JMIR Public Health and Surveillance

Impact Factor (2022): 8.5
Volume 8 (2022), Issue 11 ISSN 2369-2960 Editor in Chief: Travis Sanchez, PhD, MPH

Contents

Original Papers

Review

Original Paper

Performance of the Swiss Digital Contact-Tracing App Over Various SARS-CoV-2 Pandemic Waves: Repeated Cross-sectional Analyses

Paola Daniore[1,2], MSc; Vasileios Nittas[3], PhD; Tala Ballouz[3], MD; Dominik Menges[3], MD, PhD; André Moser[4], PhD; Marc Höglinger[5], Prof Dr; Petra Villiger[6], MSc; Krisztina Schmitz-Grosz[6], MD; Viktor Von Wyl[1,2,3], Prof Dr

[1]Institute for Implementation Science in Healthcare, University of Zurich, Zurich, Switzerland
[2]Digital Society Initiative, University of Zurich, Zurich, Switzerland
[3]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland
[4]Clinical Trials Unit, University of Bern, Bern, Switzerland
[5]Winterthur Institute of Health Economics, Zurich University of Applied Sciences, Winterthur, Switzerland
[6]Medgate Aktiengesellschaft, Basel, Switzerland

Corresponding Author:
Viktor Von Wyl, Prof Dr
Institute for Implementation Science in Healthcare
University of Zurich
Universitatstrasse 84
Zurich, 8006
Switzerland
Phone: 41 044 634 37 62
Email: viktor.vonwyl@uzh.ch

Abstract

Background: Digital proximity-tracing apps have been deployed in multiple countries to assist with SARS-CoV-2 pandemic mitigation efforts. However, it is unclear how their performance and effectiveness were affected by changing pandemic contexts and new viral variants of concern.

Objective: The aim of this study is to bridge these knowledge gaps through a countrywide digital proximity-tracing app effectiveness assessment, as guided by the World Health Organization/European Center for Prevention and Disease Control (WHO/ECDC) indicator framework to evaluate the public health effectiveness of digital proximity-tracing solutions.

Methods: We performed a descriptive analysis of the digital proximity-tracing app SwissCovid in Switzerland for 3 different periods where different SARS-CoV-2 variants of concern (ie, Alpha, Delta, and Omicron, respectively) were most prevalent. In our study, we refer to the indicator framework for the evaluation of public health effectiveness of digital proximity-tracing apps of the WHO/ECDC. We applied this framework to compare the performance and effectiveness indicators of the SwissCovid app.

Results: Average daily registered SARS-CoV-2 case rates during our assessment period from January 25, 2021, to March 19, 2022, were 20 (Alpha), 54 (Delta), and 350 (Omicron) per 100,000 inhabitants. The percentages of overall entered authentication codes from positive tests into the SwissCovid app were 9.9% (20,273/204,741), 3.9% (14,372/365,846), and 4.6% (72,324/1,581,506) during the Alpha, Delta, and Omicron variant phases, respectively. Following receipt of an exposure notification from the SwissCovid app, 58% (37/64, Alpha), 44% (7/16, Delta), and 73% (27/37, Omicron) of app users sought testing or performed self-tests. Test positivity among these exposure-notified individuals was 19% (7/37) in the Alpha variant phase, 29% (2/7) in the Delta variant phase, and 41% (11/27) in the Omicron variant phase compared to 6.1% (228,103/3,755,205), 12% (413,685/3,443,364), and 41.7% (1,784,951/4,285,549) in the general population, respectively. In addition, 31% (20/64, Alpha), 19% (3/16, Delta), and 30% (11/37, Omicron) of exposure-notified app users reported receiving mandatory quarantine orders by manual contact tracing or through a recommendation by a health care professional.

Conclusions: In constantly evolving pandemic contexts, the effectiveness of digital proximity-tracing apps in contributing to mitigating pandemic spread should be reviewed regularly and adapted based on changing requirements. The WHO/ECDC framework allowed us to assess relevant domains of digital proximity tracing in a holistic and systematic approach. Although the Swisscovid app mostly worked, as reasonably expected, our analysis revealed room for optimizations and further performance

improvements. Future implementation of digital proximity-tracing apps should place more emphasis on social, psychological, and organizational aspects to reduce bottlenecks and facilitate their use in pandemic contexts.

(JMIR Public Health Surveill 2022;8(11):e41004) doi:10.2196/41004

KEYWORDS

digital contact tracing; exposure notification; COVID-19; SARS-CoV-2; public health; surveillance; digital proximity; contact-tracing app; mobile app; Switzerland; variant of concern; SwissCovid app; digital tool

Introduction

To contribute to mitigation efforts against the spread of SARS-CoV-2, digital proximity-tracing apps were developed and widely adopted in multiple countries. This gave rise to a novel research area within digital public health, which aims to assess the possible contribution of such apps toward disease control. Prominent examples of digital proximity-tracing apps in Europe include the United Kingdom's National Health Service's (NHS) COVID-19 app, the German Corona-Warn-App, and the SwissCovid app from Switzerland [1-3]. In Switzerland, smartphone ownership exceeding 90% [4] across all socioeconomic groups presented an opportunity for the SwissCovid app to be widely adopted and complement manual contact-tracing efforts. Conducted in the form of interviews, manual contact tracing is labor intensive and prone to errors due to its reliance on people's abilities to recall proximity contacts [5]. The SwissCovid app promised to deliver exposure notifications at a faster rate, with broader reach and greater scalability [6,7]. However, it was essential that exposure notifications be sent quickly and without interruptions, ultimately providing a time advantage over manual contact tracing [8].

There is growing interest in further evaluating the effectiveness of digital proximity-tracing apps. However, effectiveness analyses face multiple challenges [7,9]. First, the outcome of interest, which is the prevention of SARS-CoV-2 transmission, is not observable. Second, the privacy-preserving architecture of digital proximity-tracing apps, particularly those that follow the Decentralized Privacy-Preserving Proximity Tracing (DP-3T) blueprint [10], provides only limited, nonidentifiable data for conducting effectiveness analyses. Lastly, additional relevant data generated, for example, through manual contact tracing, information hotlines, and testing centers, henceforth described as "points of contact for app users," are often dispersed across different systems and not readily available due to privacy regulations [11].

Empirical evaluations of the effectiveness of digital proximity-tracing apps remain scarce [12]. Recent evaluations have mainly produced mixed results, ranging from substantial [13-15] to moderate [16,17] or disappointing [18] findings. There is also a large heterogeneity of analytical methods and data used for these analyses, which makes a direct comparison of their results difficult. To foster standardization, the World Health Organization (WHO) and the European Center for Disease Prevention and Control (ECDC) recently developed a framework outlining the most relevant data and monitoring indicators for digital contact-tracing apps (henceforth referred to as the "WHO/ECDC framework") [19]. To the best of our knowledge, however, this framework has not yet been applied to a systematic, countrywide analysis, and its utility for effectiveness analyses remains to be explored.

The aim of this study is to bridge these knowledge gaps through a countrywide digital proximity-tracing app effectiveness assessment, as guided by the WHO/ECDC framework. Specifically, we performed a descriptive analysis of the digital proximity-tracing app in Switzerland for 3 different periods where different SARS-CoV-2 variants of concern (ie, Alpha, Delta, and Omicron, respectively) were most prevalent. We performed this analysis by applying the WHO/ECDC framework to individual and public-level data, which we complemented with additional indicators of mitigative actions taken by app users after receiving an exposure notification. Accordingly, our analysis applies the WHO/ECDC framework indicators in the greater pandemic context to inform future indicator-based app monitoring and effectiveness assessment efforts.

Methods

SwissCovid Digital Proximity-Tracing App

Switzerland was 1 of the first countries that launched a digital proximity-tracing app (SwissCovid) based on the DP-3T architecture on June 25, 2020 [20]. The DP-3T architecture works by sending low-energy Bluetooth beacons with a pseudonymized, regularly changing user identification number to other SwissCovid app users in its surroundings. Here, the Bluetooth signal strength serves as a proxy for the physical distance between 2 smartphones. Copies of a user's own identification numbers, as well as those of recent proximity encounters with other apps, are then stored locally on the users' smartphones.

The SwissCovid app worked through an exposure notification cascade system to identify and isolate possible SARS-CoV-2 cases of interest. The exposure notification cascade started when a user received a positive polymerase chain reaction (PCR) test result for SARS-CoV-2. This triggered the first step in the cascade (illustrated in Supplementary Figure 1 in Multimedia Appendix 1), in which the user was issued an authentication code. Users subsequently entered their authentication code in the app, leading to the release of their own pseudonymized identification numbers to a central server. The SwissCovid app regularly downloaded identification numbers and searched locally registered identification numbers from proximity encounters. An exposure notification was triggered by the app if contact exposure between 2 or more individuals met predefined proximity and time thresholds (proximity of ≤1.5 m to an infected person for ≥15 minutes). This message included further instructions for the exposed individuals, such as the

phone number for a SwissCovid infoline and a link to a risk self-assessment web form (from December 2020). Exposure-notified SwissCovid app users were advised to call the infoline number and to seek free-of-charge SARS-CoV-2 testing.

During its operational period, until its deactivation on April 1, 2022, the SwissCovid app reached approximately 1.9 million users, corresponding to 26.1% of all Swiss inhabitants aged 16 years and older [20]. In total, 205,000 positive test results triggered exposure notifications through the SwissCovid app, and 141,000 infoline calls or web forms were completed. Further details on how digital proximity-tracing apps work [11] as well as existing evidence of SwissCovid app effectiveness in pandemic mitigation for Switzerland have been presented in detail elsewhere [17].

Data Collection

Our study's approach was guided by the WHO/ECDC framework. In brief, this framework provides a set of key indicators to guide the monitoring and evaluation of digital proximity-tracing apps, as well as to measure the performance and effectiveness of the corresponding exposure notification cascade in preventing onward transmission of SARS-CoV-2 (see Supplementary Table 1 in Multimedia Appendix 2).

We used data from public and nonpublic sources. Public monitoring data for the SwissCovid app [20] and the SARS-CoV-2 pandemic [21] were retrieved from the website of the Swiss Federal Office of Public Health. Data on the Oxford measurement of stringency of COVID-19 measures were retrieved from the respective website [22]. We also used data provided by the company that operated SwissCovid Infoline (Medgate Aktiengesellschaft) for aggregated daily counts of generated upload authentication codes, infoline calls, and self-assessment web entries. Additionally, we used longitudinal individual-level data, collected through surveys within the COVID-19 Social Monitor study, to provide additional indicators of interest regarding the mitigative actions taken by individuals upon receiving an exposure notification [23]. Further details on indicator definitions and data sources are presented in Multimedia Appendix 3.

Statistical Analysis

Longitudinal analyses of SARS-CoV-2–monitoring indicators, defined in Supplementary Table 1 in Multimedia Appendix 2, were conducted for the entire study period from January 25, 2021, to March 19, 2022. Daily count values were averaged over 7 days or over the entire study period. Comparisons of SwissCovid app effectiveness indicators were conducted for stratified periods based on the 3 predominant SARS-CoV-2 variants of concern [21] and were aligned with the COVID-19 Social Monitor survey data collection phases: (1) Alpha variant (January 25-June 17, 2021, survey waves 13-17), (2) Delta variant (August 30-December 16, 2021, survey waves 18-20), and (3) Omicron BA.1 variant (January 24-March 19, 2022, survey waves 21-22); see Supplementary Figure 2 in Multimedia Appendix 4.

Our analysis focused on 3 of the WHO/ECDC framework indicators: (a) adoption of the SwissCovid app and frequency of exposure notifications, (b) successfulness of digital proximity-tracing apps in detecting contacts at risk of infection, and (c) whether digital proximity-tracing apps are faster in notifying contacts than conventional contact tracing. Specifically, all assessments in our analyses are linked to SwissCovid app users in their individual uptake and engagement with the app. The indicators further assess the performance and effectiveness of the SwissCovid app in mitigating onward viral transmission based on user responses to exposure notifications (ie, in forms of mitigative actions or noncompliance). To further provide context to the development of the indicators assessed in this study, we retrieved Oxford stringency index values for Switzerland, which quantify the strictness of countrywide lockdown policies during the SARS-CoV-2 pandemic [22].

To evaluate possible gaps in compliance with recommended measures, we defined a theoretical upper ceiling estimate for app users testing positive for SARS-CoV-2 infection. This upper ceiling estimate was calculated as the number of individuals who tested positive multiplied by the percentage of app users in the general population. Additional indicators were calculated based on mitigative actions taken by SwissCovid app users and by using individual-level data from the COVID-19 Social Monitor: (1) having been tested for SARS-CoV-2, (2) having tested positive for SARS-CoV-2, (3) having been in isolation or in quarantine ordered by a physician or manual contact tracing, and (4) having received an exposure notification (see Supplementary Figure 3 and Supplementary Table 4 in Multimedia Appendix 5).

Analyses were performed in Stata version 16.1 (StataCorp LLC). All data were analyzed descriptively as counts and percentages. Selected indicators were visualized using 3 topical radar plots. Reporting was informed by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist (Multimedia Appendix 6) [24].

Ethical Considerations

For the COVID-19 Social Monitor study, the Cantonal Ethics Commission of Zurich concluded that our study did not fall within the scope of the Human Research Act (BASEC-Nr. Req-2020-00323). All other data did not require ethics approval.

Results

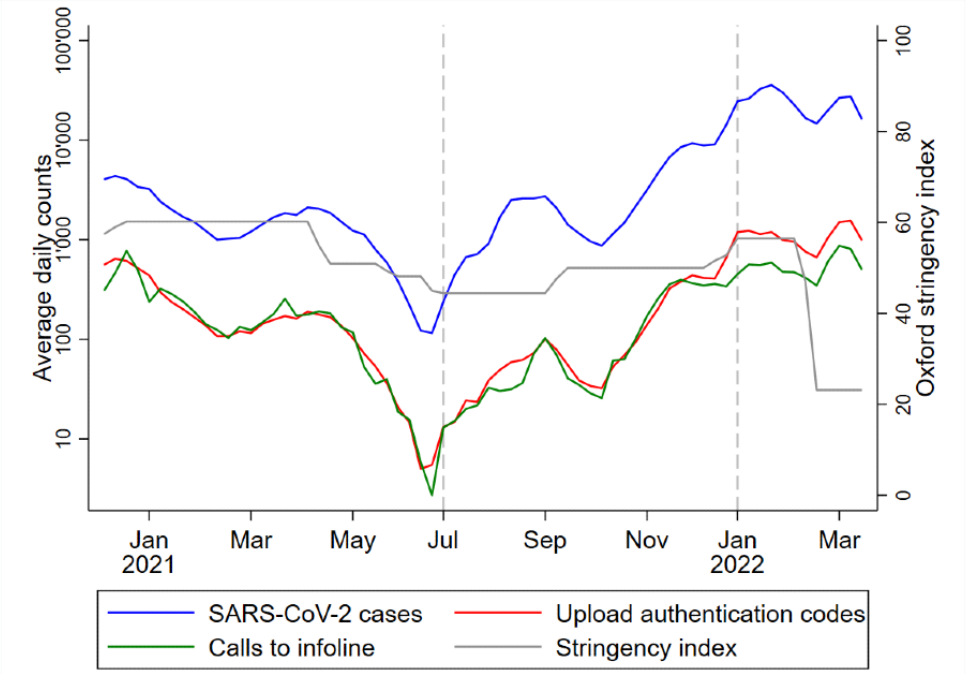
Longitudinal Analysis of Monitoring Indicators From Official Public Health Sources

Figure 1 depicts the evolution of measured indicators across the 3 pandemic waves of the SARS-CoV-2 variants of concern. The blue line represents the counts of positive SARS-CoV-2 tests in Switzerland. The trend here suggests several incidence peaks in January 2021, which were due to the Alpha variant, and January 2022, which marks the transition of predominance from the Delta to the Omicron variant. The average daily cases over the study period were 20 (Alpha), 54 (Delta), and 350 (Omicron) per 100,000 inhabitants. The gray line illustrates the Oxford measure of stringency of COVID-19 measures, which ranges from 0 (lowest stringency) to 100 (highest stringency). In our observation period, the stringency of measures was highest between January and April 2021. This coincided with

the Alpha variant phase, where measures such as home office and prohibition of gatherings were mandated by the Swiss Federal Office of Public Health. The stringency measure was also high during the final Delta variant phase and the beginning of the Omicron variant phase. Almost all mitigation measures were removed in February 2022.

The red and green lines illustrate the number of entered authentication codes by SARS-CoV-2–positive SwissCovid app users and calls to the infoline or completion of a self-assessment form upon receipt of an exposure notification, respectively. In the assessed period, the counts of these user-driven actions closely followed the incidence curve. Furthermore, they occurred in an almost stable 1:1 ratio, with 1 infoline call or completed web form per shared positive test result for the majority of the study period. However, there was a shift in this ratio deviating toward fewer user actions taken by exposed contacts during the Omicron variant phase.

Figure 1. Longitudinal description of key indicators (7-day averages). The dashed vertical lines delineate different pandemic phases that were dominated by the Alpha, Delta, or Omicron SARS-CoV-2 variants of concern.



Indicator Comparisons Across Pandemic Phases

Indicators of Exposure Notification Cascade Performance

Indicators from the WHO/ECDC framework and selected complementary indicators from the COVID-19 Social Monitor data are illustrated in radar plots (Figures 2-4, data in Supplementary Tables 1 and 2 in Multimedia Appendix 2). Figure 2 illustrates indicators that relate to the performance of the exposure notification cascade (ie, completeness and speed of events). Starting with the top indicator and moving clockwise, indicator 1 shows the average weekly SARS-CoV-2 incidence from daily values (rescaled as percentage from the peak incidence). The maximum of daily case numbers was reached during the Omicron variant phase and the lowest daily case numbers during the Alpha variant phase. Indicator 2 shows that around 1 in 4 (1,779,546/7,280,501, 24.4%) Swiss individuals aged 16 years and older were active SwissCovid app users during the Alpha variant phase, while the percentage of SwissCovid app users decreased slightly during the Delta (1,624,946/7,280,501, 22.3%) and Omicron (1,568,104/7,280,501, 21.5%) variant phases.

Indicator 3 represents the number of authentication codes that were shared with the SwissCovid app as a fraction of the total number of individuals with a positive SARS-CoV-2 test. This percentage was 9.9% (20,273/204,741) during the Alpha variant phase and then declined to 3.9% (14,372/365,846) and 4.6% (72,324/1,581,506) during the Delta and Omicron variant phases, respectively. Indicator 4 reflects the ratio of authentication codes entered into the SwissCovid app over issued authentication codes. Here, we observed a nearly twice as large proportion of entered codes during the Alpha variant phase (20,273/31,658, 64%) compared with the Delta (14,372/44,455, 32.3%) and Omicron variant phases (72,324/269,700, 26.8%). Indicator 5 represents the timing of authentication code upload into the Swisscovid app from symptom onset or positive test date if the app user was asymptomatic at the time of testing. This indicator suggests that between 50% and 56% of all entered codes were uploaded within 48 hours after symptom onset, with lower percentages observed in the following 2 variant phases.

Lastly, indicator 6 represents the proportion of SwissCovid app users who completed the provided web form and called an infoline after receiving an exposure notification. Here, we observed that between 23% and 28% of exposure-notified app users contacted the infoline or completed the web form within 48 hours after the exposure date, which is provided in the exposure notification message.

Figure 2. Indicators reflecting the performance of the exposure notification cascade. The colored lines represent the Alpha (blue), Delta (red), and Omicron (green) variant phases. The plot ranges from 0 (center) to 1 and illustrates the proportions and ratios of the relevant indicators. Indicator definitions and data sources are provided in Supplementary Table 1 in Multimedia Appendix 2. Prop.: proportion.

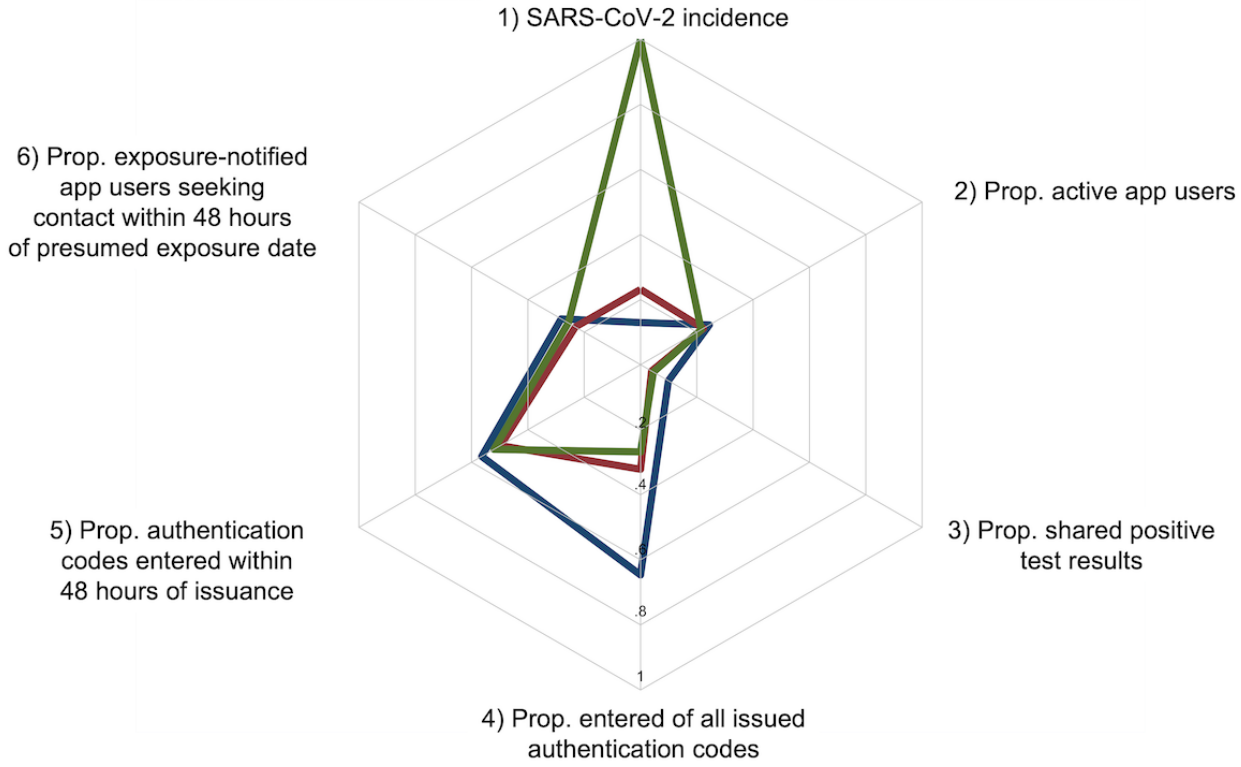


Figure 3. Indicators reflecting the proportion of exposure notifications or individuals who tested positive. The colored lines represent the Alpha (blue), Delta (red), and Omicron (green) variant phases. The plot ranges from 0 (center) to 0.5 and illustrates the proportions and ratios of the relevant indicators. Indicator definitions and data sources are provided in Supplementary Table 1 in Multimedia Appendix 2. Prop.: proportion.

