes predict for outcome in HPV-positive oropharyngeal cancer. Br J Cancer 2014 Jan 21;110(2):489-500 [FREE Full text] [doi: 10.1038/bjc.2013.639] [Medline: 24169344]
41. Schache A, Liloglou T, Risk J, Filia A, Jones T, Sheard J, et al. Evaluation of human papilloma virus diagnostic testing in oropharyngeal squamous cell carcinoma: sensitivity, specificity, and prognostic discrimination. Clin Cancer Res 2011 Oct 01;17(19):6262-6271 [FREE Full text] [doi: 10.1158/1078-0432.CCR-11-0388] [Medline: 21969383]
42. Putten L, Bree R, Doornaert PA, Buter J, Eerenstein SEJ, Rietveld DHF, et al. Salvage surgery in post-chemoradiation laryngeal and hypopharyngeal carcinoma: outcome and review. Acta Otorhinolaryngol Ital 2015 Jun;35(3):162-172 [FREE Full text] [Medline: 26246660]

Abbreviations

AJCC: American Joint Committee on Cancer
HNCSS: head and neck cancer-specific survival
HNSCC: head and neck squamous cell carcinoma
HPV: human papillomavirus
HR: hazard ratio
HSCC: hypopharyngeal squamous cell carcinoma
OS: overall survival
OSCC: oropharyngeal squamous cell carcinoma
PSM: propensity score matching
SEER: Surveillance, Epidemiology, and End Results

Edited by A Mavragani, G Eysenbach; submitted 09.06.22; peer-reviewed by JY Sun, L Liu; comments to author 09.09.22; revised version received 23.09.22; accepted 18.10.22; published 16.12.22.

Please cite as:
Yang SP, Lin XY, Hu M, Cai CF
The Prognostic and Predictive Effects of Human Papillomavirus Status in Hypopharyngeal Carcinoma: Population-Based Study
JMIR Public Health Surveill 2022;8(12):e40185
URL: https://publichealth.jmir.org/2022/12/e40185
doi:10.2196/40185
PMID:36525304

Short Paper

The Importance of Incorporating At-Home Testing Into SARS-CoV-2 Point Prevalence Estimates: Findings From a US National Cohort, February 2022

Saba A Qasmieh[1,2], MPH; McKaylee M Robertson[1], MPH, PhD; Madhura S Rane[1], ScM, PhD; Yanhan Shen[1,2], MS; Rebecca Zimba[1,2], MHS; Camila A Picchio[3], MPH; Angela M Parcesepe[1,4,5], PhD; Mindy Chang[1], DVM, MPH; Sarah G Kulkarni[1], MPH; Christian Grov[1,2], MPH, PhD; Denis Nash[1,2], MPH, PhD

[1]Institute for Implementation Science in Population Health, City University of New York, New York, NY, United States

[2]Graduate School of Public Health and Health Policy, City University of New York, New York, NY, United States

[3]Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, University of Barcelona, Barcelona, Spain

[4]Department of Maternal and Child Health, Gillings School of Public Health, University of North Carolina, Chapel Hill, NC, United States

[5]Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Saba A Qasmieh, MPH
Institute for Implementation Science in Population Health
City University of New York
6th Floor
55 W 125th St
New York, NY, 10027
United States
Phone: 1 646 364 9618
Email: saba.qasmieh@sph.cuny.edu

Abstract

Background: Passive, case-based surveillance underestimates the true extent of active infections in the population due to undiagnosed and untested cases, the exclusion of probable cases diagnosed point-of-care rapid antigen tests, and the exclusive use of at-home rapid tests which are not reported as part of case-based surveillance. The extent in which COVID-19 surveillance may be underestimating the burden of infection is likely due to time-varying factors such as decreased test-seeking behaviors and increased access to and availability of at-home testing.

Objective: The objective of this study is to estimate the prevalence of SARS-CoV-2 based on different definitions of a case to ascertain the extent to which cases of SARS-CoV-2 may be underestimated by case-based surveillance.

Methods: A survey on COVID-19 exposure, infection, and testing was administered to calculate point prevalence of SARS-CoV-2 among a diverse sample of cohort adults from February 8, 2022, to February 22, 2022. Three-point prevalence estimates were calculated among the cohort, as follows: (1) proportion positives based on polymerase chain reaction (PCR) and rapid antigen tests; (2) proportion positives based on testing exclusively with rapid at-home tests; and (3) proportion of probable undiagnosed cases. Test positivity and prevalence differences across booster status were also examined.

Results: Among a cohort of 4328, there were a total of 644 (14.9%) cases. The point prevalence estimate based on PCR or rapid antigen tests was 5.5% (95% CI 4.8%-6.2%), 3.7% (95% CI 3.1%-4.2%) based on at-home rapid tests, and 5.7% (95% CI 5.0%-6.4%) based on the case definition of a probable case. The total point prevalence across all definitions was 14.9% (95% CI 13.8%-16.0%). The percent positivity among PCR or rapid tests was 50.2%. No statistically significant differences were observed in prevalence between participants with a COVID-19 booster compared to fully vaccinated and nonboosted participants except among exclusive at-home rapid testers.

Conclusions: Our findings suggest a substantial number of cases were missed by case-based surveillance systems during the Omicron B.1.1.529 surge, when at-home testing was common. Point prevalence surveys may be a rapid tool to be used to understand SARS-CoV-2 prevalence and would be especially important during case surges to measure the scope and spread of active infections in the population.

(JMIR Public Health Surveill 2022;8(12):e38196) doi:10.2196/38196

KEYWORDS

COVID-19 prevalence; at-home rapid SARS-CoV-2 tests; population-based surveys; COVID-19; surveillance; public health; rapid test; Omicron variant; point prevalence

Introduction

Since the first US case of SARS-CoV-2 Omicron variant, B.1.1.529 (BA.1), was announced in December 2021 [1], its high transmissibility and immunogenetic characteristics led to dramatic increases in new cases and reinfections [2-4]. The rapid surge gave rise to community-wide spread across the country, straining testing capacities. In March 2022, the Centers for Disease Control and Prevention (CDC) updated their guidelines for monitoring community COVID-19 levels by tracking incident cases and hospital admissions and deaths to inform community prevention measures [5]; yet the number of new cases and the proportion positive among SARS-CoV-2 testers (percent positivity) are still used as local metrics to monitor SARS-CoV-2 transmission.

Both the number of reported cases and percent positivity are useful in monitoring changes in SARS-CoV-2 transmission; however, they inadequately capture the extent and spread of SARS-CoV-2 epidemic in the population due to the exclusion of undiagnosed and untested cases by standard surveillance [6-10]. To our knowledge, there is currently no mechanism in place in state and local jurisdictions in the United States for systematically capturing rapid at-home tests as part of a population-level indicator of SARS-CoV-2 spread. In Australia and the United Kingdom, for example, health departments put in place a reporting mechanism for individuals to report their rapid antigen test results. The extent to which the number of active SARS-CoV-2 infections is underestimated is likely to vary by geographic, sociodemographic, and economic factors associated with community and self-testing, in addition to temporal factors that drive test-seeking behaviors during a surge [11,12].

The objective of this analysis was to identify the extent to which cases of SARS-CoV-2 may be incomplete in standard case-based surveillance during the recent surge of the Omicron BA.1 variant. Using data from the national and longitudinal CHASING COVID cohort study, we compared point prevalence of SARS-CoV-2 infections captured by case-based surveillance based on polymerase chain reaction (PCR) and rapid antigen testing to a point prevalence estimated exclusively using rapid at-home SARS-CoV-2 tests as well as probable COVID-19 cases among nontesters. We also examined whether point prevalence differed by SARS-CoV-2 vaccine booster status.

Methods

Recruitment

The CHASING COVID Cohort study is a geographically and sociodemographically diverse sample of adults (18 and older), residing in the United States or its territories and enrolled into a prospective follow-up [13]. Study participants were originally recruited during the emergence of the US COVID-19 pandemic (March-April 2020) via social media (eg, Facebook) or via referral. Details of cohort recruitment and follow-up have been described elsewhere [13], but briefly, cohort participants have been prospectively followed with surveys occurring approximately every 3 months to capture a variety of measures, including COVID-19 symptoms, testing, hospitalizations, and adoption of nonpharmaceutical interventions. Survey materials and the timing of each survey are accessible on our website.

Ethical Considerations

Informed consent was obtained at study enrollment. Participants receive US $10-15 in compensation for every standard study interaction and are entered into drawings for US $100 with 10 winners awarded. For brief study engagements, participants were entered into drawings with ten US $100 gift cards awarded. Study data are deidentified before analysis, and identifiable information remains on a secure server with limited access. The study protocol was approved by the Institutional Review Board at the City University of New York (protocol 2020-0256-PHHP).

Point Prevalence Estimation

A questionnaire on recent COVID-19 exposure, infection, and testing was administered as the Omicron BA.1 surge was subsiding in the United States, in February 8-22, 2022. The questionnaire asked about the type and result of viral diagnostic tests taken in the past 7 days; the viral tests included PCR, rapid antigen, and rapid at-home tests. The survey collected information on experience in the previous 10 days with any COVID-19 symptoms for self, household, and close contacts, as well as exposure to a confirmed or probable COVID-19 case. COVID-19 symptoms were defined as having at least one of the following: fever of 100 degrees Fahrenheit or greater, cough, runny nose or nasal congestion, shortness of breath, sore throat, fatigue, muscle or body aches, headaches, loss of smell or taste, nausea, as well as vomiting or diarrhea [14].

We calculated 3 mutually exclusive prevalence estimates. First, prevalence was calculated as the proportion of participants reporting a positive result detected by PCR or rapid antigen tests and captured by case-based surveillance. Second, we calculated prevalence as the proportion of participants reporting a positive result using at-home rapid tests and who did not seek further testing, as well as prevalence of probable cases. A probable case, based on the Council of State and Territorial Epidemiologists case definition, did not receive any diagnostic test but reported SARS-CoV-2 symptoms and had an epidemiological linkage, either with a household member or close contact with infection [15]. We calculated the percent positivity as the proportion of positive cases among all testers.

Finally, we ascertained differences in point prevalence by booster status for the 3 case definitions. Booster status was measured as having received a SARS-CoV-2 booster between September 2, 2021, and January 11, 2022 [16]. Among

participants who did not receive a booster dose, we further classified participants as fully, partially, or nonvaccinated with the SARS-CoV-2 vaccine.

Statistical Analysis

Sociodemographic and health behaviors were described for testers and nontesters and by testing outcome. Pearson chi-squared test of independence was performed to assess group differences between testers and nontesters. To assess the effect of booster status on prevalence, we used a log-binomial model and presented adjusted prevalence ratios, adjusted for age, race or ethnicity, education, employment, smoking, essential worker status, and comorbidities. Analyses were performed using SAS, version 9.4 (SAS Institute).

Results

A total of 4328 cohort participants (80% response rate among 5441 participants responding in 2021) completed the point prevalence questionnaire. Among the 841 testers, 396 (47.1%) had tested for SARS-CoV-2 on any diagnostic test (PCR, rapid antigen, or at-home rapid test; Table 1). Among the 3484 nontesters, 248 (7.1%) were probable cases. Testers were more likely to be >39 years old, gender nonbinary, college graduates, employed, and symptomatic, and to report close contact with a case, to have children in their households, to be in households with income above US $70,000, to have a prior SARS-CoV2 infection, to be at high risk for severe COVID-19 outcomes, and to have received a booster vaccine.

There was a total of 644 cases, among which 237 (36.8%) were positive based on point-of-care PCR or rapid antigen tests, 159 (24.7%) cases that were identified exclusively with at-home rapid tests, and 248 (38.5%) cases were probable cases. The prevalence estimate based on confirmed point-of-care PCR or rapid antigen tests was 5.5% (95% CI 4.8%-6.2%), of which 1.1% (95% CI 0.8%-1.4%) was based on rapid antigen tests only, 1.7% (95% CI 1.3%-2.2%) based on PCR tests only, and 2.6% (95% CI 2.2%-3.1%) based on both PCR and rapid antigen tests. The point prevalence based on those testing exclusively via rapid at-home tests was 3.7% (95% CI 3.1%-4.2%) and was 5.7% (95% CI 5.0%-6.4%) for probable cases. The total point prevalence was 14.9% (95% CI 13.8%-16.0%). The percent positivity among PCR or rapid antigen tests was 50.2%. Differences in SARS-CoV-2 prevalence among participants who had a COVID-19 booster versus those fully vaccinated and nonboosted participants were not statistically significant, except those diagnosed using at-home tests (adjusted prevalence ratio: 2.2, 95% CI 1.4%-3.4%; Table 2).

Table 1. Cohort characteristics by testing status and by test type (N=4328).

Characteristics	Total, n (%)	Nontesters, n (%)	Testers (any), n (%)	POC[a] PCR[b] test only, n (%)	POC rapid antigen test only, n (%)	With provider (POC) and at-home testers, n (%)	At-home rapid test only, n (%)	*P* value[c]
Total	4328	3487 (80.6)	841 (19.4)	167 (3.9)	89 (2.1)	216 (5.0)	369 (8.5)	—[d]
SARS-CoV-2 positive	644 (14.8)	—	—	—	—	—	—	—
POC PCR or rapid antigen test cases	237 (5.5)	—	—	—	—	—	—	—
Exclusive at-home test cases	159 (3.7)	—	—	—	—	—	—	—
Probable cases	248 (5.7)	—	—	—	—	—	—	—
Age range								<.001
18-29	826 (19.1)	636 (18.2)	190 (22.6)	56 (33.5)	15 (16.9)	51 (23.6)	68 (18.4)	
30-39	1217 (28.1)	946 (27.1)	271 (32.2)	46 (27.5)	23 (25.8)	87 (40.3)	115 (31.2)	
40-49	808 (18.7)	650 (18.6)	158 (18.8)	16 (9.6)	18 (20.2)	42 (19.4)	82 (22.2)	
50-64	941 (21.7)	794 (22.8)	147 (17.5)	28 (16.8)	23 (25.8)	21 (9.7)	75 (20.3)	
>65	536 (12.4)	461 (13.2)	75 (8.9)	21 (12.6)	10 (11.2)	15 (6.9)	29 (7.9)	
Gender								.04
Male	1913 (44.2)	1538 (44.1)	375 (44.6)	68 (40.7)	41 (46.1)	97 (44.9)	169 (45.8)	
Female	2294 (53.0)	1862 (53.4)	432 (51.4)	90 (53.9)	46 (51.7)	112 (51.9)	184 (49.9)	
Gender nonbinary	121 (2.8)	87 (2.5)	34 (4.0)	9 (5.4)	2 (2.3)	7 (3.2)	16 (4.3)	
Race or ethnicity								.54
Hispanic	657 (15.2)	527 (15.1)	130 (15.5)	28 (16.8)	19 (21.4)	40 (18.5)	43 (11.7)	
Black non-Hispanic	385 (8.9)	308 (8.8)	77 (9.2)	11 (6.6)	21 (23.6)	21 (9.7)	24 (6.5)	
Asian American or Pacific Islander	302 (7.0)	233 (6.7)	69 (8.2)	18 (10.8)	5 (5.6)	20 (9.3)	26 (7.1)	
White non-Hispanic	2824 (65.5)	2287 (65.6)	537 (63.9)	102 (61.1)	43 (48.3)	129 (59.7)	263 (71.3)	
Other	160 (3.4)	132 (3.8)	28 (3.3)	8 (4.8)	1 (1.1)	6 (2.8)	13 (3.5)	
Income (US $)								.009
<35,000	1115 (25.8)	937 (26.9)	178 (21.2)	40 (24.0)	23 (25.8)	48 (22.2)	67 (18.2)	
35,000-49,000	479 (11.1)	389 (11.2)	90 (10.7)	19 (11.4)	8 (9.0)	30 (13.9)	33 (8.9)	
50,000-69,000	638 (14.7)	513 (14.7)	125 (14.9)	26 (15.6)	14 (15.7)	38 (17.6)	47 (12.7)	
70,000-99,000	737 (17.0)	592 (17.0)	145 (17.2)	28 (16.8)	22 (24.7)	25 (11.6)	70 (19.0)	
>100,000	1236 (28.6)	961 (27.6)	275 (32.7)	45 (27.0)	21 (23.6)	67 (31.0)	142 (38.5)	
Missing or unknown	123 (2.8)	95 (2.7)	28 (3.3)	9 (5.4)	1 (1.1)	8 (3.7)	10 (2.7)	
Education								.03
<High school	59 (1.4)	51 (1.5)	8 (1.0)	3 (1.8)	1 (1.1)	2 (0.9)	2 (0.5)	
High school	383 (8.9)	324 (9.3)	59 (7.0)	12 (7.2)	6 (6.7)	20 (9.3)	21 (5.7)	
Some college	1089 (25.2)	892 (25.6)	197 (23.4)	38 (22.8)	31 (34.8)	52 (24.1)	76 (20.6)	
College graduate	2797 (64.6)	2220 (63.7)	577 (68.6)	114 (68.3)	51 (57.3)	142 (65.7)	270 (73.2)	
Employment								.001
Employed	1704 (39.4)	1343 (38.5)	361 (42.9)	65 (38.9)	35 (39.3)	91 (42.1)	170 (46.1)	
Out of work	615 (14.2)	522 (15.0)	93 (11.1)	13 (7.8)	11 (12.4)	22 (10.2)	47 (12.7)	
Student	250 (5.8)	187 (5.4)	63 (7.5)	25 (15.0)	4 (4.5)	19 (8.8)	15 (4.1)	
Other or unknown	1759 (40.6)	1435 (41.2)	324 (38.5)	64 (38.3)	39 (43.8)	84 (38.9)	137 (37.1)	

Characteristics	Total, n (%)	Nontesters, n (%)	Testers (any), n (%)	POC[a] PCR[b] test only, n (%)	POC rapid antigen test only, n (%)	With provider (POC) and at-home testers, n (%)	At-home rapid test only, n (%)	*P* value[c]
Children in household								.57
Yes	1163 (26.9)	915 (26.2)	248 (29.5)	32 (19.2)	29 (32.6)	85 (39.4)	102 (27.6)	
No	3165 (73.1)	2572 (73.8)	593 (70.5)	135 (80.8)	60 (67.4)	131 (60.7)	267 (72.4)	
Vaccination status								<.001
Boosted	2810 (64.9)	2191 (62.8)	619 (73.6)	121 (72.5)	51 (57.3)	154 (71.3)	293 (79.4)	
Fully vaccinated	1029 (23.8)	875 (25.1)	154 (18.3)	35 (21.0)	25 (28.1)	40 (18.5)	54 (14.6)	
Partially vaccinated	81 (1.9)	63 (1.8)	18 (2.1)	1 (0.6)	4 (4.5)	6 (2.8)	7 (1.9)	
Not vaccinated	408 (9.4)	358 (10.3)	50 (6.0)	10 (6.0)	9 (10.1)	16 (7.4)	15 (4.1)	
Prior COVID-19 infection								.01
Yes	696 (16.1)	545 (15.6)	151 (18.0)	26 (15.6)	23 (25.8)	56 (25.9)	46 (12.5)	
No	3632 (83.9)	2942 (84.4)	690 (82.1)	141 (84.4)	66 (74.2)	160 (47.2)	323 (87.5)	
COVID-19–like symptoms								<.001
Yes	760 (17.6)	434 (12.5)	326 (38.8)	44 (26.4)	25 (28.1)	116 (53.7)	141 (38.2)	
No	3568 (82.4)	3053 (87.6)	515 (61.2)	123 (73.7)	64 (71.9)	100 (46.3)	228 (61.8)	
High risk status[e]								.003
Yes	2191 (50.6)	1804 (51.7)	387 (46.0)	74 (44.3)	53 (59.3)	101 (46.8)	159 (43.1)	
No	2137 (49.4)	1683 (48.3)	454 (54.0)	93 (55.7)	36 (40.5)	115 (53.2)	210 (56.9)	
Close contact with confirmed case								<.001
Yes	630 (14.6)	336 (9.6)	294 (35.0)	51 (30.5)	28 (31.5)	102 (47.2)	113 (30.6)	
No	3698 (85.4)	3151 (90.4)	547 (65.0)	116 (69.5)	61 (68.5)	114 (52.8)	256 (69.4)	

[a]POC: point-of-care.

[b]PCR: polymerase chain reaction.

[c]*P* value corresponds to cohort group differences between testers and nontesters.

[d]Not applicable.

[e]Essential worker, >60 years old, smoker, and reported comorbidities.

Table 2. Point prevalence estimates by vaccination status, February 2-22, 2022 (N=4328).

Variable	Point prevalence							
	Cases identified with PCR[a] or rapid antigen tests		Cases identified with at-home rapid tests		Probable cases		Total prevalence	
	N	% (95% CI)	N	% (95% CI)	N	% (95% CI)	N	% (95% CI)
Total	237	5.5 (4.8-6.2)	159	3.7 (3.1-4.3)	248	5.7 (5.0-6.4)	644	14.9 (13.8-15.9)
Boosted	150	5.3 (4.5-6.2)	120	4.3 (3.5-5.0)	132	4.7 (3.9-5.4)	402	14.3 (13.0-15.6)
Nonboosted or fully vaccinated	59	5.7 (4.3-7.2)	22	2.1 (1.3-3.0)	67	6.5 (5.0-8.0)	148	14.3 (12.2-16.5)
Nonboosted or partially vaccinated	7	8.6 (2.4-14.9)	4	4.9 (0.1-9.8)	6	7.4 (1.6-13.2)	17	21.0 (11.9-30.0)
No vaccine or unknown	21	5.1 (3.0-7.3)	13	3.2 (1.5-4.9)	43	10.5 (7.5-13.5)	77	18.9 (15.1-22.7)
Boosted vs fully vaccinated[b]	237	1.1 (0.84-1.56)[c]	159	2.2 (1.4-3.4)[c]	248	0.8 (0.6-1.1)[c]	644	1.1 (1.0-1.4)[c]

[a]PCR: PCR: polymerase chain reaction.

[b]Model adjusted for race or ethnicity, age, education, employment, smoking, essential worker status, and comorbidities.

[c]Adjusted prevalence ratio.

Discussion

Principal Findings

Our findings showed a high prevalence of SARS-CoV-2 in our cohort during the decline of the Omicron BA.1 wave in the United States in February 2022. Our results are not directly comparable to US national estimates as CDC's COVID-19 tracker only captures test positive results based on PCR tests and does not include point-of-care antigen tests as done at some local or state levels [9]. Our study suggests a substantial proportion of cases would be missed by standard case-based surveillance systems during the Omicron BA.1 wave, when at-home testing was common [17]. The number of cases detected by case-based surveillance was lower than the total number of cases in our cohort, while the percent positivity was higher than the total prevalence based on all definitions. The underestimated case burden and overestimated percent positivity illustrates the limitations of case-based surveillance, and the extent to which current metrics used to monitor SARS-CoV-2 infection may be incomplete. In addition, we found the characteristics among testers differed considerably from nontesters, underscoring the limitations around case-based surveillance data for understanding the epidemiology and any disparities around SARS-CoV-2 burden and community transmission.

The CDC issued recommendations that shifted away from positivity rates and toward the use of hospital admission and death rate. While hospital admission and death rates better capture disease severity, they lag community transmission by weeks and are of limited use in providing early warning for active community infection. By contrast, and while state and local health departments continue to use metrics such as incident cases and test positivity, population-based surveys may be deployed frequently to capture spread and susceptibility to inform more effective mitigation measures.

We found no statistically significant differences in SARS-CoV-2 prevalence by booster status among those who tested exclusively using at-home rapid tests. These findings may be driven by higher testing frequency as was observed among boosted adults compared to those nonboosted but fully vaccinated adults. In general, our findings align with evidence from studies that show that standard SARS-CoV-2 vaccines plus the additional booster dose offer limited additional protection against symptomatic and asymptomatic infection from the Omicron BA.1 variant; however, boosters have been shown to be effective at reducing severe outcomes such as COVID-19 hospitalizations and deaths, which we did not assess [18,19].

Limitations

Our method had key limitations. First, we measured infection and testing outcomes with self-report, which is prone to misclassification bias. In lieu of biomarker data, we classified an undiagnosed and untested case based on any self-reported COVID-19 symptoms and on contact with a confirmed or probable case, which might lead to an overestimation of true infection status. Furthermore, the latest booster status information on participants was collected before January 11, 2022, potentially missing booster information on those who received a booster between January 11 and the survey date. Additionally, our results for booster dose effectiveness did not adjust for the timing of the booster or consider previous infection history.

Our survey questionnaire consisted of fewer than 20 questions and required less than 10 minutes to complete. Our survey was not intended to be representative of the US population as it aimed to capture the extent of which surveillance data are incomplete and representative, and probability-based point prevalence surveys may be used in tandem with surveillance metrics to rapidly understand local spread and to measure the scope of active infections in the population [20-22] and other highly pertinent epidemiological information. At this stage of the pandemic, the application of low-cost and low-resource intensive tools such as routine population-based surveys may have a large impact on effectively informing the control and prevention of community spread of SARS-CoV-2.

Acknowledgments

Funding for this project is provided by The National Institute of Allergy and Infectious Diseases, award number 3UH3AI133675-04S1 (DN and CG), the City University of New York (CUNY) Institute for Implementation Science in Population Health and the COVID-19 Grant Program of the CUNY Graduate School of Public Health and Health Policy, and National Institute of Child Health and Human Development grant P2C HD050924 (Carolina Population Center). The National Institute of Health played no role in the production of this manuscript nor necessarily endorses the findings.

The authors wish to thank the participants of the CHASING COVID Cohort Study. We are grateful to you for your contributions to the advancement of science around the SARS-CoV-2 pandemic. CAP acknowledges support from the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya and the European Social Fund as an AGAUR-funded PhD fellow.

Conflicts of Interest

None declared.

References

1. First confirmed case of Omicron variant detected in the United States. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/media/releases/2021/s1201-omicron-variant.html [accessed 2022-11-15]

2. Wolter N, Jassat W, Walaza S, Welch R, Moultrie H, Groome M, et al. Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: a data linkage study. The Lancet 2022 Jan;399(10323):437-446. [doi: 10.1016/s0140-6736(22)00017-4]
3. Nyberg T, Ferguson NM, Nash SG, Webster HH, Flaxman S, Andrews N, et al. Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 Omicron (B.1.1.529) and Delta (B.1.617.2) variants in England. SSRN Journal 2022:1-15. [doi: 10.2139/ssrn.4025932]
4. Iuliano AD, Brunkard JM, Boehmer TK, Peterson E, Adjei S, Binder AM, et al. Trends in disease severity and health care utilization during the early Omicron variant period compared with previous SARS-CoV-2 high transmission periods - United States, December 2020-January 2022. MMWR Morb Mortal Wkly Rep 2022 Jan 28;71(4):146-152 [FREE Full text] [doi: 10.15585/mmwr.mm7104e4] [Medline: 35085225]
5. COVID-19 community levels : a measure of the impact of COVID-19 illness on health and healthcare systems. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/coronavirus/2019-ncov/science/community-levels.html [accessed 2022-11-15]
6. Wu SL, Mertens AN, Crider YS, Nguyen A, Pokpongkiat NN, Djajadi S, et al. Substantial underestimation of SARS-CoV-2 infection in the United States. Nat Commun 2020 Sep 09;11(1):4507 [FREE Full text] [doi: 10.1038/s41467-020-18272-4] [Medline: 32908126]
7. Angulo FJ, Finelli L, Swerdlow DL. Estimation of US SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths using seroprevalence surveys. JAMA Netw Open 2021 Jan 04;4(1):e2033706 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.33706] [Medline: 33399860]
8. Nash D, Geng E. Goal-aligned, epidemic intelligence for the public health response to the COVID-19 pandemic. Am J Public Health 2020 Aug;110(8):1154-1156. [doi: 10.2105/ajph.2020.305794]
9. COVID data tracker. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker/ [accessed 2022-11-15]
10. 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]
11. Bien-Gund C, Dugosh K, Acri T, Brady K, Thirumurthy H, Fishman J, et al. Factors associated with US public motivation to use and distribute COVID-19 self-tests. JAMA Netw Open 2021 Jan 04;4(1):e2034001 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.34001] [Medline: 33471114]
12. Sherratt K, Abbott S, Meakin S, Hellewell J, Munday J, Bosse N. Exploring surveillance data biases when estimating the reproduction number: with insights into subpopulation transmission of COVID-19 in England. Philos Trans R Soc Lond B Biol Sci. 2021 2021;376(1829):20200283. [doi: 10.1098/rstb.2020.0283]
13. Robertson MM, Kulkarni SG, Rane M, Kochhar S, Berry A, Chang M, et al. Cohort profile: a national, community-based prospective cohort study of SARS-CoV-2 pandemic outcomes in the USA—the CHASING COVID Cohort study. BMJ Open 2021 Sep 21;11(9):e048778 [FREE Full text] [doi: 10.1136/bmjopen-2021-048778] [Medline: 34548354]
14. Mirzayi C, Waldron L, Berry A, Chang M, Kochhar S, Zimba R, et al. The CHASING COVID Cohort Study: A national, community-based prospective cohort study of SARS-CoV-2 pandemic outcomes in the USA. Zenodo. URL: https://zenodo.org/record/6127735#.Yi9rnBDML0p2022 [accessed 2022-11-15]
15. CSTE interim position statement: Update to COVID-19 case definition. Council of State and Territorial Epidemiologists. URL: https://www.cste.org/news/520707/CSTE-Interim-Position-Statement-Update-to-COVID-19-Case-Definition.htm [accessed 2022-11-15]
16. Fast HE, Zell E, Murthy BP, Murthy N, Meng L, Scharf LG, et al. Booster and additional primary dose COVID-19 vaccinations among adults aged ≥65 years - United States, August 13, 2021-November 19, 2021. MMWR Morb Mortal Wkly Rep 2021 Dec 17;70(50):1735-1739 [FREE Full text] [doi: 10.15585/mmwr.mm7050e2] [Medline: 34914672]
17. Rader B, Gertz A, Iuliano AD, Gilmer M, Wronski L, Astley CM, et al. Use of At-Home COVID-19 Tests - United States, August 23, 2021-March 12, 2022. MMWR Morb Mortal Wkly Rep 2022 Apr 01;71(13):489-494 [FREE Full text] [doi: 10.15585/mmwr.mm7113e1] [Medline: 35358168]
18. Butt A, Talisa V, Yan P, Shaikh O, Omer S, Mayr F. Vaccine Effectiveness of 3 Versus 2 Doses of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) mRNA Vaccines in a High-Risk National Population. Clin Infect Dis 2022 Aug 24;75(1):e579-e584 [FREE Full text] [doi: 10.1093/cid/ciac178] [Medline: 35245940]
19. Abu-Raddad LJ, Chemaitelly H, Ayoub HH, AlMukdad S, Yassine HM, Al-Khatib HA, et al. Effect of mRNA Vaccine Boosters against SARS-CoV-2 Omicron Infection in Qatar. N Engl J Med 2022 May 12;386(19):1804-1816. [doi: 10.1056/nejmoa2200797]
20. Dean N. Tracking COVID-19 infections: time for change. Nature 2022 Feb 08;602(7896):185-185. [doi: 10.1038/d41586-022-00336-8] [Medline: 35136224]
21. Clark SJ, Turner AN. Monitoring epidemics: Lessons from measuring population prevalence of the coronavirus. Proc Natl Acad Sci U S A 2021 Mar 02;118(9):1-3 [FREE Full text] [doi: 10.1073/pnas.2026412118] [Medline: 33627409]
22. Dowd JB. The UK's covid-19 data collection has been "world beating"-let's not throw it away. BMJ 2022 Feb 25;376:o496. [doi: 10.1136/bmj.o496] [Medline: 35217522]

Abbreviations

BA.1: B.1.1.529
CDC: Centers for Disease Control and Prevention
PCR: polymerase chain reaction

Edited by A Mavragani; submitted 23.03.22; peer-reviewed by A Perez-Navarro, T Fahey, R Guo; comments to author 12.08.22; revised version received 30.09.22; accepted 13.10.22; published 27.12.22.

Please cite as:
Qasmieh SA, Robertson MM, Rane MS, Shen Y, Zimba R, Picchio CA, Parcesepe AM, Chang M, Kulkarni SG, Grov C, Nash D
The Importance of Incorporating At-Home Testing Into SARS-CoV-2 Point Prevalence Estimates: Findings From a US National Cohort, February 2022
JMIR Public Health Surveill 2022;8(12):e38196
URL: https://publichealth.jmir.org/2022/12/e38196
doi:10.2196/38196
PMID:36240020

Original Paper

Natural Language Processing for Improved Characterization of COVID-19 Symptoms: Observational Study of 350,000 Patients in a Large Integrated Health Care System

Deborah E Malden[1,2*], MSc, DPhil; Sara Y Tartof[2,3*], MPH, PhD; Bradley K Ackerson[4], MD; Vennis Hong[2], MPH; Jacek Skarbinski[5,6], MD; Vincent Yau[7], MA, PhD; Lei Qian[2], MS, PhD; Heidi Fischer[2], MS, PhD; Sally F Shaw[2], MPH, DrPH; Susan Caparosa[2], MA; Fagen Xie[2], PhD

[1]Epidemic Intelligence Service, Centers for Disease Control and Prevention, Atlanta, GA, United States

[2]Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA, United States

[3]Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, United States

[4]Southern California Permanente Medical Group, Harbor City, CA, United States

[5]The Permanente Medical Group, Kaiser Permanente Northern California, Oakland, CA, United States

[6]Division of Research, Kaiser Permanente Northern California, Oakland, CA, United States

[7]Genentech, a Member of the Roche Group, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:
Deborah E Malden, MSc, DPhil
Department of Research & Evaluation
Kaiser Permanente Southern California
100 S. Los Robles, 2nd Floor
Pasadena, CA, 91101
United States
Phone: 1 310 456 4324
Email: debbie.e.malden@kp.org

Abstract

Background: Natural language processing (NLP) of unstructured text from electronic medical records (EMR) can improve the characterization of COVID-19 signs and symptoms, but large-scale studies demonstrating the real-world application and validation of NLP for this purpose are limited.

Objective: The aim of this paper is to assess the contribution of NLP when identifying COVID-19 signs and symptoms from EMR.

Methods: This study was conducted in Kaiser Permanente Southern California, a large integrated health care system using data from all patients with positive SARS-CoV-2 laboratory tests from March 2020 to May 2021. An NLP algorithm was developed to extract free text from EMR on 12 established signs and symptoms of COVID-19, including fever, cough, headache, fatigue, dyspnea, chills, sore throat, myalgia, anosmia, diarrhea, vomiting or nausea, and abdominal pain. The proportion of patients reporting each symptom and the corresponding onset dates were described before and after supplementing structured EMR data with NLP-extracted signs and symptoms. A random sample of 100 chart-reviewed and adjudicated SARS-CoV-2–positive cases were used to validate the algorithm performance.

Results: A total of 359,938 patients (mean age 40.4 [SD 19.2] years; 191,630/359,938, 53% female) with confirmed SARS-CoV-2 infection were identified over the study period. The most common signs and symptoms identified through NLP-supplemented analyses were cough (220,631/359,938, 61%), fever (185,618/359,938, 52%), myalgia (153,042/359,938, 43%), and headache (144,705/359,938, 40%). The NLP algorithm identified an additional 55,568 (15%) symptomatic cases that were previously defined as asymptomatic using structured data alone. The proportion of additional cases with each selected symptom identified in NLP-supplemented analysis varied across the selected symptoms, from 29% (63,742/220,631) of all records for cough to 64% (38,884/60,865) of all records with nausea or vomiting. Of the 295,305 symptomatic patients, the median time from symptom onset to testing was 3 days using structured data alone, whereas the NLP algorithm identified signs or symptoms approximately

1 day earlier. When validated against chart-reviewed cases, the NLP algorithm successfully identified signs and symptoms with consistently high sensitivity (ranging from 87% to 100%) and specificity (94% to 100%).

Conclusions: These findings demonstrate that NLP can identify and characterize a broad set of COVID-19 signs and symptoms from unstructured EMR data with enhanced detail and timeliness compared with structured data alone.

(JMIR Public Health Surveill 2022;8(12):e41529) doi:10.2196/41529

KEYWORDS

natural language processing; NLP; COVID-19; symptoms; disease characterization; artificial intelligence; symptoms; application; data; cough; fever; headache; surveillance

Introduction

COVID-19, the infection caused by the novel coronavirus, SARS-CoV-2 [1], has accounted for more than 623 million cases and more than 6.5 million deaths globally as of October 2022 [2]. SARS-CoV-2 primarily affects the respiratory system but can also affect the cardiovascular, gastrointestinal, neurologic, and other systems [3-6]. The most common signs and symptoms include fever, cough, shortness of breath, fatigue, muscle aches, headaches, loss of taste or smell, sore throat, congestion, nausea or vomiting, and diarrhea [7]. However, prevalence estimates for each sign or symptom have been inconsistent, with most being derived from studies relying on self-reported surveys that are more subjective than electronic medical records (EMR) [4,8,9]. Of the studies using EMR for disease characterization, most are restricted to subgroups of patients (ie, hospitalized patients) who may have distinct symptom profiles [3,10,11]. An improved understanding of signs and symptoms of COVID-19 can inform patient care and improve population screening and disease surveillance.

Signs and symptoms can be documented in EMR by health care providers in four primary forms, broadly defined as “structured” and “unstructured,” which are as follows: (1) structured COVID-19 lab test order–related questionnaires; (2) structured diagnosis codes; (3) structured clinical notes (which may include self-reported information); and (4) unstructured free-text clinical notes. However, of the few large-scale studies using EMR, most are limited to structured data alone, particularly International Classification of Diseases (ICD) diagnoses, which have demonstrated low concordance with self-reported information due to incomplete documentation during physician visits [12]. Natural language processing (NLP) is a subfield of artificial intelligence devoted to the understanding and generation of language and can be used to supplement structured data fields with data extracted from unstructured health care provider notes across different EMR data sources [13]. In short, NLP algorithms can be designed to convert information residing in natural language into structured formats for medical research, public health surveillance, and clinical decision support [14]. During the COVID-19 pandemic, NLP has mostly been used to extract key information on COVID-19 from scientific publications [15], media articles [16], or social media platforms [17]. However, despite containing rich information on signs and symptoms of COVID-19, limited NLP-based tools have been developed for COVID-19 information extraction from unstructured EMR data. The highest-quality study thus far used an NLP-based tool termed “COVID-19 SignSym” to extract signs or symptoms from a small subset of clinical notes and performed a small validation study using data collected from 3 institutions in the United States [18]. However, the real-world application and overall usefulness of NLP for this purpose has not been assessed at scale in a large population.

Large integrated health care systems with access to complete EMR data provide a unique resource to investigate the value of NLP algorithms in the extraction of additional information from unstructured text fields. This paper describes the distribution and time of the onset of COVID-19 signs and symptoms before and after supplementing structured EMR with an NLP algorithm among more than 350,000 members of a large integrated health care system. In addition, we performed a validation substudy to assess the accuracy of the NLP algorithm in identifying COVID-19 signs and symptoms.

Methods

Study Setting

Kaiser Permanente Southern California (KPSC) is one of the largest integrated health care systems in the United States providing medical services to over 4.7 million members. KPSC’s comprehensive EMR data contains individual-level structured data (including diagnosis codes, procedure codes, self-assessment health forms, medications, immunization records, and laboratory results) and unstructured data (including free-text clinical notes, radiology reports, and pathology reports) covering all medical visits. Therefore, the EMR represents a standardized data collection method across all health care settings (ie, all outpatient services, hospitals, emergency department, and virtual care encounters). Care delivered to members outside of the KPSC system is also captured, as outside providers must submit detailed claims to KPSC for reimbursement. KPSC has a diverse member population that is largely representative of all residents in Southern California with health insurance [19]. As of December 2018, persons of Hispanic or Latino race or ethnicity make up the largest proportion of KPSC members (43%), followed by Non-Hispanic White (35%), Non-Hispanic Asian or Pacific Islander (12%), Non-Hispanic Black or African American (9%), and Other (1%).

Study Population

This is a retrospective cohort study of KPSC patients of all ages with positive SARS-CoV-2 laboratory tests from March 2020 to May 2021. SARS-CoV-2 tests of all types (ie, PCR and antigen tests) across all care settings were included. Participants were included in the analysis if they had at least 6 months of

continuous KPSC membership (allowing for a 45-day administrative enrollment gap between memberships) prior to the date of their first positive COVID-19 test.

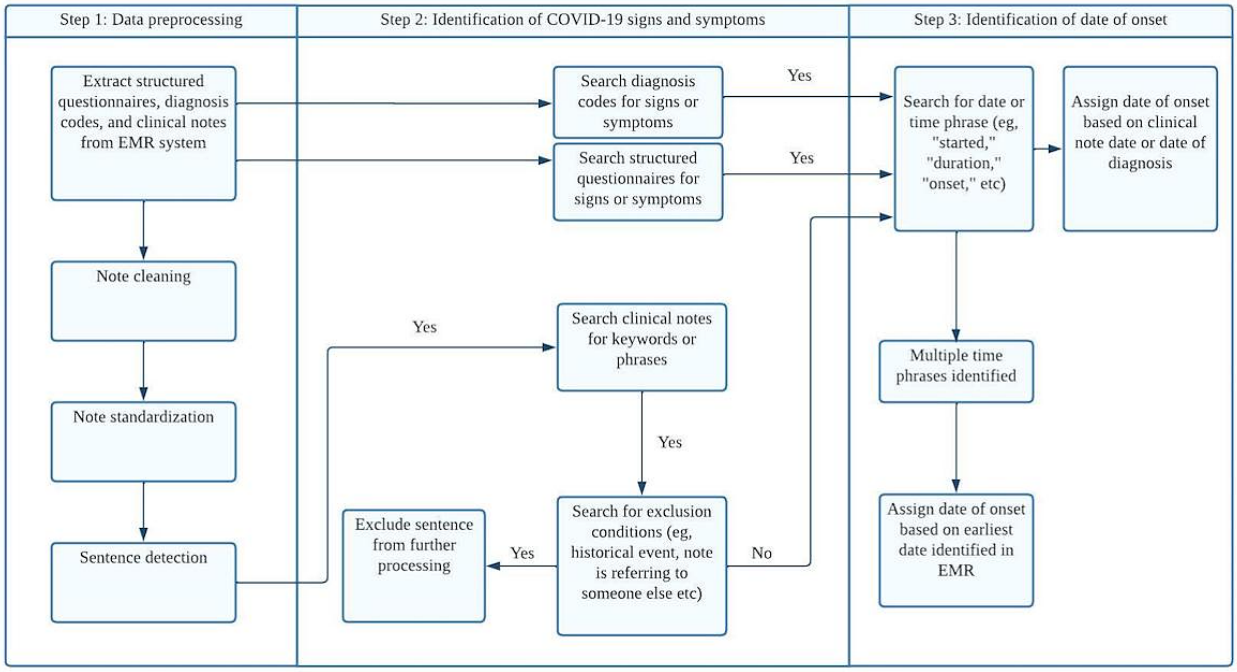
Signs or Symptoms of COVID-19

All EMR records were searched for 12 prespecified signs and symptoms within 30 days prior to and following the positive COVID-19 lab test order date. Signs and symptoms included fever, cough, headache, fatigue, dyspnea, chills, sore throat, myalgia, anosmia, diarrhea, vomiting or nausea, and abdominal pain, consistent with the Centers for Disease Control and Prevention (CDC) definitions [7,20]. If none of the above signs or symptoms were detected in the EMR, the patient was categorized as asymptomatic. Signs or symptoms were identified from the following three primary sources in the EMR: (1) ICD-10 diagnosis codes; (2) keywords or phrases in medical charts; or (3) COVID-19 lab order–related questionnaires. Keywords for signs and symptoms were predetermined in consultation with trained clinicians. The complete list of ICD-10 diagnosis codes and keywords or phrases used to identify signs and symptoms can be found in Table S1 in Multimedia Appendix 1.

NLP Algorithm Development

An NLP algorithm was developed to identify signs and symptoms of COVID-19 and to determine their corresponding onset dates from the EMR. The algorithm development process was implemented using a rule-based approach via Python 3.6 (Python Software Foundation). This was an iterative process in which the developed algorithm was refined to align with the reference standards derived through medical chart review and adjudication. The stages of NLP algorithm development are described below and summarized in Figure 1.

Figure 1. Flow diagram describing the natural language processing algorithm for detecting signs and symptoms of COVID-19. EMR: electronic medical records.



Step 1: Data Preprocessing

Clinical notes and structured data (diagnosis codes and symptom related questionnaires) within 30 days prior to or following the order date of the positive SARS-CoV-2 lab test were extracted from the KPSC EMR system. The extracted clinical notes were preprocessed through letter lowercase conversion, misspelled word correction, abbreviated word standardization, sentence separation, and tokenization (ie, segmenting text into linguistic units such as words and punctuation) [13].

Step 2: Identification of Signs and Symptoms

Patients were categorized as "Yes" for a particular symptom of interest under a set of prespecified situations (eg, if EMR notes contained a keyword or phrase related to a sign or symptom of interest, or if the patient answered "Yes" to a KPSC-administered medical questionnaire regarding COVID-19 symptoms). Keywords and phrases related to the 12 symptoms of interest were compiled by searching additional diagnosis terms and ontologies in the Unified Medical Language System [21] and were enriched by experienced clinicians and the training data set. Potential variants, abbreviations, and misspellings were also identified during algorithm development and manual chart review. For example, "shortness of breath" can be abbreviated as "sob" and "nausea/vomiting" as "n/v." Further misspellings and abbreviations are included in Table S1 in Multimedia Appendix 1. A regular expression was constructed to search and exclude sentences that contained a combination of preselected terms (eg, when notes refer to a *lack* of signs or symptoms or a *historical* medical event or indicate that signs or symptoms were experienced by someone else). A complete list of predefined sentence exclusion scenarios as well as "Yes" criteria for all signs and symptoms are provided in Table S2 in Multimedia Appendix 1.

Step 3: Date of Symptom Onset Determination

For each instance of identified signs or symptoms, the corresponding onset date was determined as either the clinical note date or by extracting the date from clinical notes under prespecified conditions, for example, where a date was detected with the symptom or followed with a phrase of "symptom (first) started," "Date of symptoms (onset):," "symptom onset date:," and "onset:" in unstructured notes. Specific examples of prespecified conditions are included in Table S2 and Table S3 in Multimedia Appendix 1. If signs or symptoms were identified from multiple clinical notes or structured data elements, the earliest date of symptom on record was assigned as the date of onset.

NLP Algorithm Validation

A sample of 100 randomly selected patients was used to assess the accuracy of the NLP algorithm in identifying each of the 12 signs or symptoms from unstructured EMR data, excluding patients used for the original algorithm development. Information on the presence or absence as well as the onset date of signs or symptoms were abstracted from EMR by trained chart abstractors using an abstraction manual. Patients for whom the sign or symptom complaint or onset date could not be clearly determined by the abstractors were further reviewed and adjudicated by a collaborating research physician. For this validation substudy, the manual chart review plus adjudicated results were deemed as the reference standard. The proportions of true positive, false positive, true negative, and false negative patients were used to estimate the sensitivity, specificity, positive predictive value (PPV), negative predictive value, and overall *F* score for each preselected sign or symptom of interest [22].

Sensitivity was defined as the proportion of patients correctly classified by the computerized NLP algorithm as experiencing the symptom of interest among patients identified with the sign or symptom by manual chart review. Specificity was the proportion of patients correctly classified as not experiencing the sign or symptom among individuals identified as not experiencing the sign or symptom according to chart review. PPV was the proportion of patients correctly classified as experiencing the sign or symptom of interest among those who were classified as experiencing the sign or symptom based on the NLP algorithm. Negative predictive value was the proportion of patients correctly classified as not experiencing the sign or symptom of interest among patients classified as not experiencing the sign or symptom based on the NLP algorithm. The *F* score for each comparison was calculated as (2 × PPV × sensitivity) / (PPV + sensitivity).

Statistical Analysis

We described patient characteristics and COVID-19 symptoms by mean, SD, median, and quartiles for continuous variables, and by frequency and percentage for categorical variables. Proportions of each symptom reported using structured EMR data were compared against proportions of each symptom identified through NLP-supplemented methods. Signs and symptoms were grouped into the following four categories according to the affected body system: respiratory (cough, sore throat, and dyspnea), systemic (fever, fatigue, chills, and myalgia), gastrointestinal (diarrhea, nausea or vomiting, and abdominal pain), and neurologic (headache and anosmia). We assessed the association between characteristics of interest and inconsistencies between traditional EMR analysis using structured data and NLP supplemented analysis. All analyses were performed using Python version 3.6 and SAS statistical software version 9.4 (SAS Institute).

Ethical Considerations

The study was reviewed by the CDC and was conducted consistent with applicable federal law and CDC policy—45 C.F.R. part 46.102(l)(2), 21 C.F.R. part 56; 42 U.S.C. Sect. 241(d); 5 U.S.C. Sect. 552a; 44 U.S.C. Sect. 3501 et seq. The study protocol was reviewed and approved by the KPSC Institutional Review Board (#12395) with a waiver of requirement for informed consent. Only authorized persons were provided access to individual-level patient data.

Results

Study Population

The study cohort included 359,938 patients with a positive SARS-CoV-2 laboratory test during March 2020-May 2021. Most patients were Hispanic (219,751/359,938, 61.0%), the mean age was 40.1 (SD 19.2) years, and approximately half (191,630/359,938, 53.2%) were female participants (Table 1). The most common comorbidities were hyperlipidemia (49,743/359,938, 13.8%), hypertension (48,637/359,938, 13.5%), and diabetes (41,591/359,938, 11.6%). The majority (252,869/359,938, 70.3%) of patients lived in census tracts with a median household income of less than US $80,000. Overall, 11.5% (41,307/359,938) of patients were enrolled in Medicaid.

Table 1. Baseline characteristics of the study population.

Characteristics	Values (N=359,938)
Sex, n (%)	
Female	191,630 (53.2)
Male	168,308 (46.8)
Race or ethnicity, n (%)	
Non-Hispanic White	72,705 (20.2)
Hispanic	219,751 (61.1)
Non-Hispanic Black	21,541 (6.0)
Non-Hispanic Asian	21,723 (6.0)
Non-Hispanic Pacific Islander	2362 (0.7)
Non-Hispanic Native American or Alaskan	639 (0.2)
Other or unknown	21,217 (5.9)
Age (years) at time of SARS-CoV-2 test, n (%)	
0-17	44,915 (12.5)
18-64	274,932 (76.4)
>65	40,091 (11.1)
Age (years), mean (SD)	40.4 (19.2)
Age (years), median (IQR)	40.0 (26.0, 55.0)
BMI, kg/m^2, n (%)	
<18.5	20,778 (5.8)
18.5-24.9	72,642 (20.2)
25.0-29.9	102,078 (28.4)
30.0-34.9	79,394 (22.1)
35.0-39.9	40,617 (11.3)
40.0-44.9	17,746 (4.9)
≥45.0	11,828 (3.3)
Missing	14,855 (4.1)
Tobacco use status, n (%)	
Current	9701 (2.7)
Former	50,013(13.9)
Never	226,518 (62.9)
Unknown	73,706 (20.5)
Comorbidities, n (%)	
Hyperlipidemia	49,743 (13.8)
Hypertension	48,637 (13.5)
Diabetes	41,591 (11.6)
Chronic pulmonary disease	21,254 (5.9)
Renal disease	10,298 (2.9)
Cancer	5401 (1.5)
Stroke	2937 (0.8)
Median annual household income [a] (US $), n (%)	
<40,000	41,352 (11.5)

Characteristics	Values (N=359,938)
40,000-79,999	211,517 (58.8)
≥80,000	106,886 (29.7)
Missing	183 (0.1)
Insurance, n (%)	
Medicaid	41,307 (11.5)
Medicare	36,013 (10.0)
Calendar period of SARS-CoV-2 test, n (%)	
March-May 2020	9138 (2.5)
June-August 2020	51,406 (14.3)
September-November 2020	54,936 (15.3)
December 2020-February 2021	233,707 (64.9)
March-May 2021	10,751 (3.0)

[a]Measured at the census tract level.

COVID-19 Signs and Symptoms

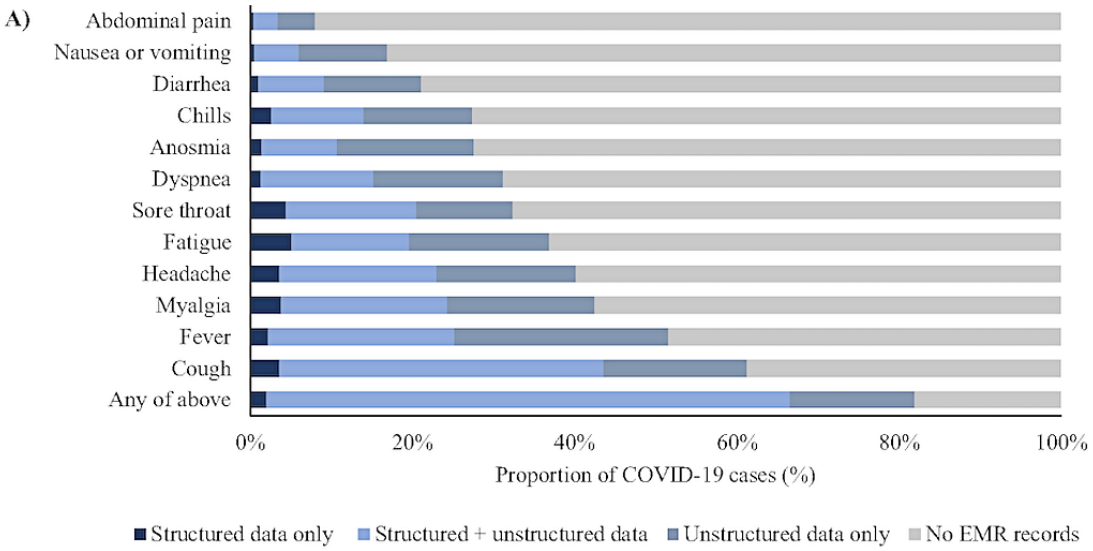
Supplementing structured EMR data with unstructured EMR data identified 55,568 additional symptomatic infections that were previously defined as asymptomatic based on structured data alone, representing 15.4% (55,568/359,938) of all infections. This proportion of additional identified symptomatic infections did not vary substantially by sex, age group, or race and ethnicity (Table S4 in Multimedia Appendix 1). However, there was an apparent decrease in the relative proportion of symptomatic infections identified with unstructured data during June-August 2020, whereby a higher proportion of all symptomatic cases (47,630/51,406, 92.7%) were identified via structured data compared to other time periods (60% [6456/10,751] to 80% [7336/9138]). In NLP-supplemented analyses, the symptoms ranged in frequency of reporting, from 8.0% (28,713/359,938) for abdominal pain to 61.3% (220,631/359,938) for cough. After cough, the most common symptoms identified in EMRs using NLP-supplemented analyses were fever (185,618/359,938, 51.6%), myalgia (154,042/359,938, 42.5%), headache (144,705/359,938, 40.2%), and fatigue (132,834/359,938, 36.9%; Figure 2A). NLP-supplemented analyses identified persons reporting each symptom that otherwise would not have been identified using structured data alone. For example, the proportion of SARS-CoV-2–positive persons reporting nausea and vomiting more than doubled, from 6.1% (21,981/359,938) in analysis restricted to structured data to 16.9% (60,865/359,938) in analyses supplementing this with NLP-derived fields from unstructured data.