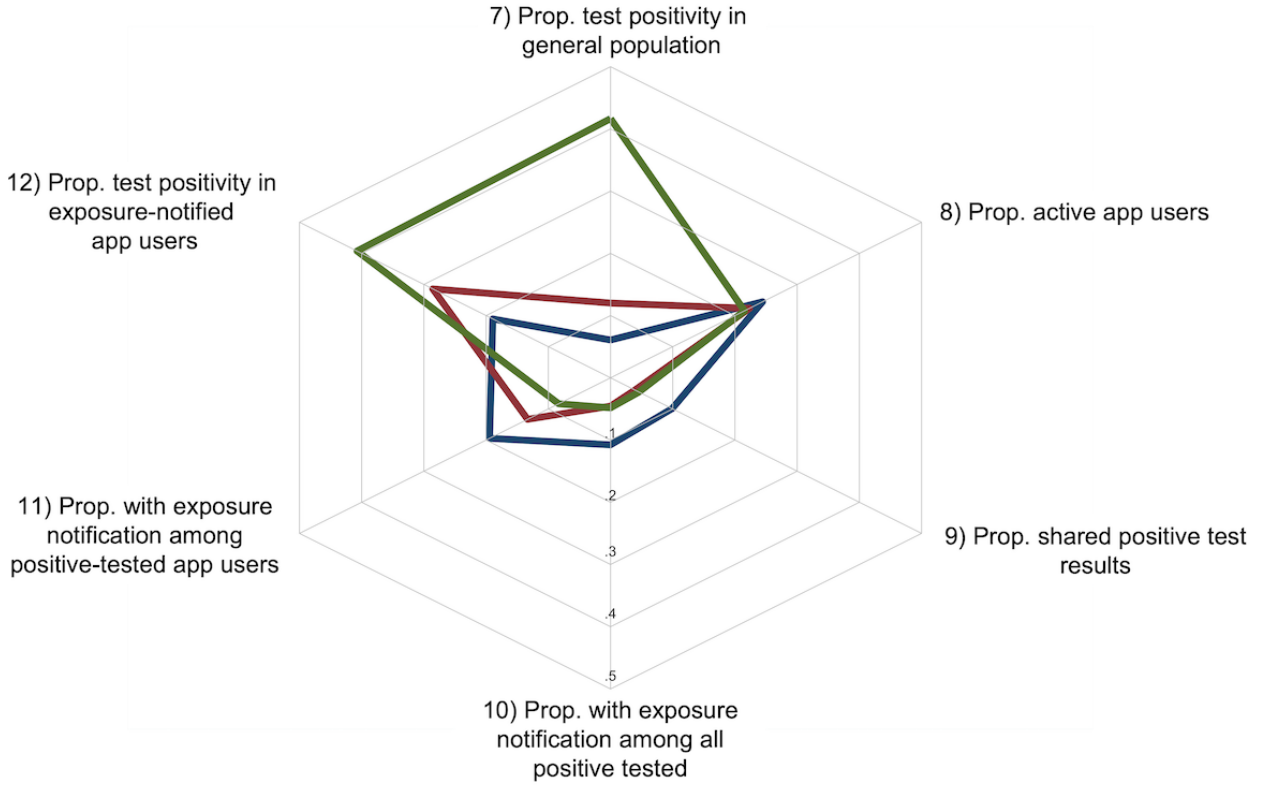
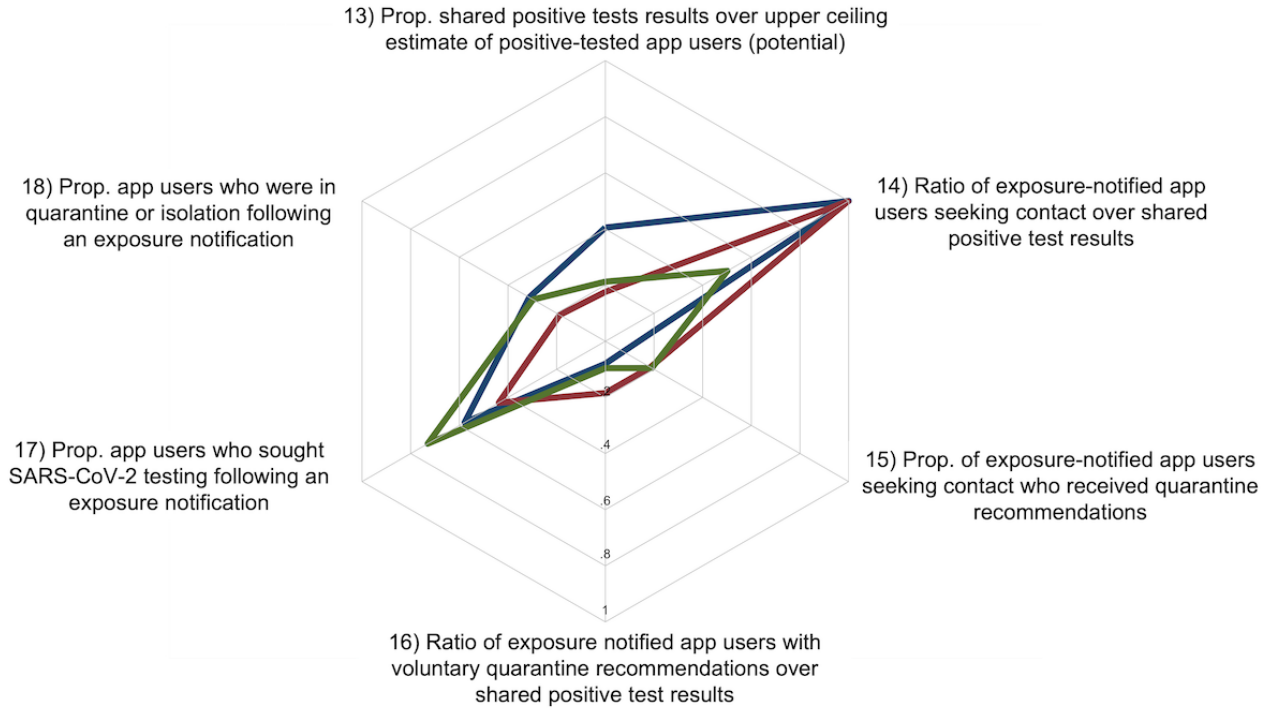


Figure 4. Indicators reflecting the probability of app user actions following exposure notifications or positive test results. The colored lines represent the Alpha (blue), Delta (red), and Omicron (green) variant phases. The plot ranges from 0 (center) to 1 (indicator 14 values were censored at 1, even though they were slightly higher; more information is available in Supplementary Table 2 in Multimedia Appendix 4) and illustrates the proportions and ratios of the relevant indicators. Indicator definitions and data sources are provided in Supplementary Table 1 in Multimedia Appendix 2. Prop.: proportion.



Indicators Reflecting Test Positivity Following Exposure Notifications

Test positivity following receipt of an exposure notification is considered a proxy to assess the precision of exposure detection in notifying affected individuals. Figure 3 summarizes the key indicators in this context, although in a more refined scale, which ranges from 0 (0%) to 0.5 (50%). Indicator 7 illustrates test positivity in the general population, which was close to 10% in the first 2 variant phases (228,103/3,755,205, 6.1%, and 413,685/3,443,364, 12%, respectively) and increased to around 41.7% (1,784,951/4,285,549) in the Omicron variant phase. Indicators 7 and 8 are equivalent to indicators 2 and 3 in Figure 2. Indicator 8 represents the percentage of active app users, and indicator 9 represents the percentage of app users among individuals who tested positive, based on generated upload authentication codes. Indicator 10 illustrates the percentage of app users who received an exposure notification among all individuals with a positive test. This value was approximately 11% (7/65) in the Alpha variant phase and around 5% in the later 2 variant phases (2/44 and 11/228, respectively). Indicator 11 represents the percentage of app users who received an exposure notification among all app users who tested positive (calculated for indicator 10). Here, they were 19% (7/36) in the Alpha variant phase, 13% (2/15) in the Delta variant phase, and 8.3% (11/132) in the Omicron variant phase. Finally, indicator 12 illustrates test positivity among app users who received an exposure notification. This value was 19% (7/37) in the Alpha variant phase, 29% (2/7) in the Delta variant phase, and 41% (11/27) in the Omicron variant phase compared to 6.1% (228,103/3,755,205), 12% (413,685/3,443,364), and 41.7% (1,784,951/4,285,549) in the general population, respectively.

Indicators Reflecting User Actions Following Exposure Notifications

The third set of indicators illustrates the extent of mitigative actions taken by SwissCovid app users following receipt of exposure notifications. Figure 4 summarizes the key indicators in this context in a scale that ranges from scores 0 to 1. Indicator 13 illustrates the proportion of authorization codes entered into the SwissCovid app from individuals who tested positive by the upper ceiling estimate, which were 40.5% (20,273/50,044) for the Alpha variant, 17.6% (14,372/81,654) for the Delta variant, and 21.2% (72,324/340,631) for the Omicron variant.

Indicator 14 illustrates the ratio of users seeking contact through the infoline or completing the web form per shared positive test result. This value decreased over the course of the pandemic from 1.08 user contacts per code during the Alpha variant phase to 1.00 during the Delta variant phase and 0.50 during the Omicron variant phase. Indicator 15 illustrates the exposure risk assessment following contact with the infoline or via a web form, as well as a voluntary quarantine recommendation following receipt of exposure notifications. The proportion of quarantine recommendations per user contact was 7.4% (1622/21,976) during the Alpha variant phase and increased to 18.5% during the Delta and 19.1% during the Omicron variant phases (2652/14,313 and 6931/36,279, respectively).

Indicator 16 illustrates the standardized voluntary quarantine recommendations by the number of shared positive test results. Here, there were approximately 8 recommendations per 100

tests in the Alpha variant phase, 18 recommendations per 100 tests during the Delta variant phase, and 10 recommendations per 100 tests during the Omicron variant phase. Indicator 17 illustrates data from the COVID-19 Social Monitor and indicates that 58% (37/64), 44% (7/16), and 73% (27/37) app users sought testing or performed self-tests following an exposure notification during the Alpha, Delta and Omicron variant phases, respectively. Lastly, indicator 18 reveals that 31% (20/64, Alpha), 19% (3/16, Delta), and 30% (11/37, Omicron) of individuals who received exposure notifications also reported to have received mandatory quarantine orders by manual contact tracing or through a recommendation by a health care professional.

Discussion

Principal Findings

Our study presented various digital proximity-tracing app performance indicators for Switzerland. These were guided by and built upon the WHO/ECDC framework for the assessment of digital proximity-tracing apps' public health effectiveness in mitigating onward transmission of SARS-CoV-2. Our analysis extends the current knowledge in the field of digital proximity tracing by comparing various pandemic periods that were characterized by different SARS-CoV-2 variants of concern, as well as by changes in public perceptions of the pandemic and public health responses. Our study further contributes to effectiveness assessments on a methodological level by introducing further indicators of interest from panel survey data that assess mitigative strategies taken by individuals following receipt of exposure notifications. To the best of our knowledge, this is the first countrywide application of the WHO/ECDC performance assessment framework.

A first set of indicators explored the exposure notification cascade performance throughout the 3 variant phases. A substantially higher SARS-CoV-2 incidence was observed during the Omicron variant phase, while active SwissCovid app use steadily declined between the Alpha and the Omicron variant phases. Compared to the peak use of the SwissCovid app in early 2021 with nearly 2 million active app users, the numbers decreased by approximately 600,000 users in March 2022. Furthermore, the early months of 2022 were marked by not only the highest SARS-CoV-2 incidence in Switzerland but also the highest absolute numbers of shared positive test results throughout the whole pandemic. This led to capacity issues in Switzerland, since an insufficient number of SARS-CoV-2 tests were available to meet such high demands. Combined with the public perception of a lower disease severity of Omicron, these 2 factors have likely contributed to the lower percentage of shared test results in later pandemic phases. A further notable difference between the 3 variant phases was that a comparatively lower proportion of issued authentication codes were entered into the app with variants of concern that appeared later in the pandemic. This may have resulted from changes in authentication code–issuing practices throughout the pandemic phases (eg, by increasingly relying on automated delivery processes), as well as possibly by a decreased acceptance of the SwissCovid app [25].

The second set of indicators focused on general test positivity in Switzerland and the proportion of individuals who tested positive for SARS-CoV-2 upon receiving an exposure notification from the SwissCovid app. The indicators illustrated a close link between test positivity and the overall SARS-CoV-2 incidence in Switzerland throughout the different phases of the pandemic. Specifically, SARS-CoV-2 case numbers and test positivity were relatively low during the Alpha variant phase but increased during the Omicron variant phase. Our individual-level analyses suggested that test positivity after receiving an exposure notification was 2-3 times higher than in the general population in the Alpha and Delta variant phases and of similar magnitude (although at very high levels) during the Omicron variant phase. Even though this assessment is based on a relatively small sample size, the observed high test positivity is plausible in a wider context since the SwissCovid app operates on more conservative Bluetooth attenuation signal thresholds compared to the apps from other countries.

The third set of indicators suggests that the mitigative actions taken by app users following the receipt of an exposure notification from the SwissCovid app may have changed over the course of the pandemic. During the Omicron variant phase, fewer people contacted the infoline or completed web forms in comparison to the Alpha and Delta variant phases. This decrease in contact attempts also resulted in relatively fewer voluntary quarantine recommendations. In the Alpha and Omicron variant phases, the proportion of reports of entering into mandatory quarantine upon receiving an exposure notification was of similar magnitude. In contrast, a higher proportion of exposure-notified app users reported to have gotten tested throughout the earlier variant phases. This may have likely been due to shifts in public perceptions regarding the disease severity of SARS-CoV-2 over time. Furthermore, it could have been a response to changing public health strategies during the Omicron variant phase, such as removing mandatory quarantine for exposed contacts in Switzerland on February 17, 2022. As suggested by the high general test positivity of 40% during the Omicron variant phase, many symptomatic or exposed individuals also relied less on SARS-CoV-2 PCR testing but, rather, self-tested or just stayed at home. Since SARS-CoV-2–infected individuals who did not get tested at official testing centers did not receive upload authentication codes, they could consequently not share their test results with proximity contacts via the SwissCovid app.

The indicators also provide insights into the possible contribution of digital proximity-tracing apps, such as SwissCovid, in mitigating viral spread. For example, the ratio of shared positive test results over the upper ceiling estimate of positive tests among app users suggest that between 60% (Alpha variant phase) and 80% (Delta and Omicron variant phases) of estimated app users who tested positive did not or were unable to share their test results. The reasons for this may include that a lower number of issued authentication codes were entered into the SwissCovid app or that there were delays in issuing authentication codes. The latter can negatively affect the potential for digital proximity tracing if exposed contacts are informed faster through other means (eg, if the number of potential contacts is small or well known and can be reached

efficiently by manual contact tracing). Nevertheless, the SwissCovid app has been shown to have advantages in timeliness and efficacy in users taking mitigative actions over manual contact tracing in recent studies. For example, 1 study revealed that app users who received an exposure notification from the SwissCovid app entered quarantine, on average, 1 day earlier than contacts who did not receive an exposure notification [16]. A simulation conducted in another study similarly found that 5% of people in manual contact tracing–mandated quarantine entered isolation after receiving a voluntary quarantine recommendation from an exposure notification [8]. The usefulness of both strategies to enable effective contact tracing can be, however, diminished by incomplete user actions. This was not observed in our study, where we found that relatively few app users who received exposure notifications ignored the exposure warning. Most of these app users undertook at least 1 recommended mitigative step in response to the notification, such as calling the infoline or completing the web form, which is in line with other studies from Switzerland [26,27].

Furthermore, relevant actions for transmission prevention were also quite frequently reported, as almost 3 out of 4 exposure-notified SwissCovid app users reported getting tested or having entered quarantine during the Omicron variant phase. These estimates fall in line with other studies using the same [28] or different Swiss survey databases [29]. However, they could be prone to reporting biases, such as social desirability bias, characterized as the tendency of survey respondents to answer questions in a manner that will be viewed favorably by others. In addition, an apparent lack of response to exposure notifications may also be due to the timing of the notification or the exposed app users' varying individual assessments of possible exposure settings and severity of transmission risks. For example, detailed reports from a Swiss study demonstrated that delayed notification, within-household exposures, or the application of preventive measures at time of exposure may be reasons for not responding to exposure notifications (Zurich Coronavirus Cohort [ZSAC]) [8].

Overall, our study contributes to the accumulating evidence of the possible contribution of digital proximity-tracing apps toward pandemic mitigation through quantitative evidence within an established public health indicator framework. However, our study also indicates various shortcomings of digital proximity-tracing apps that interfere with their ability to function at their full potential. In the case of the SwissCovid app, the flow of information along the exposure notification cascades was limited by various bottlenecks, such as delayed code delivery for test result sharing, complex user interfaces, or misaligned incentives for subsequent mitigative actions. This was observed with the SwissCovid app use visibly decreasing over time despite increasing prevalence with the more recent SARS-CoV-2 variants. The bottlenecks that may have contributed to decreased use of the SwissCovid app were recently illustrated by a study where case-contact pairs fulfilled all necessary conditions to enable exposure notifications (ie, use of the SwissCovid app, sharing of test results), but only 6 of 10 exposed contacts ended up receiving exposure notifications [26]. To enable future large-scale implementations of digital proximity-tracing apps, further testing of such apps under higher-capacity requirements, as well as co-design processes in app development, may be beneficial.

Limitations

Our study bears some limitations. The data and assessment methods used in this analysis cannot provide evidence for causality between digital proximity-tracing app use and transmission prevention. Due to a lack of clinical outcomes data, our findings are also not suited to extrapolate the population-level impact of digital proximity-tracing apps, such as avoided hospitalizations or deaths due to a lack of clinical outcome data. Moreover, despite drawing on an extensive database that includes almost 2700 individuals and 23,500 assessments, the number of recorded events of interest (ie, exposure notifications, positive SARS-CoV-2 tests, quarantine mandates) was still relatively low. This is a common issue of population-based surveys, where the probability of occurrence at any time point remains small and thus rather represents a general methodological challenge in such research. Finally, survey-driven studies may be prone to different reporting biases, including over- or underreporting of mitigative behaviors, such as noncompliance with rules and social norms. However, this was to a degree mitigated by the longitudinal nature of our data collection and repeated surveying of SwissCovid app use and outcomes, which allowed for various quality checks and did not reveal indications for systematic reporting biases.

Conclusion

Our study provides a comprehensive countrywide assessment of key indicators for the SwissCovid digital proximity-tracing app based on the WHO/ECDC framework and highlights the importance of considering the overall pandemic context in the assessment of the performance and effectiveness of such apps. For example, test positivity upon receipt of an exposure notification from the SwissCovid app was at least as high as (Omicron variant phase) or higher than (Alpha and Delta variant phases) general test positivity, with a high percentage of app users taking mitigative actions upon receiving an exposure notification. Furthermore, more than 200,000 individuals shared positive test results with the app over the course of the pandemic. Nevertheless, our indicator assessment also suggests room for improvement, including improving the speed and completeness of the exposure notification cascade or establishing stronger incentives for app use and test result sharing. Future implementations of digital proximity-tracing apps should place more emphasis on the social, psychological, and organizational aspects of the exposure notification cascade to improve their effectiveness in mitigating pandemic spread. In the context of constantly evolving requirements across different pandemic waves, the implementation of digital proximity-tracing apps should be regularly reviewed and revised.

Acknowledgments

This study was partially funded by the Digital Society Initiative (DSI). The COVID-19 Social Monitor project was funded by the Federal Office of Public Health and Health Promotion Switzerland.

Data Availability

Data used from the COVID-19 Social Monitor study and from Medgate Aktiengesellschaft are available upon request from the corresponding author. All other data sources are publicly available.

Authors' Contributions

PD revised different versions of the manuscript and approved the final manuscript. VN, TB, and DM revised and approved the final manuscript. AM and MH collected and analyzed data and revised and approved the final manuscript. VvW designed the study, interpreted the data, wrote the first draft of the manuscript, and approved the final manuscript.

Conflicts of Interest

VvW had a mandate by the Swiss Federal Office of Public Health to evaluate the SwissCovid app; however, this study was planned and executed independently, without any involvement of the Swiss Federal Office of Public Health.

Multimedia Appendix 1
Exposure notification cascade in Switzerland and related indicators.
[DOCX File , 1065 KB - publichealth_v8i11e41004_app1.docx]

Multimedia Appendix 2
Description and assessments of indicators.
[DOCX File , 31 KB - publichealth_v8i11e41004_app2.docx]

Multimedia Appendix 3
Description of data sources, assessments, and pandemic context.
[DOCX File , 26 KB - publichealth_v8i11e41004_app3.docx]

Multimedia Appendix 4
Study population and participant characteristics.
[DOCX File , 150 KB - publichealth_v8i11e41004_app4.docx]

Multimedia Appendix 5
Description of Venn diagram and subpopulations.
[DOCX File , 161 KB - publichealth_v8i11e41004_app5.docx]

Multimedia Appendix 6
Description and assessments of indicators.
[PDF File (Adobe PDF File), 88 KB - publichealth_v8i11e41004_app6.pdf]

References

1. National Health Service (NHS). NHS COVID-19 app. URL: https://www.gov.uk/government/collections/nhs-covid-19-app [accessed 2022-05-28]
2. Die Bundesregierung. Die Corona-Warn-App: Unterstützt uns im Kampf gegen Corona. URL: https://www.bundesregierung.de/breg-de/themen/corona-warn-app [accessed 2022-05-28]
3. Federal Office of Public Health (FOPH). Coronavirus: SwissCovid App. URL: https://www.bag.admin.ch/bag/en/home/krankheiten/ausbrueche-epidemien-pandemien/aktuelle-ausbrueche-epidemien/novel-cov/swisscovid-app-und-contact-tracing.html [accessed 2022-05-28]
4. Deloitte Switzerland. Smartphones Are Becoming the Control Centre of People's Lives – Only 8% of Swiss Do Not Have One. URL: https://www2.deloitte.com/ch/en/pages/press-releases/articles/deloitte-in-switzerland-smartphones-become-control-centre.html [accessed 2022-09-25]
5. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Global Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]

6. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936. [doi: 10.1126/science.abb6936]
7. von Wyl V, Bonhoeffer S, Bugnion E, Puhan MA, Salathé M, Stadler T, et al. A research agenda for digital proximity tracing apps. Swiss Med Wkly 2020 Jul 13;150:w20324 [FREE Full text] [doi: 10.4414/smw.2020.20324] [Medline: 32672340]
8. Ballouz T, Menges D, Aschmann HE, Domenghino A, Fehr JS, Puhan MA, et al. Adherence and association of digital proximity tracing app notifications with earlier time to quarantine: results from the Zurich SARS-CoV-2 cohort study. Int J Public Health 2021 Aug 16;66:1603992 [FREE Full text] [doi: 10.3389/ijph.2021.1603992] [Medline: 34471402]
9. Colizza V, Grill E, Mikolajczyk R, Cattuto C, Kucharski A, Riley S, et al. Time to evaluate COVID-19 contact-tracing apps. Nat Med 2021 Mar 15;27(3):361-362. [doi: 10.1038/s41591-021-01236-6] [Medline: 33589822]
10. Troncoso C, Bogdanov D, Bugnion E, Chatel S, Cremers C, Gürses S, et al. Deploying decentralized, privacy-preserving proximity tracing. Commun ACM 2022 Sep;65(9):48-57 [FREE Full text] [doi: 10.1145/3524107]
11. Lueks W, Benzler J, Bogdanov D, Kirchner G, Lucas R, Oliveira R, et al. Toward a Common Performance and Effectiveness Terminology for Digital Proximity Tracing Applications. Front Digit Health 2021;3:677929 [FREE Full text] [doi: 10.3389/fdgth.2021.677929] [Medline: 34713149]
12. Poletto C, Boëlle P. Learning from the initial deployment of digital contact tracing apps. Lancet Public Health 2022 Mar;7(3):e206-e207 [FREE Full text] [doi: 10.1016/S2468-2667(22)00035-4] [Medline: 35131044]
13. Wymant C, Ferretti L, Tsallis D, Charalambides M, Abeler-Dörner L, Bonsall D, et al. The epidemiological impact of the NHS COVID-19 app. Nature 2021 Jun 12;594(7863):408-412. [doi: 10.1038/s41586-021-03606-z] [Medline: 33979832]
14. Kendall M, Milsom L, Abeler-Dörner L, Wymant C, Ferretti L, Briers M, et al. Epidemiological changes on the Isle of Wight after the launch of the NHS Test and Trace programme: a preliminary analysis. Lancet Dig Health 2020 Dec;2(12):e658-e666 [FREE Full text] [doi: 10.1016/S2589-7500(20)30241-7] [Medline: 33078140]
15. Salathé M, Althaus C, Anderegg N, Antonioli D, Ballouz T, Bugnon E, et al. Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland. Swiss Med Wkly 2020 Dec 14;150:w20457 [FREE Full text] [doi: 10.4414/smw.2020.20457] [Medline: 33327003]
16. Menges D, Aschmann HE, Moser A, Althaus CL, von Wyl V. A data-driven simulation of the exposure notification cascade for digital contact tracing of SARS-CoV-2 in Zurich, Switzerland. JAMA Netw Open 2021 Apr 01;4(4):e218184 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.8184] [Medline: 33929521]
17. Daniore P, Ballouz T, Menges D, von Wyl V. The SwissCovid digital proximity tracing app after one year: were expectations fulfilled? Swiss Med Wkly 2021 Sep 08;151(35-36):w30031 [FREE Full text] [doi: 10.4414/smw.2021.w30031] [Medline: 34495624]
18. Vogt F, Haire B, Selvey L, Katelaris AL, Kaldor J. Effectiveness evaluation of digital contact tracing for COVID-19 in New South Wales, Australia. Lancet Public Health 2022 Mar;7(3):e250-e258 [FREE Full text] [doi: 10.1016/S2468-2667(22)00010-X] [Medline: 35131045]
19. World Health Organization. Indicator Framework for the Evaluation of the Public Health Effectiveness of Digital Proximity Tracing Solutions. 2021. URL: https://apps.who.int/iris/handle/10665/341818 [accessed 2022-04-20]
20. Federal Statistical Office (FSO). SwissCovid App Monitoring. URL: https://www.experimental.bfs.admin.ch/expstat/en/home/innovative-methoden/swisscovid-app-monitoring.html [accessed 2022-05-28]
21. Federal Office of Public Health (FOPH). COVID- 19 Switzerland. URL: https://www.covid19.admin.ch/en/overview [accessed 2022-04-20]
22. Oxford Covid-19 Government Response Tracker (OxCGRT). URL: https://github.com/OxCGRT/covid-policy-tracker [accessed 2022-11-02]
23. Moser A, Carlander M, Wieser S, Hämmig O, Puhan MA, Höglinger M. The COVID-19 Social Monitor longitudinal online panel: Real-time monitoring of social and public health consequences of the COVID-19 emergency in Switzerland. PLoS One 2020 Nov 11;15(11):e0242129 [FREE Full text] [doi: 10.1371/journal.pone.0242129] [Medline: 33175906]
24. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Epidemiology 2007 Nov;18(6):805-835. [doi: 10.1097/EDE.0b013e3181577511] [Medline: 18049195]
25. Covid Norms: Monitoring und Analyse von Präventionsverhalten. URL: https://www.ikmz.uzh.ch/de/research/divisions/media-use-and-effects/projects/Covid-Norms.html [accessed 2022-11-02]
26. Ballouz T, Menges D, Aschmann HE, Jung R, Domenghino A, Fehr JS, et al. Individual-level evaluation of the exposure notification cascade in the SwissCovid digital proximity tracing app: observational study. JMIR Public Health Surveill 2022 May 19;8(5):e35653 [FREE Full text] [doi: 10.2196/35653] [Medline: 35476726]
27. Daniore P, Nittas V, Moser A, Höglinger M, von Wyl V. Using Venn diagrams to evaluate digital contact tracing: panel survey analysis. JMIR Public Health Surveill 2021 Dec 06;7(12):e30004 [FREE Full text] [doi: 10.2196/30004] [Medline: 34874890]

28. von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, et al. Drivers of acceptance of covid-19 proximity tracing apps in Switzerland: panel survey analysis. JMIR Public Health Surveill 2021 Jan 06;7(1):e25701 [FREE Full text] [doi: 10.2196/25701] [Medline: 33326411]
29. Speierer A, Chocano-Bedoya P, Anker D, Schmid A, Keidel D, Vermes T, et al. The Corona Immunitas Digital Follow-Up eCohort to Monitor Impacts of the SARS-CoV-2 Pandemic in Switzerland: Study Protocol and First Results. Int J Public Health 2022;67:1604506 [FREE Full text] [doi: 10.3389/ijph.2022.1604506] [Medline: 35295967]

Abbreviations

DP-3T: Decentralized Privacy-Preserving Proximity Tracing
ECDC: European Center for Prevention and Disease Control
PCR: polymerase chain reaction
WHO: World Health Organization

Edited by A Mavragani, T Sanchez; submitted 12.07.22; peer-reviewed by Z Zrubka, CC Udeagu; comments to author 16.09.22; revised version received 28.09.22; accepted 09.10.22; published 11.11.22.

Please cite as:
Daniore P, Nittas V, Ballouz T, Menges D, Moser A, Höglinger M, Villiger P, Schmitz-Grosz K, Von Wyl V
Performance of the Swiss Digital Contact-Tracing App Over Various SARS-CoV-2 Pandemic Waves: Repeated Cross-sectional Analyses
JMIR Public Health Surveill 2022;8(11):e41004
URL: https://publichealth.jmir.org/2022/11/e41004
doi:10.2196/41004
PMID:36219833

Original Paper

The Use of a Health Compliance Monitoring System During the COVID-19 Pandemic in Indonesia: Evaluation Study

Dewi Nur Aisyah[1,2*], PhD; Logan Manikam[1,3*], PhD; Thifal Kiasatina[2*], BSc; Maryan Naman[3*], MSc; Wiku Adisasmito[2,4], Prof Dr; Zisis Kozlakidis[5], PhD

[1]Department of Epidemiology and Public Health, Institute of Epidemiology and Health Care, University College London, London, United Kingdom
[2]Indonesia One Health University Network, Depok, Indonesia
[3]Aceso Global Health Consultants Pte Limited, Faculty of Public Health, Universitas Indonesia, Depok, Indonesia
[4]Faculty of Public Health, Universitas Indonesia, Depok, Indonesia
[5]International Agency for Research on Cancer, World Health Organization, Lyon, France
[*]these authors contributed equally

Corresponding Author:
Logan Manikam, PhD
Department of Epidemiology and Public Health
Institute of Epidemiology and Health Care
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 207 679 2000
Email: logan.manikam.10@ucl.ac.uk

Abstract

Background: COVID-19 cases are soaring in Asia. Indonesia, Southeast Asia's most populous country, is now ranked second in the number of cases and deaths in Asia, after India. The compliance toward mask wearing, social distancing, and hand washing needs to be monitored to assess public behavioral changes that can reduce transmission.

Objective: This study aimed to evaluate this compliance in Indonesia between October 2020 and May 2021 and demonstrate the use of the *Bersatu Lawan COVID-19* (BLC) mobile app in monitoring this compliance.

Methods: Data were collected in real time by the BLC app from reports submitted by personnel of military services, police officers, and behavioral change ambassadors. Subsequently, the data were analyzed automatically by the system managed by the Indonesia National Task Force for the Acceleration of COVID-19 Mitigation.

Results: Between October 1, 2020, and May 2, 2021, the BLC app generated more than 165 million reports, with 469 million people monitored and 124,315,568 locations under observation in 514 districts/cities in 34 provinces in Indonesia. This paper grouped them into 4 colored zones, based on the degree of compliance, and analyzed variations among regions and locations.

Conclusions: Compliance rates vary among the 34 provinces and among the districts and cities of those provinces. However, compliance to mask wearing seems slightly higher than social distancing. This finding suggests that policy makers need to promote higher compliance in other measures, including social distancing and hand washing, whose efficacies have been proven to break the chain of transmission when combined with masks wearing.

(JMIR Public Health Surveill 2022;8(11):e40089) doi:10.2196/40089

KEYWORDS

COVID-19, public health informatics; behavioral change; digital health; public health policy; monitoring; Asia; mask; social distance; mobile app; app; transmission; policy; health compliance

Introduction

COVID-19, caused by the SARS-CoV-2 virus, remains a major global health threat. Since it was first identified in Wuhan, China, in December 2019, the virus has spread globally. As of May 2, 2021, the number of total confirmed cases stood at 151,812,556 cases, whereas the cumulative deaths reached 3,186,817. India ranks first in Asia (19,557,457 total cases and

215,542 deaths), whereas Indonesia ranks second in the region with 1,672,880 cases and 45,652 deaths as of May 1, 2021 [1].

Public health measures, in particular nonpharmaceutical interventions (NPIs), have been used consistently to reduce the likelihood of infections and community transmission [2,3]. Such measures include case isolation, voluntary home quarantine, social distancing, stopping mass gatherings, curfews, travels ban, lockdowns, as well as personal NPIs such as mask wearing, hand washing, and other health precautions. Mask wearing has been proven to be effective in reducing the likelihood of infections [4], yet its efficacy in reducing the risk of transmission is still being evaluated [4]. However, the evidence thus far indicates that when masks wearing is combined with regular hand washing and social distancing, they generally have a positive impact toward reducing SARS-CoV-2 transmissions [4]. Community mask wearing can prevent infected persons and protect uninfected wearers, which reduced the risk of infection by 79% [5-7]. In accordance with policies for disciplinary enforcement, mandating face mask use in public have been associated with a significant decline in reducing infection rates in 15 US states when comparing before and after mask mandates [8]. Furthermore, the timing in implementing such NPIs in relation to the curve of the epidemic and the population's adaptation through behavioral changes are main factors contributing to the success of NPIs [9,10].

Within this context of pandemic control, establishing the basic reproduction number of COVID-19 is critical in predicting herd immunity targets and having a relative measure of effectiveness for public health interventions [11]. However, another critical element beyond the reproduction number is the need for rapid and widespread behavioral change that remains adaptable to the changing conditions [12]. Behavioral change allowing the implementation of the NPIs mentioned above needs to be articulated clearly and internalized collectively [13], in conjunction with socioeconomic activities that aim to allow society to remain productive and safe as a whole [14].

During the current pandemic, there are some examples where existing capacities were activated and enhanced coordination mechanisms across multiple sectors, as well as toward establishing monitoring evaluation systems, thus introducing large-scale behavioral change by using health technology [15]. For instance, South Korea has invested from the outset of the pandemic in digital health solutions as a means of strengthening surveillance capacity, aided by the use of a national smartphone app for tracing and tracking infected people by GPS and combining this information with other public health measures [16-19]. Therefore, health technology applications have started to emerge as potential key solutions in the control of the COVID-19 pandemic, beyond the tracing aspect and often including advice or recommendations on personal preventive aspects, resulting to a behavioral change that could be also monitored in terms of its health protocol compliance [17,20-22].

Indonesia is administratively divided into 34 provinces and 514 cities and regencies, with independent local governments and parliamentary bodies. Often, health policies are decided at the federal level and implemented at the provincial level, as in the case of infectious disease surveillance for zoonotic diseases. The nationwide health care infrastructure includes 10,138 public health centers (*Pusat Kesehatan Masyarakat*; primary health care facilities) and 2902 hospitals (tertiary health care facilities) spread across these provinces, of which 132 hospitals are designated as national referral centers for the treatment of COVID-19 [23]. As such, the centralization of the COVID-19 response represented a departure from the routine implementation of health care response policies.

The government of Indonesia—the world's fourth most populous nation—has taken NPIs to promote behavioral changes, collectively termed as "health protocol." The health protocol consists of mask wearing (*Menggunakan masker*), hand washing (*Mencuci tangan*), and social distancing (*Menjaga jarak*). The government promoted it consistently under the popularized "3M" acronym (from the initial of each action in the Bahasa Indonesian language) and monitored public compliance thereof. This study aimed to evaluate compliance to the health protocol in public spaces between October 2020 and March 2021, thus including the entire second wave of the pandemic in Indonesia. Importantly, the data used here have been collected from the Indonesia National Task Force for the Acceleration of COVID-19 Mitigation by using the *Bersatu Lawan COVID-19* (BLC) digital monitoring app. This represents the first time in which an app for digital health, introduced nationally, produced data able to be analyzed on a real-time basis and using an integrated approach. Importantly, the system uses observer-reported compliance, thus this app is able to minimize bias from self-reported data. Additionally, this paper will describe and discuss how such data allowed the Indonesian government agencies to monitor health protocol compliance among the Indonesian public and in turn inform policy making.

Methods

The BLC Integrated System

The BLC is an integrated information system built by the National Task Force for the Acceleration of COVID-19 Mitigation. The task force was formed by the President of the Republic of Indonesia to perform, control, monitor, create, and implement strategic policies to accelerate national COVID-19 responses [24]. In performing those duties, it needed, created, and used an enhanced data reporting system to bring together and produce an in-depth analysis of the available COVID-19 information. This system aims to describe case distribution and determine the zoning of the COVID-19 transmission level, including health protocol compliance monitoring. It is first system of this kind in the country that used a big data approach, with real-time, systematic, and interoperable processes for delivering evidence-based policies [25]. The BLC system integrates data from many sectors. For example, it contains health care data (laboratory, hospital, and surveillance data) from the Ministry of Health; public transportation data; educational data from the Ministry of Education and Culture; logistics data regarding the vaccination rollout, etc. These data are obtained through the connection of different databases at ministries and agencies and are made accessible through a single interface.

The Health Protocol Compliance Monitoring System Using BLC

The BLC health protocol compliance app was developed from May to July 2020. Initially, it was designed as a means of helping Indonesian frontline public order forces (such as police and military) to move from paper-based to digital reports in monitoring compliance to the newly implemented public health restrictions (September 2020). In comparison to paper-based reports, digital reports are easier to compile and analyze. Consequently, app use included monitoring the compliance during potentially high-risk national events, such public holidays and regional elections or election campaigns. The subsequent step of app use extension (October 2020) was the inclusion of volunteers from the general public, termed behavioral change ambassadors (*Duta Perubahan Perilaku*). Behavioral change ambassador are individuals who volunteered for this role and are from a wealth of backgrounds and age groups—for instance, including from students to university lecturers, Civil Service Police Unit (*Satpol PP*) personnel, as well as from many other sectors. They are required to have digital literacy so that app use is as complete and accurate as possible and can report the data daily during their activities, particularly in monitoring the wider public health protocol adherence.

As of May 2, 2021, the app had 437,093 registered users, of which 97,598 were military personnel, 253,984 were police services personnel, and 85,511 were members of the public/ambassadors. The app itself contains a training module, showing users how to generate an account for personnel in the field, how to report data, and how to understand the dashboard's statistic results. As the monitoring can be an entry from all Indonesian levels, the account given is generated based on the regional levelling access.

The monitoring system was reported in real time using the BLC behavioral change app at public places, which tend to be crowded locations, such as markets, recreational areas, shopping malls, restaurants, places of worship, offices, train stations, bus terminals, airports, sport centers, schools, etc. Those locations were chosen based on the tendency or potential for crowds to become a place for clusters of COVID-19 transmission. Several studies have found the potential for transmission both indoors and outdoors, such as in transit places, restaurants, fitness centers, places of worship, schools, supermarkets, etc [26,27]. The reports sent include a photo of the monitoring results and an input data questionnaire by all personnel in the field. When the report data have been received, the integrated BLC system will analyze them into statistical data to determine location mapping to improve health protocol compliance. Furthermore, the information based on report data will be visualized and monitored through the BLC integrated dashboard accessible to all levels (central government, provinces, cities, districts, and subdistricts; Figure 1).

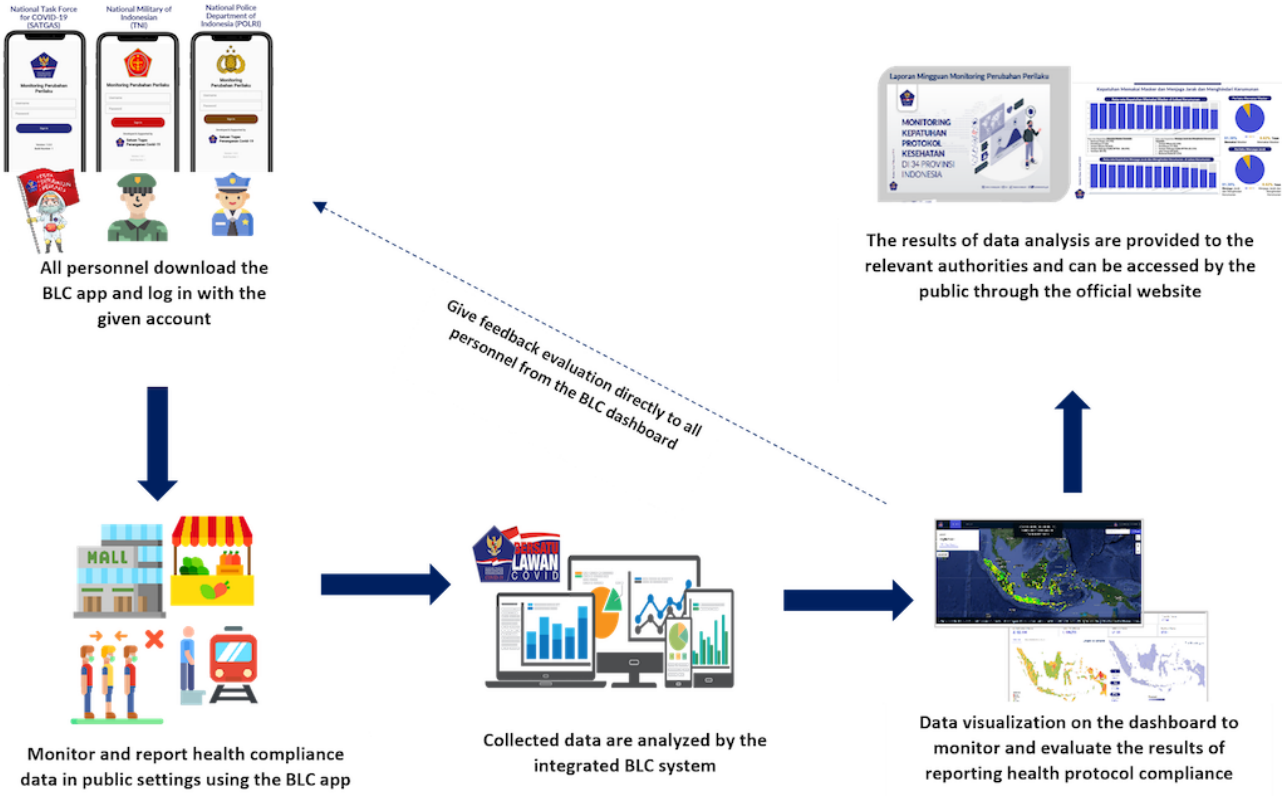
The reporting personnel from the military (*Tentara Nasional Indonesia*) and police force (*Kepolisian Negara Republik Indonesia*) are given incentives by their respective agencies, whereas the behavioral change ambassadors are community volunteers who do not receive incentives. All of them will continuously report any potential crowds and compliance levels in the local community wherever they are. In addition, there is no limit to the number of reports a person can submit per day. The emphasis is on generating objective reporting that shows compliance conditions in the field and inputting data correctly.

To ensure validation and have quality control measures, first, all personnel were provided with training on how to complete the report, and second, the report can be populated only using specific parameters and within a specific range for each variable. The report also contains mandatory fields for the collected variable; otherwise, it cannot be submitted within the system. In this way, we allow for a standardized, common, and minimum data set of information to be collected across all locations, hence allowing the real-time creation of the dashboards.

Furthermore, the quality control is conducted by having a regular randomized check by an operator at upper levels of the system (for example, reports at district levels are monitored at the provincial level). This routine monitoring process considers the number of reports collected per area, the reporting locations, the number of personnel submitting reports, as well as the quality of reports submitted. The latter is checked manually, that is, counting how many individuals in a given photo are wearing masks. If the report contains erroneous information, the person who submitted the report will be given a warning message. If a second report by the same person fails the quality control, then further reports by this person might be disqualified.

Additionally, if the personnel do not submit a complete report in the app, a warning will appear to urge the user to complete the data input. If duplications are found, the system will automatically delete data from the same villages or subdistricts (*kelurahan*).

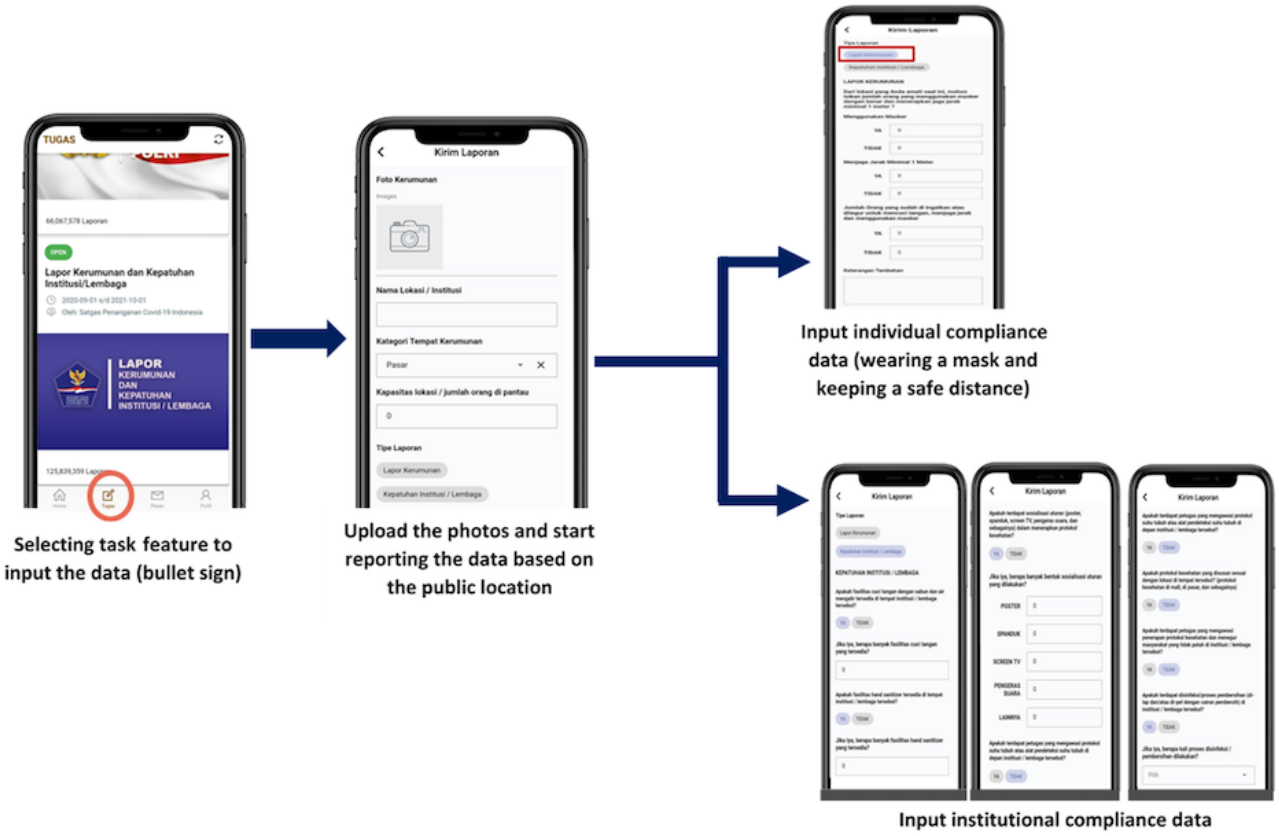
Figure 1. Overview of health protocol compliance monitoring at public places. BLC: Bersatu Lawan COVID-19. Higher-resolution version of this figure is available in Multimedia Appendix 1.



Data Collection

Data input is started by selecting the task feature in the app. There are 2 levels of compliance that are monitored, namely individual compliance and institutional compliance. Individual compliance consists of compliance with mask wearing and social distancing as well as avoiding crowds. Institutional compliance consists of monitoring hand washing facility availability, socialization of the application of health protocols, body temperature checks (using a thermo-gun or thermal body), the presence of health protocol supervisory officers, and regular disinfection activities. All personnel may frequently input data to report health protocol compliance and monitor crowd activities around them based on reporting the location using GPS (Figure 2). Some personnel can submit live report by taking photos in the field, whereas others can submit several delayed reports by the end of each day through saved pictures from their photo gallery. The National COVID Task Force actually recommends that the personnel send live reports. However, not all personnel are equipped with stable internet connection and reliable cellular phone all day long; thus, delayed reporting should still be allowed for their convenience.

Figure 2. Data input using the BLC behavioral change app. BLC: Bersatu Lawan COVID-19.



Data Analysis

The BLC system will automatically analyze the data collected to generate results and visualize the information on simplified output graphs, collectively presented on a dashboard. The data can be queried using a set of predetermined variables, based on big data analytics infrastructure. For instance, the proportion of masked individuals are calculated by dividing the number of people wearing masks by the sum of people present within a given locality (as provided by the photographic evidence). Likewise, the proportion of social distancing compliance is obtained by dividing the number of people keeping social distance by the number of people present at a given locality (the number of people who keeping social distance is calculated by using the photographic evidence).

The result also is analyzed for regional compliance zoning, both in individual and institutional compliance. These are based on the personnel's observed report on individual and institutional compliance. The map is then zoned for compliance to mask wearing and social distancing, and the data are updated in real time to show how many cities/districts have compliance levels of <60% (red), 61%-75% (orange), 76%-90% (yellow), and 91%-100% (green).

Additionally, the BLC systems also presents "institutional compliance," which refers to places or locations where crowds are likely to converge, namely markets, recreational areas, shopping malls, restaurants, places of worship, offices, train stations, bus terminals, airports, sport centers, schools, etc. Their compliance is then divided into 4 categories: "noncompliant" institutions (0%-35% compliance rate), "less compliant" institutions (35.01%-65% compliance rate), "compliant" institutions (65.01%-85% compliance rate), and "very compliant" institutions (85.01%-100% compliance rate).

Taxonomy of the Health Protocol Compliance Monitoring System

According to the typology of digital public health tools from Gasser et al [28], the typology is based on 4 main categorial variables—that is, key actors, data types, data source, and model of consent. In this system, based on the typology, our key actors are government and citizens. The data types for this system are categorized as nonsensitive, whereas the data source come from IP, GPS, and citizens. The consent is categorized as opt-in. These typologies can also identify 4 main functional categories of digital public health technologies for pandemic management, such as proximity and contact tracing, symptom monitoring, quarantine control, and flow modeling. The system for monitoring compliance presented here is the closest aligned (although not entirely overlapping) with the quarantine compliance functional category [28].

Furthermore, based on the Behavior Change Techniques taxonomy from Michie et al [29], this app was categorized in the group "Feedback and monitoring," particularly within subgroup 2.1 "Monitoring of behavior by others without feedback." Observing people in crowd locations for mask wearing and social distancing is part of data collection, with the person's knowledge being part of the behavior change strategy to reduce the risk of COVID-19 transmission [29].

In terms of ethics, especially protecting privacy, the health protocol compliance monitoring system does not collect

individual-level data. The data are collected in an aggregated format. The only individual data embedded in the system are the personnel identities of the app operators, which they have to provide so that they can complete the information input. However, this information is limited only within their own respective institutions (eg, armed or police forces) and not open to the public. In terms of preserving autonomy, the use of the compliance monitoring app is not compulsory but based on the voluntary commitment of the data providers. The app does not contain data that could be used for discrimination (eg, race, ethnic group, gender, etc); however, some areas of the country could be identified as better or worse performing at the population compliance level. This is unlikely to generate discrimination as defined by Gasser et al [28], although it might result in additional temporary restrictive measures. Finally, there are active, ongoing discussions as to a potential expiration for the collected data. However, no decision has been reached yet.

In addition, we also considered the reactivity of the subject/community during system design and development, although it was not considered an issue. The BLC app monitoring system was developed to answer the data needs related to the compliance of the Indonesian people. Indonesian frontline public order forces (such as military and police) were chosen as observers because they have the main function/duties in enforcing discipline and have already been trained in dealing with the public at large while respecting legal and ethical norms. The National police department of Indonesia (*Kepolisian Negara Republik Indonesia*) has the authority to issue warnings, fines, and social sanctions. This is in accordance with the instructions of the president and the commanders of the National Military and the National Police Chief [30].

Ethics Consideration

We declare that the data collected for this paper do not require ethical approval, because they are made available to the public by the National Task Force for the Acceleration of COVID-19 Mitigation on their website [31].

Results

Real-time Health Protocol Compliance Monitoring Report

The total number of reports gathered through BLC between October 1, 2020, and May 2, 2021, was more than 165 million, with 469 million people's behaviors monitored, observed in 124,315,568 locations in 514 districts/cities across 34 provinces in Indonesia (ie, near complete national coverage, as also explained below). Additionally, within the same period, over 508,000 institutions were observed in more than 41,235,847 locations in 504 districts/cities.

This system always received more than 680,000 reports per 24 hours as of May 2, 2021, the end of this observation period. This system also received over 2500 reports per minute and reached a peak capacity of 1894 reports per second on April 14, 2021.

The overall national figures received through BLC showed 85.89% (322,736,010/375,711,304) of the observed individuals wearing masks and 14.11% (52,975294/375,711,304) not wearing masks. Similarly, 84.13% (315,973,207/375,711,304) of people kept social distancing and 15.8% (59,738,097/375,711,304) did not, as a cumulative estimate. Figure 3 shows locations ranked according to mask-wearing and social distancing compliance.

Figure 4 demonstrates how the same information can become more granular, incorporating the relative proportion of compliant/less-compliant categories to the cumulative total. The line in the middle of graphs within Figure 4 shows the range; the longer the line in the box plot, the greater the variation in the data.

Figure 4 presents this information according to the provinces in Indonesia. It is estimated that 11 provinces have average compliance rates more than 85% (Bali, Daerah Istimewa Yogyakarta, Daerah Khusus Ibukota Jakarta, East Java, Riau Island, Central Kalimantan, East Kalimantan, North Kalimantan, West Sulawesi, North Sulawesi, and West Papua), and this rate is lower for all other provinces.

Figures 5 maps this provincial variation. Out of a total of 348 districts/cities visualized, Figure 5 shows the estimated mask-wearing compliance rate with 51 (14.66%) districts/cities in the red zone, 52 (14.94%) in the orange zone, 111 (31.9%) in the yellow zone, and 134 (38.51%) in the green zone. For social distancing compliance, Figure 5 shows that 48 (13.79%) districts/cities were in the red zone, 51 (14.66%) were in the orange zone; 126 (36.21%) were in the yellow zone; and 123 (35.34%) were in the green zone.

In terms of a wider view—and one that can be linked to NPI announcements—Figure 6 shows a weekly average of the cumulative compliance rates for the 2 categories mentioned above. Overall, estimated compliance fell in November and December (before the peak of the second wave of the pandemic), whereas it increased from January to May 2021 (during and after the second wave).

Figure 7 shows institutional compliance across Indonesian districts and cities: 126 (46.67%) districts had a high rate of noncompliant institutions, 15 (5.56%) had a smaller rate of noncompliant institutions, 7 (2.59%) had compliant institutions, and 122 (45.19%) had very compliant institutions. Figure 8 shows the association of health protocol compliance in relation to the weekly number of COVID-19 cases during the second wave in 2020 (this is the first wave where the mobile app was implemented). The graph shows that there was a lower level of compliance before the advent of the second wave and that compliance rose significantly as the wave progressed and as further public health measures were introduced and monitored.

Figure 3. The average of the compliance of (A) wearing marks and (B) social distancing at crowded locations cumulatively (reports submitted from October, 1 2020, to May 2, 2021). Locations are divided by function and identified as areas of the highest risk for COVID-19 transmission.

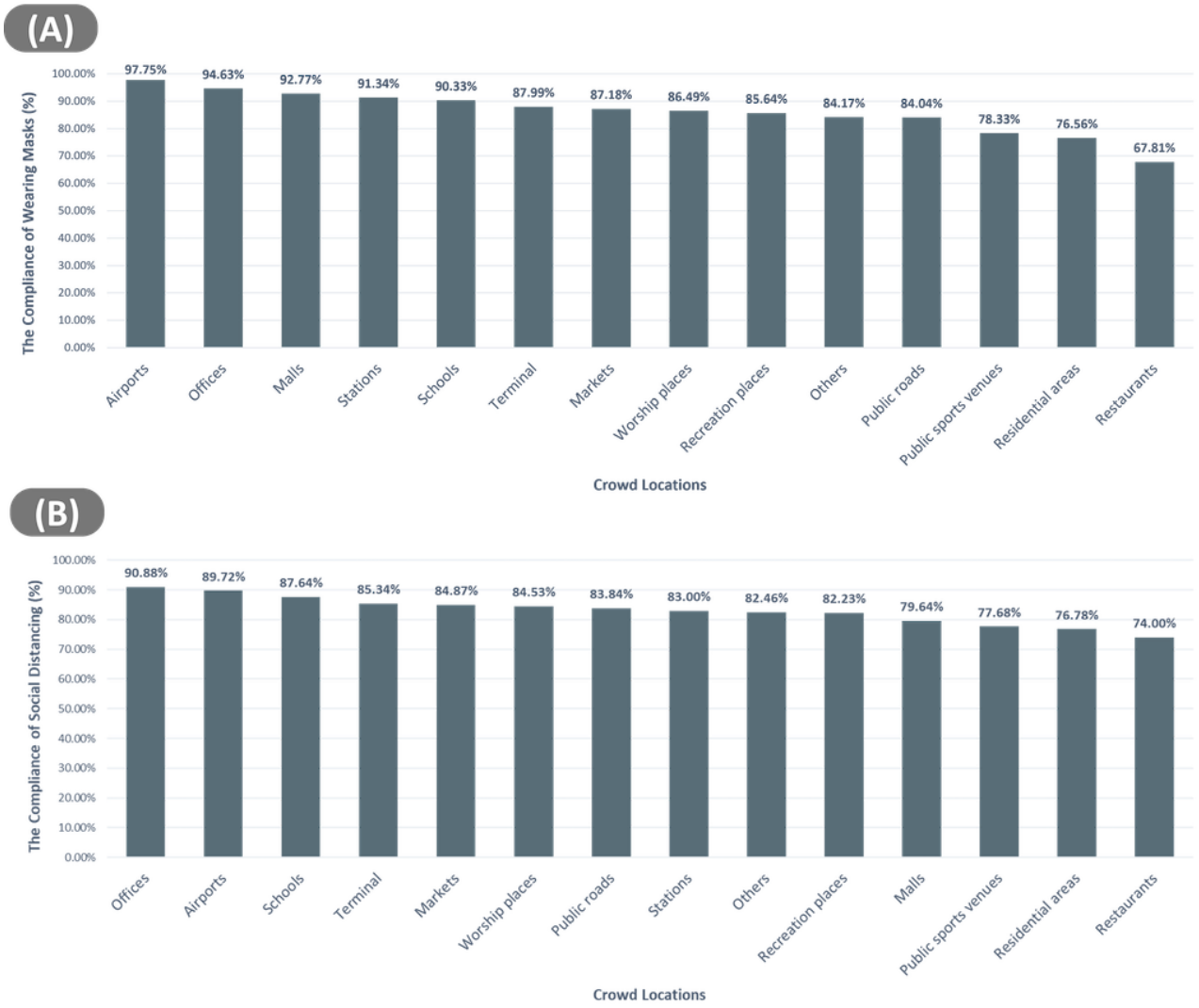


Figure 4. The number of the lowest, highest, and average compliance rates for (A) wearing masks and (B) social distancing from all districts/cities in 34 provinces, calculated in the last 7 days as of May 2, 2021. (There were no reports for the last 7 days in North Maluku Province). DI: Daerah Istimewa; DKI: Daerah Khusus Ibukota; Kep.: Kepulauan; NTB: Nusa Tenggara Barat; NTT: Nusa Tenggara Timur.

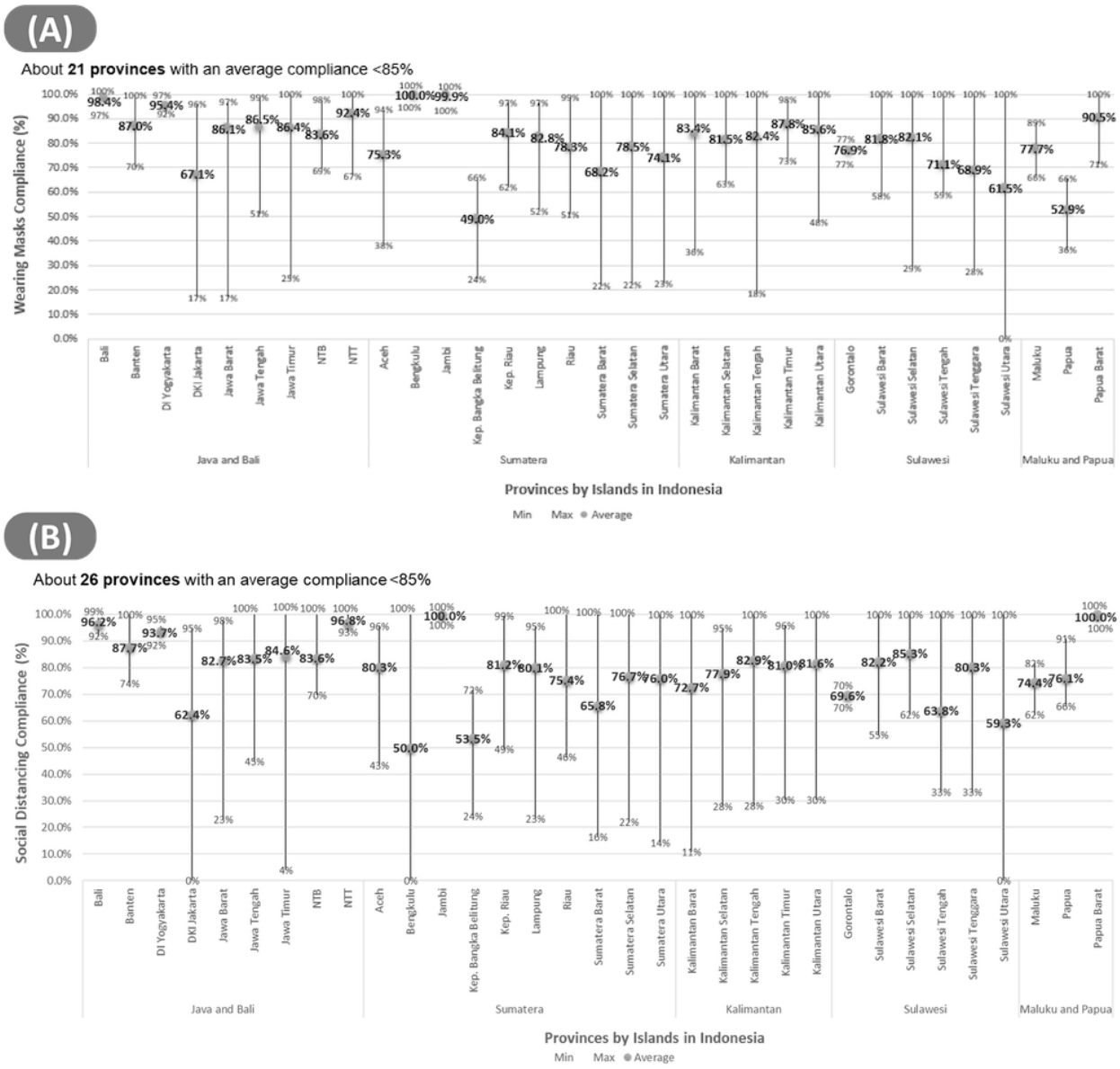


Figure 5. The zoning map of (A) wearing masks and (B) social distancing compliance, calculated in the last 7 days as of May 2, 2021.

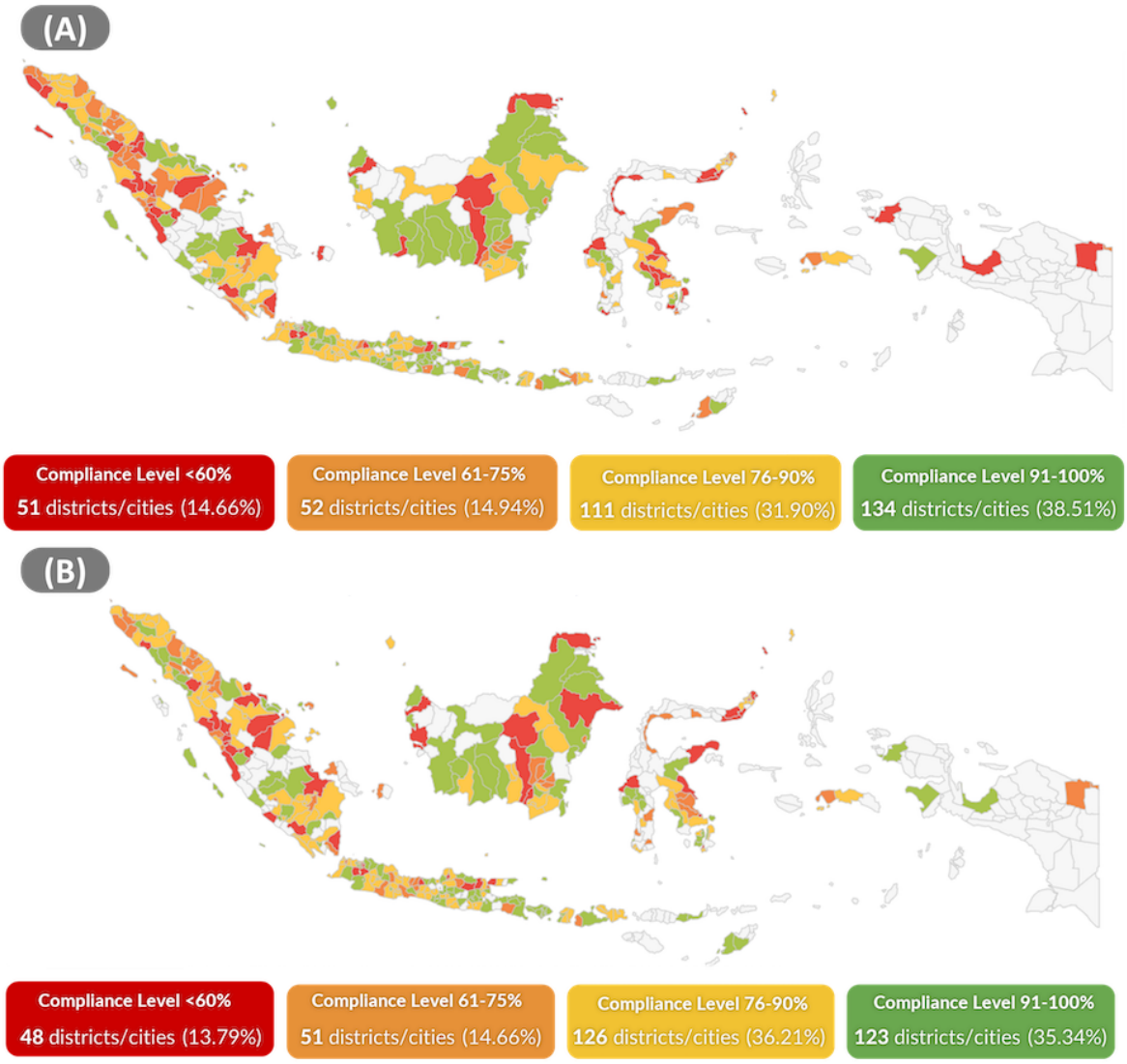


Figure 6. The compliance zoning development by the number of districts/cities in a weekly period from October 4, 2020 to May 2, 2021: (A) mask wearing and (B) social distancing.

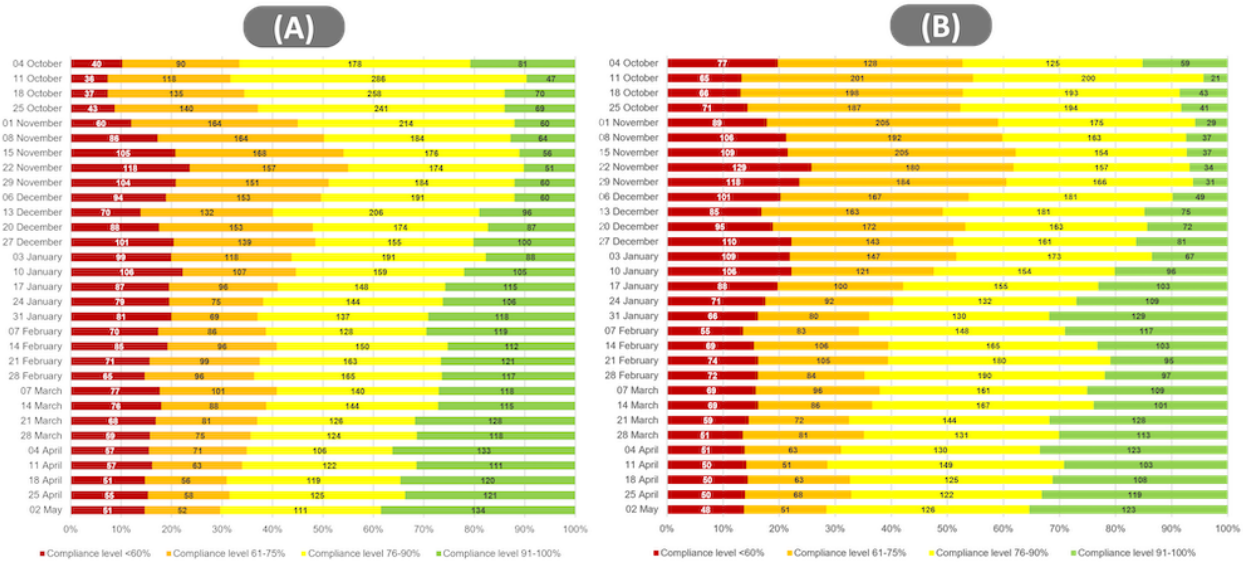


Figure 7. The zoning map of institution compliance, calculated in the last 7 days leading up to May 2, 2021.