NLP-supplemented analyses consistently identified additional signs and symptoms across all body systems relative to structured data alone, increasing the proportion of all SARS-CoV-2–positive patients identified with respiratory symptoms from 52.6% (189,146/359,938) to 69.4% (249,987/359,938), systemic symptoms from 44.4% (159,934/359,938) to 68.9% (247,988/359,938), neurological symptoms from 29.5% (106,243/359,938) to 52.1% (187,649/359,938), and gastrointestinal symptoms from 14.8% (53,193/359,938) to 31.4% (113,006/359,938; Table 2).

Among all 359,938 patients with positive SARS-CoV-2 results, 64,633 (18%) were not identified as symptomatic at any point over the study period based on the 12 preselected symptoms used in NLP-supplemented analyses (Table 2). Among all patients identified as reporting at least one symptom, the majority (252,466/295,305, 85.5%) were tested for SARS-CoV-2 following symptom onset, and 16,491 (4.6%) were tested on the same day as symptoms were reported (Table 2). Of the remaining 26,348 persons who reported symptoms after the SARS-CoV-2 test date, most (17,956/26,348, 68.1%) reported symptoms within the first 1-7 days following the SARS-CoV-2 test. Compared with structured data alone, NLP-supplemented analyses approximately doubled the proportion of identified symptomatic cases in the 6 to 30 days prior to SARS-CoV-2 sample collection (Figure 2B). The median time between the onset of first symptom and obtaining a test for SARS-CoV-2 was 3 days (IQR 1-6) for analysis restricted to traditional structured EMR data, and 4 days (IQR 2-9) for analysis supplemented with NLP algorithms.

NLP-supplemented analyses also increased the number of signs or symptoms identified per individual, often across multiple body systems. The proportion of patients reporting greater than 4 symptoms more than doubled in NLP-supplemented analysis compared to structured data alone, from 25.1% (90,202/359,938) to 53.1% (190,961/359,938) of all cases (Table 2). Similarly, the proportion of patients reporting symptoms related to 3 or more body systems increased from 22.6% (81,229/359,938) to 49.3% (177,440/359,938) after applying the NLP algorithm.

Figure 2. A comparison between structured and unstructured data. (A) Proportion of patients with SARS-CoV-2 with identified selected symptoms reported through structured and unstructured electronic medical records (EMR) data, by sign or symptom. (B) Days between testing and reported symptom onset before and after supplementing structured data with unstructured data (this includes IDC-10 codes, COVID-19 test-related questionnaires, and symptoms collected via keywords or phrases). ICD: International Classification of Diseases.



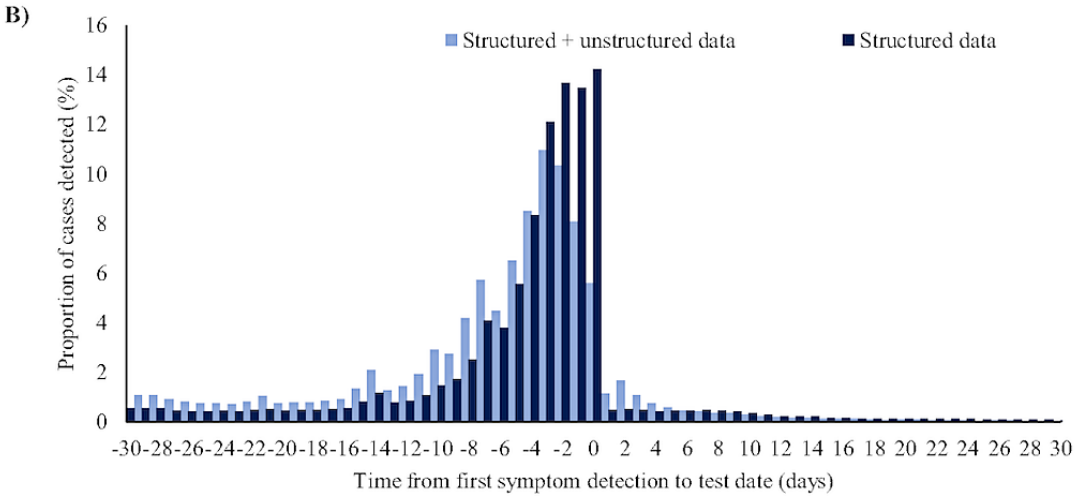


Table 2. COVID-19 characterization within 30 days prior to and after SARS-CoV-2 test date among all patients with confirmed SARS-CoV-2 infection (N=359,938), by data type.

Characteristics	Structured data	Structured and unstructured data
Days between testing and symptom onset[a], n (%)		
Testing 15-30 days after symptom onset	19,376 (5.4)	42,696 (11.9)
10-14 days after onset	12,751 (3.5)	28,317 (7.9)
7-9 days after symptom onset	19,896 (5.5)	37,325 (10.4)
4-6 days after symptom onset	42,368 (11.8)	57,569 (16.0)
1-3 days after symptom onset	94,157 (26.2)	86,559 (24.1)
Tested on same day as symptom onset	34,146 (9.5)	16,491 (4.6)
1-7 days before symptom onset	7949 (2.2)	17,956 (5.0)
8-14 days before symptom onset	5053 (1.4)	5147 (1.4)
15-30 days before symptom onset	4041 (1.1)	3245 (0.9)
No symptoms reported	120,201 (33.4)	64,633 (18.0)
Days between testing and symptom onset[a], mean (SD)	–3.96 (7.46)	–6.31 (8.49)
Days between testing and symptom onset[a], median (IQR)	–3.00 (–6.00, –1.00)	–4.00 (–9.00, –2.00)
Number of symptoms reported[a], n (%)		
None	120,201 (33.4)	64,633 (18.0)
1-3	149,535 (41.5)	104,344 (30.0)
4-6	72,929 (20.3)	111,132 (30.9)
7-9	16,164 (4.5)	65,037 (18.1)
10-12	1109 (0.3)	14,792 (4.1)
Body system Involved[a,b], n (%)		
Respiratory	189,146 (52.6)	249,987 (69.4)
Gastrointestinal	53,193 (14.8)	113,006 (31.4)
Systemic	159,934 (44.4)	247,988 (68.9)
Neurologic	106,243 (29.5)	187,649 (52.1)
Number of body systems involved[a], n (%)		
No symptoms reported	120,201 (33.4)	64,633 (18.0)
1	70,399 (19.6)	41,452 (11.5)
2	88,109 (24.5)	76,413 (21.2)
3	63,017 (17.5)	105,408 (29.3)
4	18,212 (5.1)	72,032 (20.0)

[a]Within 30 days prior to and after SARS-CoV-2 test date.

[b]Reported the percentage among the study cohort for each body system.

NLP Algorithm Validation

Compared to signs or symptoms identified using structured data only, NLP-supplemented analyses consistently returned a high proportion of true positive cases across the signs and symptoms studied, with PPV values of >95% for all symptoms except abdominal pain (75%). Sensitivity ranged from 87% for nausea or vomiting to 100% for cough, fever, anosmia, and abdominal pain (Table 3). Specificity ranged from 94.1% for chills to 100% (7 symptoms). *F* scores ranged from 0.86 to 1.00, with the majority being over 0.90. Regarding validation of onset time, 87% of onset dates identified by NLP were within +/- 3 days of those found by chart review; 70% were the same date (Table S5 in Multimedia Appendix 1).

Table 3. Performance measurements of natural language processing (NLP) algorithm to identify COVID-19 signs or symptoms, as compared with chart-confirmed validation data.

Sign or symptom	Chart review, (n/N)	TP[a] by NLP	TN[b] by NLP	FN[c] by NLP	FP[d] by NLP	Sensitivity[e] (%)	Specificity[f] (%)	PPV[g] (%)	NPV[h] (%)	*F* score[i]
Cough	76/100	76	23	0	1	100.0	100.0	98.7	95.8	1.00
Fever	73/100	73	23	0	4	100.0	100.0	94.8	85.2	0.97
Body ache	67/100	64	33	3	0	95.5	100.0	100.0	91.7	0.98
Headache	54/100	50	46	4	0	92.6	100.0	100.0	92.0	0.96
Fatigue	48/100	44	50	4	2	91.7	96.2	95.7	92.6	0.94
Dyspnea	40/100	38	60	2	0	95.0	100.0	100.0	96.8	0.97
Sore throat	49/100	46	51	3	0	93.9	100.0	100.0	94.4	0.97
Anosmia	35/100	35	65	0	0	100.0	100.0	100.0	100.0	1.00
Chills	36/100	32	64	4	0	88.9	94.1	100.0	100.0	0.94
Diarrhea	29/100	28	70	1	1	96.6	98.6	96.6	98.6	0.97
Nausea or vomiting	23/100	20	76	3	1	87.0	98.7	95.2	96.2	0.91
Abdominal pain	9/100	9	88	0	3	100.0	96.7	75.0	100.0	0.86

[a]TP: true positive.

[b]TN: true negative.

[c]FN: false negative.

[d]FP: false positive.

[e]The proportion of symptoms correctly classified by the computerized algorithm (TP) among all cases (TP+FN) ascertained by chart review.

[f]The proportion of cases correctly classified as absence of symptoms by the computerized algorithm (TN) among all individuals without symptom (TN+FP) according to chart review.

[g]PPV: positive predictive value—the proportion of symptom cases correctly classified (TP) among all those classified by the computerized algorithm (TP+FP).

[h]NPV: negative predictive value—the proportion of cases correctly classified as nonsymptom (TN) among all nonsymptom cases classified by the computerized algorithm (TN+FN).

[i]The overall accuracy of NLP algorithm in identifying each sign or symptom calculated as (2×PPV×sensitivity)/(PPV+sensitivity).

Discussion

Overview

Among more than 350,000 patients, this paper demonstrates that NLP algorithms can be used to extract unstructured data from EMR on COVID-19 signs and symptoms with enhanced detail and timeliness compared with structured data alone. To the authors' knowledge, this analysis represents the largest population study to date using NLP-based methods for identification and characterization of COVID-19 signs and symptoms.

Principal Findings

Overall, we observed that up to 60% of information on signs and symptoms may only be documented in the clinical narrative; however, this proportion varied widely between the conditions studied. Hence, previous real-world population studies that were limited to classical epidemiological methods (ie, using structured EMR data alone) may have underestimated the complexity and diversity of COVID-19 symptoms. This finding has important implications for patient care by improving our understanding of the whole spectrum and pathophysiology of COVID-19. This appeared particularly relevant for respiratory and gastrointestinal symptoms, whereby our data indicate that a significant proportion of symptomatic patients (24% and 53%, respectively) are overlooked when data are limited to structured components alone.

Comparison With Prior Work

Prior studies have noted similar improvements in COVID-19 case detection when clinical notes, ICD-10 diagnosis codes, and temperature fields have been used together, particularly for gastrointestinal conditions, rash or fever, and influenza-like illness syndromes, reporting almost double the sensitivity of detection [23,24]. The highest-quality evidence describing COVID-19 signs and symptoms to date has been derived from large meta-analyses that combine data from different study populations. In a large-scale meta-analysis including EMR data from over 4.5 million patients diagnosed with COVID-19 across 23 real-world health care databases [25], of the 6 signs or symptoms studied, cough, fever, and dyspnea were the most commonly identified. In general, this pattern was similar to the results presented in this paper; however, the proportions reported per symptom were significantly lower than those identified in this study with NLP-supplemented analyses. For example, whereas 32% was the highest proportion of patients identified

with a cough in the large meta-analysis, this study identified a total of 61% with cough in NLP-supplemented analyses.

Compared to a systematic review including EMR and self-reported symptom data pooled from 24,410 cases across 148 studies in 9 countries [10], we identified similar estimates for some signs and symptoms in this paper using NLP-supplemented analyses, such as cough (61% in this study vs 57%, respectively), fatigue (37% vs 31%), and anosmia (28% vs 25%). However, we observed a higher proportion of cases reporting most other prespecified symptoms, including dyspnea (31% vs 23%), sore throat (32% vs 12%), diarrhea (21% vs 10%), nausea or vomiting (17% vs 10%), abdominal pain (8% vs 4%), and headache (40% vs 13%). Importantly, gastrointestinal symptoms are increasingly being recognized as part of the COVID-19 spectrum, yet prior meta-analyses underestimate their prevalence compared with our work. One meta-analysis of 47 studies estimated diarrhea and nausea or vomiting in 7.7% and 7.8% patients with COVID 19 infection, respectively [26], and another analysis of 78 studies estimated a weighted pooled prevalence of 12.4% (95% CI, 8.2% to 17.1%) for diarrhea, 9.0% (95% CI, 5.5% to 12.9%) for nausea or vomiting, and 6.2% (95% CI, 2.6% to 10.3%) for abdominal pain [27]. In our study, approximately 21% (75,911/359,938) of patients with confirmed SARS-CoV-2 infection reported diarrhea, 17% (60,865/359,938) reported nausea or vomiting, and 8% (28,713/359,938) reported abdominal pain, all of which are higher estimates than have been reported in previous studies. Gastrointestinal involvement has been associated with delays in diagnosis compared with patients without digestive symptoms and hence may have been overlooked previously [28,29].

The observed discrepancies between this paper and prior evidence may be the direct result of the contribution of NLP algorithms when identifying COVID-19 signs and symptoms from EMR in this study, whereas prior studies have relied on structured components of EMR alone, such as ICD-10 diagnosis codes [25]. Among survey-based studies, results may be systematically biased due to responder bias or recall bias [30,31]. Importantly, study populations contributing to large meta-analyses and systematic reviews are heterogeneous with respect to their study populations and methodologies, with some restricted to symptomatic hospitalized patients [26,27,32]. Indeed, prior EMR- and survey-based studies restricted to hospitalized cases report higher frequencies of symptom complaints compared to this study [33,34]. This paper includes structured and unstructured EMR data from all care settings among a single diverse patient population of all ages, substantially expanding the scope compared with prior work.

Together, the findings presented here demonstrate the complexity of COVID-19, which often manifests as multiple diverse signs or symptoms across different body systems. With most prior large-scale real-world studies lacking unstructured EMR data, this observation may have been overlooked previously. As well as informing clinicians to guide patient care, understanding the complete array of signs or symptoms associated with COVID-19 could enhance population-level screening efforts. In addition, we found that NLP-supplemented analyses identified an earlier date of onset of potential COVID-19 signs and symptoms compared to traditional structured EMR data. Importantly, most of the transmission occurs within the first 5 days after symptom onset [35]. Therefore, by possibly facilitating identification of an earlier date of onset relative to test positivity at the population level, NLP methods could enhance public health surveillance systems, potentially informing preventive strategies to reduce community transmission.

Limitations

This study has at least 5 limitations, some of which are ubiquitous and unavoidable in observational research. First, while we capture symptoms occurring within 30 days of a COVID-19–positive test, it is possible that the reported symptoms detected in the EMR were due to other causes. However, chart review verified that the identified symptoms occurring within 20 days of testing were attributable to COVID-19 in the overwhelming majority of cases. Nevertheless, a comprehensive assessment of the overall usefulness of NLP would have involved a comparison with symptom reports in a SARS-CoV-2–negative population. Second, SARS-CoV-2 diagnostic tests were restricted to certain populations at differing points over the study period corresponding to periods of limited availability. As such, our estimates largely represented patients with symptomatic COVID-19 who sought medical care, and therefore it is likely that asymptomatic individuals were underrepresented in our analysis. Third, we defined symptomatic COVID-19 according to 12 conditions established as signs or symptoms of COVID-19 in the scientific literature; hence, it is possible that symptomatic cases reporting conditions outside of this established list are not counted as symptomatic. Fourth, the validation data set used in this paper included a relatively small sample size, which may have led to spurious findings. However, despite the small sample, the NLP algorithm performed well when identifying COVID-19 symptoms, producing similar sensitivity, *F* statistics, and PPV values to previously developed algorithms for symptom identification and COVID-19 characterization [18,36,37]. Lastly, this study was limited to insured individuals residing in Southern California from March 2020 to May 2021. Therefore, the findings may not be representative of or generalizable to other populations or to infections attributable to SAR-CoV-2 variants such as Delta or Omicron. However, the findings reported in this paper remain internally valid over the study period in demonstrating the overwhelming advantage of applying NLP to EMR for enhanced disease characterization across multiple clinical conditions.

Conclusions

This paper demonstrates that NLP can identify and characterize a broad set of COVID-19 signs and symptoms from medical records, with enhanced detail and timeliness, compared with prior EMR-based studies. These findings provide clear evidence that structured EMR data alone are incomplete for symptom capture, and NLP can enhance our understanding of the whole spectrum of disease pathophysiology. Further, as a scalable and timely method for disease characterization, NLP could strengthen COVID-19 surveillance beyond conventional surveillance systems.

Acknowledgments

The authors thank the patients of Kaiser Permanente for helping to improve care using information collected across electronic health record systems. We would like to thank the Kaiser Permanente Information Technology team for their assistance throughout. We would like to thank Michael Sheppard and Amanda Smith for offering their expertise on NLP.

Data Availability

COVID-19 cases were obtained from Kaiser Permanente Southern California (KPSC)'s electronic medical records with approval from KPSC Institutional Review Board. The data sets analyzed during this study are not publicly available due to their confidential nature.

Authors' Contributions

All authors contributed to the conception and design of the study; SYT, FX, BKA, and JS contributed to the development of the natural language processing algorithm; VH, LQ, HF, FX, SYT, VY, and JS contributed to acquisition, analysis, and interpretation of data; DM, SYT, BKA, VH, JS, VY, LQ, HF, SFS, SC, and FX contributed to drafting the work; DM, SYT, BKA, VH, JS, VY, LQ, HF, SFS, SC, and FX reviewed and contributed to the development of the final draft.

Conflicts of Interest

SYT received a grant from Roche/Genentech, Inc. to support this work. SYT, BKA, VH, JS, VY, LQ, HF, SFS, SC, and FX received support for research time with this funding. VY works for Roche-Genentech. The funder had no role in the design, conduct, or analysis of this study, or to manuscript development.

Multimedia Appendix 1
Supporting information.
[PDF File (Adobe PDF File), 246 KB - publichealth_v8i12e41529_app1.pdf]

References

1. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
2. WHO coronavirus (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2022-12-11]
3. Mao R, Qiu Y, He J, Tan J, Li X, Liang J, et al. Manifestations and prognosis of gastrointestinal and liver involvement in patients with COVID-19: a systematic review and meta-analysis. The Lancet Gastroenterology & Hepatology 2020 Jul;5(7):667-678 [FREE Full text] [doi: 10.1016/S2468-1253(20)30126-6] [Medline: 32405603]
4. Tenforde MW, Billig Rose E, Lindsell CJ, Shapiro NI, Files DC, Gibbs KW, CDC COVID-19 Response Team. Characteristics of Adult Outpatients and Inpatients with COVID-19 - 11 Academic Medical Centers, United States, March-May 2020. MMWR Morb Mortal Wkly Rep 2020 Jul 03;69(26):841-846 [FREE Full text] [doi: 10.15585/mmwr.mm6926e3] [Medline: 32614810]
5. Varatharaj A, Thomas N, Ellul MA, Davies NWS, Pollak TA, Tenorio EL, CoroNerve Study Group. Neurological and neuropsychiatric complications of COVID-19 in 153 patients: a UK-wide surveillance study. Lancet Psychiatry 2020 Oct;7(10):875-882 [FREE Full text] [doi: 10.1016/S2215-0366(20)30287-X] [Medline: 32593341]
6. Zheng Y, Ma Y, Zhang J, Xie X. COVID-19 and the cardiovascular system. Nat Rev Cardiol 2020 May 05;17(5):259-260 [FREE Full text] [doi: 10.1038/s41569-020-0360-5] [Medline: 32139904]
7. Symptoms of COVID-19. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html [accessed 2022-12-12]
8. Allen WE, Altae-Tran H, Briggs J, Jin X, McGee G, Shi A, et al. Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing. Nat Hum Behav 2020 Sep;4(9):972-982 [FREE Full text] [doi: 10.1038/s41562-020-00944-2] [Medline: 32848231]
9. 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;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
10. Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, et al. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. PLoS One 2020;15(6):e0234765 [FREE Full text] [doi: 10.1371/journal.pone.0234765] [Medline: 32574165]

11. Meystre SM, Heider PM, Kim Y, Davis M, Obeid J, Madory J, et al. Natural language processing enabling COVID-19 predictive analytics to support data-driven patient advising and pooled testing. J Am Med Inform Assoc 2021 Dec 28;29(1):12-21 [FREE Full text] [doi: 10.1093/jamia/ocab186] [Medline: 34415311]
12. Crabb BT, Lyons A, Bale M, Martin V, Berger B, Mann S, et al. Comparison of International Classification of Diseases and Related Health Problems, Tenth Revision Codes With Electronic Medical Records Among Patients With Symptoms of Coronavirus Disease 2019. JAMA Netw Open 2020 Aug 03;3(8):e2017703 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.17703] [Medline: 32797176]
13. Loper E, Bird S. NLTK: The Natural Language Toolkit. 2002 Presented at: Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics; July 07, 2002; Philadelphia, Pennsylvania, USA p. 63-70. [doi: 10.3115/1118108.1118117]
14. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 May 02;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]
15. Verspoor K, Cohen KB, Dredze M, Ferrara E, May J, Munro R, et al. Introduction to the 1st Workshop on Natural Language Processing for COVID-19 at ACL 2020. 2020 Presented at: Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020; July 9-10, 2020; Online.
16. Shams AB, Hoque Apu E, Rahman A, Sarker Raihan MM, Siddika N, Preo RB, et al. Web Search Engine Misinformation Notifier Extension (SEMiNExt): A Machine Learning Based Approach during COVID-19 Pandemic. Healthcare (Basel) 2021 Feb 03;9(2):156 [FREE Full text] [doi: 10.3390/healthcare9020156] [Medline: 33546110]
17. Matharaarachchi S, Domaratzki M, Katz A, Muthukumarana S. Discovering Long COVID Symptom Patterns: Association Rule Mining and Sentiment Analysis in Social Media Tweets. JMIR Form Res 2022 Sep 07;6(9):e37984 [FREE Full text] [doi: 10.2196/37984] [Medline: 36069846]
18. Wang J, Abu-El-Rub N, Gray J, Pham HA, Zhou Y, Manion FJ, et al. COVID-19 SignSym: a fast adaptation of a general clinical NLP tool to identify and normalize COVID-19 signs and symptoms to OMOP common data model. J Am Med Inform Assoc 2021 Jun 12;28(6):1275-1283 [FREE Full text] [doi: 10.1093/jamia/ocab015] [Medline: 33674830]
19. Koebnick C, Langer-Gould AM, Gould MK, Chao CR, Iyer RL, Smith N, et al. Sociodemographic characteristics of members of a large, integrated health care system: comparison with US Census Bureau data. Perm J 2012;16(3):37-41 [FREE Full text] [doi: 10.7812/TPP/12-031] [Medline: 23012597]
20. Burke RM, Killerby ME, Newton S, Ashworth CE, Berns AL, Brennan S, Case Investigation Form Working Group. Symptom Profiles of a Convenience Sample of Patients with COVID-19 - United States, January-April 2020. MMWR Morb Mortal Wkly Rep 2020 Jul 17;69(28):904-908 [FREE Full text] [doi: 10.15585/mmwr.mm6928a2] [Medline: 32673296]
21. Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. J Am Med Inform Assoc 1998;5(1):1-11 [FREE Full text] [doi: 10.1136/jamia.1998.0050001] [Medline: 9452981]
22. Goutte C, Gaussier E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. 2005 Presented at: ECIR 2005: Advances in Information Retrieval; March 21-23, 2005; Compostela, Spain p. 345-359. [doi: 10.1007/978-3-540-31865-1_25]
23. Bouchouar E, Hetman BM, Hanley B. Development and validation of an automated emergency department-based syndromic surveillance system to enhance public health surveillance in Yukon: a lower-resourced and remote setting. BMC Public Health 2021 Jun 29;21(1):1247 [FREE Full text] [doi: 10.1186/s12889-021-11132-w] [Medline: 34187423]
24. Ising A, Travers DA, Travers J, Kipp A, Waller AE. Triage note in emergency department-based syndromic surveillance. Advances in Disease Surveillance 2006;1:34 [FREE Full text]
25. Prieto-Alhambra D, Kostka K, Duarte-Salles T, Prats-Uribe A, Sena A, Pistillo A, et al. Unraveling COVID-19: a large-scale characterization of 4.5 million COVID-19 cases using CHARYBDIS. Res Sq 2021 Mar 01:369-384 [FREE Full text] [doi: 10.21203/rs.3.rs-279400/v1] [Medline: 33688639]
26. Sultan S, Altayar O, Siddique SM, Davitkov P, Feuerstein JD, Lim JK, AGA Institute. AGA Institute Rapid Review of the Gastrointestinal and Liver Manifestations of COVID-19, Meta-Analysis of International Data, and Recommendations for the Consultative Management of Patients with COVID-19. Gastroenterology 2020 Jul;159(1):320-334.e27 [FREE Full text] [doi: 10.1053/j.gastro.2020.05.001] [Medline: 32407808]
27. Tariq R, Saha S, Furqan F, Hassett L, Pardi D, Khanna S. Prevalence and Mortality of COVID-19 Patients With Gastrointestinal Symptoms: A Systematic Review and Meta-analysis. Mayo Clin Proc 2020 Aug;95(8):1632-1648 [FREE Full text] [doi: 10.1016/j.mayocp.2020.06.003] [Medline: 32753138]
28. Han C, Duan C, Zhang S, Spiegel B, Shi H, Wang W, et al. Digestive Symptoms in COVID-19 Patients With Mild Disease Severity: Clinical Presentation, Stool Viral RNA Testing, and Outcomes. Am J Gastroenterol 2020 Jun;115(6):916-923 [FREE Full text] [doi: 10.14309/ajg.0000000000000664] [Medline: 32301761]
29. Pan L, Mu M, Yang P, Sun Y, Wang R, Yan J, et al. Clinical Characteristics of COVID-19 Patients With Digestive Symptoms in Hubei, China: A Descriptive, Cross-Sectional, Multicenter Study. Am J Gastroenterol 2020 May;115(5):766-773 [FREE Full text] [doi: 10.14309/ajg.0000000000000620] [Medline: 32287140]

30. Lan F, Filler R, Mathew S, Buley J, Iliaki E, Bruno-Murtha LA, et al. COVID-19 symptoms predictive of healthcare workers' SARS-CoV-2 PCR results. PLoS One 2020;15(6):e0235460 [FREE Full text] [doi: 10.1371/journal.pone.0235460] [Medline: 32589687]
31. Couper MP, Gremel G, Axinn W, Guyer H, Wagner J, West BT. New options for national population surveys: The implications of internet and smartphone coverage. Soc Sci Res 2018 Jul;73:221-235. [doi: 10.1016/j.ssresearch.2018.03.008] [Medline: 29793688]
32. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
33. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, ISARIC4C investigators. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. BMJ 2020 May 22;369:m1985 [FREE Full text] [doi: 10.1136/bmj.m1985] [Medline: 32444460]
34. Talavera B, García-Azorín D, Martínez-Pías E, Trigo J, Hernández-Pérez I, Valle-Peñacoba G, et al. Anosmia is associated with lower in-hospital mortality in COVID-19. J Neurol Sci 2020 Dec 15;419:117-163 [FREE Full text] [doi: 10.1016/j.jns.2020.117163] [Medline: 33035870]
35. Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. Lancet Microbe 2021 Jan;2(1):e13-e22 [FREE Full text] [doi: 10.1016/S2666-5247(20)30172-5] [Medline: 33521734]
36. Alhussayni KH, Alshamery ES. Automated COVID-19 Dialogue System Using a New Deep Learning Network. PEN 2021 Apr 13;9(2):667. [doi: 10.21533/pen.v9i2.1862]
37. Ancochea J, Izquierdo JL, Soriano JB. Evidence of Gender Differences in the Diagnosis and Management of Coronavirus Disease 2019 Patients: An Analysis of Electronic Health Records Using Natural Language Processing and Machine Learning. J Womens Health (Larchmt) 2021 Mar;30(3):393-404. [doi: 10.1089/jwh.2020.8721] [Medline: 33416429]

Abbreviations

CDC: Centers for Disease Control and Prevention
EMR: electronic medical records
ICD: International Classification of Diseases
KPSC: Kaiser Permanente Southern California
NLP: natural language processing
PPV: positive predictive value

Edited by A Mavragani, T Sanchez; submitted 29.07.22; peer-reviewed by T Koleck, D Chrimes, Y Li; comments to author 18.10.22; revised version received 07.11.22; accepted 29.11.22; published 30.12.22.

Please cite as:
Malden DE, Tartof SY, Ackerson BK, Hong V, Skarbinski J, Yau V, Qian L, Fischer H, Shaw SF, Caparosa S, Xie F
Natural Language Processing for Improved Characterization of COVID-19 Symptoms: Observational Study of 350,000 Patients in a Large Integrated Health Care System
JMIR Public Health Surveill 2022;8(12):e41529
URL: https://publichealth.jmir.org/2022/12/e41529
doi:10.2196/41529
PMID:36446133

Original Paper

Crowdsourced Perceptions of Human Behavior to Improve Computational Forecasts of US National Incident Cases of COVID-19: Survey Study

David Braun[1], PhD; Daniel Ingram[2], MBA; David Ingram[2], CERA, FRM, PRM; Bilal Khan[3], PhD; Jessecae Marsh[1], PhD; Thomas McAndrew[4], PhD

[1]Department of Psychology, Lehigh University, Bethlehem, PA, United States
[2]Actuarial Risk Management, Austin, TX, United States
[3]Computer Science and Engineering, Lehigh University, Bethlehem, PA, United States
[4]College of Health, Lehigh University, Bethlehem, PA, United States

Corresponding Author:
David Braun, PhD
Department of Psychology
Lehigh University
17 Memorial Dr E
Bethlehem, PA, 18015
United States
Phone: 1 6107583000
Email: dab414@lehigh.edu

Abstract

Background: Past research has shown that various signals associated with human behavior (eg, social media engagement) can benefit computational forecasts of COVID-19. One behavior that has been shown to reduce the spread of infectious agents is compliance with nonpharmaceutical interventions (NPIs). However, the extent to which the public adheres to NPIs is difficult to measure and consequently difficult to incorporate into computational forecasts of infectious diseases. Soliciting judgments from many individuals (ie, crowdsourcing) can lead to surprisingly accurate estimates of both current and future targets of interest. Therefore, asking a crowd to estimate community-level compliance with NPIs may prove to be an accurate and predictive signal of an infectious disease such as COVID-19.

Objective: We aimed to show that crowdsourced perceptions of compliance with NPIs can be a fast and reliable signal that can predict the spread of an infectious agent. We showed this by measuring the correlation between crowdsourced perceptions of NPIs and US incident cases of COVID-19 1-4 weeks ahead, and evaluating whether incorporating crowdsourced perceptions improves the predictive performance of a computational forecast of incident cases.

Methods: For 36 weeks from September 2020 to April 2021, we asked 2 crowds 21 questions about their perceptions of community adherence to NPIs and public health guidelines, and collected 10,120 responses. Self-reported state residency was compared to estimates from the US census to determine the representativeness of the crowds. Crowdsourced NPI signals were mapped to 21 mean perceived adherence (MEPA) signals and analyzed descriptively to investigate features, such as how MEPA signals changed over time and whether MEPA time series could be clustered into groups based on response patterns. We investigated whether MEPA signals were associated with incident cases of COVID-19 1-4 weeks ahead by (1) estimating correlations between MEPA and incident cases, and (2) including MEPA into computational forecasts.

Results: The crowds were mostly geographically representative of the US population with slight overrepresentation in the Northeast. MEPA signals tended to converge toward moderate levels of compliance throughout the survey period, and an unsupervised analysis revealed signals clustered into 4 groups roughly based on the type of question being asked. Several MEPA signals linearly correlated with incident cases of COVID-19 1-4 weeks ahead at the US national level. Including questions related to social distancing, testing, and limiting large gatherings increased out-of-sample predictive performance for probabilistic forecasts of incident cases of COVID-19 1-3 weeks ahead when compared to a model that was trained on only past incident cases.

Conclusions: Crowdsourced perceptions of nonpharmaceutical adherence may be an important signal to improve forecasts of the trajectory of an infectious agent and increase public health situational awareness.

(JMIR Public Health Surveill 2022;8(12):e39336) doi:10.2196/39336

KEYWORDS

crowdsourcing; COVID-19; forecasting; human judgment

Introduction

Forecasting the transmission of infectious agents can support decisions made by public health officials and key decision makers [1,2]. Past forecasts of seasonal influenza, Ebola, dengue, chikungunya, and Zika have helped officials take short-term action to stymie the spread and burden of disease and draft policy decisions [3-8]. The COVID-19 pandemic has further highlighted the importance that forecasts play in support of public health situational awareness [9-11].

The majority of forecasts of an infectious disease are generated by computational models; however, past work has shown that human judgment is also capable of making accurate predictions of a diverse number of phenomena [12,13], including infectious agents [14-17].

Work in human judgment predictions can be categorized into direct and indirect predictions. Direct predictions are collected by asking humans to estimate the probability of a future event of interest. Researchers have used various methods to solicit direct predictions from a lay, expert, or mixed crowd by varying the format humans use to submit predictions and training different algorithms to combine individual forecasts [18-21]. Structured elicitation formalizes how a prediction should be collected to minimize potential biases or undue influences, and a researcher could use several different protocols to rigorously collect predictions [19,20,22].

Past work has found middling performance when asking those with subject matter expertise to make direct predictions [23,24]. As with experts, the performance of predictions made by lay people has been mixed, and the variability in predictive performance is likely due to cues in the environment that are related to the event of interest [25], as well as people's reliance on heuristics to make fast decisions with little information [26-29]. Humans are subject to several cognitive biases that negatively impact our ability to make sound judgments [30,31]. That said, there are many examples where predictions based on mental heuristics outperformed computational models [32].

Work on aggregating direct human judgment predictions has focused on adjusting for correlated predictions between individuals, assessing the number of individual predictions to combine, and determining how to appropriately weight individuals based on past predictive performance [18,21]. Direct predictions take advantage of a human's ability to build a prediction from available structured data and information typically unavailable to a computational model, such as subjective information, intuition, and expertise [33].

Indirect predictions of a future event are collected by (1) extracting human judgment data from a passive source such as social media [34-37], (2) actively asking a crowd about covariates that may be related to the target of interest, or (3) asking a crowd to take actions in a prediction market, which can be mapped to probabilistic predictions [38,39]. Indirect predictions offer an opportunity to train a statistical model on both measured objective data and subjective data.

Past work that incorporated social media data in a model often mapped behaviors to a set of random variables and included these random variables in a statistical model [34-37,40]. Most studies have framed these human and social media sources as passive signals that can be mined to contribute to more accurate forecasts. For example, a recent study leveraged mobility data gathered from Twitter to improve forecasts of incident COVID-19 cases at multiple geographic levels [41]. Digital interaction and engagement data beyond social media may be useful predictive signals as well, as a recent study found that Google search trends related to COVID-19 symptoms improved both nowcasting and forecasting of COVID-19 incident cases and deaths [42]. Compartmental models have also been proposed that take into account human behavior by estimating the contact network between individuals, and the reproductive and recovery rates, or by building a more complicated function between disease states that takes into account human behavior [43]. Prediction markets are another approach for aggregating human judgment, which ask a pool of participants to place bets on the potential of future events with an incentive for each participant to optimize their total earnings [38,39]. The goal of creating a prediction market is not to link behavior to outcomes of interest but to take advantage of an individual's ability to extract alternative data sources that are not accessible to computational models and respond to the aggregate behaviors of a market. Models that include indirect predictions report improved performance compared to models that do not include indirect predictions; however, performance varies by the infectious agent and type of data collected. Human behavior and perceptions can also be used to predict social media engagement and community behavior that might benefit decision-making of policy makers and community leaders. For example, past work has looked at which types of messages from organizations shared on Twitter foster the strongest public engagement [44], as well as which sources for health-related information are likely to be sought out based on demographics and how these factors contribute to adherence to social distancing guidelines [45].

In this work, we study how crowdsourced questions related to nonpharmaceutical interventions (NPIs) in one's community can contribute to an improved forecast of COVID-19 incident cases at the national level. We posed 21 questions related to NPIs to a representative sample from the United States over a period of 36 weeks. These crowdsourced data were used to estimate the association between perceptions of adherence to NPIs and incident cases at the US national level 1-4 weeks in advance. In addition, we fit a predictive model and showed that adding crowdsourced data on perceptions of adherence improves forecast accuracy for incident cases when compared to a control that does not include perceptive data.

To the current literature, we contribute a novel data stream of community-scale perceptive information [46] that shows (1) strong associations with incident cases 1-4 weeks ahead at the national level and (2) improved predictive accuracy of out-of-sample predictions 1-3 weeks ahead when included in a computational model.

Methods

Ethical Considerations

We obtained retroactive clearance from Lehigh University's institutional review board (IRB) to publish the data (#1808500-1). The IRB determined obtaining informed consent was not necessary because the data were recorded in such a manner that the identity of human subjects cannot be readily ascertained directly or through identifiers linked to the subjects. Data that have been made publicly available are similarly deidentified [46]. Participants completed surveys either (1) on a volunteer basis or (2) in exchange for compensation. Compensated participants earned credits from the survey platform that could be redeemed for gift cards or donated to charity.

Survey Logistics

Participants and Recruitment

There were 10,852 responses to the survey over the course of 36 weeks starting August 30, 2020, and ending April 28, 2021 (281 responses per week on average with an SD of 119). Paid participants were initially recruited through the *SurveyMonkey* platform (4405/10,852, 40.5%) from September 23, 2020, through February 15, 2021. SurveyMonkey is a survey platform with access to more than 140 million participants globally. The platform requires a fee per service and comes with assurance that paid participants will be a representative sample from the locale of interest. A survey can be sent to a set of participants who meet specific criteria (called a targeted audience), such as country of origin, age, socioeconomic factors (income, marital status, and employment), etc. Participants in this study were required to reside in the United States and be at least 18 years old. Survey design, distribution, and data collection were managed via SurveyMonkey software.

From February 16, 2021, to April 27, 2021, participants were recruited from the *Pollfish* survey platform (3295/10,852, 30.4%). This change was made due to SurveyMonkey delivering a highly variable number of responses per week and, in some weeks, failing to deliver the number of responses ordered. Pollfish is another fee per response survey platform that allows the researcher to specify a targeted audience and guarantees a representative number of responses. The goals and services of SurveyMonkey and Pollfish are similar, though Pollfish software collects higher resolution spatial data about respondents. The Pollfish platform collected responses from participants who met the same criteria as those for SurveyMonkey.

Compensated respondents from SurveyMonkey and Pollfish accounted for approximately 70% (7700/10,852, 71.0%) of the responses, and the final approximately 30% (3152/10,852, 29.0%) of participants were recruited as volunteers and participated through the SurveyMonkey platform from August 30, 2020, to April 28, 2021. These volunteers were mostly recruited via word of mouth and social media.

We removed participant responses from the analysis if (1) more than half of the questions (ie, 11 of the 21 questions) were left blank or had a response of "Don't know" (4.7% [511/10,852] of responses) or (2) a participant gave the same response to every question (2.3% [331/10,852] of responses). All blank and "Don't know" responses were excluded from the analysis (9.7% [20,569/214,200] of total question responses [ie, $N_{Participants}$×21]).

Survey Timeline

A total of 36 weekly surveys were sent to participants beginning on September 6, 2020, and ending on April 30, 2021. Surveys were distributed to unique participants each Monday, Wednesday, and Friday, and surveys were closed on Sundays. Surveys were not sent to the same participant more than once in a week.

SurveyMonkey surveys were open to participants for compensation from the 4th week of the survey period (September 2020) to the 21st consecutive week of the survey (February 2021), and SurveyMonkey surveys were open to volunteers over the entire 36-week survey period. Pollfish surveys were open to participants from the 21st week of the survey period (February 2021) until the 36th consecutive week of data collection (the end of the survey period; April 2021).

In July and August 2020, surveys were sent to participants to (1) fill out the survey and (2) solicit feedback about whether the questions asked in the survey were worded clearly. Feedback from these first 2 pilot surveys was used to update and finalize surveys sent between September 2020 and April 2021.

Survey Content and Questions

Surveys between September 2020 and April 2021 asked participants to answer the same set of 21 "core" questions (see Textbox 1 for a list of core questions). Core questions asked participants about their perceptions of their community members' adherence to NPIs, such as mask wearing, and their adherence to public health guidelines related to testing, quarantine, and large gatherings. Participants gave responses to survey questions on a Likert scale with the following options: "None/not adopted," "Few/20%," "Some/40%," "Many/80%," "All/100%," and "Don't know."

In addition to the 21 core questions, several weeks included topical questions asking participants about their perceptions of behavior during specific events (eg, the size of holiday gatherings). Because these questions were not consistent throughout the duration of the study, we chose not to include them in the analyses. At the end of the survey, participants were also asked for optional thoughts and feedback about how COVID-19 is being addressed in their community and how the survey may be improved in the future (for summary reports of the data composed in real time, see a previous report [47]).

The order in which questions were presented was randomized across all 21 questions in the Pollfish surveys, and SurveyMonkey questions were randomized within 5 categories

that asked participants about individual NPI behaviors, adherence to guidelines associated with community businesses, testing and quarantine, awareness, and restrictions or policies related to educational institutions (see Multimedia Appendix 1).

Textbox 1. List of the 21 "core" questions that were presented to participants in every survey from September 6, 2020, to April 30, 2021.

Questions
What percent of people in your community do you notice are usually:
1. Wearing a mask in public
2. Maintaining social distance
3. Staying at home
How common is it in your community for:
4. Restaurants to have reduced seating
5. Businesses to be closed – work from home only
6. Hairdressers and barbers to be open with restrictions
7. Visitors to senior living facilities to be restricted
8. Commonly touched surfaces to be sanitized
9. Hospitals to have special protection in areas that treat COVID patients
In your community, how common is it for people to follow recommendations or requirements to:
10. Get tested for active virus
11. Get antibody testing to detect prior infection
12. Quarantine people who have been in close contact with people with positive tests
13. Quarantine people with positive tests
14. Quarantine travelers from higher infection places
15. Limit large gatherings of people
How many people in your community are aware of:
16. Local level of COVID infections
17. Statewide targets for reducing COVID spread
18. Local approach to limiting COVID spread
In your state, what percent of:
19. Colleges are closed or holding only remote classes
20. Schools (K-12) are closed or holding only remote classes
21. Violations of COVID restrictions result in fines or police enforcement

Data Acquisition and Availability

Survey data were acquired retrospectively from a team of actuaries (Daniel Ingram and David Ingram) who were interested in the study of human behavior, crowdsourcing, and how perceptions may be predictive of the spread of SARS-CoV-2. There were several limitations to survey collection: (1) participant identifiers were not collected longitudinally and so we cannot track individuals who contributed to the survey, and (2) the wording of survey instructions was slightly different across the SurveyMonkey and Pollfish platforms, which could bias responses.

Individual respondent data of all 21 questions for all 36 weeks are available in a previous report [46]. The data are in wide format where each row represents a single survey response, and columns are present for the date the survey was completed and the 21 answers to survey questions.

We obtained approval from Lehigh University's IRB to publish these data on an open-source platform.

Epidemiological Data

Incident cases per epidemiological week (epidemic week) at the national level were collected from the Johns Hopkins University CSSE GitHub repository [48]. This repository stores cumulative cases per day from January 22, 2020, to the present for all 50 states and a set of 5 territories. To compute incident cases for day *D*, we subtracted cumulative cases at day *D* from cumulative cases at day *D*+1. We computed incident cases for day *D* at the national level by summing incident cases for all 50 states and all 5 territories. Daily incident cases at the national level were summed to arrive at incident cases per epidemic week, where an epidemic week began on Sunday and ended on Saturday.

Assessing Whether the Crowd was Representative of the US Population

We assessed graphically whether our sample was representative of the US population by plotting for all states (s) the pair (r_s,e_s), where r_s is the total number of observed participants for state s and e_s is the estimated expected number of responses from state s.

Our estimate e_s assumes that r_s was drawn from a random variable $R_s \sim \text{Bin}(N,\theta_s)$, where N is the total number of participants across all surveys and θ_s is the probability of choosing at random a citizen registered in state s. We estimated θ_s,  , as the census estimate for state s divided by the sum of census estimates for all states. The value e_s is  .

We included an estimated correlation coefficient between the observed and expected number of participants sampled across all states. For each state, we also compared the relative difference between the observed and expected proportions of participants (Multimedia Appendix 2).

Statistical Setup

We suppose a survey response to question q from participant i, at time t, $x_{t,i,q}$, was generated from a random variable $X_{t,i,q}$ which has support $supp(X_{t,i,q}) = \{0,1,2,3,4\}$ corresponding to 5 different levels of adherence. The value 0 corresponds to no adherence or adherence not adopted in the community, and the value 4 corresponds to complete adherence (the response "All/100%" on the survey). Random variables at time t for question q between 2 participants are considered independent.

Mean perceived adherence (MEPA) is defined for a specific question q and at a specific time t as the average of $x_{t,i,q}$ over participants, or

where N is the number of responses for question q at time t. $\text{MEPA}_{q,t}$ is intended to measure an aggregated adherence to a specific type of NPI. Though individual responses are discrete, $\text{MEPA}_{q,t}$ is a continuous value. If we define the random variable $\text{MEPA}_{q,t}$ as the average of N independent random variables with finite variance, then we expect $\text{MEPA}_{q,t}$ to have a bell-curved distribution that resembles the normal distribution restricted to the closed interval from 0 to 4.

Incident US national COVID-19 cases at epidemiological week t, (c_t), are assumed to be generated from a corresponding random variable C_t, and we make no additional assumptions about this time series.

Estimating the Correlation Between MEPA and Incident Cases

For each survey question, we estimated the correlation coefficient between MEPA at epidemiological week t and US national incident cases at epidemiological week t, t+1, t+2, t+3, and t+4. Line lists of the estimated correlation coefficient at each week-ahead time point and 95% CIs are available in Multimedia Appendix 3.

Clustering Questions

We fit a hierarchical clustering algorithm to all 21 MEPA time series for 2 through 10 clusters. Dissimilarity between 2 time series was computed using the Euclidean distance. The Silhouette coefficient was used to assess the quality of fitting 2 clusters, 3 clusters, and so on (up to 10 clusters) [49]. A dendrogram was plotted to visualize the clustering, and MEPA time series were grouped and plotted over the epidemiological week.

Forecast Models With and Without Crowdsourced Perceptions

SIR Plus Vector Autoregression Moving Average

An SIR (susceptible, infected, and removed) model was fit to the number of US incident cases to produce an estimated number of incident cases I_t, and residuals ($_t=c_t-I_t$) were modeled with a vector autoregression moving average (VARMA) model that included one or more MEPA time series.

The SIR model estimates at time t the number of individuals existing in the susceptible (S_t), infected (I_t), and removed (R_t) compartment according to

with initial values S_0, I_0, and R_0, and parameters $\beta>0$ and $\gamma>0$. We chose S_0 equal to the number of individuals in the United States, according to the most recent census. The initial value I_0 was set equal to the reported number of infections for the first epidemiological week in which survey data were collected (August 30, 2020, to September 05, 2020), and R_0 was set to 0. The initial value problem above was integrated by the Runge-Kutta-Fehlberg method, and parameters β and γ were estimated by minimizing the least squares solution between I_t and the reported number of incident cases (estimates of the SIR model at 4 different time points can be found in Multimedia Appendix 4).

Residuals were generated as $e_t=c_t-I_t$, and we assumed that these residuals together with one of the MEPA time series can be modeled as a VARMA model. VARMA assumes the residuals, and the MEPA time series M_q follows

$$\theta(L)Y_t = \psi(L)U_t$$

where $Y_t=[\ _t,\ m_{q,t}]'$, U_t is a random vector following a white noise process or $U_t \sim N(0,\Sigma)$, the operator $\theta(L)=B_1L+B_2L^2+\cdots$ and B_k is a matrix of coefficients, the operator $\psi(L)=A_1L+A_2L^2+\cdots$ and A_k is a matrix of coefficients, and the operator L^j is the lag operator or $L^jY_t=Y_{t-j}$. We assumed the covariance between any Y_s and Y_t is fixed and equal to Σ.

The optimal number of lags for θ and for ψ was estimated every week through each of the 36 weeks by computing the Akaike information criterion (AIC) for models fit with all combinations

of 1 through 3 lags for θ and 1 through 3 lags for ψ. The combination that resulted in the lowest AIC was picked.

SIR Plus Random Forest Plus VARMA

To incorporate all MEPA time series into a model, we first fit an SIR model to the original time series and computed the residuals $e_t=c_t-I_t$. Next, we trained a random forest regression f with 5000 trees, where the desired output is $_t$ as a function of e_{t-1}, and all the MEPA time series values, smoothed using LOWESS, with a lag of 1. The residuals $\delta_t=e5;_t-f(e_{t-1},\hat{M}_{1,t-1},\hat{M}_{2,t-1},\cdots,\hat{M}_{21,t-1})$, where $\hat{M}_{q,t}$ is the LOWESS smoothed MEPA time series value for question q at time t, were computed and were assumed to follow an autoregressive integrated moving average (ARIMA) process, or $\theta(L)\delta_t=\psi(L)u_t$. Lags were chosen at each week based on the AIC in the same manner as with the above SIR plus VARMA model.

Control Model

Our control model followed the same SIR "detrending" of the original incident case time series and then fit an ARIMA to the residuals. The ARIMA followed a similar approach as the VARMA model when modeling

$$Y_t \sim e_t$$

$$\Theta(L)Y_t=\psi(L)u_t$$

where $u_t \sim N(0, \sigma^2)$. The only addition to this model is that we may "difference" Y_t by successively subtracting the values of Y at time t-1 from the values of Y at time t for all times. The difference computes $d_t=\nabla Y_t=Y_t-Y_{t-1}$, fits the model above, generates forecasts of d_{t+1}, d_{t+2},⋯, and then recovers $Yt+l$ by computing $Y(t+l)-1 + d(t+l)$.

The ARIMA process is a first attempt model in many time series applications. If models that include MEPA variables cannot improve upon the above SIR plus ARIMA model, then MEPA may not add any predictive value over using lagged values of incident cases alone.

The above VARMA and ARIMA models were fit using the *statsmodels* package in Python [50].

Predictive Scoring

Forecasts were scored using the weighted interval score (WIS) over K central quantiles [51].

where the interval score ($IS_{\alpha k}$) is

and where F is a predictive cumulative distribution function, $1(x)$ is an indicator function, the value u represents the $(1-\alpha/2)$ quantile of F, l represents the $\alpha/2$ quantile of F, m represents the median or 0.50 quantile, and c is the eventually reported truth [52]. Moreover, weight w_0 equals 1/2 and $w_k=\alpha_k/2$.

The WIS and interval score are negatively sensed, with larger values indicating worse predictive performance compared to smaller values. The best possible WIS is 0, and the worst possible WIS is positive infinity.

Results

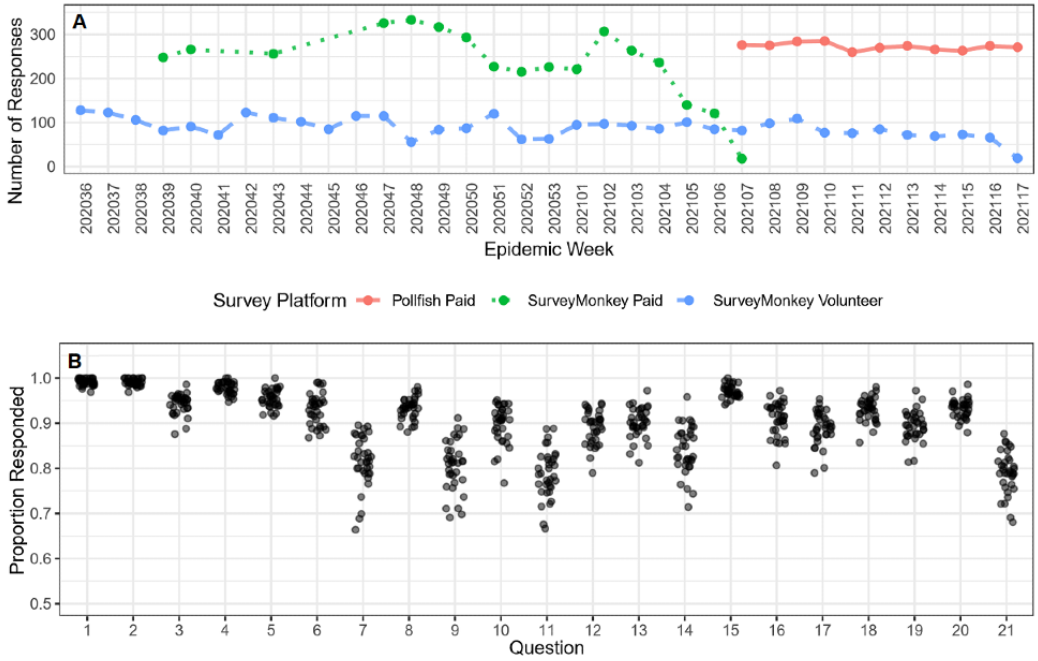
Overview

Comparison of the response rates across the 2 survey platforms (ie, SurveyMonkey and Pollfish) revealed that sample sizes each week were consistently higher following the switch to Pollfish. The sample was mostly geographically representative of the US population with slight oversampling in the Northeast. MEPA values were more variable at the beginning of the survey period than at the end, suggesting either that responses became more consistent over time or that larger sample sizes throughout the survey period resulted in lower response variability. A clustering analysis revealed that survey questions could be clustered into 4 groups based on question type, suggesting that future surveys might be more efficient by targeting these question types using fewer questions. A correlation analysis revealed reasonably strong correlations between several MEPA time series and incident COVID-19 cases 1-4 weeks ahead. Several MEPA time series also increased the predictive accuracy of a forecasting model of incident COVID-19 cases 1-4 weeks ahead.

Survey Platform Response Rates

SurveyMonkey surveys received an average of 236.06 (SD 81.14) compensated responses per week and an average of 88.80 (SD 22.68) volunteer responses per week, revealing that response rates for paid surveys were higher but more variable across weeks than volunteer survey responses. Pollfish surveys received an average of 272.55 (SD 7.80) compensated responses per week, and volunteer responses were not collected on the Pollfish system. Overall, sample sizes each week were consistently higher following the switch to Pollfish (Figure 1A).

Figure 1. (A) The number of participant responses per epidemic week for the Pollfish platform (red) and for those who submitted responses on SurveyMonkey who were compensated (green) and who were volunteers (blue). (B) The proportion of participants who responded to each question in a given epidemic week. Volunteers made consistent contributions each week as did the Pollfish participants who were compensated, while the number of compensated participant contributions on the SurveyMonkey platform varied. Questions with a lower proportion of responses corresponded to those questions that asked about nonpharmaceutical intervention behaviors that were more difficult to observe, such as visitation rules at senior living facilities (question 7), whether members of the community received antibody testing (question 11), and quarantine of recent travelers (question 14).



Question Response Rates

The mean percentage of questions that a participant answered was 87.89% (SD 6.15%) (Figure 1B). Questions 1 through 5 and question 15 were answered on average 94.98% (SD 1.47%) of the time, while questions 7, 9, 11, 14, and 21 had the lowest probability of responses, with an average response rate of 78.63% (SD 2.09%).

Representative Sampling

States from which most responses were collected included California (956/10,120, 9.5%), New York (876/10,120, 8.7%), Pennsylvania (678/10,120, 6.7%), Texas (645/10,120, 6.4%), and Florida (456/10,120, 4.5%).