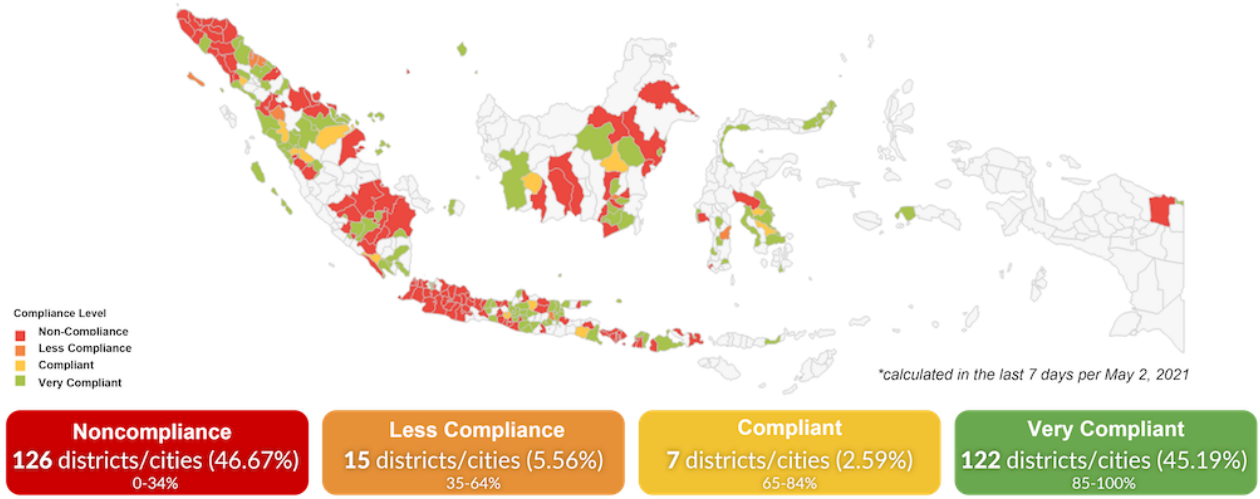
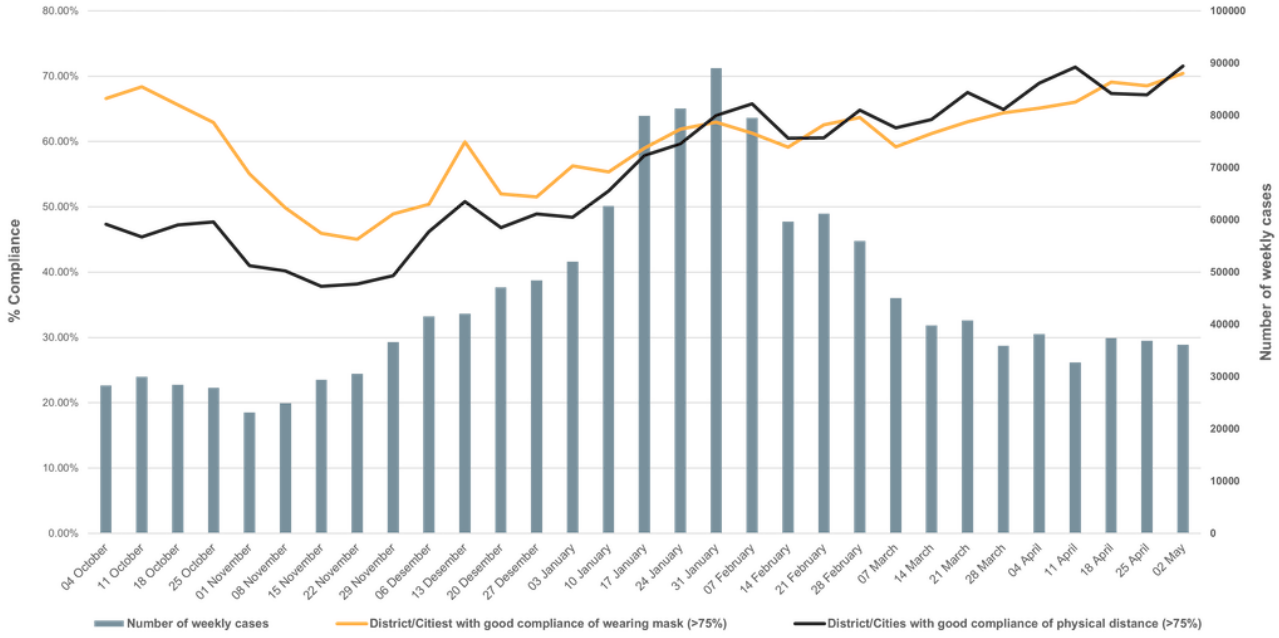


Figure 8. Proportion of health protocol compliance in relation to the weekly number of COVID-19 cases.



Discussion

Principal Findings

Despite variation between provinces and among districts and cities, which are illustrated using 4 colored zones, the overall majority of districts/cities demonstrate medium to high compliance for both mask wearing and social distancing. Compliance to mask wearing is higher than to social distancing, as would be expected for being an easier element to be controlled at an individual level. These results demonstrate good correspondence to the survey results from the Indonesian National Bureau of Statistics, conducted in September 2020, which showed 91.98% of respondents always wearing masks when leaving their homes, as opposed to 73.53% of respondents who stated that they always keep social distancing when leaving their homes [32]. Furthermore, the Indonesian National Bureau of Statistics results showed that more than half of the respondents self-stating their own noncompliance thought there were no penalties from the authorities for such noncompliance; more than a third did not comply as they could not see or hear a COVID-19 case in their immediate familial environment; and nearly all noncompliant individuals perceived the protocols as disturbing for them in performing their jobs [32].

This seemingly higher compliance with mask wearing compared to social distancing is an interesting finding despite the complication and hesitancy of mask wearing observed across the globe [33] and the limited evidence available during the time of the reported observations to claim its efficacy in breaking the chain of transmission [4,19]. Seeing the relatively high reported compliance that countries, such as Indonesia, not used to wearing masks routinely are able to do so is a positive sign for the penetration of the public health messages. However, compliance is variable between different types of activities, and as such, the messaging might have to be nuanced to promote other measures, such as hand washing, that are the most effective at a population level when combined with masks wearing [7,34].

The results have shown the variation of compliance rate between provinces and among the districts and cities in provinces. However, Bali and Daerah Istimewa Yogyakarta provinces have average compliance rates of more than 85% in mask wearing and social distancing. This finding might be due to the high number of field personnel from the police and military who are deployed in these provinces to ensure health protocol compliance [35,36]. These provinces are among the most popular tourist destinations in Indonesia [37], with a higher likelihood of crowd gathering and thus attracting a higher level of policing.

For future references, the insights on health protocol compliance monitoring across all 34 provinces are updated regularly on the National Task Force for the Acceleration of COVID-19 Mitigation website [31], with the latest one posted on September 25, 2022 [38].

Strengths and Limitations

Therefore, the number of reports generated by this BLC behavioral change app might be restricted by the numbers and locations of reporters. Nevertheless, this study has revealed the insights from a digital reporting system that can benefit policy makers in monitoring behavioral changes when the reporting is done comprehensively and using big data analytics. One of the factors supporting this monitoring's success is its real-time data collection at a micro-scale, based on cloud technology. This enables data interconnection among districts, cities, and provinces, which can be analyzed altogether by the Indonesia National Task Force for the Acceleration of COVID-19 Mitigation. In a large and decentralized country such as Indonesia, data interconnection is key to obtaining national analysis and informing effective evidence-based policies.

Police and military forces have made major contributions to supplying these real-time data. Although military forces involvement in a health crisis remains a contested idea [39], this case can be an additional example of the essential roles of the police and military in COVID-19 response within Indonesia's large territory [40].

However, this study also has certain limitations. First, the app is provided only for users who have Android smartphones. Second, human errors are still found in the reports, such as irrelevant pictures being uploaded to the system. Third, the reporters are limited to personnel and ambassadors in several public spaces. In the future, this app might expand the reporters to the wider public to generate reports from more categories of public spaces.

Conclusion

To conclude, this paper has demonstrated the importance of promoting NPIs to prevent COVID-19 transmission and case surge. These interventions require public behavioral changes to wear masks, keep social distancing, and wash hands frequently. This paper discovers that the need to monitor these behavioral changes can be done through a mobile app. Therefore, this paper discusses the example of the BLC behavioral change app as used in Indonesia, the most populous country in Southeast Asia, whose COVID-19 cases are ranked second in Asia, after India, to date.

This paper discusses the multisectoral coordination behind the development and report submissions to this app, which includes police officers, military personnel, and community ambassadors. It further discovers how the big data analytics have been used to analyze these reports on a weekly basis to provide updates to policy makers and inform government COVID-19 response policies through the Indonesia National Task Force for the Acceleration of COVID-19 Mitigation.

Based on the data gathered through the app during the period from October 1, 2020, to May 2, 2021, it is apparent that compliance rate varies among the 34 provinces and among the districts and cities of those provinces. However, it is interesting to find that compliance to mask wearing seems to be slightly higher than social distancing. Although this can be a positive finding on behavioral change promotion, policy makers need to promote higher compliance in other measures, including social distancing and hand washing, whose efficacies have been proven to break the chain of transmission when combined with mask wearing. Nevertheless, this app has provided data that can inform public behavior patterns, which can inform policy makers to take the necessary actions to prevent a surge in COVID-19 cases.

Acknowledgments

The authors would like to thank the Indonesia National Task Force for the Acceleration of COVID-19 Mitigation for providing access and insights to the *Bersatu Lawan COVID-19* behavioral change app data and development process; the National Military of Indonesia (*Tentara Nasional Indonesia* [TNI]); National Police Department of Indonesia (*Kepolisian Negara Republik Indonesia* [POLRI]); and the community ambassadors who have supported the reporting processes.

No author received funding for the creation of this manuscript

Although authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization (WHO), the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or, views of the International Agency for Research on Cancer/WHO.

Authors' Contributions

DNA created the paper concept, main structure, and methodology. TK accessed and analyzed the data. WA provided access to the data. DNA and TK created the first draft of the paper. ZK, MN, and LM reviewed and edited the paper to meet sufficient academic standards. DNA is the guarantor of this paper.

Conflicts of Interest

Although LM and MN are affiliated with Aceso Global Health Consultants Pte Limited, which is a private company, we declare that this research project does not receive funding from Aceso Global Health Consultants. The company does not have a role in the study design, data collection, and analysis; decision to publish; or preparation of the manuscript. LM is the director of the company, and MN is a consultant of the company. However, both of them contributed to this paper on a pro bono basis. All authors declared no other conflicts of interest.

Multimedia Appendix 1
Higher resolution version of Figure 1. Overview of health protocol compliance monitoring at public places. BLC: Bersatu Lawan COVID-19.
[PNG File , 907 KB - publichealth_v8i11e40089_fig.png]

References

1. Weekly epidemiological update on COVID-19 - 4 May 2021. World Health Organization. 2021 May 04. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---4-may-2021 [accessed 2021-05-06]
2. Perra N. Non-pharmaceutical interventions during the COVID-19 pandemic: a review. Phys Rep 2021 May 23;913:1-52 [FREE Full text] [doi: 10.1016/j.physrep.2021.02.001] [Medline: 33612922]
3. Ferguson NM, Cummings DA, Cauchemez S, Fraser C, Riley S, Meeyai A, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 2005 Sep 08;437(7056):209-214. [doi: 10.1038/nature04017] [Medline: 16079797]
4. Rader B, White LF, Burns MR, Chen J, Brilliant J, Cohen J, et al. Mask-wearing and control of SARS-CoV-2 transmission in the USA: a cross-sectional study. Lancet Digit Health 2021 Mar;3(3):e148-e157 [FREE Full text] [doi: 10.1016/S2589-7500(20)30293-4] [Medline: 33483277]
5. Brooks JT, Butler JC. Effectiveness of mask wearing to control community spread of SARS-CoV-2. JAMA 2021 Mar 09;325(10):998-999 [FREE Full text] [doi: 10.1001/jama.2021.1505] [Medline: 33566056]
6. Wang Y, Tian H, Zhang L, Zhang M, Guo D, Wu W, et al. Reduction of secondary transmission of SARS-CoV-2 in households by face mask use, disinfection and social distancing: a cohort study in Beijing, China. BMJ Glob Health 2020 May 28;5(5):e002794 [FREE Full text] [doi: 10.1136/bmjgh-2020-002794] [Medline: 32467353]
7. Burnett ML, Sergi CM. Face masks are beneficial regardless of the level of infection in the fight against COVID-19. Disaster Med Public Health Prep 2020 Oct;14(5):e47-e50 [FREE Full text] [doi: 10.1017/dmp.2020.320] [Medline: 32900420]
8. Lyu W, Wehby GL. Community use of face masks and COVID-19: evidence from a natural experiment of state mandates in the US. Health Aff (Millwood) 2020 Aug 01;39(8):1419-1425. [doi: 10.1377/hlthaff.2020.00818] [Medline: 32543923]
9. Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, RADAR-CNS Consortium. Using smartphones and wearable devices to monitor behavioral changes during COVID-19. J Med Internet Res 2020 Sep 25;22(9):e19992 [FREE Full text] [doi: 10.2196/19992] [Medline: 32877352]
10. Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, Usher Network for COVID-19 Evidence Reviews (UNCOVER) group. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. Lancet Infect Dis 2021 Feb;21(2):193-202 [FREE Full text] [doi: 10.1016/S1473-3099(20)30785-4] [Medline: 33729915]
11. Linka K, Peirlinck M, Kuhl E. The reproduction number of COVID-19 and its correlation with public health interventions. Comput Mech 2020 Jul 28;66(4):1035-1050 [FREE Full text] [doi: 10.1007/s00466-020-01880-8] [Medline: 32836597]
12. Betsch C, Wieler LH, Habersaat K, COSMO group. Monitoring behavioural insights related to COVID-19. Lancet 2020 Apr 18;395(10232):1255-1256 [FREE Full text] [doi: 10.1016/S0140-6736(20)30729-7] [Medline: 32247323]
13. Abdulkareem SA, Augustijn EW, Filatova T, Musial K, Mustafa YT. Risk perception and behavioral change during epidemics: comparing models of individual and collective learning. PLoS One 2020 Jan 6;15(1):e0226483 [FREE Full text] [doi: 10.1371/journal.pone.0226483] [Medline: 31905206]
14. Chang Y, Chien C, Shen LF. Telecommuting during the coronavirus pandemic: future time orientation as a mediator between proactive coping and perceived work productivity in two cultural samples. Pers Individ Dif 2021 Mar;171:110508 [FREE Full text] [doi: 10.1016/j.paid.2020.110508] [Medline: 33191964]
15. Responding to community spread of COVID-19: interim guidance, 7 March 2020. World Health Organization. 2020 Mar 07. URL: https://apps.who.int/iris/handle/10665/331421 [accessed 2021-01-12]

16. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
17. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health 2020 Aug;2(8):e435-e440 [FREE Full text] [doi: 10.1016/S2589-7500(20)30142-4] [Medline: 32835201]
18. Park SW, Sun K, Viboud C, Grenfell BT, Dushoff J. Potential role of social distancing in mitigating spread of coronavirus disease, South Korea. Emerg Infect Dis 2020 Nov;26(11):2697-2700 [FREE Full text] [doi: 10.3201/eid2611.201099] [Medline: 32795385]
19. Bae SY, Chang P. The effect of coronavirus disease-19 (COVID-19) risk perception on behavioural intention towards 'untact' tourism in South Korea during the first wave of the pandemic (March 2020). Current Issues in Tourism 2020 Jul 27;24(7):1017-1035. [doi: 10.1080/13683500.2020.1798895]
20. Kim JH, Choi WS, Song JY, Yoon YK, Kim MJ, Sohn JW. The role of smart monitoring digital health care system based on smartphone application and personal health record platform for patients diagnosed with coronavirus disease 2019. BMC Infect Dis 2021 Feb 27;21(1):229 [FREE Full text] [doi: 10.1186/s12879-021-05898-y] [Medline: 33639861]
21. Scott BK, Miller GT, Fonda SJ, Yeaw RE, Gaudaen JC, Pavliscsak HH, et al. Advanced digital health technologies for COVID-19 and future emergencies. Telemed J E Health 2020 Oct 01;26(10):1226-1233. [doi: 10.1089/tmj.2020.0140] [Medline: 32456560]
22. Jose T, Warner DO, O'Horo JC, Peters SG, Chaudhry R, Binnicker MJ, et al. Digital health surveillance strategies for management of coronavirus disease 2019. Mayo Clin Proc Innov Qual Outcomes 2021 Feb;5(1):109-117 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2020.12.004] [Medline: 33521582]
23. Minister of Health. Minister of Health Decree No. HK.01.07/MENKES/169/2020 regarding the Appointment of Referral Hospitals for Selected Emerging Infectious Disease. Indonesia: Minister of Health; Mar 10, 2020.
24. President of the Republic of Indonesia. Peraturan Presiden Republik Indonesia nomor 82 tahun 2020 tentang Komite Penanganan Corona Virus Disease 2019 (COVID-19) dan Pemulihan Ekonomi Nasional. Database Peraturan. 2020 Jul 20. URL: https://peraturan.bpk.go.id/Home/Details/141403/perpres-no-82-tahun-2020 [accessed 2022-11-08]
25. Younes Y. Benchmarking and sizing your Elasticsearch cluster for logs and metrics Internet. Elastic. 2020 Oct 29. URL: https://www.elastic.co/blog/benchmarking-and-sizing-your-elasticsearch-cluster-for-logs-and-metrics [accessed 2021-05-06]
26. Transmission of SARS-CoV-2: implications for infection prevention precautions Internet. World Health Organization. 2020 Jul 09. URL: https://tinyurl.com/wutbd6m5 [accessed 2021-07-09]
27. Leclerc Q, Fuller NM, Knight LE, CMMID COVID-19 Working Group, Funk S, Knight GM. What settings have been linked to SARS-CoV-2 transmission clusters? Wellcome Open Res 2020 Jun 05;5:83 [FREE Full text] [doi: 10.12688/wellcomeopenres.15889.2] [Medline: 32656368]
28. Gasser U, Ienca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. Lancet Digit Health 2020 Aug;2(8):e425-e434 [FREE Full text] [doi: 10.1016/S2589-7500(20)30137-0] [Medline: 32835200]
29. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
30. President of the Republic of Indonesia. Regulation of the President of the Republic of Indonesia number 82 of 2020 regarding the for COVID-19 Handling and National Economic Recovery Committee. Cabinet Secretariat of the Republic of Indonesia. 2020 Jul 20. URL: https://setkab.go.id/en/govt-issues-regulation-on-covid-19-mitigation-national-economic-recovery/ [accessed 2022-11-17]
31. COVID-19 National Task Force. Monitoring Kepatuhan Protokol Kesehatan. 2021 May 02. URL: https://covid19.go.id/monitoring-kepatuhan-protokol-kesehatan [accessed 2022-06-22]
32. Community behavior during the COVID-19 pandemic: results of a survey of community behavior during the COVID-19 pandemic. Article in Indonesian. Central Bureau of Statistic. URL: https://tinyurl.com/bdhwkdpk [accessed 2020-11-04]
33. Haelle T. Should everyone wear a mask in public? maybe—but it's complicated. Forbes. 2020 Apr 01. URL: https://tinyurl.com/36yd3bsp [accessed 2021-03-18]
34. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020 Jun 27;395(10242):1973-1987 [FREE Full text] [doi: 10.1016/S0140-6736(20)31142-9] [Medline: 32497510]
35. Wijana E, Weadcaksana H. Awasi PTKM, Satpol-PP DIY Siapkan 6 Tim Gabungan TNI dan Polri. Article in Indonesian. Suara. 2021 Jan 11. URL: https://jogja.suara.com/read/2021/01/11/131753/awasi-ptkm-satpol-pp-diy-siapkan-6-tim-gabungan-tni-dan-polri?page=all [accessed 2022-11-08]
36. Sulistyowati A. Polisi Bali Awasi Kerumunan di Malam Tahun Baru. Kompas. URL: https://nasional.kompas.com/read/2011/12/29/13432120/polisi.bali.awasi.kerumunan.di.mal [accessed 2021-12-29]

37. Tanjung I. 5 most popular destinations in Indonesia for holidaymakers. The Jakarta Post. 2016 May 08. URL: https://www.thejakartapost.com/travel/2016/05/08/5-most-popular-destinations-in-indonesia-for-holidaymakers.html [accessed 2021-03-12]
38. Monitoring Kepatuhan Protokol Kesehatan Tingkat Nasional (Update per 25 September 2022). Article in Indonesian. Bidang Data dan IT Satuan Tugas Penanganan COVID-19. 2021 Sep 25. URL: https://covid19.go.id/artikel/2022/09/30/monitoring-kepatuhan-protokol-kesehatan-tingkat-nasional-update-25-september-2022 [accessed 2021-11-17]
39. Gibson-Fall F. Military responses to COVID-19, emerging trends in global civil-military engagements. Rev Int Stud 2021 Jan 21;47(2):155-170. [doi: 10.1017/s0260210521000048]
40. Djalante R, Lassa J, Setiamarga D, Sudjatma A, Indrawan M, Haryanto B, et al. Review and analysis of current responses to COVID-19 in Indonesia: period of January to March 2020. Prog Disaster Sci 2020 Apr;6:100091 [FREE Full text] [doi: 10.1016/j.pdisas.2020.100091] [Medline: 34171011]

Abbreviations

BLC: Bersatu Lawan COVID-19
NPI: nonpharmaceutical intervention

Edited by A Mavragani, G Eysenbach; submitted 27.06.22; peer-reviewed by SM Lei, T Scherr; comments to author 25.07.22; revised version received 31.08.22; accepted 09.10.22; published 22.11.22.

Please cite as:
Aisyah DN, Manikam L, Kiasatina T, Naman M, Adisasmito W, Kozlakidis Z
The Use of a Health Compliance Monitoring System During the COVID-19 Pandemic in Indonesia: Evaluation Study
JMIR Public Health Surveill 2022;8(11):e40089
URL: https://publichealth.jmir.org/2022/11/e40089
doi:10.2196/40089
PMID:36219836

Review

Underestimated Prevalence of HIV, Hepatitis B Virus (HBV), and Hepatitis D Virus (HDV) Triple Infection Globally: Systematic Review and Meta-analysis

Sisi Chen[1,2*], MM; Feng Ren[1*], PhD; Xiaojie Huang[3*], PhD; Ling Xu[1], PhD; Yao Gao[1], PhD; Xiangying Zhang[1], PhD; Yaling Cao[1], MM; Zihao Fan[1], PhD; Yuan Tian[1], PhD; Mei Liu[2*], PhD

[1]Beijing Institute of Hepatology, Beijing Youan Hospital, Capital Medical University, Beijing, China

[2]Department of Oncology, Beijing Youan Hospital, Capital Medical University, Beijing, China

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

[*]these authors contributed equally

Corresponding Author:
Mei Liu, PhD
Department of Oncology
Beijing Youan Hospital
Capital Medical University
No. 8 You An Men Wai Street
Fengtai District
Beijing, 100069
China
Phone: 86 13581980530
Email: liumei@ccmu.edu.cn

Abstract

Background: Hepatitis delta virus (HDV) is a satellite RNA virus that relies on hepatitis B virus (HBV) for transmission. HIV/HBV/HDV coinfection or triple infection is common and has a worse prognosis than monoinfection.

Objective: We aimed to reveal the epidemiological characteristics of HIV/HBV/HDV triple infection in the global population.

Methods: A systematic literature search in PubMed, Embase, and the Cochrane Library was performed for studies of the prevalence of HIV/HBV/HDV triple infection published from January 1, 1990, to May 31, 2021. The Der Simonian-Laird random effects model was used to calculate the pooled prevalence.