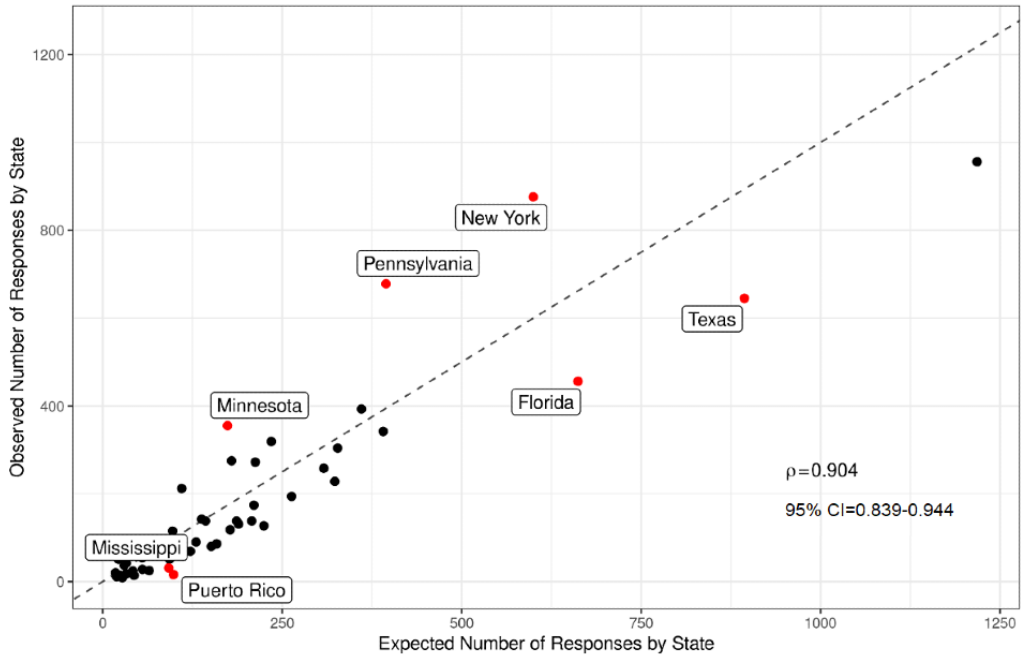
The correlation between the observed frequency of responses and expected frequency was 0.90 (95% CI 0.84-0.94; $P<.001$) and suggested that the response rates were proportional to the population at the state level. We compared for each state the proportion of observed responses to the proportion of individuals in that state according to the census (see Multimedia Appendix 2 for the observed proportion, expected proportion, and relative difference).

Seven states deviated from the expected response rates by more than 9 SDs. Four states were underrepresented (Mississippi, Puerto Rico, Florida, and Texas), and 3 states were overrepresented (Minnesota, Pennsylvania, and New York) (Figure 2). Pennsylvania was the most overrepresented state.

When both compensated and volunteer responses were included, the response frequency in Pennsylvania was 10 SDs above the expectation and when volunteer responses were removed the response frequency decreased to 3.5 SDs below the expectation.

To assess how switching survey platforms in the midst of data collection may have impacted the results, we analyzed whether the representativeness of the sample changed depending on the survey platform. We computed the average relative difference between expected and observed responses across all states, and compared this measure across survey platforms. This analysis revealed that the state residency of paid participants (ie, not volunteers) from SurveyMonkey was more representative of the US population (mean −0.599, SE 0.015) compared with the state residency of paid participants from Pollfish (mean −0.751, SE 0.019; $t_{51}=7.58$; $P<.001$).

Figure 2. The number of observed responses to the surveys summed over the survey period (vertical axis) compared to the expected number of total responses according to the census (horizontal axis). The dashed line indicates if the observed and expected numbers of responses equal one another. Some states are oversampled and undersampled.



MEPA Over Time

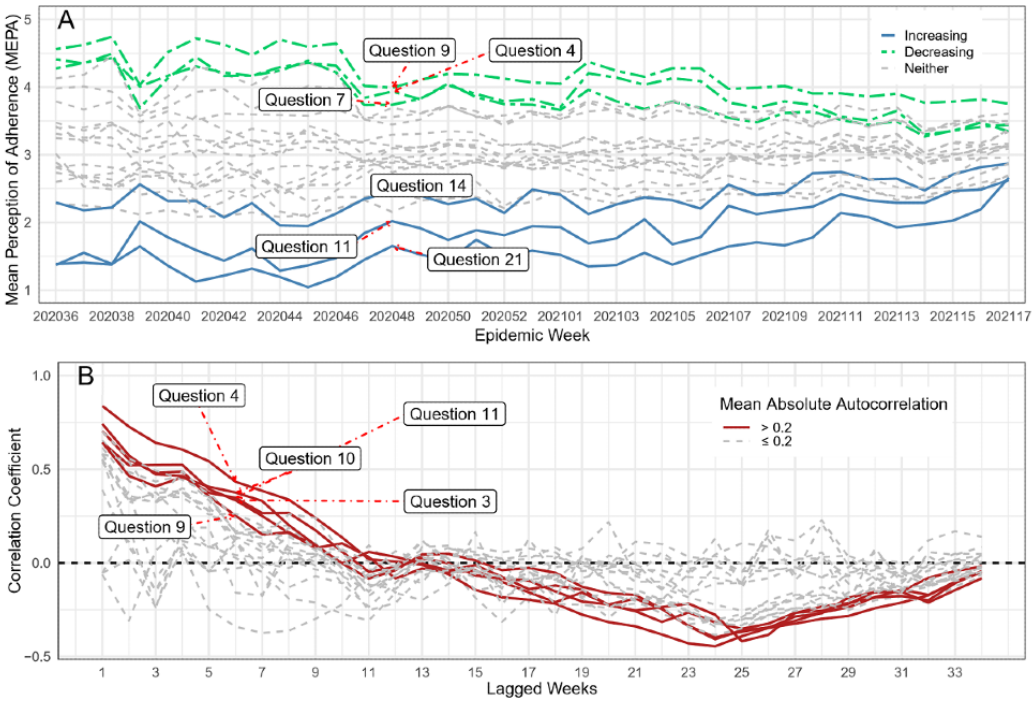
MEPA increased the most from the start to the end of the survey for the following 3 questions: question 21 ($\Delta mean_{21}=mean_{21,week36} - mean_{21,week1}=1.29$) that asked participants about knowledge of their state policies and whether "violations of COVID restrictions result in fines or police enforcement;" question 11 ($\Delta mean_{11}=1.24$) that asked how frequently community members follow recommendations to seek "[...] antibody testing to detect prior infection;" and question 14 ($\Delta mean_{14}=0.57$) that asked participants how frequently members of their community quarantine after traveling (Figure 3A).

MEPA decreased the most from the start to the end of the survey for the following 3 questions: question 7 ($\Delta mean_{7}=-1.07$) that asked participants how frequently restrictions are placed on visiting senior living facilities; question 4 ($\Delta mean_{4}=-0.83$) that asked how frequently restaurants have reduced seating capacity; and question 9 ($\Delta mean_{9}=-0.81$) that asked about the frequency of special protection in hospitals when treating patients with COVID-19.

The SD between MEPA values at the beginning of the survey period ($SD_{beginning}=0.89$) was larger than the SD between MEPA values at the end of the survey period ($SD_{end}=0.33$) (Figure 3A). The mean MEPA value over all 21 questions remained similar over the course of the survey ($mean_{beginning}=3.15$, $mean_{end}=3.14$). This result could be due to either a convergence in perceptions over time or reduced variability due to increased sample sizes throughout the survey period.

The estimated correlation between MEPA values at time t and $t-l$ was greater than 0.35 for lags of up to 4 weeks ($l=4$) for a majority of MEPA time series (Figure 3B) and suggested that many MEPA time series contain more structure than a random walk. Responses to the following 5 survey questions had a mean absolute autocorrelation greater than 0.2: question 3 ([...] staying at home), question 4 ([...] restaurants complying with Centers for Disease Control and Prevention [CDC] recommendations to have reduced seating), question 9 ([...] special protection in hospital areas that treat COVID patients), question 10 ([...] get tested for active virus), and question 11 ([...] get antibody testing to detect prior infection). The mean absolute autocorrelation for these 5 questions across 34 lagged weeks was above 0.2. A more detailed view of autocorrelation for a lag of 1 week has been provided in Multimedia Appendix 5.

Figure 3. (A) Mean perception of adherence (MEPA) for 21 questions asked over the survey period. (B) Autocorrelation for all 21 MEPA time series for a lag of 1 to 34 weeks. Perceptions of adherence for questions that asked about state policies (question 21) and antibody testing practices (question 11) show an increase over the survey period, while perceptions of adherence for questions that asked about restrictions placed on senior living facilities (question 7) and restaurants (question 4) show a decrease. Mean absolute autocorrelations for 5 questions across 34 lagged weeks are above 0.2. The estimated correlation between MEPA values at time t and t−l is greater than 0.35 for lags of up to 4 weeks (l=4) for a majority of MEPA time series. MEPA time series appear to contain more structure than a random walk, suggesting that crowdsourced perceptions may be a useful signal for predicting incident cases.



Clustering Questions According to Similarities in Responses Over Time

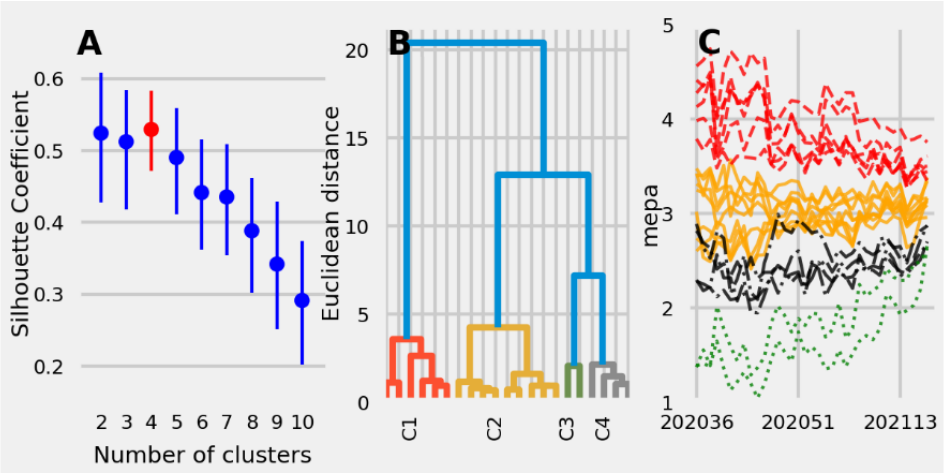
MEPA time series were grouped into the following 4 clusters (Figure 4A and B): (1) cluster of questions with values between 2.5 and 3.5 (ie, low to medium adherence; Figure 4C), (2) cluster with values that decreased over time (Figure 4C), (3) cluster with values near 2.25 at the beginning of the survey and that increased over time (Figure 4C), and (4) cluster with values near 1.25 at the beginning of the survey and that increased over time, ending above 2.50 by the end of the survey (Figure 4C).

Cluster quality as measured by the silhouette coefficient was the highest when grouping MEPA time series into 4 clusters; however, the silhouette coefficient for 4 clusters was similar to the silhouette coefficient for 2 and 3 clusters (Figure 4A). In the cluster in Figure 4C, there may exist 2 clusters—one with increasing adherence over time and another with decreasing adherence over time.

MEPA time series within the same cluster asked participants about similar adherence behaviors. Questions corresponding to avoidance behaviors (questions 2, 12, and 15) were more similar to one another than the other questions, as were questions that asked about limitations to businesses (questions 4 and 6), awareness of the high infectivity rate of the virus at a local level (questions 2 and 13), and awareness at the state level (questions 16 and 17). These results suggested that participants might have considered groups of questions in similar ways (eg, those related to avoidance), which suggests that future surveys might benefit from targeting these factors more directly.

For autocorrelations between MEPA responses 1-4 weeks ahead across the different clusters, see Multimedia Appendix 6.

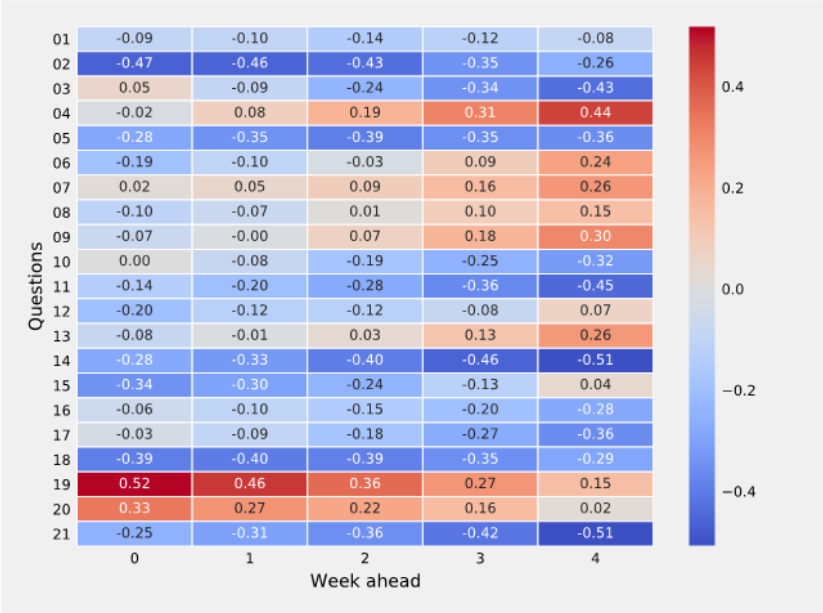
Figure 4. Hierarchical clustering of 21 mean perception of adherence (MEPA) time series using Euclidean distance as a measure of dissimilarity between 2 time series. (A) Silhouette coefficients for 2-10 clusters of MEPA time series. (B) Dendrogram that reports questions on the horizontal axis and dissimilarity between individual questions or clusters on the vertical axis. (C) MEPA time series clustered into 4 groups corresponding to the highest silhouette coefficient. Because MEPA time series can be separated into similar groups, a smaller survey may be able to capture the same patterns of the US public's perceptions of adherence to nonpharmaceutical interventions.



Correlation Between Perceptions of Adherence and Reported Incident Cases

The estimated correlation (ρ) between the MEPA time series representing responses to the question "What percent of people in your community do you notice are usually maintaining social distance?" and incident cases 1 week ahead was −0.46 (95% CI −0.69 to −0.15). Moreover, the correlation (ρ) was −0.3 (95% CI −0.67 to −0.12) for incident cases 2 weeks ahead, −0.35 (95% CI −0.61 to −0.02) for those 3 weeks ahead, and −0.26 (95% CI −0.55 to 0.08) for those 4 weeks ahead (Figure 5). The MEPA time series for the question "In your state, what percent of colleges are closed or holding only remote classes?" had an estimated correlation (ρ) of 0.46 (95% CI 0.15 to 0.69) for cases 1 week ahead. Moreover, the correlations (ρ) were 0.36 (95% CI 0.04 to 0.62), 0.27 (95% CI −0.07 to 0.55), and 0.15 (95% CI −0.19 to 0.46) for reported incident cases 2 weeks, 3 weeks, and 4 weeks ahead, respectively, at the US national level (Figure 5, row 19). Correlation coefficients and 95% CIs for each question are available in Multimedia Appendix 3. Taken together, these results show that changes in the perceptions of NPI compliance (ie, MEPA time series) are associated with changes in COVID-19 incident cases.

Figure 5. Linear correlation between 21 mean perception of adherence (MEPA) time series associated with questions about the perception of adherence and incident cases 1-4 weeks ahead at the US national level. The correlation between question 2 that asked "What percent of people in your community do you notice are usually wearing a mask in public?" and incident cases 1-4 weeks ahead was −0.26 or lower, and the correlation between question 19 that asked "In your state, what percent of colleges are closed or holding only remote classes?" and cases 1-3 weeks ahead was 0.27 or higher. Select crowdsourced perceptions of adherence to nonpharmaceutical interventions correlated with short-range and long-range reported incident cases at the national level.



Out-of-Sample Improvement in Forecasting With the Crowdsourced MEPA

Models that included both historical counts of US national incident cases and MEPA data changed the forecast trajectory and the width of prediction intervals compared to a model that only took into account the past time series of incident US national cases (Figure 6). The model that included historical counts and a random forecast regression incorporating all MEPA data proposed a similar trajectory to the ARIMA (control) model that included only case data, had wider prediction intervals before the peak of reported cases, and had a smaller prediction interval just after the peak of reported cases (Figure 6A and F).

The proportion of times a forecast that included a single MEPA time series generated a smaller (improved) WIS compared to a model that did not use MEPA, was above 50% for the majority of adherence questions for forecast horizons of 1-3 weeks ahead (Figure 7). MEPA most improved forecasts 2 weeks ahead. The MEPA time series corresponding to the questions "What percent of people usually stay home?" "How common do people follow recommendations to receive antibody testing?" and "How common do people in your community follow guidelines to limit large gatherings?" improved 76% (95% CI 58%-94%) of forecasts 2 weeks ahead. For 3 weeks ahead, the question "What percent of people usually stay home?" improved 76% (95% CI 58%-94%) of forecasts and the machine learning model that incorporated all adherence questions improved 76% (95% CI 58%-94%) of forecasts. Including MEPA data improved forecasts 4 weeks ahead minimally and for only a small set of questions.

Compared with the control model, including MEPA data improved forecast accuracy 1-4 weeks ahead (ie, reduced WIS) at and after the peak reported number of incident cases (Figure 8). Forecasts 1 week ahead showed consistent small gains in forecast accuracy over time (Figure 8A). Forecasts 2 and 3 weeks ahead showed large gains in forecast accuracy at and just after the peak number of incident cases (Figure 8B and C), and improvements in forecast accuracy 4 weeks ahead appeared near the peak number of cases (Figure 8D). Overall, these results revealed that certain perceptions of NPI compliance can be useful signals in a model predicting COVID-19 incident cases.

Figure 6. Forecasts of US national incident cases 1-4 week ahead at 6 time points throughout the survey period by first fitting an SIR (susceptible, infected, and removed) model and then modeling the residuals by (A) fitting an autoregressive model with 1 lag, (B-E) fitting a vector autoregression moving average that includes the residual time series and mean perception of adherence (MEPA) values for select questions, and (F) fitting a random forecast to residuals including MEPA values for all questions asked of participants plus an AR(1) model. AR(1): autoregression with lag of 1; ARIMA: autoregressive integrated moving average; RF: random forest.

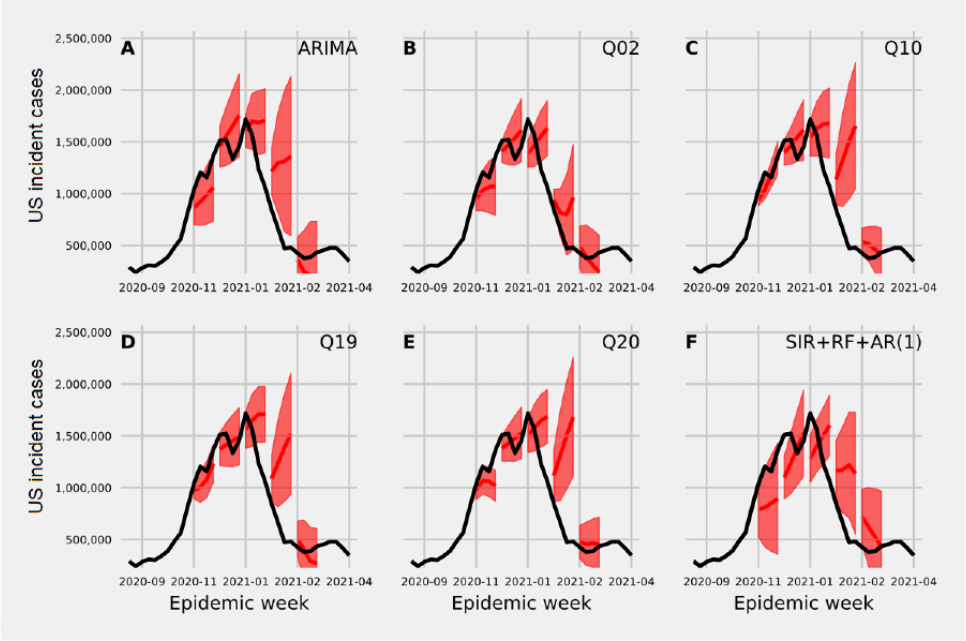


Figure 7. The proportion and 95% CIs of weighted interval scores (WISs) that were improved (smaller) for an SIR (susceptible, infected, and removed) plus vector autoregression moving average (VARMA) model that included mean perceived adherence (MEPA) time series 1-21 compared to the control SIR model without using a MEPA time series for forecasts 1-4 weeks ahead. An additional model, to the right of model number 21, is an SIR model plus a random forecast that includes all 21 MEPA time series and an ARIMA to model residuals. The majority of MEPA time series improved forecasts of incident cases 1 and 2 weeks ahead. A smaller number of MEPA time series improved forecasts 3 weeks ahead, and forecasts 4 weeks ahead were improved only modestly.

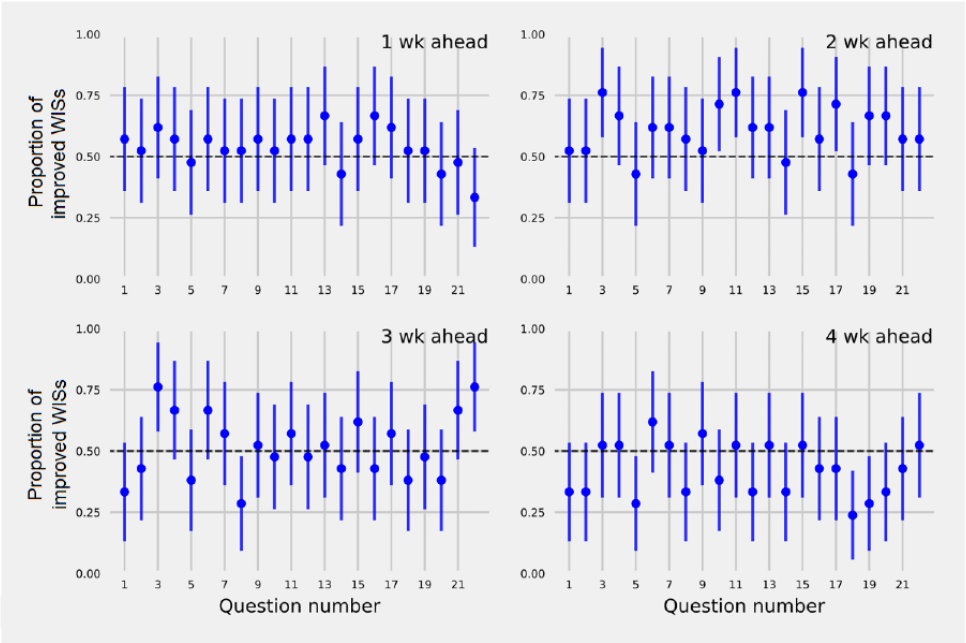


Figure 8. The differences in weighted interval scores (WISs) for forecasts of US national incident cases (A) 1 week ahead, (B) 2 weeks ahead, (C) 3 weeks ahead, and (D) 4 weeks ahead between models that included 1 mean perceived adherence time series and the control model that used only past incident case data to produce a forecast. The differences in WISs correspond to the forecasted epidemic weeks, not when the forecast was generated. The reported number of incident cases at the US national level is provided in grey. A point represents the difference in the WIS at the specific epidemic week and is colored red when a model weakens predictive performance and blue when this forecast improves upon the control model. Including perceptions of human behavior surrounding nonpharmaceutical interventions improves predictions at and after peak incident cases.

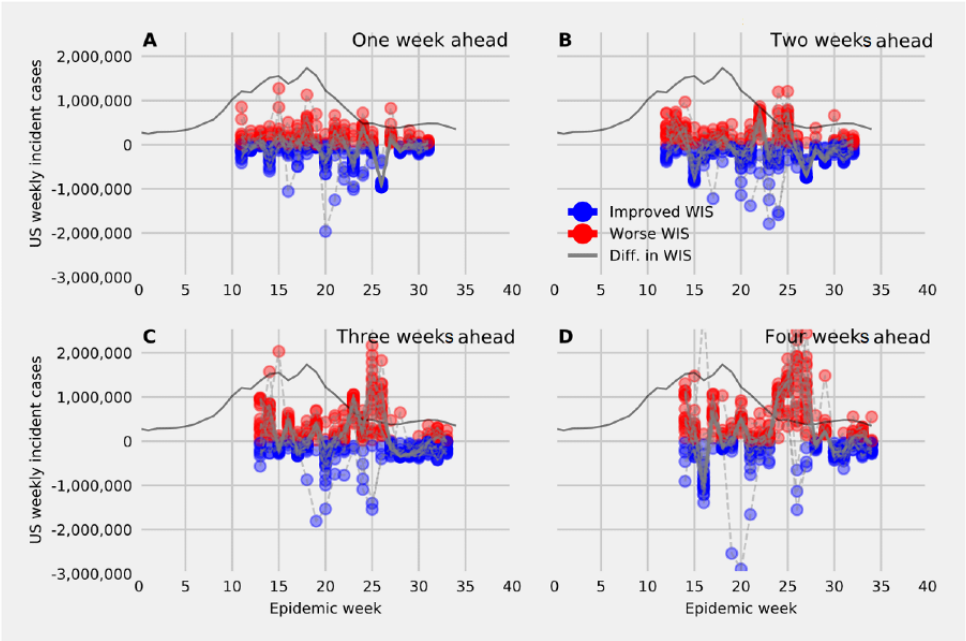


Figure 15. Inline graphic 7.

Discussion

We found that crowdsourced perceptions of adherence to NPIs correlated with incident cases 1-4 weeks ahead at the US national level and that including perceptual data into a computational model improved forecast accuracy 1-3 weeks ahead. Because responses from a crowd can be collected quickly (ie, within hours of distributing an online survey), these responses can be included into a computational model that could provide real-time weekly forecasts of epidemiological targets to organizations such as the CDC.

Since forecasts based on public perceptions are rapid and informative, these forecasts would be highly effective at times following the issuing of new NPI guidelines from state or federal agencies to assess the effectiveness of these new guidelines. Our models could reveal the extent to which people perceive public compliance with these guidelines and how changes in compliance impact the trajectory of an infectious agent, thereby informing public health officials about which interventions are able to curtail risk-seeking behaviors. These forecasts may also be valuable for policy makers and community leaders as they decide, for example, whether college classes should be held in person or remotely.

This work supports the hypothesis that a crowd may be able to assign realistic probabilities to outcomes about community adherence to NPIs in line with recent work, which has shown that lay people can elicit accurate probabilistic predictions of diverse real-world phenomena such as box-office income of a new movie or the impact of an infectious agent [13,53]; however, much more work needs to be completed to assess to what degree including human judgment perceptions improves the predictive accuracy of an infectious disease model (Multimedia Appendix 7). Past literature about lay people's ability to make accurate probabilistic predictions is mixed. Some past work suggests people may not be able to map environmental cues to accurate probabilities of outcomes [54], while other work has shown people's statistical intuitions may overlap with the statistics of their environment [53].

Evidence from this study suggests that participants were able to gauge what activities they were able to observe and predict, and at what spatial level they could make predictions. For instance, participants were given the option to reply "Don't know" or to leave questions blank. Participants responded more often to questions that were related to their environment, such as the proportion of people wearing masks, and responded less often to questions that were not related to their environment, such as restrictions on visitation to senior living centers. Survey questions during the initial pilot stage of the study asked participants to make predictions at the state level rather than community level, and many participants during this pilot stage protested that they could not make reasonable predictions at this level, suggesting that participants have some sense of how far a local community-level prediction could be extrapolated. Lastly, strong correlations between weekly responses to specific NPI questions indicated that the judgment of participants in this study was consistent (see Multimedia Appendix 5). Our results may support the idea that human judgment is predictive of incident cases because people can accurately perceive and make inferences about their surroundings.