Results: We included 14 studies with 11,852 participants. The pooled triple infection rate in the global population was 7.4% (877/11,852; 95% CI 0.73%-29.59%). The results of the subgroup analysis showed that the prevalence of triple infection was significantly higher in the Asian population (214/986, 21.4%; 95% CI 7.1%-35.8%), in men (212/5579, 3.8%; 95% CI 2.5%-5.2%), and in men who have sex with men (216/2734, 7.9%; 95% CI 4.3%-11.4%). In addition, compared with people living with HIV, the HIV/HBV/HDV triple infection rate was higher in people with hepatitis B.

Conclusions: This meta-analysis suggests that the prevalence of HIV/HBV/HDV triple infection in the global population is underestimated, and we should focus more effort on the prevention and control of HIV/HBV/HDV triple infection.

Trial Registration: PROSPERO CRD42021273949; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=273949

(JMIR Public Health Surveill 2022;8(11):e37016) doi:10.2196/37016

KEYWORDS

HIV; HBV; HDV; triple infection; epidemiology; public health

Introduction

Hepatitis D virus (HDV) is a peculiar, small, defective virus that requires the assistance of hepatitis B virus (HBV) surface antigen (HBsAg) for replication and pathogenesis [1]. Accordingly, HDV infection can occur via either coinfection with HBV or superinfection in patients with chronic hepatitis B. The main transmission routes of HDV are parenteral and sexual contact. In addition, mother-to-child transmission can occur [2]. Despite being a defective virus, HDV infection is

widely perceived as the most severe and aggressive form of human viral hepatitis. Approximately 10% to 15% of patients with hepatitis D progress to cirrhosis within 1 year to 2 years, and 70% to 80% of patients progress to cirrhosis within 5 years to 10 years [3]. Moreover, HDV infection is more prone to hepatic decompensation and is associated with a higher risk of hepatocellular carcinoma [4]. However, HDV infection has been considered a relatively rare disease over the past decades as a result of the universal promotion of HBV vaccination and the clinical neglect of HDV detection. According to recent meta-analyses, the approximate HDV infection rate is 4.5% to 14.57% in the HBsAg-positive population, affecting up to nearly 72 million individuals worldwide [5,6].

Given the shared transmission routes with HIV, HIV/HBV/HDV triple infection is relatively common [7]. HIV/HBV/HDV triple infection is not only widespread but also associated with worse outcomes than monoinfection. First, it can have a negative impact on disease progression for people living with HIV. Combination with hepatitis virus infection may promote immune activation, causing dysfunction of $CD4^+$ and $CD8^+$ T lymphocytes and natural killer cells, resulting in poor immune recovery after antiretroviral therapy, thus affecting AIDS disease progression [8,9]. Additionally, HIV combined with HBV or HDV infection significantly reduces the clearance rate of these 2 types of hepatitis virus and prolongs the course of hepatitis [10]. Meanwhile, liver fibrosis is significantly accelerated after coinfection, and patients are also at higher risk of mortality due to liver cirrhosis, hepatic decompensation, hepatocellular carcinoma, and other liver diseases [11,12]. Significantly higher rates of poorer prognosis also occur. Thus, the disease burden of HIV/HBV/HDV triple infection appears more serious than initially expected.

There have been several meta-analyses exploring the global HIV/HBV coinfection rate; however, the prevalence of HIV/HBV/HDV triple infection remains largely unknown. Chu et al [13] examined the prevalence of multiple hepatitis viruses and HIV infection among drug users in Taiwan and found that HIV/HBV/HDV infection rates were as high as 16.7% among HIV-positive drug users. Shen et al [14] pooled the HIV/HBV/HDV triple infection rate in the global population from 2002 to 2018 and estimated a triple infection rate of only 1.03% in people living with HIV. In addition, Nicolini et al [15], in 2015, tested triple infection in blood samples from the Italian general population and found a triple infection rate of 3.5%. In general, the results of current studies on HIV/HBV/HDV triple infection rates fluctuate widely, and some studies may underestimate triple infection rates to some extent due to the small number of included samples or limitations of the included population characteristics [14]. This systematic review and meta-analysis aimed to determine a high reliability estimate of the prevalence of HIV/HBV/HDV triple infection in people with HBV infection or living with HIV globally.

Methods

This meta-analysis was conducted based on the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines, and it was registered in PROSPERO (CRD42021273949).

Search Strategy

A systematic literature search in PubMed, Embase, and the Cochrane Library was performed for studies on the prevalence of HBV, HIV/AIDS, and HDV triple infection published from January 1, 1990, to May 31, 2021.

Selection Criteria

Inclusion Criteria

Studies were selected based on the following inclusion criteria: (1) study participants were HIV or HBV monoinfected or HBV and HIV coinfected; (2) the diagnosis of infection with HIV or HBV met the international uniform standards; (3) studies related to the prevalence of triple infection were cross-sectional studies, and those related to incidence and risk factors associated with triple infection were case control studies or cohort studies; (4) in the original study, coinfection with HBV was defined as HBsAg-positive, coinfection with HDV was defined as anti-HDV positive, and coinfection with HIV was confirmed by Western blot.

Exclusion Criteria

Studies were excluded if they (1) were case reports or review articles, (2) had a research sample size of less than 50 participants, (3) were duplicate studies, or (4) had incomplete or unclear study information. The study screening was carried out independently by 2 reviewers, who both read the full text and screened the studies that met the inclusion and exclusion criteria. Disagreements between reviewers about inclusion were resolved by consulting third-party experts.

Data Extraction

Two researchers independently extracted and coded data using an Excel spreadsheet. The data obtained included basic information of the included studies, including the first author, year of publication, study period, research type, study location, age or sex distribution, total number of participants, number of participants with HIV/HBV/HDV triple infection, and crude prevalence rate.

Quality Assessment

The Newcastle-Ottawa Scale (NOS; 11 items in total, out of 11 points) [16] was used to evaluate the quality of the included cohort studies, and the Agency for Healthcare Research and Quality (AHRQ) questionnaire (9 items in total, out of 10 points) [17] was used to evaluate the quality of the included cross-sectional studies. We used 3 grades: A, B, and C. Grade A corresponds to 7-10 points on the NOS scale and 8-11 “Yes” responses on the AHRQ questionnaire. Grade B corresponds to 3-6 points on the NOS scale and 4-7 “Yes” responses on the AHRQ questionnaire. Grade C corresponds to 0-2 points on the NOS scale and 0-3 “Yes” responses on the AHRQ questionnaire.

Statistical Analysis

We used Stata software for the meta-analysis. Heterogeneity was assessed statistically using the I^2 measurement. The threshold for the heterogeneity test result was 0.05, and that of

the level of goodness-of-fit test was 0.10. If heterogeneity was high (P<.10 or I^2>50%), the random effects model was used to calculate pooled prevalence estimates with 95% CIs. Subgroup analysis was conducted according to the basic disease, research continent, research country, and HIV transmission route to explore the source of heterogeneity. The Egger linear regression method combined with the observation funnel plot was used to evaluate the publication bias. We evaluated the stability of the model through a sensitivity analysis.

Results

Literature Search

The detailed flow of the literature search is shown in Figure 1. The literature search yielded 359 studies from 3 databases (Embase: 203; PubMed: 132; Cochrane Library: 24). After removing 90 duplicates, the remaining literature was screened for titles and abstracts. Of the 61 studies that underwent full-text assessment, we excluded 48 studies because they investigated treatments, were duplicate studies, had a small sample size (<50), or had unclear study information. Finally, 14 articles with a total of 11,852 participants were included [10,11,15,18-28]. Of the included studies, 9 were prospective cohort studies, and 5 were cross-sectional studies (Table 1).

Figure 1. Literature identification process.

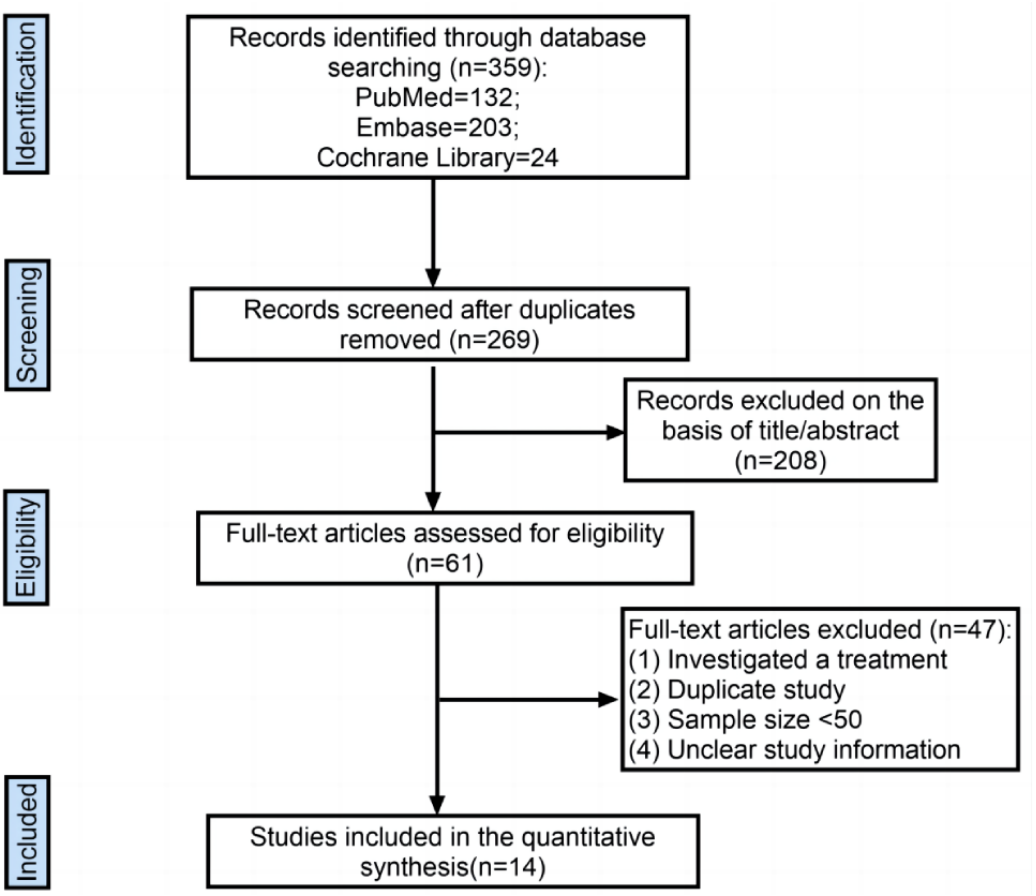


Table 1. Main characteristics of the included studies assessing the prevalence of triple infection.

Study number	First author	Type of study	Sample size, n	Male participants, n	Study population	HBV[a]/HDV[b]/HIV triple infection		Year	Continent	Country	Quality evaluation
						Sample size, n	Prevalence, %				
1	Soriano [18]	Cohort	5342	NR[c]	HIV	61	1.14	2011	Europe	NR	B[d]
2	Hønge [19]	Cross-sectional	576	180	HIV	18	3.13	2014	Africa	Guinea-Bissau	A[e]
3	Dény [20]	Cross-sectional	206	NR	HIV	19	9.22	1993	Europe	France	B
4	Coffie [21]	Cross-sectional	791	319	HIV	10	1.26	2017	Africa	NR	A
5	Ifeorah [22]	Cohort	1102	450	HIV	8	0.73	2017	Africa	Nigeria	A
6	Nicolini [15]	Cross-sectional	454	NR	HBV	16	3.52	2015	Europe	Italy	B
7	Saravanan [23]	Cohort	450	270	HBV	4	0.89	2015	Europe	Italy	A
8	Butler [24]	Cohort	1928	806	HBV	390	20.23	2018	Africa	Cameroon	A
9	Chang [25]	Cross-sectional	507	NR	HBV	150	29.59	2011	Asia	China (Taiwan)	B
10	Oprea [26]	Cohort	205	NR	HIV+HBV	21	10.24	2009	Europe	Romania	B
11	Béguelin [10]	Cohort	771	NR	HIV+HBV	117	15.18	2016	Europe	Switzerland	A
12	Sheng [11]	Cohort	104	100	HIV+HBV	26	25.00	2006	Asia	China (Taiwan)	A
13	Lee [27]	Cohort	375	363	HIV+HBV	38	10.13	2014	Asia	China (Taiwan)	B
14	Boyd [28]	Cohort	308	259	HIV+HBV	12	3.90	2009	Europe	France	A

[a]HBV: hepatitis B virus.

[b]HDV: hepatitis D virus.

[c]NR: not reported.

[d]Grade B corresponds to 3-6 points on the Newcastle-Ottawa Scale (NOS) scale and 4-7 "Yes" responses on the Agency for Healthcare Research and Quality (AHRQ) questionnaire.

[e]Grade A corresponds to 7-10 points on the NOS scale and 8-11 "Yes" responses on the AHRQ questionnaire.

Study Characteristics

The detailed characteristics of the 14 included studies are listed in Table 1. Among the 14 studies, 3 were conducted in Taiwan (China); 2 were conducted in France; 2 were conducted in Italy; 1 each was conducted in Guinea-Bissau, Nigeria, Cameroon, and Switzerland; and there were 2 multinational studies, 1 in Europe and 1 in Africa. The sample size ranged from 104 to 5342 participants, with a total sample size across the included studies of 11,852 participants. The prevalence of HIV/HBV/HDV triple infection was evaluated in 4 studies with patients with chronic hepatitis B, 5 studies with people living with HIV, and 5 studies with patients with HIV and HBV coinfection. Quality evaluations were either A or B. The demographic characteristics of the included studies are listed in Table 2. Participants ranged in age from 12 years to 61 years.

Table 2. Demographic characteristics of the included studies assessing the prevalence of triple infection.

Study number	First author	Sample size, n	Study population	HBV[a]/HDV[b]/HIV triple infection			
				Sample size, n	Prevalence, %	Age, mean (range)	Male participants, n (%)
1	Soriano [18]	5342	HIV	61	1.14	34 (NR)	44 (72.1)
2	Hønge [19]	576	HIV	18	3.13	NR	NR
3	Dény [20]	206	HIV	19	9.22	NR	NR
4	Coffie [21]	791	HIV	10	1.26	NR	NR
5	Ifeorah [22]	1102	HIV	8	0.73	NR (31-40)	5 (62.5)
6	Nicolini [15]	454	HBV	16	3.52	34.25 (6.16)[d]	14 (87.5)
7	Saravanan [23]	450	HBV	4	0.89	NR (21-40)	4 (100)
8	Butler [24]	1928	HBV	390	20.23	NR	NR
9	Chang [25]	507	HBV	150	29.59	NR	NR
10	Oprea [26]	205	HIV+HBV	21	10.24	16 (12-20)	NR
11	Béguelin [10]	771	HIV+HBV	117	15.18	34 (29-37)	92 (79)
12	Sheng [11]	104	HIV+HBV	26	25.00	35 (25-61)	25 (96.2)
13	Lee [27]	375	HIV+HBV	38	10.13	38 (NR)	36 (94.7)
14	Boyd [28]	308	HIV+HBV	12	3.90	35.2 (NR)	8 (66.7)

[a]HBV: hepatitis B virus.

[b]HDV: hepatitis D virus.

[c]NR: not reported.

[d]Median (SD).

Meta-analyses of the Data

The heterogeneity between studies in the meta-analysis of HIV/HBV/HDV triple infection rate was significant (I^2=98.995%, P<.01), and a random effects model was selected to combine the results of the included studies, which showed that the triple infection rate in the global population was 7.4% (877/11,852; 95% CI 0.73%-29.59%; Figure 2). Funnel plots and the Egger test were performed to detect publication bias. The funnel plot was basically symmetrical, and the Egger test results showed no significant statistical evidence of publication bias (t=1.13, P=.28; Figure 3). In view of the significant heterogeneity of HIV/HBV/HDV triple infection rates in different regions, countries, sexes, ages, sample sizes, basic diseases, and different transmission routes, subgroup analyses were conducted for these factors (Table 3). There was high heterogeneity among the results for all population groups.

Figure 2. Forest plot showing the prevalence of hepatitis B virus (HBV), hepatitis D virus (HDV), and HIV triple infection in the included studies. ES: effect size.

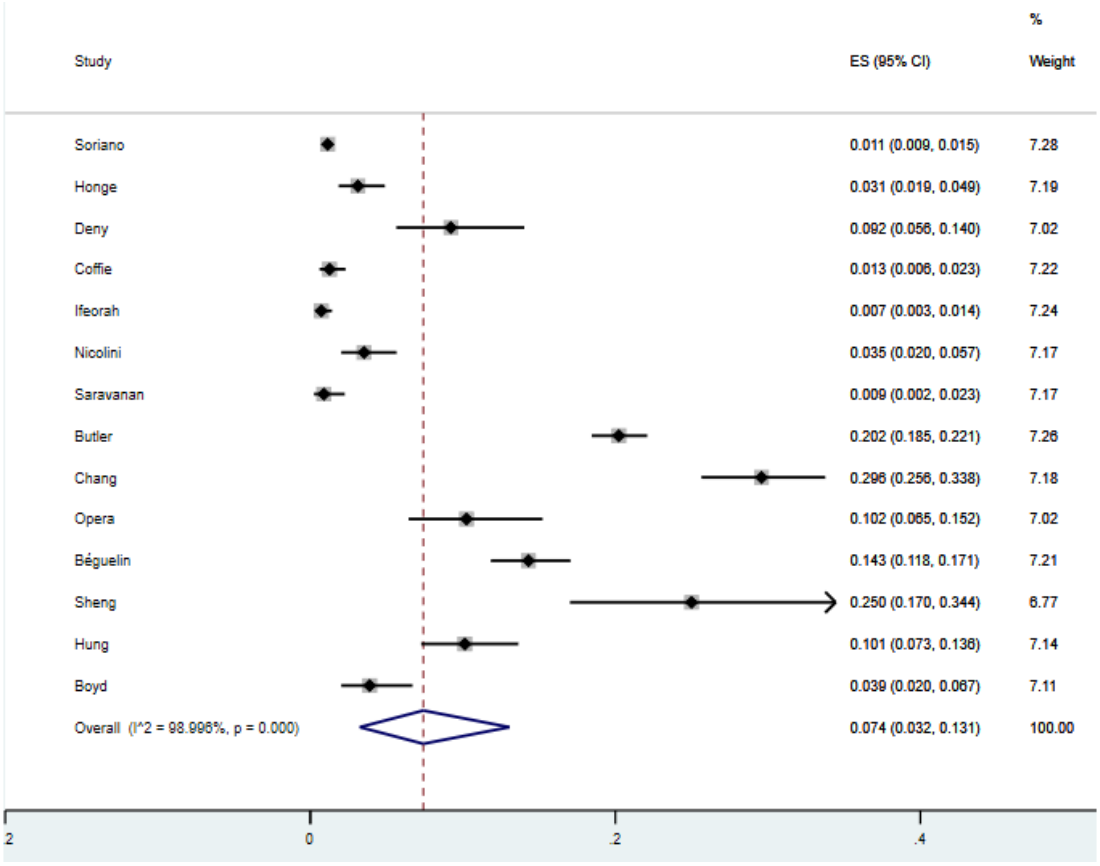


Figure 3. Funnel plot with 95% pseudo confidence limits for all included studies. ES: effect size.

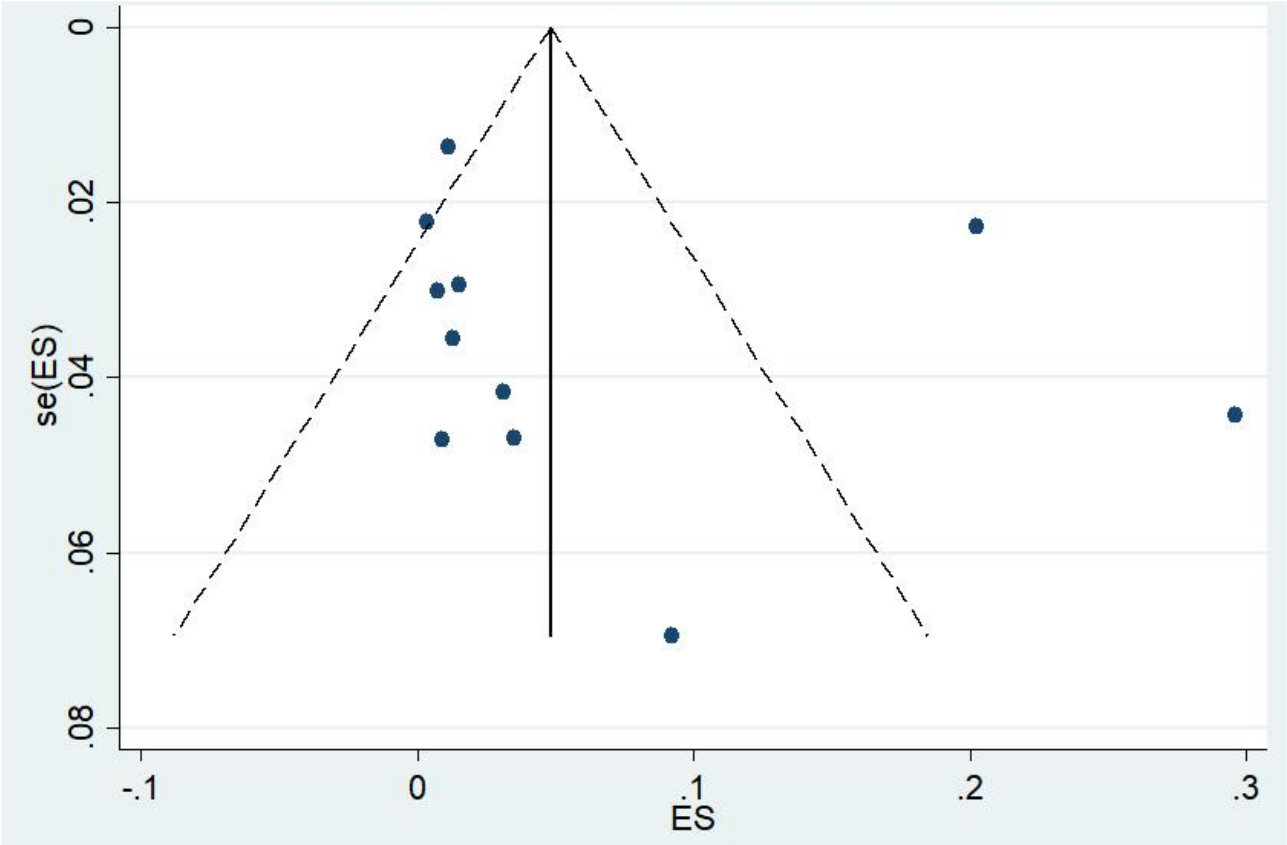


Table 3. Prevalence of HIV, hepatitis B virus (HBV), and hepatitis D virus (HDV) triple infection in different subgroups.

Subgroup	Studies, n	Triple infection, n	Prevalence, % (95% CI)	I^2, %	*P* value
Global	14	877	7.4 (3.2-13.1)	99.0	<.001
World region					
Europe	7	237	5.7 (3.2-8.2)	95.9	<.001
Africa	4	426	6.3 (0.7-11.8)	99.3	<.001
Asia	3	214	21.4 (7.1-35.8)	96.7	<.001
Country or area					
France	2	31	6.3 (1.1-11.5)	81.4	.02
Italy	2	20	2.1 (0.5-4.7)	86.4	.007
Cameroon	1	390	20.2 (18.4-22.0)	—[a]	<.001
Romania	1	21	10.2 (6.1-14.4)	—	—
Switzerland	1	117	14.3 (11.8-16.9)	—	—
China (Taiwan)	3	214	21.4 (7.1-35.8)	96.7	<.001
Guinea-Bissau	1	18	3.1 (1.7-4.5)	—	—
Nigeria	1	8	0.7 (0.2-1.2)	—	—
Sex					
Male	8	212	3.8 (2.5-5.2)	95.3	<.001
Female	7	57	0.7 (0.2-1.2)	78.0	<.001
Age (years)					
<30	1	21	5.6 (3.6-7.3)	—	—
≥30	8	269	5.1 (3.3-6.8)	96.2	<.001
Sample size					
50-500	7	136	7.7 (4.2-11.1)	93.5	<.001
500-1000	4	282	11.8 (4.0-19.7)	98.9	<.001
>1000	3	459	7.2 (2.4-11.9)	99.5	<.001
Study population					
HIV	5	116	1.7 (0.9-2.5)	84.8	<.001
HBV	4	560	13.4 (2.5-24.4)	99.4	<.001
HIV+HBV	5	201	11.9 (6.5-17.2)	92.5	<.001
Possible mode of HIV transmission					
Injection drug use	6	274	7.4 (2.9-11.9)	97.9	<.001
Men who have sex with men	6	216	7.9 (4.3-11.4)	98.1	<.001
Heterosexual transmission	6	201	6.5 (3.4-9.6)	98	<.001
Other	6	248	10.3 (6.0-14.7)	98.4	<.001

[a]Not available.

We analyzed 14 studies by regional subgroup and found that Asia had the highest pooled prevalence of triple infection at 21.4% (214/986; 95% CI 7.1%-35.8%, I^2=96.7%), and Europe had the lowest prevalence at 5.7% (237/4158; 95% CI 3.2%-8.2%, I^2=95.9%; Table 3, Multimedia Appendix 1). In addition, a subgroup analysis of the prevalence of triple infection according to country was performed, and a total of 12 studies were included, showing that the prevalence in Taiwan (China) was 21.4% (214/986; 95% CI 7.1%-35.8%, I^2=96.7%), the highest prevalence among the 8 countries included. Nigeria had the lowest prevalence at 0.7% (8/1143; 95% CI 0.2%-1.2%). The prevalence rates in the other 6 countries are shown in Table 3 and Multimedia Appendix 2.

We also performed a subgroup analysis by sex and found that men had a higher prevalence than women (228/8862, 3.8%; 95% CI 2.5%-5.2% vs 57/8412, 0.7%; 95% CI 0.3%-1.2%; Table 3, Multimedia Appendix 3). The triple infection rate was 5.6% (21/375; 95% CI 3.6%-7.3%) in those aged <30 years and

5.1% (269/5275; 95% CI 3.3%-6.8%) in those aged ≥30 years (Table 3, Multimedia Appendix 4). In studies with a sample size of 500-1000, the triple infection rate was the highest, at 11.8% (282/2390; 95% CI 4.0%-19.7%; Table 3, Multimedia Appendix 5).

Moreover, population-specific subgroup analyses were performed depending on the characteristics of the populations included in the study, and the results showed that the prevalence of triple infection varied greatly among the different populations. The prevalences of HIV/HBV/HDV triple infection were 13.4% (560/4079; 95% CI 2.5%-24.4%) in patients with chronic hepatitis B, 1.7% (116/6224; 95% CI 0.9%-2.5%) in people living with HIV, and 11.9% (201/1549; 95% CI 6.5%-17.2%) in people with HBV/HIV coinfection (Table 3, Multimedia Appendix 6). Meanwhile, since there are multiple transmission routes for HIV, such as transmission via men who have sex with men, heterosexual transmission, and transmission via injection drug use (IDU), a subgroup analysis of the transmission routes was performed. The pooled prevalence of triple infection was 7.9% (216/2734; 95% CI 4.3%-11.4%) in men who have sex with men, 6.5% (201/3092; 95% CI 3.4%-9.6%) in those with heterosexual transmission, and 7.4% (274/3703; 95% CI 2.9%-11.9%) in those with IDU (Table 3, Multimedia Appendix 7). “Other” modes of transmission include perinatal, risk not identified, and blood transfusion. In addition, given that people with IDU may be more prone to triple infections, we collapsed the data on drug-using people included in the literature and calculated a prevalence of triple infection of 20.6% among people with IDU, which is significantly higher than the prevalence in the total population (Multimedia Appendix 8).

Discussion

Since the first discovery of the HDV in 1977 [29], it is estimated that 15 million to 20 million people have been infected worldwide [30,31]. Given that HDV is a defective virus dependent on the envelope proteins of HBV for assembly and release of infectious virus particles, HDV infection occurs either with or secondary to HBV infection. People infected with HDV can also have other viral infections, and chronic HDV infection is considered to be the most severe form of viral hepatitis infection in humans [32]. HIV was first reported in 1981, and the number of people living with HIV and deaths due to illness have remained high for a long time [33]. The virus is widely prevalent worldwide. As of 2018, there were more than 37.9 million people living with HIV in the world, and a total of 35 million people have died from AIDS-related diseases [34]. HIV infection causes progressive immunodeficiency, making people living with HIV highly susceptible to coinfection with other diseases. HBV and HDV are common viruses for coinfection in people living with HIV, as all 3 share the same route of infection [14,35]. Studies have shown that HIV coinfection with HBV and HDV is widespread in various regions of the world, but the coinfection rate varies among countries and regions [36]. HIV coinfection with HBV or HDV can cause more serious damage to the body than a single infection, and the harm of triple infection is even more serious. Therefore, it is of great public health significance to actively prevent such coinfections [37,38].

The results of this meta-analysis showed that the prevalence of HIV/HBV/HDV triple infection was 7.4%. The results of the subgroup analysis showed significant regional differences in global triple infection rates. The prevalence of triple infection is significantly higher in Asia, especially in Taiwan and China, than in other countries or regions. These results may be caused by several reasons. First, the included studies included people with IDU; IDU is a high-risk factor that may promote triple infections [39]. In addition, the higher prevalence rate might be due to the small sample size of the included studies, which may not accurately reflect the real situation of triple infection in this population. We found that the characteristics of the study population greatly influenced the prevalence; for example, the prevalence of triple infection in the Cameroon region was higher because the study population was HBsAg-positive and in a general hospital, which itself confers high risk for infection. Among the different sexes, the results of this meta-analysis showed a higher rate of triple infection in men than in women, which is consistent with several other studies [40,41]. This review suggested that men with HIV or HBV infection may have a higher prevalence of triple infection, and more attention should be given to the prognosis of their triple infection.

There were significant differences in the rate of triple infection for people living with HIV from different population sources, which was similar to rates for HIV/HBV coinfection [42]; however, for triple infection, men who have sex with men and people living with HIV are particularly worthy of attention. The rate of triple coinfection in this population reached 7.9%, exceeding the rate of triple infection in other populations. Nevertheless, the literature related to these special populations suffers from the same shortcomings as aforementioned, with small numbers, regional limitations, small sample sizes, and mostly poor quality, which urgently needs to be supplemented with similar studies to help understand the current situation.

Furthermore, through the analysis of the results, we found a very interesting phenomenon. The rate of HIV/HBV/HDV triple infection was higher in people with HBV monoinfection than in people with HIV monoinfection or people with HBV/HIV coinfection. However, the credibility and reason for these findings are still unclear, and a large number of clinical studies is needed to better confirm the results.

There were several limitations in this systematic review. Although strict inclusion and exclusion criteria were established and the quality of the included literature was evaluated using the NOS or AHRQ statement entries during the search and screening processes, there was still some subjectivity in the evaluation of the literature due to the lack of accepted quality evaluation criteria, which may lead to some selection bias in the included literature. The results of the sensitivity analysis in this systematic review showed that there was a certain selection bias. In addition, the wide inclusion criteria in this study produced significant heterogeneity that could not be explained. We used a random effects model with subgroup analyses whenever possible to reduce the effect of heterogeneity. Furthermore, the population included in this meta-analysis included people with HBV infection or HIV, lacking a comparable general population; some of these patients were

drug users, which would increase the overall prevalence to some extent; and the evidence base had some shortcomings.

In summary, the prevalence of HIV/HBV/HDV triple infection in the global population is underestimated. Therefore, during the management and antiviral treatment of patients with HBV/HIV single infection or coinfection, they should be screened for HIV/HBV/HDV triple infection in a timely manner. In addition, the prevention and treatment of coinfection should be combined with antiviral treatment to provide comprehensive prevention and treatment of triple infection and improve the quality of survival for this population. Additionally, the rates of triple infection in the two special groups of men who have sex with men and people with IDU are also worthy of attention. However, because there are few relevant studies, it is impossible to accurately evaluate the current status of rates of triple infection in the global populations of men who have sex with men and people with IDU [43,44]. More research is urgently needed to provide evidence, identify high-risk populations, and guide the formulation and improvement of prevention and control strategies for HIV/HBV/HDV infection.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (81770611, 82002243); Key Projects of the Beijing Municipal Education Commission's Science and Technology Plan (KZ202010025035); Special key research project of capital health development scientific research (2020-1-1151); the Demonstrating Application and Research of Clinical Diagnosis and Treatment Technology in Beijing (Z191100006619096 and Z191100006619097); Beijing Talents Foundation, (2018000021469G289, 2018000021223ZK04, 2020A35); and Beijing Hospitals Authority Youth Program (QML20201702).

Authors' Contributions

XH, ML, and FR conceptualized the study and developed the research protocol. SC, LX, YG, and XY identified articles for full-text review and extracted data that matched the inclusion criteria. SC, ZF, and YT performed the statistical analyses. All authors contributed to the writing of the manuscript. XH and FR polished and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Forest plot showing the meta-analysis of triple infection in different regions.
[PNG File , 46 KB - publichealth_v8i11e37016_app1.png]

Multimedia Appendix 2
Forest plot showing meta-analysis of triple infection in different countries.
[PNG File , 27 KB - publichealth_v8i11e37016_app2.png]

Multimedia Appendix 3
Forest plot showing the meta-analysis of triple infection in different genders.
[PNG File , 44 KB - publichealth_v8i11e37016_app3.png]

Multimedia Appendix 4
Forest plot showing the meta-analysis of triple infection for different ages.
[PNG File , 37 KB - publichealth_v8i11e37016_app4.png]

Multimedia Appendix 5
Forest plot showing the meta-analysis of triple infection in different sample sizes.
[PNG File , 45 KB - publichealth_v8i11e37016_app5.png]

Multimedia Appendix 6
Forest plot showing the meta-analysis of triple infection in different study populations.
[PNG File , 45 KB - publichealth_v8i11e37016_app6.png]

Multimedia Appendix 7
Forest plot showing the meta-analysis of triple infection for different HIV transmission routes.
[PNG File , 46 KB - publichealth_v8i11e37016_app7.png]

Multimedia Appendix 8
Forest plot showing the meta-analysis of triple infection in people who inject drugs.
[PNG File , 23 KB - publichealth_v8i11e37016_app8.png]

References

1. Taylor JM. Infection by hepatitis delta virus. Viruses 2020 Jun 16;12(6):648 [FREE Full text] [doi: 10.3390/v12060648] [Medline: 32560053]
2. Lanini S, Ustianowski A, Pisapia R, Zumla A, Ippolito G. Viral hepatitis: etiology, epidemiology, transmission, diagnostics, treatment, and prevention. Infect Dis Clin North Am 2019 Dec;33(4):1045-1062. [doi: 10.1016/j.idc.2019.08.004] [Medline: 31668190]
3. Miao Z, Zhang S, Ou X, Li S, Ma Z, Wang W, et al. Estimating the global prevalence, disease progression, and clinical outcome of hepatitis delta virus infection. J Infect Dis 2020 Apr 27;221(10):1677-1687 [FREE Full text] [doi: 10.1093/infdis/jiz633] [Medline: 31778167]
4. Zhang Z, Urban S. New insights into HDV persistence: The role of interferon response and implications for upcoming novel therapies. J Hepatol 2021 Mar;74(3):686-699 [FREE Full text] [doi: 10.1016/j.jhep.2020.11.032] [Medline: 33276031]
5. Chen H, Shen D, Ji D, Han P, Zhang W, Ma J, et al. Prevalence and burden of hepatitis D virus infection in the global population: a systematic review and meta-analysis. Gut 2019 Mar;68(3):512-521. [doi: 10.1136/gutjnl-2018-316601] [Medline: 30228220]
6. Stockdale AJ, Kreuels B, Henrion MYR, Giorgi E, Kyomuhangi I, de Martel C, et al. The global prevalence of hepatitis D virus infection: Systematic review and meta-analysis. J Hepatol 2020 Sep;73(3):523-532 [FREE Full text] [doi: 10.1016/j.jhep.2020.04.008] [Medline: 32335166]
7. Soriano V, Sherman KE, Barreiro P. Hepatitis delta and HIV infection. AIDS 2017 Apr 24;31(7):875-884. [doi: 10.1097/QAD.0000000000001424] [Medline: 28121714]
8. Nisini R, Paroli M, Accapezzato D, Bonino F, Rosina F, Santantonio T, et al. Human CD4+ T-cell response to hepatitis delta virus: identification of multiple epitopes and characterization of T-helper cytokine profiles. J Virol 1997 Mar;71(3):2241-2251 [FREE Full text] [doi: 10.1128/JVI.71.3.2241-2251.1997] [Medline: 9032359]
9. Ferrante ND, Lo Re V. Epidemiology, natural history, and treatment of hepatitis delta virus infection in HIV/hepatitis B virus coinfection. Curr HIV/AIDS Rep 2020 Aug;17(4):405-414 [FREE Full text] [doi: 10.1007/s11904-020-00508-z] [Medline: 32607773]
10. Béguelin C, Moradpour D, Sahli R, Suter-Riniker F, Lüthi A, Cavassini M, Swiss HIV Cohort Study. Hepatitis delta-associated mortality in HIV/HBV-coinfected patients. J Hepatol 2017 Feb;66(2):297-303. [doi: 10.1016/j.jhep.2016.10.007] [Medline: 27746337]
11. Sheng WH, Hung CC, Kao JH, Chang SY, Chen MY, Hsieh SM, et al. Impact of hepatitis D virus infection on the long-term outcomes of patients with hepatitis B virus and HIV coinfection in the era of highly active antiretroviral therapy: a matched cohort study. Clin Infect Dis 2007 Apr 01;44(7):988-995. [doi: 10.1086/511867] [Medline: 17342655]
12. Fernández-Montero JV, Vispo E, Barreiro P, Sierra-Enguita R, de Mendoza C, Labarga P, et al. Hepatitis delta is a major determinant of liver decompensation events and death in HIV-infected patients. Clin Infect Dis 2014 Jun;58(11):1549-1553. [doi: 10.1093/cid/ciu167] [Medline: 24633686]
13. Chu F, Chiang S, Su F, Chang Y, Cheng S. Prevalence of human immunodeficiency virus and its association with hepatitis B, C, and D virus infections among incarcerated male substance abusers in Taiwan. J Med Virol 2009 Jun;81(6):973-978. [doi: 10.1002/jmv.21481] [Medline: 19382252]
14. Shen D, Han P, Ji D, Chen H, Cao W, Goyal H, et al. Epidemiology estimates of hepatitis D in individuals co-infected with human immunodeficiency virus and hepatitis B virus, 2002-2018: A systematic review and meta-analysis. J Viral Hepat 2021 Jul;28(7):1057-1067. [doi: 10.1111/jvh.13512] [Medline: 33877742]
15. Nicolini LA, Taramasso L, Schiavetti I, Giannini EG, Beltrame A, Feasi M, Ligurian HBV Study Group. Epidemiological and clinical features of hepatitis delta in HBsAg-positive patients by HIV status. Antivir Ther 2015;20(2):193-197. [doi: 10.3851/IMP2819] [Medline: 24963642]
16. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol 2010 Sep;25(9):603-605. [doi: 10.1007/s10654-010-9491-z] [Medline: 20652370]
17. Farquhar M. AHRQ Quality Indicators. In: Hughes RG, editor. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: Agency for Healthcare Research and Quality; 2008.
18. Soriano V, Grint D, d'Arminio Monforte A, Horban A, Leen C, Poveda E, et al. Hepatitis delta in HIV-infected individuals in Europe. AIDS 2011 Oct 23;25(16):1987-1992. [doi: 10.1097/QAD.0b013e32834babb3] [Medline: 21857493]
19. Hønge BL, Jespersen S, Medina C, Té DS, da Silva ZJ, Lewin S, Bissau HIV cohort study group. Hepatitis B and delta virus are prevalent but often subclinical co-infections among HIV infected patients in Guinea-Bissau, West Africa: a cross-sectional study. PLoS One 2014;9(6):e99971 [FREE Full text] [doi: 10.1371/journal.pone.0099971] [Medline: 24915064]

20. Dény P, Lecot C, Jeantils V, Ovaguimian L, Krivitzky A, Bréchot C. Polymerase chain reaction-based detection of hepatitis D virus genome in patients infected with human immunodeficiency virus. J Med Virol 1993 Mar;39(3):214-218. [doi: 10.1002/jmv.1890390307] [Medline: 8468565]
21. Coffie PA, Tchounga BK, Bado G, Kabran M, Minta DK, Wandeler G, et al. Prevalence of hepatitis B and delta according to HIV-type: a multi-country cross-sectional survey in West Africa. BMC Infect Dis 2017 Jul 04;17(1):466 [FREE Full text] [doi: 10.1186/s12879-017-2568-5] [Medline: 28676076]
22. Ifeorah IM, Bakarey AS, Adeniji JA, Onyemelukwe FN. Seroprevalence of hepatitis B and delta viruses among HIV-infected population attending anti-retroviral clinic in selected health facilities in Abuja, Nigeria. J Immunoassay Immunochem 2017;38(6):608-619. [doi: 10.1080/15321819.2017.1372474] [Medline: 28854102]
23. Saravanan S, Madhavan V, Velu V, Murugavel KG, Waldrop G, Solomon SS, et al. High prevalence of hepatitis delta virus among patients with chronic hepatitis B virus infection and HIV-1 in an intermediate hepatitis B virus endemic region. J Int Assoc Provid AIDS Care 2014;13(1):85-90 [FREE Full text] [doi: 10.1177/2325957413488166] [Medline: 23722085]
24. Butler EK, Rodgers MA, Coller KE, Barnaby D, Krilich E, Olivo A, et al. High prevalence of hepatitis delta virus in Cameroon. Sci Rep 2018 Aug 02;8(1):11617 [FREE Full text] [doi: 10.1038/s41598-018-30078-5] [Medline: 30072752]
25. Chang S, Yang C, Ko W, Liu W, Lin C, Wu C, et al. Molecular epidemiology of hepatitis D virus infection among injecting drug users with and without human immunodeficiency virus infection in Taiwan. J Clin Microbiol 2011 Mar;49(3):1083-1089 [FREE Full text] [doi: 10.1128/JCM.01154-10] [Medline: 21191061]
26. Oprea C, Radoi R, Ungureanu E, Tardei G, Ene L, Erhan R, et al. Hepatitis delta in HIV-1-infected Romanian adolescents. HIV Med 2010 Jul 16(10):3-3. [doi: 10.1111/j.1468-1293.2009.00789.x]
27. Lee CY, Tsai HC, Lee SS, Wu KS, Sy CL, Chen JK, et al. Higher rate of hepatitis events in patients with human immunodeficiency virus, hepatitis B, and hepatitis D genotype II infection: a cohort study in a medical center in southern Taiwan. J Microbiol Immunol Infect 2015 Feb;48(1):20-27. [doi: 10.1016/j.jmii.2013.08.001] [Medline: 24064291]
28. Boyd A, Lacombe K, Miailhes P, Gozlan J, Bonnard P, Molina J, et al. Longitudinal evaluation of viral interactions in treated HIV-hepatitis B co-infected patients with additional hepatitis C and D virus. J Viral Hepat 2010 Jan;17(1):65-76. [doi: 10.1111/j.1365-2893.2009.01153.x] [Medline: 19682317]
29. Toy M, Ahishali E, Yurdaydın C. Hepatitis delta virus epidemiology in the industrialized world. AIDS Rev 2020 Oct 26;22(4):203-212. [doi: 10.24875/AIDSRev.20000056] [Medline: 33104688]
30. Koh C, Heller T, Glenn JS. Pathogenesis of and new therapies for hepatitis D. Gastroenterology 2019 Jan;156(2):461-476.e1 [FREE Full text] [doi: 10.1053/j.gastro.2018.09.058] [Medline: 30342879]
31. Mentha N, Clément S, Negro F, Alfaiate D. A review on hepatitis D: from virology to new therapies. J Adv Res 2019 May;17:3-15 [FREE Full text] [doi: 10.1016/j.jare.2019.03.009] [Medline: 31193285]
32. Tseligka ED, Clément S, Negro F. HDV pathogenesis: unravelling Ariadne's thread. Viruses 2021 Apr 28;13(5):778 [FREE Full text] [doi: 10.3390/v13050778] [Medline: 33924806]
33. Eisinger RW, Fauci AS. Ending the HIV/AIDS pandemic. Emerg Infect Dis 2018 Mar;24(3):413-416 [FREE Full text] [doi: 10.3201/eid2403.171797] [Medline: 29460740]
34. Ghosn J, Taiwo B, Seedat S, Autran B, Katlama C. HIV. Lancet 2018 Aug 25;392(10148):685-697. [doi: 10.1016/S0140-6736(18)31311-4] [Medline: 30049419]
35. Singh KP, Crane M, Audsley J, Avihingsanon A, Sasadeusz J, Lewin SR. HIV-hepatitis B virus coinfection: epidemiology, pathogenesis, and treatment. AIDS 2017 Sep 24;31(15):2035-2052 [FREE Full text] [doi: 10.1097/QAD.0000000000001574] [Medline: 28692539]
36. Torimiro JN, Nanfack A, Takang W, Keou CK, Joyce AN, Njefi K, et al. Rates of HBV, HCV, HDV and HIV type 1 among pregnant women and HIV type 1 drug resistance-associated mutations in breastfeeding women on antiretroviral therapy. BMC Pregnancy Childbirth 2018 Dec 22;18(1):504 [FREE Full text] [doi: 10.1186/s12884-018-2120-7] [Medline: 30577760]
37. Assih M, Ouattara AK, Diarra B, Yonli AT, Compaore TR, Obiri-Yeboah D, et al. Genetic diversity of hepatitis viruses in West-African countries from 1996 to 2018. World J Hepatol 2018 Nov 27;10(11):807-821 [FREE Full text] [doi: 10.4254/wjh.v10.i11.807] [Medline: 30533182]
38. Hu J, Liu K, Luo J. HIV-HBV and HIV-HCV coinfection and liver cancer development. Cancer Treat Res 2019;177:231-250. [doi: 10.1007/978-3-030-03502-0_9] [Medline: 30523627]
39. Safaie P, Razeghi S, Rouster SD, Privitera I, Sherman KE. Hepatitis D diagnostics: utilization and testing in the United States. Virus Res 2018 May 02;250:114-117. [doi: 10.1016/j.virusres.2018.03.013] [Medline: 29596839]
40. Ziaee M, Sharifzadeh G, Namaee MH, Fereidouni M. Prevalence of HIV and hepatitis B, C, D infections and their associated risk factors among prisoners in Southern Khorasan Province, Iran. Iran J Public Health 2014 Mar;43(2):229-234 [FREE Full text] [Medline: 26060747]
41. Ramezan Ghorbani N, Qorbani M, Djalalinia S, Kazemzadeh Atoofi M, Tajbakhsh R, Mansourian M, et al. Oncogenic viral infections among Iranian hemodialysis patients: a systematic review. Int J Prev Med 2019;10:216 [FREE Full text] [doi: 10.4103/ijpvm.IJPVM_458_17] [Medline: 31929863]
42. Soares CC, Georg I, Lampe E, Lewis L, Morgado MG, Nicol AF, et al. HIV-1, HBV, HCV, HTLV, HPV-16/18, and Treponema pallidum infections in a sample of Brazilian men who have sex with men. PLoS One 2014;9(8):e102676 [FREE Full text] [doi: 10.1371/journal.pone.0102676] [Medline: 25083768]

43. Kibaya RM, Lihana RW, Kiptoo M, Songok EM, Ng'ang'a Z, Osman S, et al. Characterization of HBV among HBV/HIV-1 co-infected injecting drug users from Mombasa, Kenya. Curr HIV Res 2015;13(4):292-299. [doi: 10.2174/1570162x13666150121113217] [Medline: 25613131]
44. Sadio AJ, Gbeasor-Komlanvi FA, Konu YR, Sewu EK, Zida-Compaore W, Salou M, et al. Prevalence of HIV infection and hepatitis B and factors associated with them among men who had sex with men in Togo in 2017. Med Sante Trop 2019 Aug 01;29(3):294-301 [FREE Full text] [doi: 10.1684/mst.2019.0922] [Medline: 31573525]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
HBsAg: hepatitis B virus surface antigen
HBV: hepatitis B virus
HDV: hepatitis D virus
IDU: injection drug use
NOS: Newcastle-Ottawa Scale

Edited by H Bradley; submitted 03.02.22; peer-reviewed by AAS Sawitri, J Opoku; comments to author 10.05.22; revised version received 19.06.22; accepted 11.10.22; published 29.11.22.