However, relying on human judgment presents challenges that are absent when using computational models for prediction. Human judgment is susceptible to a wide array of biases often triggered by subtle changes in how a judgment prompt is presented [55]. Seemingly irrelevant information can have large impacts on judgment. For example, when asked to complete an irrelevant task, such as writing down the last 2 digits of their social security number before bidding on common items (such as a bottle of wine), people with higher social security numbers bid more money on wine than those with lower numbers [56]. Such findings underscore the importance of carefully crafting judgment questions to avoid activating judgment biases. Human judgment data must also be inspected for quality, as participants in this study often left one or more questions blank in a single survey and approximately 2% of participants gave the same response for every question, suggesting that they were not reading the survey items closely. Lastly, recruiting human participants demands time, effort, and money. Recruiting volunteers saves money but demands effort and implies an uncertain number of responses, which can be challenging when collecting data in response to a time-sensitive event such as an epidemic or pandemic. Participation rates in this study tended to increase throughout the data collection period, which created difficulties in assessing whether changes in MEPA over time were driven more by changes in perceived adherence or by changes in participation rates.

There are several limitations to address in future work. One limitation that we wish to overcome is that participants were not traced longitudinally, and so, we could not analyze how responses from individuals changed over time. Another limitation is that emails used to solicit volunteer participants

contained a link to a summary of the findings from previous months of data collection. While this may have added value to a participant's experience in the study, it may have biased their subsequent responses by anchoring their judgments to those summary values [56]. Another limitation arose from switching survey platforms (from SurveyMonkey to Pollfish) in the midst of data collection. The need for this switch was driven by a sudden decrease in the ability of SurveyMonkey to provide the requested number of paid responses each week (see Figure 1A). This switch seemed to have an impact on the geographical representativeness of the sample, as Pollfish provided a less representative sample than SurveyMonkey. Because switching survey platforms was confounded with both number of responses and epidemic week, the impact that switching survey platforms may have had on responses is largely unclear. Additionally, variable sampling rates across states created difficulties in estimating predictions at the state level. Oversampling from states with lower populations would ensure that a predictive model has sufficient data for estimating reliable predictions. No other demographic information was consistently collected throughout the surveys, and so, we were not able to assess whether the sample was representative for other demographic dimensions. Finally, there is evidence to suggest that self-expression may vary by geographic location [57]. Future research should consider how location and surrounding demographics may impact perceptions by, for example, leading to an overestimation of the prevalence of mask wearing in more densely populated areas.

Future research should explore whether more accurate and calibrated predictions of incident cases from human judgments can be made by matching the spatial scale of the questions posed to the crowd with the epidemiological target of interest. Instead of predicting incident cases at the national level, much stronger connections may be observed between state- or community-level judgments and state- or community-level incident cases. For example, one could investigate whether the accuracy of forecasts depends on factors such as the geographical size of the state (eg, Texas vs Delaware) or ethnic diversity (eg, California vs West Virginia). Additionally, respondents could be asked to judge compliance specifically at the level of their county, and then, these judgments could be added to a model that produces county-level predictions. Strong predictions at this local level would be valuable for community leaders when deciding, for example, whether a town hall meeting should be in person or remote. A significant challenge to estimating these local predictions is collecting enough responses from a given community over time, which, as mentioned above, can be remedied by targeting and oversampling from areas of interest to make local predictions. Future research should also explore whether perceptions of NPI compliance can predict other epidemiological targets. While we focused on incident cases in this study, our current methods should scale to other prediction outcomes of interest, such as COVID-19 hospitalizations and deaths.

Crowdsourced perceptions of human behavior, such as nonpharmaceutical adherence, may be a fast and informative signal that can improve probabilistic forecasts of the trajectory of an infectious agent and may have important implications for policy around infectious diseases.

Acknowledgments

We thank the following members of the COVID Mitigation Monitoring Project (CMMP): Dale Hall, FSA, CERA, MAAA, CFA; Robert Wolf, FSA, CFA, CERA; Kailan Shang, FSA, CFA, PRM, SCJP; and John Stark, FSA, CERA. We thank the Society of Actuaries for support in data collection. We also wish to thank all participants who offered their perceptions over the survey period.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey provided to participants to capture perceptions of adherence to nonpharmaceutical interventions.
[DOCX File , 942 KB - publichealth_v8i12e39336_app1.docx]

Multimedia Appendix 2
Observed and expected proportions of participants for each state.
[DOCX File , 26 KB - publichealth_v8i12e39336_app2.docx]

Multimedia Appendix 3
Correlation between mean perceived adherence and US national incident cases.
[DOCX File , 39 KB - publichealth_v8i12e39336_app3.docx]

Multimedia Appendix 4
SIR (susceptible, infected, and removed) model fit to US national incident cases.
[DOCX File , 68 KB - publichealth_v8i12e39336_app4.docx]

Multimedia Appendix 5
Autocorrelation of 1 week for mean perceived adherence time series.
[DOCX File , 447 KB - publichealth_v8i12e39336_app5.docx]

Multimedia Appendix 6
Bivariate relationships between question clusters and incident cases.
[DOCX File , 287 KB - publichealth_v8i12e39336_app6.docx]

Multimedia Appendix 7
Hypothesis testing across models in forecasting with crowdsourced mean perceived adherence.
[DOCX File , 20 KB - publichealth_v8i12e39336_app7.docx]

References

1. Lutz CS, Huynh MP, Schroeder M, Anyatonwu S, Dahlgren FS, Danyluk G, et al. Applying infectious disease forecasting to public health: a path forward using influenza forecasting examples. BMC Public Health 2019 Dec 10;19(1):1659 [FREE Full text] [doi: 10.1186/s12889-019-7966-8] [Medline: 31823751]
2. Pollett S, Johansson MA, Reich NG, Brett-Major D, Del Valle SY, Venkatramanan S, et al. Recommended reporting items for epidemic forecasting and prediction research: The EPIFORGE 2020 guidelines. PLoS Med 2021 Oct 19;18(10):e1003793 [FREE Full text] [doi: 10.1371/journal.pmed.1003793] [Medline: 34665805]
3. Del Valle SY, McMahon BH, Asher J, Hatchett R, Lega JC, Brown HE, et al. Summary results of the 2014-2015 DARPA Chikungunya challenge. BMC Infect Dis 2018 May 30;18(1):245 [FREE Full text] [doi: 10.1186/s12879-018-3124-7] [Medline: 29843621]
4. Evans MV, Dallas TA, Han BA, Murdock CC, Drake JM. Data-driven identification of potential Zika virus vectors. Elife 2017 Feb 28;6:e22053 [FREE Full text] [doi: 10.7554/eLife.22053] [Medline: 28244371]
5. Johansson M, Apfeldorf K, Dobson S, Devita J, Buczak A, Baugher B, et al. An open challenge to advance probabilistic forecasting for dengue epidemics. Proc Natl Acad Sci U S A 2019 Nov 26;116(48):24268-24274 [FREE Full text] [doi: 10.1073/pnas.1909865116] [Medline: 31712420]
6. Kobres P, Chretien J, Johansson MA, Morgan JJ, Whung P, Mukundan H, et al. A systematic review and evaluation of Zika virus forecasting and prediction research during a public health emergency of international concern. PLoS Negl Trop Dis 2019 Oct 4;13(10):e0007451 [FREE Full text] [doi: 10.1371/journal.pntd.0007451] [Medline: 31584946]
7. Reich NG, McGowan CJ, Yamana TK, Tushar A, Ray EL, Osthus D, et al. Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. PLoS Comput Biol 2019 Nov 22;15(11):e1007486 [FREE Full text] [doi: 10.1371/journal.pcbi.1007486] [Medline: 31756193]
8. Viboud C, Sun K, Gaffey R, Ajelli M, Fumanelli L, Merler S, RAPIDD Ebola Forecasting Challenge group. The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. Epidemics 2018 Mar;22:13-21 [FREE Full text] [doi: 10.1016/j.epidem.2017.08.002] [Medline: 28958414]
9. Biggerstaff M, Slayton RB, Johansson MA, Butler JC. Improving Pandemic Response: Employing Mathematical Modeling to Confront Coronavirus Disease 2019. Clin Infect Dis 2022 Mar 09;74(5):913-917 [FREE Full text] [doi: 10.1093/cid/ciab673] [Medline: 34343282]
10. Cramer E, Ray E, Lopez V, Bracher J, Brennen A, Castro Rivadeneira A, et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. Proc Natl Acad Sci U S A 2022 Apr 12;119(15):e2113561119 [FREE Full text] [doi: 10.1073/pnas.2113561119] [Medline: 35394862]
11. Evan LR, Nutcha W, Jarad N, Abdul HK, Katie H, Estee YC, et al. Ensemble Forecasts of Coronavirus Disease 2019 (COVID-19) in the U.S. MedRXiv. 2020. URL: https://www.medrxiv.org/content/10.1101/2020.08.19.20177493v1 [accessed 2022-11-16]
12. Hanea AM, Hemming V, Nane GF. Uncertainty Quantification with Experts: Present Status and Research Needs. Risk Anal 2022 Feb;42(2):254-263. [doi: 10.1111/risa.13718] [Medline: 33629402]
13. McAndrew T, Wattanachit N, Gibson GC, Reich NG. Aggregating predictions from experts: a review of statistical methods, experiments, and applications. Wiley Interdiscip Rev Comput Stat 2021 Jun 16;13(2):e1514 [FREE Full text] [doi: 10.1002/wics.1514] [Medline: 33777310]
14. Bosse NI, Abbott S, Bracher J, Hain H, Quilty BJ, Jit M, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Comparing human and model-based forecasts of COVID-19 in Germany and Poland. PLoS Comput Biol 2022 Sep;18(9):e1010405 [FREE Full text] [doi: 10.1371/journal.pcbi.1010405] [Medline: 36121848]
15. Farrow DC, Brooks LC, Hyun S, Tibshirani RJ, Burke DS, Rosenfeld R. A human judgment approach to epidemiological forecasting. PLoS Comput Biol 2017 Mar 10;13(3):e1005248 [FREE Full text] [doi: 10.1371/journal.pcbi.1005248] [Medline: 28282375]

16. McAndrew T, Reich NG. An expert judgment model to predict early stages of the COVID-19 pandemic in the United States. PLoS Comput Biol 2022 Sep;18(9):e1010485 [FREE Full text] [doi: 10.1371/journal.pcbi.1010485] [Medline: 36149916]
17. Recchia G, Freeman ALJ, Spiegelhalter D. How well did experts and laypeople forecast the size of the COVID-19 pandemic? PLoS One 2021 May 5;16(5):e0250935 [FREE Full text] [doi: 10.1371/journal.pone.0250935] [Medline: 33951092]
18. Clemen RT, Winkler RL. Combining Probability Distributions From Experts in Risk Analysis. Risk Analysis 2006 May 29;19(2):187-203. [doi: 10.1111/j.1539-6924.1999.tb00399.x]
19. Grigore B, Peters J, Hyde C, Stein K. EXPLICIT: a feasibility study of remote expert elicitation in health technology assessment. BMC Med Inform Decis Mak 2017 Sep 04;17(1):131 [FREE Full text] [doi: 10.1186/s12911-017-0527-0] [Medline: 28870196]
20. Hemming V, Burgman MA, Hanea AM, McBride MF, Wintle BC. A practical guide to structured expert elicitation using the IDEA protocol. Methods Ecol Evol 2017 Sep 05;9(1):169-180. [doi: 10.1111/2041-210x.12857]
21. Winkler RL, Grushka-Cockayne Y, Lichtendahl KC, Jose VRR. Probability Forecasts and Their Combination: A Research Perspective. Decision Analysis 2019 Dec;16(4):239-260. [doi: 10.1287/deca.2019.0391]
22. Cooke RM. Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford, United Kingdom: Oxford University Press; 1991.
23. Burgman MA, McBride M, Ashton R, Speirs-Bridge A, Flander L, Wintle B, et al. Expert status and performance. PLoS One 2011 Jul 29;6(7):e22998 [FREE Full text] [doi: 10.1371/journal.pone.0022998] [Medline: 21829574]
24. Spence MT, Brucks M. The Moderating Effects of Problem Characteristics on Experts' and Novices' Judgments. Journal of Marketing Research 2018 Dec 20;34(2):233-247. [doi: 10.1177/002224379703400204]
25. Stewart TR, Roebber PJ, Bosart LF. The Importance of the Task in Analyzing Expert Judgment. Organizational Behavior and Human Decision Processes 1997 Mar;69(3):205-219. [doi: 10.1006/obhd.1997.2682]
26. Kahneman D, Tversky A. Subjective probability: A judgment of representativeness. Cognitive Psychology 1972 Jul;3(3):430-454. [doi: 10.1016/0010-0285(72)90016-3]
27. Kahneman D, Tversky A. On the psychology of prediction. Psychological Review 1973 Jul;80(4):237-251. [doi: 10.1037/h0034747]
28. Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. Cognitive Psychology 1973 Sep;5(2):207-232. [doi: 10.1016/0010-0285(73)90033-9]
29. Tversky A, Kahneman D. Evidential impact of base rates. DTIC. 1981. URL: https://apps.dtic.mil/sti/pdfs/ADA099501.pdf [accessed 2022-11-16]
30. Gigerenzer G, Gaissmaier W. Heuristic decision making. Annu Rev Psychol 2011 Jan 10;62(1):451-482. [doi: 10.1146/annurev-psych-120709-145346] [Medline: 21126183]
31. Shah AK, Oppenheimer DM. Heuristics made easy: an effort-reduction framework. Psychol Bull 2008 Mar;134(2):207-222. [doi: 10.1037/0033-2909.134.2.207] [Medline: 18298269]
32. Wübben M, Wangenheim FV. Instant Customer Base Analysis: Managerial Heuristics Often "Get It Right". Journal of Marketing 2008 May;72(3):82-93. [doi: 10.1509/jmkg.72.3.82]
33. Ibrahim R, Kim S, Tong J. Eliciting Human Judgment for Prediction Algorithms. Management Science 2021 Apr;67(4):2314-2325. [doi: 10.1287/mnsc.2020.3856]
34. Bae S, (Christine) Sung E, Kwon O. Accounting for social media effects to improve the accuracy of infection models: combatting the COVID-19 pandemic and infodemic. European Journal of Information Systems 2021 Feb 25;30(3):342-355. [doi: 10.1080/0960085x.2021.1890530]
35. Hirose H, Wang L. Prediction of Infectious Disease Spread Using Twitter: A Case of Influenza. 2012 Presented at: Fifth International Symposium on Parallel Architectures, Algorithms and Programming; December 17-20, 2012; Taipei, Taiwan. [doi: 10.1109/PAAP.2012.23]
36. McDonald DJ, Bien J, Green A, Hu AJ, DeFries N, Hyun S, et al. Can auxiliary indicators improve COVID-19 forecasting and hotspot prediction? Proc Natl Acad Sci U S A 2021 Dec 21;118(51):e2111453118 [FREE Full text] [doi: 10.1073/pnas.2111453118] [Medline: 34903655]
37. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]
38. Polgreen PM, Nelson FD, Neumann GR. Use of prediction markets to forecast infectious disease activity. Clin Infect Dis 2007 Jan 15;44(2):272-279. [doi: 10.1086/510427] [Medline: 17173231]
39. Sell TK, Warmbrod KL, Watson C, Trotochaud M, Martin E, Ravi SJ, et al. Using prediction polling to harness collective intelligence for disease forecasting. BMC Public Health 2021 Nov 20;21(1):2132 [FREE Full text] [doi: 10.1186/s12889-021-12083-y] [Medline: 34801014]
40. 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]

41. Zeng C, Zhang J, Li Z, Sun X, Olatosi B, Weissman S, et al. Spatial-Temporal Relationship Between Population Mobility and COVID-19 Outbreaks in South Carolina: Time Series Forecasting Analysis. J Med Internet Res 2021 Apr 13;23(4):e27045 [FREE Full text] [doi: 10.2196/27045] [Medline: 33784239]
42. Rabiolo A, Alladio E, Morales E, McNaught AI, Bandello F, Afifi AA, et al. Forecasting the COVID-19 Epidemic by Integrating Symptom Search Behavior Into Predictive Models: Infoveillance Study. J Med Internet Res 2021 Aug 11;23(8):e28876 [FREE Full text] [doi: 10.2196/28876] [Medline: 34156966]
43. Funk S, Salathé M, Jansen VAA. Modelling the influence of human behaviour on the spread of infectious diseases: a review. J R Soc Interface 2010 Sep 06;7(50):1247-1256 [FREE Full text] [doi: 10.1098/rsif.2010.0142] [Medline: 20504800]
44. Singhal A, Baxi MK, Mago V. Synergy Between Public and Private Health Care Organizations During COVID-19 on Twitter: Sentiment and Engagement Analysis Using Forecasting Models. JMIR Med Inform 2022 Aug 18;10(8):e37829 [FREE Full text] [doi: 10.2196/37829] [Medline: 35849795]
45. Fridman I, Lucas N, Henke D, Zigler CK. Association Between Public Knowledge About COVID-19, Trust in Information Sources, and Adherence to Social Distancing: Cross-Sectional Survey. JMIR Public Health Surveill 2020 Sep 15;6(3):e22060 [FREE Full text] [doi: 10.2196/22060] [Medline: 32930670]
46. Braun D, Ingram D, Ingram D, Khan B, Marsh J, McAndrew T. Incorporating crowdsourced perceptions of human behavior into computational forecasts of US national incident cases of COVID-19. OSF Preprints. 2022. URL: https://osf.io/7vrmy [accessed 2022-11-16]
47. Ingram D, Ingram D. COVID-19 Mitigations in the U.S. September 2020 to April 2021. Society of Actuaries. URL: https://www.soa.org/globalassets/assets/files/resources/research-report/2021/covid-eight-month-report.pdf [accessed 2022-11-16]
48. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
49. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 1987 Nov;20:53-65. [doi: 10.1016/0377-0427(87)90125-7]
50. Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. SciPy. URL: https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf [accessed 2022-11-16]
51. Bracher J, Ray EL, Gneiting T, Reich NG. Evaluating epidemic forecasts in an interval format. PLoS Comput Biol 2021 Feb;17(2):e1008618 [FREE Full text] [doi: 10.1371/journal.pcbi.1008618] [Medline: 33577550]
52. Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. J Royal Statistical Soc B 2007 Apr;69(2):243-268. [doi: 10.1111/j.1467-9868.2007.00587.x]
53. Griffiths TL, Tenenbaum JB. Optimal predictions in everyday cognition. Psychol Sci 2006 Sep 17;17(9):767-773. [doi: 10.1111/j.1467-9280.2006.01780.x] [Medline: 16984293]
54. McDowell M, Jacobs P. Meta-analysis of the effect of natural frequencies on Bayesian reasoning. Psychol Bull 2017 Dec;143(12):1273-1312. [doi: 10.1037/bul0000126] [Medline: 29048176]
55. Tversky A, Kahneman D. Rational Choice and the Framing of Decisions. In: Karpak B, Zionts S, editors. Multiple Criteria Decision Making and Risk Analysis Using Microcomputers. NATO ASI Series, vol 56. Berlin, Heidelberg: Springer; 1989:81-126.
56. Ariely D, Loewenstein G, Prelec D. "Coherent Arbitrariness": Stable Demand Curves Without Stable Preferences. The Quarterly Journal of Economics 2003 Feb 01;118(1):73-106. [doi: 10.1162/00335530360535153]
57. 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]

Abbreviations

AIC: Akaike information criterion
ARIMA: autoregressive integrated moving average
CDC: Centers for Disease Control and Prevention
IRB: institutional review board
MEPA: mean perceived adherence
NPI: nonpharmaceutical intervention
SIR: susceptible, infected, and removed
VARMA: vector autoregression moving average
WIS: weighted interval score

Edited by A Mavragani, G Eysenbach; submitted 06.05.22; peer-reviewed by R Gore, S Yang; comments to author 12.09.22; revised version received 08.10.22; accepted 09.10.22; published 30.12.22.

Please cite as:
Braun D, Ingram D, Ingram D, Khan B, Marsh J, McAndrew T
Crowdsourced Perceptions of Human Behavior to Improve Computational Forecasts of US National Incident Cases of COVID-19: Survey Study
JMIR Public Health Surveill 2022;8(12):e39336
URL: https://publichealth.jmir.org/2022/12/e39336
doi:10.2196/39336
PMID:36219845

Original Paper

Factors Influencing Adoption and Use of Telemedicine Services in Rural Areas of China: Mixed Methods Study

Yumeng Du[1*], MSc; Qiru Zhou[2*], MD; Weibin Cheng[1,3*], MD; Zhang Zhang[4,5], MSc; Samantha Hoelzer[4], MSc; Yizhi Liang[6,7], MSc; Hao Xue[8], PhD; Xiaochen Ma[9], PhD; Sean Sylvia[4,5], PhD; Junzhang Tian[1], MD; Weiming Tang[1,6,10], MD, PhD

[1]Institute for Healthcare Artificial Intelligence Application, Guangdong Second Provincial General Hospital, Guangzhou, China

[2]Internet Hospital, Guangdong Second Provincial General Hospital, Guangzhou, China

[3]School of Data Science, City University of Hong Kong, Hong Kong, China

[4]Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[5]Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[6]University of North Carolina at Chapel Hill Project-China, Guangzhou, China

[7]Department of Global Health and Population, Harvard University, Boston, MA, United States

[8]Stanford Center for China's Economy and Institutions, Stanford University, Stanford, CA, United States

[9]China Center for Health Development Studies, Peking University, Beijing, China

[10]Institute for Global Health and Infectious Disease, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[*]these authors contributed equally

Corresponding Author:
Weiming Tang, MD, PhD
Institute for Healthcare Artificial Intelligence Application
Guangdong Second Provincial General Hospital
No.466 Xingangzhong Road
Haizhu District
Guangzhou, 510317
China
Phone: 86 15920567132
Email: weiming_tang@med.unc.edu

Abstract

Background: The shortage of medical resources in rural China reflects the health inequity in resource-limited settings, whereas telemedicine could provide opportunities to fill this gap. However, evidence of patient acceptance of telemedicine services from low- and middle-income countries is still lacking.

Objective: We aimed to understand the profile of patient end-user telemedicine use and identify factors influencing telemedicine service use in rural China.

Methods: Our study followed a mixed methods approach, with a quantitative cross-sectional survey followed by in-depth semistructured interviews to describe telemedicine use and its associated factors among rural residents in Guangdong Province, China. In the quantitative analysis, explanatory variables included environmental and context factors, household-level factors, individual sociodemographic factors, access to digital health care, and health needs and demand factors. We conducted univariate and multivariate analyses using Firth logistic regression to examine the correlations of telemedicine uptake. A thematic approach was used, guided by the Social Cognitive Theory for the qualitative analysis.

Results: A total of 2101 households were recruited for the quantitative survey. With a mean age of 61.4 (SD 14.41) years, >70% (1364/2101, 72.94%) of the household respondents were male. Less than 1% (14/2101, 0.67%) of the respondents reported experience of using telemedicine. The quantitative results supported that villagers living with family members who had a fever in the past 2 weeks (adjusted odds ratio 6.96, 95% CI 2.20-21.98; *P*=.001) or having smartphones or computers (adjusted odds ratio 3.71, 95% CI 0.64-21.32; *P*=.14) had marginally higher telemedicine uptake, whereas the qualitative results endorse these findings. The results of qualitative interviews (n=27) also supplemented the potential barriers to telemedicine use from the lack of knowledge, trust, demand, low self-efficacy, and sufficient physical and social support.

Conclusions: This study found extremely low use of telemedicine in rural China and identified potential factors affecting telemedicine uptake. The main barriers to telemedicine adoption among rural residents were found, including lack of knowledge, trust, demand as well as low self-efficacy, and insufficient physical and social support. Our study also suggests strategies to facilitate telemedicine engagement in low-resource settings: improving digital literacy and self-efficacy, building trust, and strengthening telemedicine infrastructure support.

(JMIR Public Health Surveill 2022;8(12):e40771) doi:10.2196/40771

KEYWORDS

telemedicine; telehealth; rural residents; mixed methods; China; mobile phone

Introduction

Background

Limited medical resources in rural areas remain a considerable challenge in China, which worsens the health inequity in resource-limited settings [1]. With the shortage of licensed doctors nationwide, there were only 1.56 village doctors and assistants per 1000 rural residents on duty in 2020 [2]. Furthermore, access to quality rural health care services is predicted to degenerate owing to the retirement of current practitioners nearing retirement and emerging opportunities for new health care workers in urban areas [3]. However, booming mobile internet communication and expansion of internet medical services in China suggest a direction for future health care services [4,5]. Telemedicine has the potential to partly fill this huge health care services gap in the rural areas.

As developments of telemedicine in primary care have been boosted since the COVID-19 pandemic crisis [6], the World Health Organization has launched the Global Strategy on Digital Health 2020 to 2025, highlighting the application of digital health technologies for consumers, health professionals, health care providers, and industry to strengthen health systems [7]. The World Health Organization defines telemedicine as “the delivery of health care services using information and communication technologies for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education, all in the interests of advancing the health of individuals and their communities” [8]. Owing to its ability to overcome geographic obstacles to high-quality health care while reducing time and financial costs [5], telemedicine service development is proposed to address medical resource maldistribution in rural areas, especially in the era of the COVID-19 pandemic [9,10]. The high coverage of smartphones and rural telecommunications infrastructure in China have been regarded as favorable conditions for telemedicine accessibility to reach the marginalized and underserved populations [4,11].

Review of Previous Literature

Understanding the overview of telemedicine use among rural residents is essential for identifying the opportunities and challenges of the adoption of eHealth programs in rural areas. Existing systematic reviews from countries other than China provide evidence regarding factors affecting telemedicine service use from both qualitative and quantitative perspectives [12-14]. Potential barriers such as sociodemographic factors (eg, older age, females, low income, less educated, and physical or mental disability), knowledge or cognitive factors (eg, lack of awareness and information and communication technology skills, lack of demand or motivation, lack of trust in web-based services, and perceived cost), and contextual factors (eg, lack of access to equipment, internet connection, and social support) have been noted [12-14]. However, data on telemedicine or telehealth use from low- and middle-income countries (LMICs) in rural areas were underrepresented [12-14]. In China, most telemedicine programs remained in the pilot stages in metropolitan areas [15,16], and the surveys on telemedicine service use were typically collected from medical professionals or small convenience samples [17,18]. There is a lack of opportunity to study patient end-user acceptance of telemedicine service and its associated factors in rural China.

Objectives

This study had the following aims: first, to understand the profile of patient end-user telemedicine use among Chinese rural residents, and second, to identify factors influencing telemedicine service use by quantitative and qualitative approaches in rural China.

Methods

Study Settings and Design

The prefecture-level cities of Meizhou and Heyuan were selected as study sites. The gross domestic product (GDP) per capita for these cities ranked last and third to last in 2020, respectively, among all cities in Guangdong Province [19]. The GDP per capita of Meizhou City was ¥31,011 (US $4496) in 2020, whereas that of Heyuan was ¥38,802 (US $5625). For comparison, the World Bank classifies economies with per-capita gross national incomes between US $4096 and US $12,695 in 2020 as upper middle income [20]. To address the problem of insufficient primary health care services in rural areas, the Guangdong Provincial Health Commission started to provide smart health monitoring equipment packages to village clinics in 2277 designated “low-income” villages across Guangdong Province in 2019 [21]. The package included telemedicine equipment (including tablets installed with telemedicine software and internet access) and medical devices to conduct examinations (Multimedia Appendix 1). The use of telemedicine services among villagers provided by the Guangdong Provincial Health Commission depends on the patient’s needs and the village doctor’s decision: villagers can request telemedicine consultation when visiting the village clinic with a clear goal of accessing telemedicine service, then the village doctor would conduct remote consultation for the

villagers. In addition, the telemedicine provided by Guangdong Provincial Health Commission can serve as an alternative to in-person visits to remote doctors when the village doctor decides to use telemedicine to gain suggestions for patients' problems. However, villagers can still conduct direct telemedicine consultation on their own through internet-connected devices that they can access (smartphones, computers, etc). The parent project of this study was a village-based cluster randomized controlled trial (CRT)—trial registration number: ChiCTR2100053872—which aimed to increase rural health care use and patient satisfaction, decrease out-of-pocket costs, and improve health outcomes by providing telemedicine platform access and training support to village doctors. Details in treatment and randomization of the CRT can be found in Multimedia Appendix 2.

As part of the baseline research of the CRT in rural areas of Guangdong Province, China, this study followed a mixed methods study approach. A sequential explanatory design was adapted, with a quantitative cross-sectional survey analysis followed by a qualitative thematic analysis of semistructured interviews to identify and explore associated factors of telemedicine use among rural residents.

Quantitative Approach

Sampling and Participant Eligibility

The survey was conducted between July and August 2021. Among all 187 counties in Meizhou and Heyuan cities, 3 counties (Meijiang in Meizhou and Dongxin and Yuangcheng in Heyuan) were first dropped because of a small number of townships. Second, Heping county covering 17 townships in Heyuan was also excluded as a related village doctor training program had recently taken place. Therefore, 167 townships (96 in Meizhou and 71 in Heyuan) were included in the sampling frame. Of these, 144 townships were randomly selected. Villages within these 144 townships would be eligible if (1) they were on the list of 2277 "low-income" villages that were provided smart health equipment packages by the Guangdong Provincial Health Commission; (2) they had at least fifteen households; and (3) the village doctors were willing to receive medical training of the parent CRT. One village was randomly selected from among all eligible villages in the 144 townships. A sampling of 15 households per village was targeted according to health management rosters in each village clinic, including 5 from hypertension and diabetes rosters, 3 with children aged 0 to 6 years and 2 with pregnant or lying-in women. Therefore, a sample of 2160 households were expected. One individual (usually the head or the one most familiar with the household) was suggested by each selected household to respond to the survey. Additional inclusion criteria for respondents of the household survey included (1) a resident in the selected villages in the 2 cities; (2) at least one household member who lived in the village for >3 months in the past year; and (3) willingness to participate in the survey.

Data Source and Measurements

The quantitative data were drawn from 2 sources, including the regional economic development data (GDP per capita of each county in 2020) from the government report [22,23] and an interviewer-administered survey questionnaire applied to household respondents.

The binary outcome variable of quantitative data is measured by whether the household's respondent has ever used the telemedicine platform. To make villagers aware of their experience of telemedicine use, we have explained the term *telemedicine* to villagers by using understandable sentences and phrases. When asking whether they have ever used telemedicine service, we described *using telemedicine* as a process of *seeing a doctor on the web, consulting or asking doctors about health issues on the web, remotely consulting or asking doctors about health issues through the internet*, or *web-based communications about health issues with the health care professional*. All these situations count as telemedicine use. The explanatory variables included environmental and contextual factors, household-level factors, individual sociodemographic factors, access to digital health care, and health needs and demand factors. The environmental and contextual factors included regional economic factors (GDP per capita of each county in 2020), geographic factors (distance from the village to the town hospital and the most frequently visited hospital), social network factors (having a close relative as a village doctor), and health system factors (measured by the number of house calls by village doctors in the past year). At the household level, the family size (number of family members), financial situation of the household (wealth index calculated by principal component analysis of household assets), and whether the family is in the poverty registration were included. Individual sociodemographic factors included whether they are the head of the household, age, sex education, and participation in insurance. Access to telemedicine services, smartphone or computer ownership, and others' help using the internet were investigated. Health needs and demand factors of family members include the prevalence of chronic disease (hypertension and diabetes), acute disease and symptoms (diarrhea, cough, or runny nose in the past year, and fever in the past 2 weeks), and health-seeking behaviors (frequency of visits to the village clinic, township hospital, and county hospital).

Quantitative Analysis

The quantitative data were exported to Stata SE 16.0 (StataCorp) for statistical analysis. Descriptive statistics such as frequencies (n), proportions (%) with 95% CIs, and numerical summary measures (including means and SDs) were used to describe the data. Cluster adjustment for SE was used in CI estimation of characteristic proportions to modify the estimation in our cluster-sampling observational study. Variance inflation factor was calculated to assess multicollinearity between variables [24]. Age and gender were defined as prior confounder and forced variables that entered the regression model. To address data *separation* and minimize bias caused by conventional logistic model maximum likelihood estimation resulted from the extremely low rate of telemedicine use in this study [25,26], Firth logistic regression that was proposed as an ideal strategy for rare events in 2020 SAS Global Forum based on penalized likelihood strategy was adapted [27]. Both univariate and multivariate analyses were conducted using regression models presented as crude and adjusted odds ratios (aORs) with 95%

CIs. The coefficient of discrimination D (also known as Tjur R^2) was calculated to indicate goodness-of-fit [28]. The receiver operating characteristic of the area under the curve (ROCAUC) was adapted to evaluate the predictive power of the Firth logistic regression [29]. The level of statistical significance was declared at P values of <.05.

Qualitative Approach

Sampling and Participant Eligibility

Semistructured face-to-face interviews were conducted for qualitative research between January and February 2022. A total of 8 local interviewers who could communicate with villagers in a local accent (ie, Hakka) and lived in 8 rural villages were recruited. The list of the 8 villages can be found in Multimedia Appendix 3. To gain a comprehensive understanding of barriers to and facilitators of telemedicine use among different groups of people through qualitative study, a purposive sampling strategy was used during interviewee selection [30]. For key topics of behavioral factors (eg, experience and practice) in the Social Cognitive Theory (SCT) would mainly be explored among telemedicine users and extremely low telemedicine use (14/2101, 0.67%) informed by quantitative survey, we purposefully sampled participants with and without experience of telemedicine use. Each interviewer was fully trained before they conducted the interviews. The interviewees' eligibility for qualitative interviews was similar to that of the quantitative survey: (1) lived in the village in Meizhou or Heyuan for >3 months in the past year; (2) lived with family members in the roster of hypertension, diabetes, pregnant or lying-in women, or children aged 0 to 6 years from the village clinic; and (3) agreed to participate in the interview.

Conceptual Framework and Data Collection

Topic guides were developed for the semistructured in-depth interviews based on SCT and an extensive review of the existing literature [12-14]. SCT is a framework that includes 3 main components that interact with each other bidirectionally: cognition and personal factors, behavior factors, and environmental factors [31]. The conceptual framework using the adapted SCT is shown in Figure 1. Considering that SCT has been used in research on health behavior and information system adoption behavior [32,33] and helped to obtain a comprehensive understanding of the mechanism that affected behavioral intentions to use telemedicine, SCT would be an appropriate theoretical basis to guide future studies. The key topics of each SCT main component were developed according to the previous research (Table 1) [12-14]. From the cognitive perspective, personal understanding and knowledge, attitudes and perceptions, and expectations related to telemedicine were asked. Regarding behavioral factors, topics focused on self-efficacy, asking about how confident the respondents thought they were able to use telemedicine to meet their health-related demands. Experience and past practice was also asked of those who already had a history of telemedicine use. Both physical and social environmental factors affecting the use of telemedicine services were included. The specific interview questions for each key topic can be found in Multimedia Appendix 4.

Figure 1. Adapted social cognitive model.

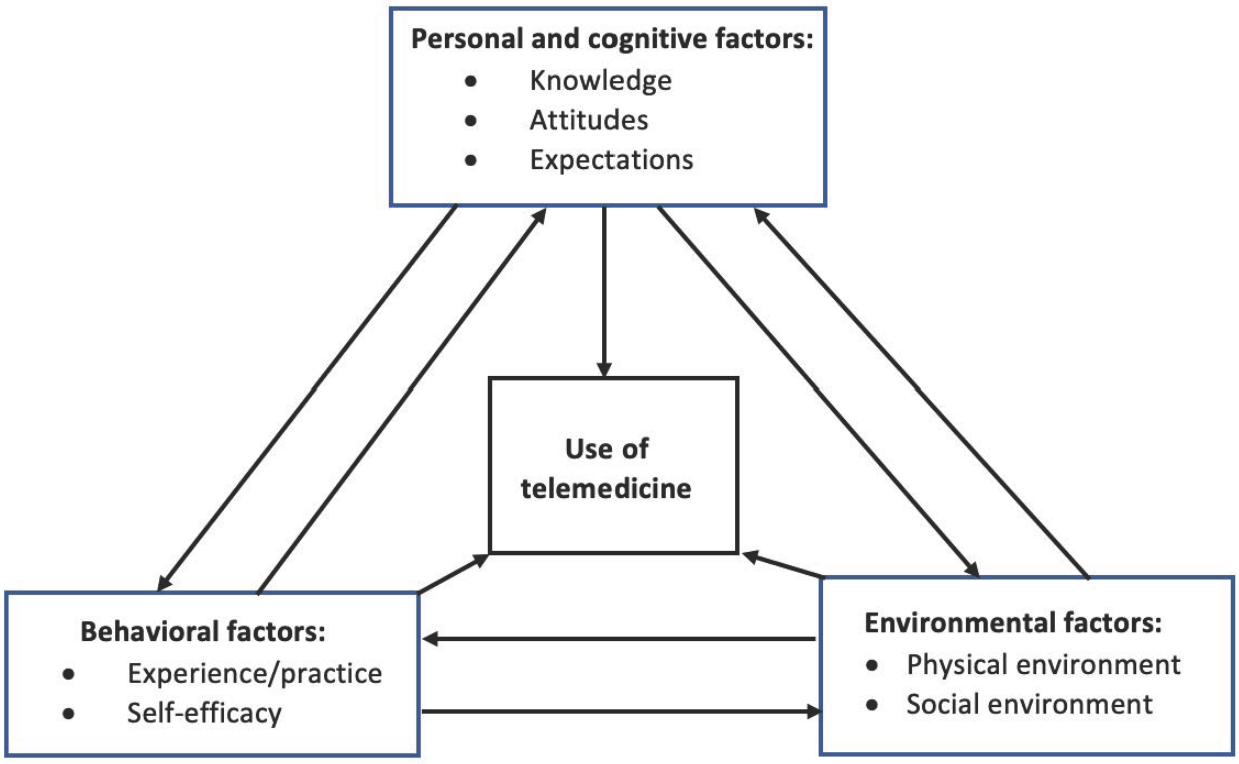


Table 1. Operational definitions and key topics examined for the Social Cognitive Theory (SCT) components explored during in-depth interviews with rural residence in Guangdong Province, China.

SCT components and operational definition	Key topics explored
Cognitive factors	
Knowledge: knowledge to perform telemedicine consultation	Information source, awareness, and operation
Attitude: positive and negative attitude toward telemedicine service	Necessity, confidence, cost, convenience, benefit, and risk
Expectation: expectations for telehealth services and health outcomes	Expected assist, service function, and health goal
Behavior factors	
Experience and practice: experience and past practice of telemedicine use	Difficulty, improvement, and doctor assessment
Self-efficacy: the extent to which the respondents believe they were able to use telemedicine to meet their health-related demands	Ability to access information or operate device or communicate on the web or seek assistance
Environmental factors	
Physical support: equipment and facility regarding telemedicine service	Internet connection, smartphone, and computer
Social support: social support that may facilitate telemedicine use	Training, financial reimbursement, professional recommendation, and external assist

The interview is based on a semistructured interview guide based on SCT. Communications were conducted in Mandarin or Hakka (for villagers who could not speak or understand Mandarin). Oral consent was sought and audio recorded before the formal interview. Anonymity and confidentiality were stressed. We continued the interview until data saturation, which refers to the stage when the interviewer keeps hearing repeated information or the interviewee cannot provide new information during interview data collection, was reached [34]. Each interview was also fully audio recorded and transcribed. Interviews that were conducted in Hakka were translated into Mandarin for transcription. One of the authors verified the accuracy of the transcripts.

Qualitative Analysis

Data were analyzed using a thematic approach with the assistance of NVivo (version 12). An inductive coding process was conducted and a codebook was set up. A total of 2 researchers independently read and coded text materials and double-coded interview transcripts using a line-by-line, open-coding process to check for consistency and accuracy. Divergences were then discussed until a consensus was reached. Afterward, codes were categorized into different themes or subthemes according to the adapted SCT model (Figure 1). Quotes were translated from Chinese to English for the write-up.

Ethical Considerations

We obtained multicenter ethics approval from Guangdong Second Provincial General Hospital (Institutional Review Board [IRB] approval number: GD2H-KY IRB-AF-SC.07-01.1), Peking University (IRB approval number: IRB00001052-21007-免), and the University of North Carolina at Chapel Hill (IRB approval number: 21-0549). It has also been registered in the clinical trial registry in China (ChiCTR2100053872). Written informed consent was obtained from all participants in the quantitative survey. Oral informed consent was obtained and recorded from the participants for qualitative interviews. Confidentiality and privacy was assured for every respondent in the survey and the interviews. All stored data were encrypted in computer-based files, and any information used for research purposes was deidentified. Data stored in computer-based files are only accessible to those with proper clearance. Only study members can access identifiable data. All survey participants received a towel worth ¥15 (US $2) as compensation for their time and inconvenience of participation in the research.

Results

Quantitative Findings

Descriptive Analysis

During the baseline investigation, 2101 households and 141 village clinics were examined. Figure 2 shows a diagram of the Strengthening the Reporting of Observational Studies in Epidemiology presenting the stages of the analytic process. We were unable to reach 45 households from 3 village clinics (villages) because of the SARS-CoV-2 infection prevention and control policy implemented during the investigation. In addition, 14 households were excluded for not meeting the eligibility criteria or failing to participate in the interview. Among the information from 2101 households being collected, 379 individuals with missing values of age, gender, or education of the main respondent did not enter multivariate analysis but were included in the descriptive and univariate analyses.

Among all the respondents (N=2101), only 14 (0.67%) reported having ever used telemedicine services. Among participants who had used telemedicine services before (n=14), 3 (21%) of them reported having registered for medical services on the web and 2 (14%) of them reported having a web-based service for doctors' follow-up visits.

Table 2 presents the characteristics of the participants (N=2101) categorized by various factors: environment and context, household, individual sociodemographic, access, and health needs and demand. With a mean age of 61.4 (SD 14.41) years; >70% (1364/2101, 72.94%) of the respondents were male. More

than 80% of the respondents were the heads of households. More than half (1021/2101, 58.98%) of the participants had a degree from senior high school or technical secondary school. In terms of access to digital health care, 67.06% (1409/2101) of participants reported ownership of a smartphone or computer, 62.54% (1314/2101) had internet connection, and 37.65% (791/2101) said they had someone who could help them use the internet in their family. Regarding health needs and demands of family members, hypertension (1405/2101, 66.87%) was the most reported disease, whereas fever in the past 2 weeks (134/2101, 6.38%) was the least reported.

Figure 2. STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) diagram for progress through stages of analyses.

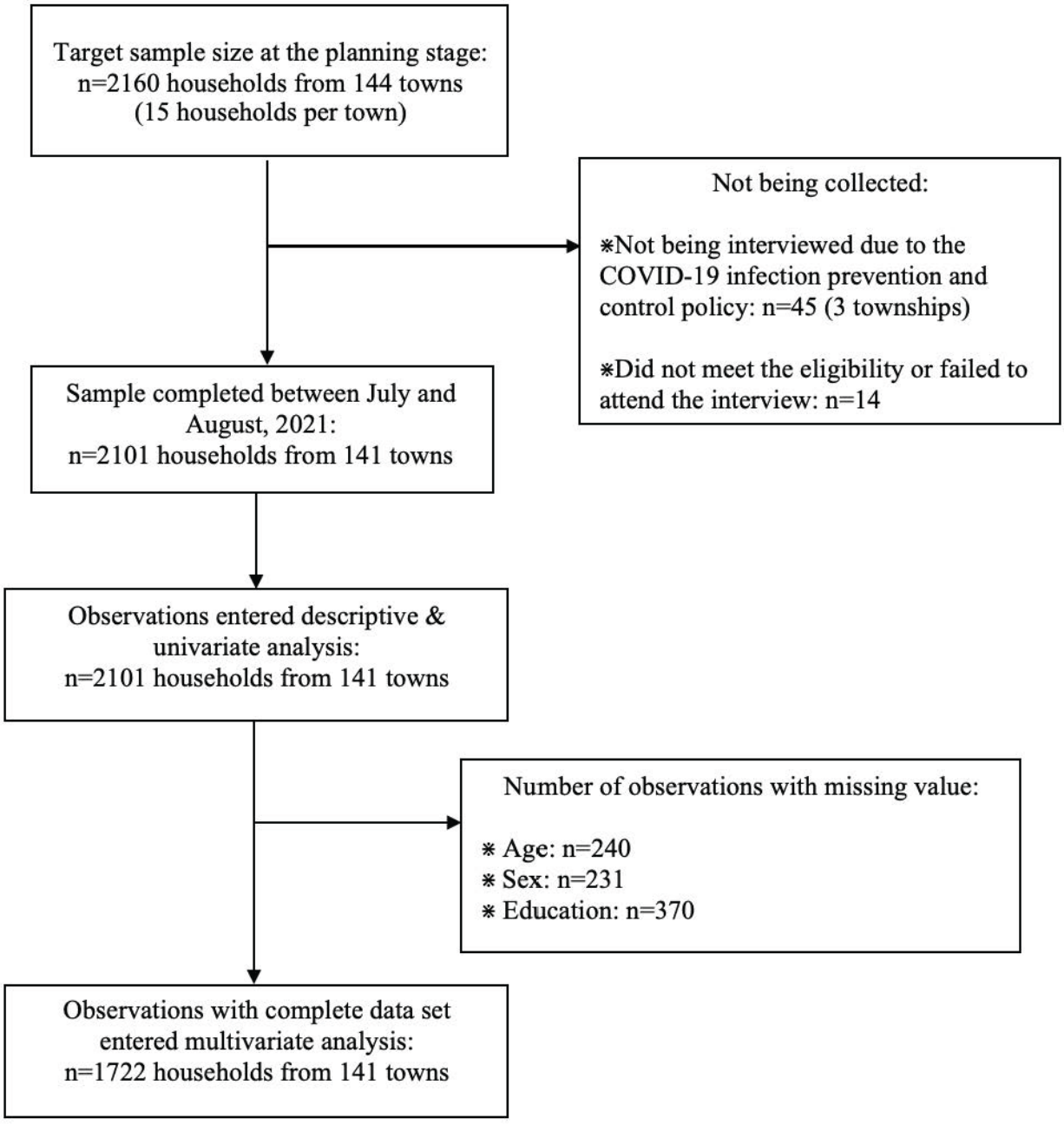


Table 2. Characteristics of rural residents in Guangdong Province, China (N=2101).

Characteristic and category	Observations	95% CI[a] of the percentage or mean
Environmental and contextual factor, n (%)		
Per capital GDP[b] of each county in 2020 (Chinese yuan renminbi)		
<25,000 (US $3623)	986 (46.93)	38.74-55.29
≥25,000 (US $3623)	1115 (53.07)	44.71-61.26
Distance from village to the town hospital (km)		
<5.0	1029 (48.78)	40.74-57.27
≥5.0	1072 (51.02)	42.73-59.26
Distance from village to the most frequently visited county hospital (km)		
<35.0	1026 (48.83)	40.58-57.15
≥35.0	1075 (51.17)	42.85-59.42
Being close relative of the village doctor		
No	1964 (93.48)	92.14-94.60
Yes	137 (6.52)	5.40-7.86
Number house calls by village doctors in the past year		
None	982 (46.74)	42.48-51.05
1-3	484 (23.04)	20.77-25.48
≥4	635 (30.22)	26.74-33.95
Household-level factor, n (%)		
Family size (number of family members)		
1	203 (9.66)	8.34-11.16
2	545 (25.94)	23.93-28.05
3-5	626 (29.80)	27.73-31.94
>5	727 (34.60)	32.05-37.25
Financial situation (wealth index)		
<20%	420 (19.99)	17.66-22.54
20%-40%	411 (19.56)	17.85-21.39
40%-60%	429 (20.42)	18.64-22.32
60%-80%	419 (19.94)	18.16-21.85
≥80%	422 (20.09)	17.93-22.43
Family in poverty registration		
No	1735 (82.66)	80.76-84.41
Yes	364 (17.34)	15.59-19.24
Individual sociodemographic factor		
Householder, n (%)		
No	288 (15.40)	13.37-17.68
Yes	1582 (84.60)	82.32-86.63
Age (years), mean (SD)		
Continuous	61.4 (14.41)	60.57-62.15
Sex, n (%)		
Male	1364 (72.94)	70.46-75.29
Female	506 (27.06)	24.71-29.54

Characteristic and category	Observations	95% CI[a] of the percentage or mean
Education level, n (%)		
Junior high school or below	710 (41.02)	38.09-44.00
Senior high school or above	1021 (58.98)	56.00-61.91
Use of any medical insurance programs , n (%)		
No	89 (4.24)	3.32-5.38
Yes	2012 (95.76)	94.62-96.67
Access to digital health care, n (%)		
Having smartphone or computer		
No	692 (32.94)	30.73-35.22
Yes	1409 (67.06)	64.78-69.27
Can be connected to the internet		
No	95 (37.46)	35.15-39.83
Yes	1314 (62.54)	60.17-64.85
Someone can help to use the internet in the family		
No	1310 (62.35)	59.80-64.83
Yes	791 (37.65)	35.17-40.20
Health needs and demand of family members, n (%)		
Hypertension		
No	696 (33.13)	31.27-35.03
Yes	1405 (66.87)	64.97-68.73
Diabetes		
No	1236 (58.83)	56.98-60.65
Yes	856 (41.17)	39.35-43.02
Diarrhea in the past year		
No	1249 (59.48)	56.65-62.19
Yes	852 (40.55)	37.81-43.34
Cough or runny nose in the past year		
No	905 (43.07)	40.76-45.42
Yes	1196 (56.93)	54.58-59.24
Fever in the past 2 weeks		
No	1967 (93.62)	92.39-94.67
Yes	134 (6.38)	5.33-7.61
Frequency of the county hospital visit in the past year		
None	1328 (63.21)	60.61-65.74
1 time	343 (16.33)	14.76-18.02
≥2 times	430 (20.47)	18.51-22.5
Frequency of the town hospital visit in the past year		
None	1028 (48.93)	46.02-51.85
1 time	274 (13.04)	11.48-14.78
≥2 times	799 (38.03)	35.19-40.95
Frequency of the village clinic visit in the past year		
None	893 (42.50)	38.79-46.31

Characteristic and category	Observations	95% CI[a] of the percentage or mean
1-3 times	318 (15.14)	13.61-16.80
≥3 times	890 (42.36)	39.02-45.78

[a]CI of the percentage estimation was adjusted by SE of cluster sampling.
[b]GDP: gross domestic product.

Correlation of Telemedicine Use

Table 3 shows the results of the crude and aORs in using telemedicine services with a 95% CI. In univariate analysis, participants who had family members who developed a fever in the past 2 weeks (OR 6.42, 95% CI 2.10-19.69; *P*=.001) were more likely to use telemedicine services than those who did not. In addition, smartphone or computer ownership (OR 1.47, 95% CI −0.31 to 3.25; *P*=.08) also tended to use telemedicine compared with their counterparts. Participants living with family members who had hypertension (OR 0.38, 95% CI 0.14-1.05; *P*=.06) or fever in the past 2 weeks (OR 6.42, 95% CI 2.10-19.69; *P*=.001) were less likely to accept telemedicine services. However, after adjusting for age and gender, only people with family members who developed a fever in the past 2 weeks (aOR 6.96, 95% CI 2.20-21.98; *P*=.001) had significantly higher uptake rates of telemedicine services. Evidence regarding the association between other independent factors and telemedicine uptake is insufficient. The variance inflation factor of each explanatory variable shown in Multimedia Appendix 5 signifies the absence of multicollinearity in the model. Multimedia Appendix 6 shows Tjur R^2 and ROCAUC for each multivariate model. The multivariate model including independent variables of fever history in the past 2 weeks, age, and gender indicated that >58% of the variance for telemedicine use was explained by independent variables (Tjur R^2=0.582) and fair predictive power of the fitted model (ROCAUC=0.718).

Table 3. Firth logistic regression analysis of correlation of telemedicine use for rural residents in Guangdong Province, China (N=2101).

Characteristic and category	Crude OR[a] (95% CI)	*P* value[b]	aOR[c] (95% CI)[d]	*P* value[e]
Macrocontext factor				
Per-capita GDP[f] of each county (Chinese yuan renminbi)				
<25,000 (US $3623)	Ref[g]	.78	N/A[h]	.92
≥25,000 (US $3623)	1.16 (0.41-3.22)	.78	0.95 (0.33-2.75)	.92
Distance from village to the town hospital (km)				
<5.0	Ref	.94	N/A	.80
≥5.0	0.96 (0.35-2.64)	.94	1.13 (0.40-3.29)	.80
Distance from village to the most frequently visited county hospital (km)				
<35.0	Ref	.27	N/A	.15
≥35.0	0.55 (0.19-1.58)	.27	0.43 (0.14-1.34)	.15
Being close relative of the village doctor				
No	Ref	.62	N/A	.63
Yes	0.49 (0.03-8.25)	.62	0.50 (0.03-8.41)	.63
Number house calls by village doctors in the past year				
None	Ref	.27	N/A	.20
1-3	1.58 (0.39-6.42)	.27	1.12 (0.24-5.31)	.20
≥4	2.59 (0.80-8.41)	.27	2.69 (0.83-8.76)	.20
Household-level factor				
Family size (number of family members)				
1	Ref	.47	N/A	.76
2	2.64 (0.14-50.91)	. 47	2.36 (0.12-46.06)	.76
3-5	2.29 (0.12-44.27)	. 47	2.01 (0.10-39.65)	.76
>5	4.81 (0.28-83.93)	. 47	3.42 (0.19-62.80)	.76
Financial situation (wealth index)				
<20%	Ref	.09	N/A	.11
20%-40%	4.48 (0.76-26.56)	.09	4.44 (0.74-26.31)	.11
40%-60%	0.33 (0.01-8.00)	.09	0.29 (0.01-7.17)	.11
60%-80%	1.00 (0.10-9.68)	.09	0.21 (0.09-8.33)	.11
≥80%	4.34 (0.73-25.79)	.09	3.06 (0.49-19.30)	.11
Family in poverty registration				
No	Ref	.54	N/A	.42
Yes	1.45 (0.44-4.81)	.54	1.63 (0.49-5.53)	.42
Individual sociodemographic factor				
Householder				
No	Ref	.80	N/A	.86
Yes	0.84 (0.21-3.32)	.80	5.53 (0.18-7.69)	.86
Age (years)				
Continuous	0.98 (0.95-0.99)	.16	0.98 (0.25-1.01)	.15
Sex				
Male	Ref	.86	N/A	.80
Female	0.90 (0.27-3.03)	.86	0.85 (0.25-2.89)	.80

Characteristic and category	Crude OR[a] (95% CI)	*P* value[b]	aOR[c] (95% CI)[d]	*P* value[e]
Education level				
Junior high school or below	Ref	.50	N/A	.73
Senior high school or above	1.48 (0.48-4.53)	.50	1.23 (0.38-4.01)	.73
Use of any medical insurance programs				
No	Ref	.29	N/A	.27
Yes	0.40 (0.07-2.18)	.29	2.64 (0.07-2.10)	.27
Access to digital health care				
Having smartphone or computer				
No	Ref	.08	N/A	.14
Yes	4.43 (0.83-24.05)	.08	3.71 (0.64-21.32)	.14
Can be connected to the internet				
No	Ref	.13	N/A	.22
Yes	2.89 (0.73-11.36)	.13	2.46 (0.58-10.38)	.22
Someone can help to use the internet in the family				
No	Ref	.65	N/A	.56
Yes	1.27 (0.45-3.53)	.65	1.36 (0.48-3.90)	.56
Health needs and demand of family members				
Hypertension				
No	Ref	.06	N/A	.17
Yes	0.38 (0.14-1.05)	.06	0.46 (0.15-1.40)	.17
Diabetes				
No	Ref	.38	N/A	.26
Yes	0.61 (0.20-1.82)	.38	0.49 (0.14-1.68)	.26
Diarrhea in the past year				
No	Ref	.46	N/A	.36
Yes	1.46 (0.53-4.06)	.46	1.63 (0.57-4.71)	.36
Cough or runny nose in the past year				
No	Ref	.59	N/A	.46
Yes	0.76 (0.27-2.10)	.59	0.67 (0.23-1.92)	.46
Fever in the past 2 weeks				
No	Ref	.001	N/A	.001
Yes	6.42 (2.10-19.69)	.001	6.96 (2.20-21.98)	.001
Frequency of the county hospital visit in the past year				
None	Ref	.34	N/A	.60
Once	0.68 (0.12-3.86)	.34	0.70 (0.12-4.01)	.60
≥Twice	1.52 (0.49-4.76)	.34	1.62 (0.51-5.10)	.60
Frequency of the town hospital visit in the past year				
None	Ref	.44	N/A	.48
Once	2.41 (0.63-9.21)	.44	2.27 (0.59-8.76)	.48
≥Twice	N/A	.44	1.26 (0.38-4.14)	.48
Frequency of the village clinic visit in the past year				
None	Ref	.29	N/A	.41

Characteristic and category	Crude OR[a] (95% CI)	*P* value[b]	aOR[c] (95% CI)[d]	*P* value[e]
1-3 times	0.15 (0.01-2.51)	.29	0.17 (0.01-2.89)	.41
≥3 times	0.57 (0.20-1.67)	.29	0.68 (0.23-2.01)	.41

[a]OR: odds ratio.

[b]*P* value for penalized likelihood-ratio test of univariate model.

[c]aOR: adjusted odds ratio.

[d]Adjusted odds ratio: marginal effect adjusted for age and sex in the multivariate logistic model.

[e]*P* value for likelihood-ratio test of multivariate model.

[f]GDP: gross domestic product.

[g]Ref: Reference value.

[h]N/A: not applicable.

Qualitative Findings

Overview

A total of 27 semistructured individual interviews were conducted with villagers from 8 villages. Background characteristic numbers (P1-P27) of each interviewee are shown in Multimedia Appendix 3. The average time length of 27 interviews was 25.6 (SD 11.2) minutes. The mean age of the interviewees was 53.4 (SD 16.7) years, and 56% (15/27) were male. Of these, 22% (6/27) said they had never heard of telemedicine services before, and 44% (12/27) said they had heard of telemedicine but never used it. Of all the interviewees, 33% (9/27) reported having experience with telemedicine use. A total of 10 themes, classified into 3 categories (personal and cognitive, behavioral, and environmental factors), were extracted from the interview content according to the adapted SCT (Table 4). Major themes that emerged as barriers to telemedicine service uptake among villagers included lack of knowledge and understanding, lack of trust, lack of demand, low self-efficacy, lack of physical support, and social support.

Table 4. Themes and subthemes of qualitative analysis on telemedicine use among rural residence in Guangdong Province, China (N=27).

Theme and subtheme	People mentioned, n (%)	Coding
Personal and cognitive factors		
Knowledge and understanding		
Web-based process recognized	16 (59)	Web-based searching, appointment, registration, consultation, diagnosis, prescription, and payment
Operation	18 (67)	Lack of technical ability and lack of comprehension ability
Attitude		
Perceived convenience	27 (100)	Avoid queuing up, avoid COVID-19 prevention policy limit, and overcome traffic barrier and distance
Perceived cost	25 (93)	Financial cost, time cost and efficiency, and manpower
Lack of trust	19 (70)	Concern about fraud or false information, personal information security, government authority and supervision, doctor's certificate, doctor's seniority, and reputation of the hospital
Lack of demand	27 (100)	Complacent about self-condition, local clinic and doctor's indoor visit as alternatives, auxiliary and reference for offline visit, chronic disease, minor aliment, and rush hour or weekends
Expectation		
Health expectation	16 (59)	Chronic disease management and address minor disease
Service expectation	21 (78)	Guide for healthy lifestyle and knowledge promotion
Behavioral factors		
Experience and practice		
Web-based process experience	6 (22)	Appointment, registration, prescription, and taking photos for doctors on the web
Difficulty or improvements	5 (19)	Web-based communications, lack of information interoperability and integrality across platforms
Assessment of the doctor	8 (30)	Web-based feedback, doctor's title, and doctor's resume,
Self-efficacy	27 (100)	Web illiterate, aging, memory loss, dull-witted, illiteracy, visually impaired, dialect barriers, and out of step with the times
Environmental factors		
Lack of physical support		
Equipment	11 (41)	Internet connection and smartphone or computer devices
Technical limitations	13 (48)	Lack of physical examination, lack of physiological tests, lack of medical imaging tests, and low video or photo resolution
Delivery system	10 (37)	Lack of direct home delivery service, lack of express pickup points in the village, and delay of drug delivery
Social support		
Training support	21 (78)	Desire for teaching and training in the village
Role of primary health care provider	25 (93)	Improve ability of village doctors, provide assistance for villagers, and formalists

Themes and Quotes

Lack of Knowledge and Understanding

The knowledge theme indicated limited understanding and knowledge of telemedicine among the interviewees. Most (18/27, 67%) of the respondents said they did not know how to use or were not good at operating telemedicine services. However, some provided examples of the web-based process (eg, appointment, register, consultation, and prescription):

> *We basically don't know how to use smartphones. I can't even take a taxi with a mobile phone.* [P27, female, 27 years]

Lack of Trust

Many (10/19, 53%) villagers showed negative perceptions about reliability, concern about internet scammers, and personal information leakage of the telemedicine technology. In addition, some (5/19, 26%) demonstrated conditional trust, meaning that they would trust telemedicine services if conditions such as

doctors' certificates, supervision or regulations of government, and reputations of doctors and hospitals were in place:

> *Now there are so many scammers, the Internet has the most of them. If you can't find reliable one, I'm afraid no one will take the risk for you. Does Alipay have a platform called Huzhubao? Is such a big platform gone now?* [P20, female, 25 years]

> *There is such a possibility, I am not so relieved. If the doctor does something bad using our ID card information, we will not be able to figure it out. People are unpredictable, especially those who have only met once.* [P3, female, 25 years]

> *I believe that the Internet could be reliable to a certain extent. It should have a certificate issued by the nation, involving a relatively well-known hospital or doctors, etc.* [P7, female, 48 years]

Lack of Demand

Many (15/27, 56%) responders did not think telemedicine was a necessity. Several (4/27, 15%) villagers said that they were complacent about their health condition or would prioritize physical hospitals (7/27, 26%). Some participants (14/27, 52%) mentioned that they would use telemedicine only if they encountered a *minor ailmen*t (eg, cold or fever). This finding was consistent with our quantitative findings that showed higher telemedicine service uptake among villagers who had family members with fever in the past 2 weeks (aOR 6.96, 95% CI 2.20-21.98; *P*=.001). Our qualitative findings also indicated that offline hospitals were preferred when dealing with a *serious disease*; respondents tended to regard the web-based results as a reference for seriousness:

> *It is not necessary for me (to use a telemedicine platform). I'm in good health now.*

> *Maybe for the time being, doctors on the Internet are still relatively unfamiliar to me. There is no need to use this approach. We all can go to the local hospital for treatment. My idea is that if we encounter a problem that the local doctor can't solve, we could then take this approach (telemedicine).* [P22, male, 60 years]

> *It depends, some simple disease such as cold or fever, I would go online for consultation. It's OK to just have a look (in an Internet hospital). If it's a serious illness, I won't go online. I will only regard the online results as a reference.* [P11, male, 29 years]

Low Self-efficacy

Most (16/27, 59%) of the participants expressed negatively about self-efficacy, for they thought themselves as aging, illiterate, dull, and forgetful people and did not believe they had the ability to use telemedicine:

> *We elderly have never used it. We are illiterate, have never been on the Internet. We can't surf the Internet, our eyes are not very good, and we can't read a few words.* [P25, male, 68 years]

> *I'm old, with limited operational abilities, dullness, and memory loss.* [P15, male, 72 years]

Lack of Physical Support

The lack of smartphone or computer or internet connection could be a barrier to telemedicine adoption in our interviews. This was also supported by quantitative findings suggesting that smartphone or computer ownership (aOR 3.71, 95% CI 0.64-21.32; *P*=.14) among villagers was marginally correlated with telemedicine uptake:

> *My old-fashioned cell phone does not have such functions. Some middle-aged people can accept it. Of course, for young people it's needless to say. It (internet platform) is very popular among them.* [P22, male, 60 years]

Besides equipment, technical limitations (13/27, 48%) and delivery systems (10/27, 37%) were also identified as insufficient physical support. Concerns about technical limitations mainly come from a lack of medical examination instruments or physical contact with doctors. Interviewees did not think doctors could make an accurate diagnosis by simply asking on the web or uploading photos without any laboratory tests or feeling the pulse (traditional Chinese medicine). Issues with the delivery system were said to cause inconvenience for drug delivery (10/27, 37%); a pharmacy pickup point located in the village or rapid home delivery was needed:

> *Laboratory tests are also required in the hospital, if the doctor purely questions the patient, it may not be so accurate. In traditional Chinese medicine, seeing a doctor in person is like seeing, hearing, asking, and feeling the pulse. That is definitely not so good if you visit the doctor online* [P4, female, 46 years]

> *If there is no pharmacy in the village for Internet hospitals, and I have to wait for 3-4 days for the medicine, and then go to the town or county to get the medicine, then I would rather go to the village doctor. It is not very cost-effective to see a doctor online and wait for three or four days for the medicine.* [P3, male, 60 years]

Social Support

The social environment described the social interaction with the villagers that may influence their telemedicine use, including training support and the roles of primary health care providers. Most (19/21, 90%) people expressed their desire or positive view of training support for telemedicine:

> *I may need to see a doctor online in the future, but only if someone teaches me. If someone teaches me, I will be more willing to go to see a doctor online.* [P16, male, 53 years]

Most (14/25, 56%) interviewees held positive views on the use of telemedicine among primary health care providers. They also thought that primary health care providers could improve their abilities and skills by using telemedicine and help to guide and oversee the process of seeking medical treatment from a superior hospital:

> *It is good (when village doctors using telemedicine), because after all, the elderly can't operate it. The elderly could be familiar with the operation through*

the village doctor, and he can use his knowledge points to express what he needs to go to the health service platform. This may achieve better results. If it is transmitted upward through the village doctor, and the village doctor takes the medicine, it will not cause something like wrong prescription delivery. He will check these things, which is also better for us farmers. [P6, male, 49 years]

Discussion

Principal Findings

Understanding telemedicine use in the rural areas of China is essential for understanding the services gap and future program planning. Our study extended the existing literature by exploring the factors associated with telemedicine service use among rural residents in Guangdong Province, China. With a quantitative survey supplemented with qualitative document analysis, this study provided multiple perspectives on potential barriers to or facilitators of telemedicine use in rural China. We found that extremely few (14/2101, 0.67%) participants in the rural areas used telemedicine services. Villagers in households with family members who had fever in the past 2 weeks were more likely to use telemedicine services. In addition, smartphone or computer ownership was marginally associated with telemedicine use. The qualitative findings supported the quantitative findings. The qualitative approach mainly supplemented the results from the knowledge, trust, demand, self-efficacy, and physical and social support perspectives.

In this study, <1% (14/2101, 0.67%) of telemedicine use was found among the villagers. This finding was lower than that of a nationwide survey that included both urban and rural areas in China from 2016 to 2017, which indicated that 10% of the population used web-based health care communication or management [35]. As rural or urban residence was proven to be the most significant predictor of telemedicine use and Chinese rural residents had a distinctly lower use than urban residents [35], this finding could be explained by the disparity between rural and urban areas. Moreover, most of the respondents in rural areas were older adults, whereas most young men or labor workers moved to urban areas in China [36]. Aging has been proven to be a significant barrier to telemedicine service use [12,35]. Future programs to promote telemedicine uptake among older people would have shared value when targeting rural residents. Another important reason is that owing to chronic disease management through primary health care taking place in rural Chinese communities, which has been proven to be correlated with fewer villagers' specialist visits and inpatient admissions [37], additional needs of the residents to seek care for these chronic diseases would be low, and generally they did not need to seek care through telemedicine.

Both quantitative and qualitative findings identified a lack of equipment (eg, smartphone or computer) or decent internet connection as barriers, which was in line with the existing findings [13,14]. Although previous studies have shown a huge mobile phone market and high network coverage among the Chinese national population [4,11], our data suggested that 67.06% (1409/2101) of smartphone or computer ownership could contribute to the underuse of telemedicine in rural areas.

Furthermore, developing fever in the past 2 weeks was associated with telemedicine service use in both quantitative and qualitative results. This could be explained by the effect of the COVID-19 pandemic outbreak on telemedicine service use. As fever is the most common symptom in patients with COVID-19, teleconsultation systems have been used to collect patient information, such as fever and cough, to control the pandemic [38]. Recent research has also found increased interest in and demand for telehealth services worldwide during the COVID-19 pandemic crisis [38]. In addition, this finding may suggest the low demand to prioritize telemedicine as a conventional health service from the patient side, for qualitative interviews described fever as a minor and common illness and mentioned physical hospitals as a prior and more reliable alternative, especially for serious or urgent conditions. Previous studies have reported available alternatives for receiving health care services [12], citing patients' perceived low value of telemedicine or lack of motivation as barriers to public engagement with digital health services [13,14].

Lack of knowledge and skills was the most coded factor in the qualitative interviews. This is consistent with previous studies [12-14]. Our study also identified the desire for operational training support. Digital health education and training have been an important element in the process of implementing telemedicine services. Previous studies suggested limited digital literacy would increase disparities in health care access for vulnerable populations (eg, rural residents) when scaling up telemedicine implementation [39]. Without addressing this, telemedicine programs would risk excluding vulnerable groups. Similar to previous research that indicated skepticism as a barrier and self-efficacy as an enabling factor for the learning and use of eHealth technologies [14], we also found a lack of trust in telemedicine resulting from concerns about information security and quality of service and low self-efficacy of telemedicine acceptance among rural residents. This suggests additional efforts to build trust and improve self-efficacy could enhance telemedicine uptake. Furthermore, our qualitative findings also highlighted the positive impact of telemedicine practice among primary health care providers (eg, village doctors) on villagers adoption. Existing literature indicated the importance of patient-provider relationship in medication regimen adherence based on telemedicine [40], although research on the influence of telemedicine practice among primary health care providers on patients' use is limited. Lack of accessibility and time efficiency of medication delivery services in rural areas were also identified as a barrier to telemedicine use. Research from Singapore found that besides web-based consultation, telemedicine-based medication delivery services have ramped up medical logistics supply without doctor-patient contact during the COVID-19 pandemic [41]. However, medication delivery services based on telemedicine are limited in LMICs. The practice model of remote pharmacy services relying on web-based consultation in China is still underexplored [42]; more evidence is needed to inform the development of an effective medication delivery service based on telemedicine.

Implications for China and Other LMICs

This study had several implications for telemedicine scaling up in China and other LMICs. First, training sessions for digital literacy improvement of potential users are essential before scaling up of telemedicine services to avoid deteriorating health care inequalities, especially in rural areas of LMICs. Self-efficacy can be achieved through educational sessions. Second, similar to the findings of previous literature, lack of equipment or infrastructure has been a barrier in low-resource settings [43]; therefore, investment in rural construction could be considered as technical equipment support. However, further research on the cost-effectiveness of telemedicine program construction should be conducted. Third, our findings indicated that concerns about inadequate supervision or regulations to ensure information security and quality of web-based medical services may result in a lack of trust in telemedicine services. Thus, we recommended formulating regulations and improving the information transparency of telemedicine platforms in China and other LMICs to strengthen trust and facilitate telemedicine use. Finally, further research on establishing sustainable patient-provider relationships for telemedicine engagement and developing effective medication delivery services based on telemedicine systems is needed, especially for LMICs where related planning or regulations might not occur.

Limitations

This study had a few limitations. First, the observational nature of the study had limitations in examining causality, but the quantitative analysis supplemented with qualitative analysis enabled us to triangulate the information to explore the correlates of telemedicine uptake in rural settings in China. Second, the proportion of people who had ever used telemedicine in quantitative analysis is very low, which limited the ability of the statistical model to detect possible associated factors such as age, gender, financial situation, smartphone or computer ownership, and geographic isolation [12-14].

Conclusions

This study reported the profile of telemedicine use among rural patient end users in Guangdong Province, China. Less than 1% of telemedicine use was found with potential barriers, including lack of knowledge, trust, demand, low self-efficacy, and insufficient physical and social support. Our study suggested that efforts to improve digital literacy and self-efficacy, build trust, and strengthen telemedicine infrastructure support could enhance telemedicine scaling up in rural China and other LMICs.

Acknowledgments

The authors sincerely thank the Medical College of Jiaying University, Health Bureau of Meizhou, and Health Bureau of Heyuan as well as the study participants. This study was funded by the National Key R&D Program of China (grant 2021YFC2009400).

Data Availability

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

Conflicts of Interest

None declared.

Multimedia Appendix 1
Items in smart health monitoring equipment packages (July 5, 2022).
[DOCX File , 2438 KB - publichealth_v8i12e40771_app1.docx]

Multimedia Appendix 2
Randomization and treatments of the parent cluster randomized controlled trial (July 5, 2022).
[DOCX File , 30 KB - publichealth_v8i12e40771_app2.docx]

Multimedia Appendix 3
Background characteristics of rural interviewees in Guangdong Province, China (N=27).
[DOCX File , 18 KB - publichealth_v8i12e40771_app3.docx]

Multimedia Appendix 4
Interview questions based on components of the Social Cognitive Theory.
[DOCX File , 24 KB - publichealth_v8i12e40771_app4.docx]

Multimedia Appendix 5
Multicollinearity diagnosis with variance inflation factor.
[DOCX File , 15 KB - publichealth_v8i12e40771_app5.docx]

Multimedia Appendix 6
Tjur R2 and receiver operating characteristic of the area under the curve of multivariate analysis.
[DOCX File , 16 KB - publichealth_v8i12e40771_app6.docx]

References

1. Han J, Wu M, Yang T. Challenge of China's rural health. BMJ 2016 Apr 12;353:i2003. [doi: 10.1136/bmj.i2003] [Medline: 27072539]
2. China Statistical Yearbook 2021. National Bureau of Statistics of China. 2021. URL: http://www.stats.gov.cn/tjsj/ndsj/2021/indexeh.htm [accessed 2022-10-26]
3. Xue H, Shi Y, Medina A. Who are rural China's village clinicians? China Agri Econ Rev 2016 Nov 07;8(4):662-676. [doi: 10.1108/caer-12-2015-0181]
4. Lv Q, Jiang Y, Qi J, Zhang Y, Zhang X, Fang L, et al. Using mobile apps for health management: a new health care mode in China. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e10299 [FREE Full text] [doi: 10.2196/10299] [Medline: 31162131]
5. Tu J, Wang C, Wu S. The internet hospital: an emerging innovation in China. Lancet Global Health 2015 Aug;3(8):e445-e446. [doi: 10.1016/S2214-109X(15)00042-X]
6. Pagliari C. Digital health and primary care: past, pandemic and prospects. J Glob Health 2021 Jul 02;11:01005 [FREE Full text] [doi: 10.7189/jogh.11.01005] [Medline: 34221352]
7. Global Strategy on Digital Health 2020-2025. World Health Organization. 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2022-10-26]
8. A health telematics policy in support of WHO's health-for-all strategy for global health development: report of the who group consultation on health telematics, 11-16 December, Geneva, 1997. World Health Organization. 1998. URL: https://apps.who.int/iris/bitstream/handle/10665/63857/WHO_DGO_98.1.pdf?sequence=1&isAllowed=y [accessed 2022-10-26]
9. He C, Zhou Q, Chen W, Tian J, Zhou L, Peng H, et al. Using an internet-based hospital to address maldistribution of health care resources in rural areas of Guangdong province, China: retrospective and descriptive study. JMIR Med Inform 2018 Dec 21;6(4):e51 [FREE Full text] [doi: 10.2196/medinform.9495] [Medline: 30578195]
10. Marcin JP, Shaikh U, Steinhorn RH. Addressing health disparities in rural communities using telehealth. Pediatr Res 2016 Jan 14;79(1-2):169-176 [FREE Full text] [doi: 10.1038/pr.2015.192] [Medline: 26466080]
11. 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]
12. Hage E, Roo JP, van Offenbeek MA, Boonstra A. Implementation factors and their effect on e-Health service adoption in rural communities: a systematic literature review. BMC Health Serv Res 2013 Jan 12;13:19 [FREE Full text] [doi: 10.1186/1472-6963-13-19] [Medline: 23311452]
13. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
14. Airola E. Learning and use of eHealth among older adults living at home in rural and nonrural settings: systematic review. J Med Internet Res 2021 Dec 02;23(12):e23804 [FREE Full text] [doi: 10.2196/23804] [Medline: 34860664]
15. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet hospitals in China: cross-sectional survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
16. Tian M, Zhang J, Luo R, Chen S, Petrovic D, Redfern J, et al. mHealth interventions for health system strengthening in china: a systematic review. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e32 [FREE Full text] [doi: 10.2196/mhealth.6889] [Medline: 28302597]
17. Liu Y, Ren W, Qiu Y, Liu J, Yin P, Ren J. The use of mobile phone and medical apps among general practitioners in Hangzhou city, eastern China. JMIR Mhealth Uhealth 2016 May 24;4(2):e64 [FREE Full text] [doi: 10.2196/mhealth.4508] [Medline: 27220417]
18. Zhang H, Zhang H, Wang X, Yang Z, Zhao Y. Analysis of requirements for developing an mHealth-based health management platform. JMIR Mhealth Uhealth 2017 Aug 03;5(8):e117 [FREE Full text] [doi: 10.2196/mhealth.5890] [Medline: 28778840]
19. Guangdong Statistical Yearbook 2020. China Year Book. 2020 Aug. URL: http://stats.gd.gov.cn/gdtjnj/content/post_3557537.html [accessed 2022-10-26]
20. Hamadeh N, Rompaey C, Metreau E. New World Bank country classifications by income level: 2021-2022. World Bank. 2021 Jul 1. URL: https://blogs.worldbank.org/opendata/new-world-bank-country-classifications-income-level-2021-2022 [accessed 2022-10-26]
21. 2,277 provincially designated poverty-stricken villages in Guangdong have achieved "full coverage" of telemedicine and AI doctors. People's Government of Guangdong Province. 2019 Sep 24. URL: http://www.gd.gov.cn/zwgk/zdlyxxgkzl/ylws/content/post_2635697.html [accessed 2022-10-26]
22. 2021 Economic Data Release of Heyuan City. Heyuan.gov.cn. URL: http://www.heyuan.gov.cn/zwgk/tjsj/tjfx/content/post_478018.html [accessed 2022-10-26]

23. Meizhou Statistical Yearbook 2021. Meizhou Municipal Government. URL: https://www.meizhou.gov.cn/attachment/0/115/115103/2261170.pdf [accessed 2022-10-26]
24. Detecting multicollinearity using variance inflation factors. The Pennsylvania State University. URL: https://online.stat.psu.edu/stat462/node/180/ [accessed 2022-10-26]
25. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med 2002 Aug 30;21(16):2409-2419. [doi: 10.1002/sim.1047] [Medline: 12210625]
26. FIRTH D. Bias reduction of maximum likelihood estimates. Biometrika 1993;80(1):27-38. [doi: 10.1093/biomet/80.1.27]
27. Karabon P. Rare events or non-convergence with a binary outcome? The power of firth regression in PROC LOGISTIC. SAS Global Forum. 2020. URL: https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2020/4654-2020.pdf [accessed 2022-10-26]
28. Tjur T. Coefficients of determination in logistic regression models—a new proposal: the coefficient of discrimination. Am Stat 2009 Nov;63(4):366-372. [doi: 10.1198/tast.2009.08210]
29. Seshan V, Gönen M, Begg C. Comparing ROC curves derived from regression models. Stat Med 2013 Apr 30;32(9):1483-1493 [FREE Full text] [doi: 10.1002/sim.5648] [Medline: 23034816]
30. Robinson OC. Sampling in interview-based qualitative research: a theoretical and practical guide. Qual Res Psychol 2013 Nov 18;11(1):25-41. [doi: 10.1080/14780887.2013.801543]
31. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. London, England: Pearson; 1985.
32. Wu D, Gu H, Gu S, You H. Individual motivation and social influence: a study of telemedicine adoption in China based on social cognitive theory. Health Policy Technol 2021 Sep;10(3):100525. [doi: 10.1016/j.hlpt.2021.100525]
33. Tsai C. Integrating social capital theory, social cognitive theory, and the technology acceptance model to explore a behavioral model of telehealth systems. Int J Environ Res Public Health 2014 May 07;11(5):4905-4925 [FREE Full text] [doi: 10.3390/ijerph110504905] [Medline: 24810577]
34. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018 Mar;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
35. Hong YA, Zhou Z. A profile of eHealth behaviors in China: results from a national survey show a low of usage and significant digital divide. Front Public Health 2018 Sep 26;6:274 [FREE Full text] [doi: 10.3389/fpubh.2018.00274] [Medline: 30320054]
36. Li Q, Huang S, Luo R, Liu C. China's labor transition and the future of China's rural wages and employment. China World Econ 2013 May 19;21(3):4-24. [doi: 10.1111/j.1749-124X.2013.12019.x]
37. Ding H, Chen Y, Yu M, Zhong J, Hu R, Chen X, et al. The effects of chronic disease management in primary health care: evidence from rural China. J Health Econ 2021 Dec;80:102539 [FREE Full text] [doi: 10.1016/j.jhealeco.2021.102539] [Medline: 34740053]
38. Aslani N, Garavand A. The role of telemedicine to control COVID-19. Arch Clin Infect Dis 2020 Mar 30;15(COVID-19):e102949. [doi: 10.5812/archcid.102949]
39. Nouri S, Khoong EC, Lyles CR, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catalyst Innovations in Care Delivery. 2020 May 4. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0123 [accessed 2022-10-26]
40. Bollmeier SG, Stevenson E, Finnegan P, Griggs SK. Direct to consumer telemedicine: is healthcare from home best? Mo Med 2020;117(4):303-309 [FREE Full text] [Medline: 32848261]
41. Lai SH, Tang CQ. Telemedicine and COVID-19: beyond just virtual consultations – the Singapore experience. Bone Joint Open 2020 Jun;1(6):203-204. [doi: 10.1302/2046-3758.16.bjo-2020-0042.r1]
42. Li H, Zheng S, Li D, Jiang D, Liu F, Guo W, et al. The establishment and practice of pharmacy care service based on internet social media: telemedicine in response to the COVID-19 pandemic. Front Pharmacol 2021 Oct 1;12:707442 [FREE Full text] [doi: 10.3389/fphar.2021.707442] [Medline: 34658854]
43. Mahmoud K, Jaramillo C, Barteit S. Telemedicine in low- and middle-income countries during the COVID-19 pandemic: a scoping review. Front Public Health 2022;10:914423 [FREE Full text] [doi: 10.3389/fpubh.2022.914423] [Medline: 35812479]

Abbreviations

aOR: adjusted odds ratio
CRT: cluster randomized controlled trial
GDP: gross domestic product
IRB: Institutional Review Board
LMIC: low- and middle-income country
ROCAUC: receiver operating characteristic of the area under the curve
SCT: Social Cognitive Theory

Edited by A Mavragani; submitted 05.07.22; peer-reviewed by J Ye, N Aldawoodi; comments to author 09.09.22; revised version received 06.10.22; accepted 18.10.22; published 23.12.22.

Please cite as:
Du Y, Zhou Q, Cheng W, Zhang Z, Hoelzer S, Liang Y, Xue H, Ma X, Sylvia S, Tian J, Tang W
Factors Influencing Adoption and Use of Telemedicine Services in Rural Areas of China: Mixed Methods Study
JMIR Public Health Surveill 2022;8(12):e40771
URL: https://publichealth.jmir.org/2022/12/e40771
doi:10.2196/40771
PMID:36563026

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org