Please cite as:
Chen S, Ren F, Huang X, Xu L, Gao Y, Zhang X, Cao Y, Fan Z, Tian Y, Liu M
Underestimated Prevalence of HIV, Hepatitis B Virus (HBV), and Hepatitis D Virus (HDV) Triple Infection Globally: Systematic Review and Meta-analysis
JMIR Public Health Surveill 2022;8(11):e37016
URL: https://publichealth.jmir.org/2022/11/e37016
doi:10.2196/37016
PMID:36445732

Original Paper

Patterns of HIV or AIDS Mortality Among Older People From 1990 to 2019 in China: Age-Period-Cohort Analysis

Ningjun Ren[1*], MPH; Yuansheng Li[1*], MPH; Zhengwei Wan[2*], PhD; Ruolan Wang[1], MPH; Wenxin Zhang[1], MPH; Emmanuel Enoch Dzakah[3], PhD; Junhui Zhang[1], PhD; Ailing Li[1], MSc; Song Fan[1], PhD

[1]School of Public Health, Southwest Medical University, Luzhou, China

[2]Department of Health Management Center & Institute of Health Management, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, China

[3]Department of Molecular Biology and Biotechnology, School of Biological Sciences, College of Agriculture and Natural Sciences, University of Cape Coast, Cape Coast, Ghana

[*]these authors contributed equally

Corresponding Author:
Song Fan, PhD
School of Public Health
Southwest Medical University
No.1, Section 1, Xianglin Road
Longmatan District
Luzhou, 64600
China
Phone: 86 8303175813
Email: fansong@swmu.edu.cn

Abstract

Background: With the increasing effectiveness of antiretroviral therapy and shifting demographics, the problem of older people with HIV or AIDS is increasingly grim in China, and neglecting infection among them may cause more serious social problems, exacerbate the difficulty of controlling HIV or AIDS transmission, and increase the risk of death.

Objective: We investigated the variations in the trends of Chinese mortality by age, period, and cohort, from 1990 to 2019, to reveal the relationship between age, period, cohort, and HIV burden, as well as providing guidance for resource allocation to prevent HIV-related deaths in vulnerable target populations.

Methods: We extracted the HIV or AIDS mortality data from the Global Burden of Disease. The joinpoint regression model was applied to detect changes in HIV or AIDS trends. The age-period-cohort model was used to explore the age, period, and cohort effects.

Results: The trends in age-standardized mortality rates in HIV or AIDS were increased in both genders, from 0.50 to 4.54/105 individuals for males, and from 0.19 to 1.43/105 individuals for females. Joinpoint regression model showed the average annual percentage change of age-standardized mortality rates was 7.0 for male and 6.4 for female individuals, showing an increasing trend. The age effect of male HIV or AIDS mortality showed a net increase of 0.59 (–0.21 to 0.38) from the ages 50-79 years. There is a gradual upward trend in the change in risk of death from HIV or AIDS for the period effect among the older population, lowest at ages 50-54 years (–0.80 for male and –0.78 for female individuals) and highest at ages 75-79 years (0.86 for male and 0.69 for female individuals). The variation of cohort effects was complex, but both genders had a nearly consistent tendency; people born in 1920-1929 had the lowest cohort effect, and those born in 1950-1954 had the highest values.

Conclusions: Our study showed a marked rise in HIV mortality for both genders in China from 1990 to 2019. Aging is an important issue in current HIV prevention and control. There is an urgent need to promote HIV testing and health education. Our findings will help predict future HIV or AIDS mortality changes and identify age-specific priority populations for intervention.

(JMIR Public Health Surveill 2022;8(11):e35785) doi:10.2196/35785

KEYWORDS

HIV; AIDS; aging; mortality; trends; age-period-cohort model; APC

Introduction

HIV and AIDS have been prevalent in China for more than 30 years since the first case of HIV was reported in 1985 [1]. Due to the substantial number of deaths attributed to the virus, HIV or AIDS has become the most severe notifiable infectious disease accounting for the most deaths, over 18,819, in 2020 in China [2]. With the increasing effectiveness of antiretroviral therapy (ART), people with HIV are living longer [3], and HIV or AIDS has transformed from a near-uniformly fatal infection to a chronic condition [4]. As a result, some estimates indicate that nearly 50% of persons with HIV in the United States are aged 50 years and older [5]. The situation of AIDS among older people in Europe is far from satisfactory, with estimates from some European countries predicting a "silver tsunami" within the HIV community, mirroring that of the general population, with those aged 50 years or older accounting for nearly 70% of people with HIV by 2030 [6-8]. Overlooking the risk of HIV or AIDS infection among older people is a mistake. The public perception that older people were not susceptible to HIV or AIDS infection, coupled with the lack of proper sexual education [9], exacerbates HIV among the older people. This problem is increasingly grim in China.

Furthermore, previous studies mainly focused on the age distribution of morbidity or mortality, with few studies considering both time and cohort effects [10,11]. However, period effects are also crucial in influencing the onset of disease. Period effects can also be understood as the role of social and epidemiological conditions in influencing some events, including policies, medical technology, screening tools, and even disease classification criteria. Tarone et al [12] found that the rise in the incidence of breast cancer in North America in the 1980s was due to the mass use of diagnostic mammography techniques, which increased diagnostic accuracy and thus the incidence of breast cancer. Ma et al [13] and Zhang et al [11] noted that the "Four Free and One Care" policy enacted in mainland China, which expanded HIV or AIDS screening and increased attention to HIV or AIDS, led to an increase in the incidence of HIV or AIDS and a decrease in the death rate.

Moreover, the cohort effect is because people in the same birth cohort will experience the same events at the same age. Birth cohorts that experience different events at different stages of their life course have different levels of exposure to economic, behavioral, policy, and environmental risks. Nevertheless, trends in Chinese HIV or AIDS deaths by age, among older people, remain unclear, as does the relative risk due to time and cohort effects [14]. The age-period-cohort (APC) model analyzes the age, period, and cohort effect for a comprehensive analysis to clarify the answers to these questions. This study examined elderly HIV or AIDS mortality trends by age, period, and cohort. A statistical analysis of the HIV or AIDS mortality of 50-79 years old in China from 1990 to 2019 was performed. Those effects were estimated by the APC model combined with the Intrinsic Estimator (IE) algorithm [15].

Studying HIV or AIDS mortality trends in older Chinese may reveal new information about the risk factors associated with HIV or AIDS. The finding reveals the relationship between age-period-cohort, on the one hand, and HIV or AIDS burden, on the other. It also provides guidance for resource allocation to prevent HIV-related deaths in vulnerable target populations.

Methods

China's HIV or AIDS mortality data were extracted from the Institute for Health Metrics and Evaluation. To examine temporal trends in HIV (coded in the International Classification of Diseases, 10th Revision) mortality over the past 30 years, we used data from the Institute for Health Metrics and Evaluation, an independent global health research center at the University of Washington in the United States. Many scientists from dozens of countries around the world wrote the Global Burden of Diseases (GBD) Injuries and Risk Factor Study (GBD 2019 [16]), which used the Bayesian disease modeling meta-regression to collect data comprehensively and accurately [17]. To standardize the mortality of different observation ages, we collected the population data of each age group from the Statistical Yearbook of Population and Employment of China from 1990 to 2019. Elderly HIV or AIDS was defined according to the United Nations program on HIV or AIDS (UNAIDS) "AIDS and aging" standards [18,19].

For the requirements of the APC model, we divided the age range of 50-79 years into 6 age groups at intervals of 5 years. Individuals younger than 50 years and older than 80 years were ruled out (>80 years old already exceeds life expectancy per capita in China, and the inclusion of a population with a complex cause of death and high mortality from reduced resistance may affect the accuracy of the model). Since the purpose of our study was aimed at older patients with HIV or AIDS, after excluding Chinese patients with HIV or AIDS who are younger than 50 years and older than 79 years, the data used in our study were from the age groups of 50-54 years old to 75-79 years old. The time range of data was from 1990-2019 (with 5 years per period) for computing the age-standardized mortality rates (ASMR) and period mortality rates.

The APC models represent a classic epidemiological approach for extracting historical morbidity and mortality risk changes from cross-sectional data, termed the cohort effect [20]. As there is a linear relationship between the age, period, and cohort, it is difficult to estimate the unique setting for every age, period, and cohort effect, referred to as the unidentification problem [21,22]. Many statistical analysis algorithms were designed to solve the unidentification problem [23-26]. Fu [15,27] applied the estimable functions and the singular value decomposition of matrices to approach the estimator of the APC model, which is the most effective for the unidentification problem, named the IE.

Finally, we described the magnitude of the rates  as a function of age (a), period (p), and birth cohort (c) using a log-linear model, with Poisson distribution and with the log of the person-years at risk defined as an offset of the IE method. D_{ap} indicates the number of incidences in the "a" age group in the "p" period; P_{ap} denotes the total number of persons in the age group "a" in period "p."

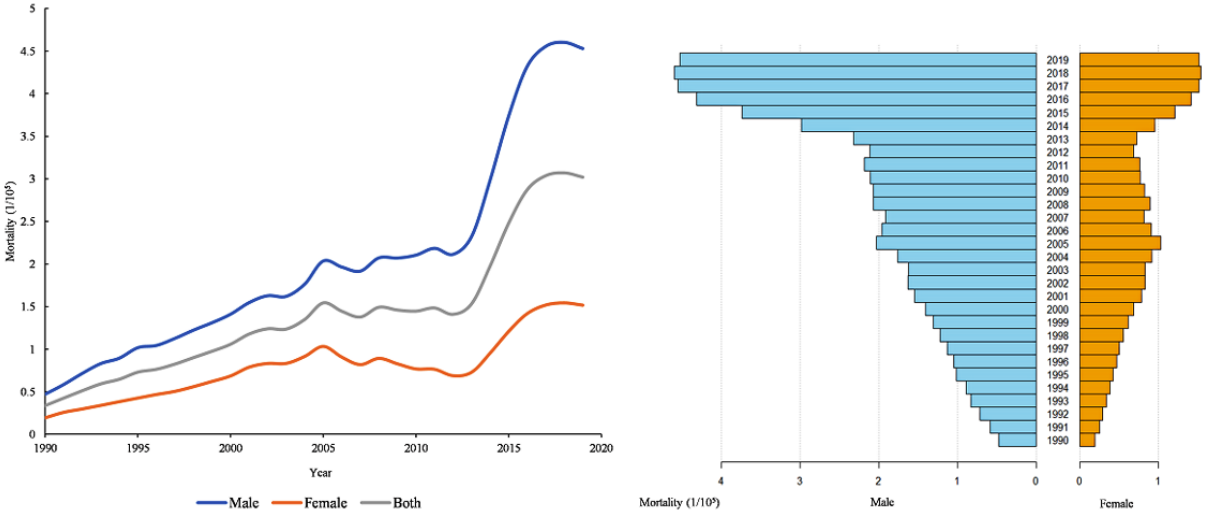
In this paper, the joinpoint regression models were performed by the Joinpoint Regression Program (version 4.3.1.0), and the age-period-cohort model analyses and graphs were conducted using APC fit in R, version 3.6.0 (R foundation for Statistical Computing). Fitting deviance, R^2, and adjust R^2 were used to evaluate the model; the closer the value to 1, the better the test performance.

Results

Mortality of HIV or AIDS in Older Chinese People

In Figure 1, ASMR for HIV or AIDS by gender from 1990 to 2019 was shown. HIV or AIDS ASMRs showed increasing trends from 0.50 to 4.51/10^5 individuals for male and 0.19/10^5 to 1.45/10^5 for female populations, slightly decreasing after 2018. ASMR increased 8.91-fold for male and 7.31-fold for female populations over the past 30 years. Our results also indicated that the gap between the mortality rates for older male and female individuals was enormous, with a maximum of 3.36 times that of female individuals in 2013.

Figure 1. Trends in the HIV age-standardized mortality rates per 100,000 population by gender from 1990 to 2019 using the Statistical Yearbook of Population and Employment of China from 1990 to 2019 for the age-standardized population.



HIV or AIDS Mortality Trend Variation in the Age, Period, and Cohort

The HIV or AIDS mortality trend variation among 50-79 years age groups of different genders in China between 1990 and 2019 is shown in Figure 2. Regardless of the period, almost all groups had insignificant changes, especially females. Only males between 2010-2014 and 2015-2019 groups showed increased HIV or AIDS mortality with age.

The variations in the HIV or AIDS mortality rates of different age groups during the decades from 1990-2019 are shown in Figure 3. There was a significant increase in HIV or AIDS mortality regardless of age or gender. The 75-79 years age groups showed the highest mortality rate (5.66/10^5) in males, and all groups of older males with AIDS had higher mortality rates than female groups. The male ASMR (5.66/10^5) was over 5 times more than that of the female groups（1.08/10^5） at ages 75-79 years from 2015-2019.

The effect of birth cohort on ASMR of HIV or AIDS among Chinese of different age groups is shown in Figure 4. The earlier the birth cohort, the higher the HIV or AIDS mortality rate. Across all cohorts, HIV mortality fluctuated more with the birth cohort, especially for males.

Figure 2. HIV mortality rates of different age groups (50-54 years old to 75-79 years old) in each period (1990-1994, 1995-1999, 2000-2004, 2005-2009, 2010-2014, and 2015-2019) are shown (male populations represented by blue lines and female populations by red).

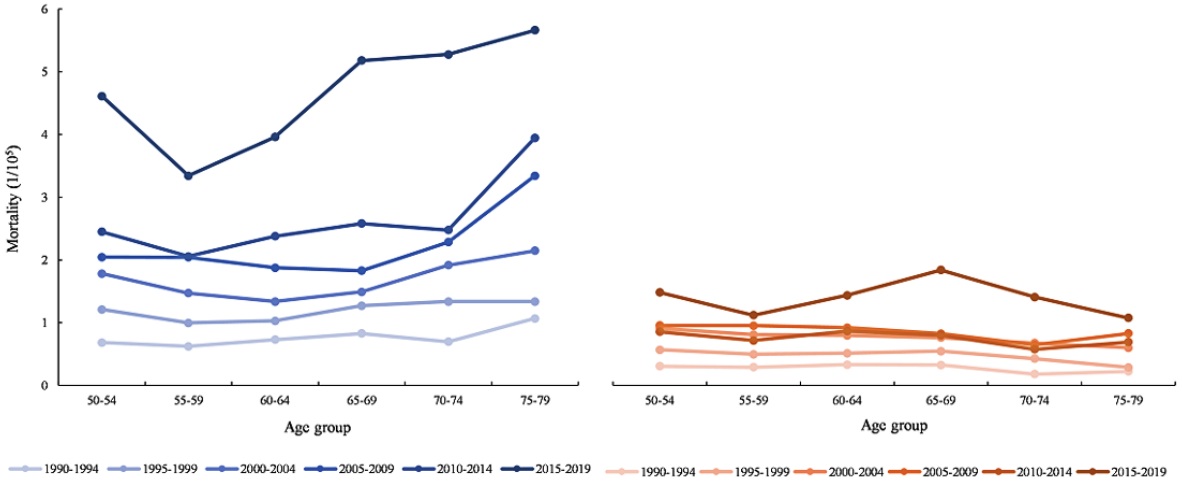


Figure 3. Age-adjusted mortality rates of HIV-infected persons per 100,000 person-years among men and women by age groups, 1990-2019 (adjusted to the data of the 6th population census of China in 2010 as the standard population).

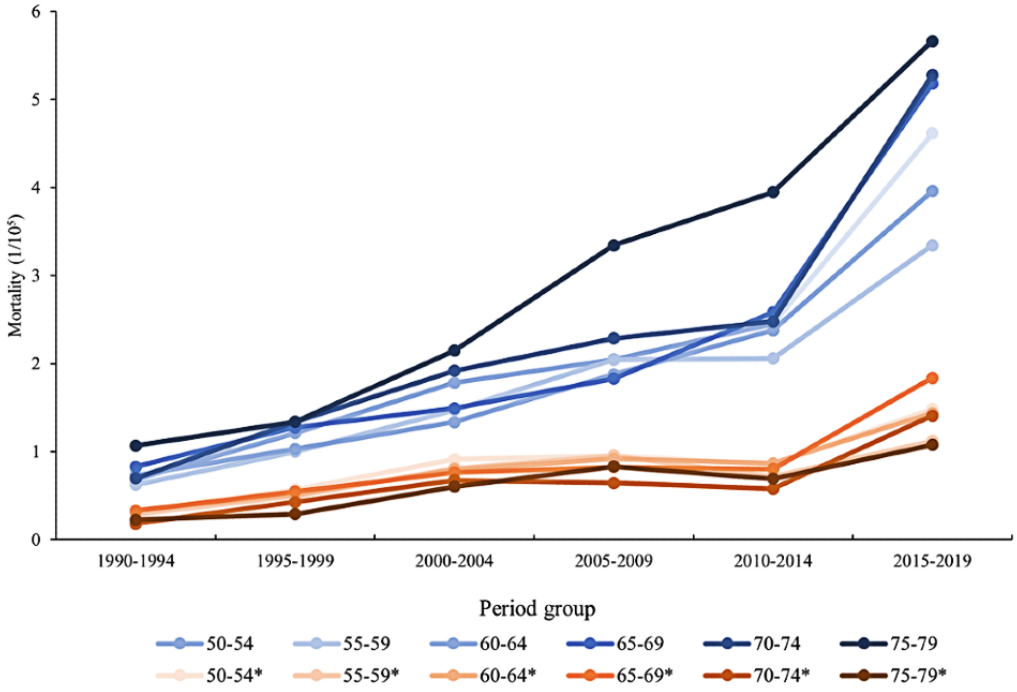
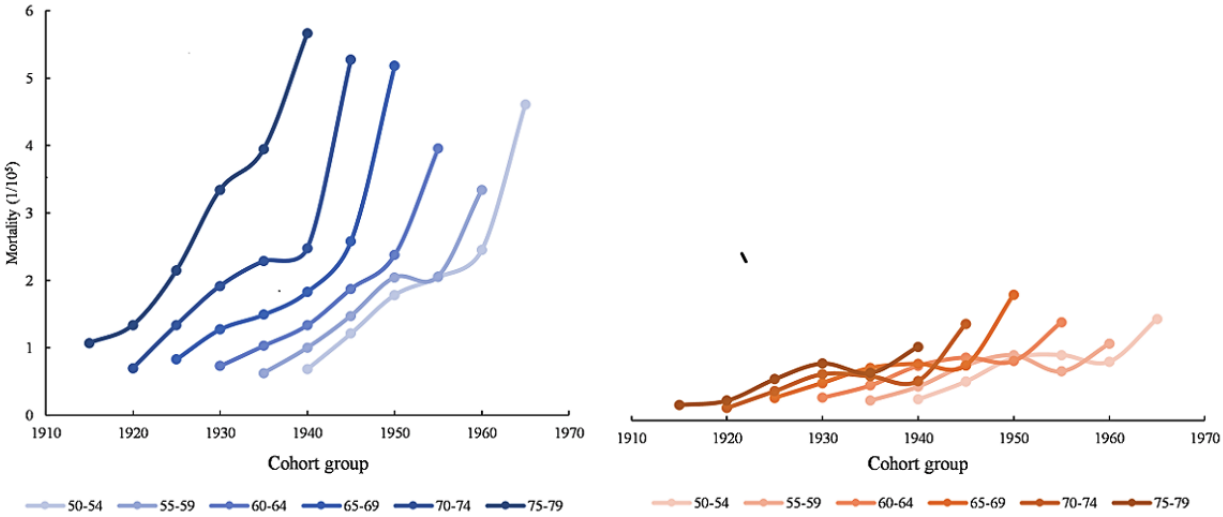


Figure 4. Age-specific HIV mortality in different Chinese cohorts aged over 50 years (per 100,000 people): HIV mortality rates of different age groups (50-54 to 75-79 years old) in each cohort (1915-1919, 1920-1924, 1925-1929, 1930-1934, 1935-1939, 1940-1944, 1945-1949, 1950-1954, 1955-1959, 1960-1964, and 1965-1969).



Trends in the Joinpoint Regression Analysis Result

Table 1 shows the joinpoint regression analysis results of the changing trend of the death rate of patients with HIV or AIDS in China, whose age was older than 50 years in different genders, by age group during the observation period. The trends, size, and statistical significance of the mortality of HIV or AIDS in different age groups during different observation periods are described.

Over the monitoring period, trends in mortality in different age groups can be broadly divided by gender into 2 categories. Older male HIV or AIDS mortality increased over time in all age groups, with slight differences in the rate of increase between periods and mortality rates stabilizing after 2016 in most groups. All age groups saw the most significant increase from 2012 to 2016, with the 70-74 years age group exhibiting the highest APC of 30.3%, whereas the trend for older female individuals increased, then decreased, and increased again.

The first period of growth was roughly 1990-2004 with an APC of around 10%, whereas the 70-79 years age group grew by more than 15%. From 2004 to 2013, the HIV or AIDS mortality rate for older female groups decreased with an APC of around –4%. The second segment grew more significantly than the first. Similar to male groups, all APC was greater than 18% over the period 2013-2016, particularly among the 65-74 years age group, where the average annual percentage change was significant, more than 35%.

Table 1. The trend in HIV mortality age-standardized mortality rates for the age>50 years in all genders during 1990-2019.

Cohort and age range	Joinpoint regression analysis											
	Trend 1		Trend 2		Trend 3		Trend 4		Trend 5		AAPC[a]	
	year	APC[b]	year	APC	year	APC	year	APC	year	APC	(%)	95% CI
Male												
adjusted	1990~2005	7.8[c]	2005~2013	2.2[c]	2013~2016	23.4[c]	2016~2019	0.4	[d]	—	7.0[c]	(5.9~8.0)
50-54 years	1990~2002	10.7[c]	2002~2012	1.1	2012~2016	20.9[c]	2016~2019	2.4	—	—	7.8[c]	（6.6~8.9）
55-59 years	1990~2006	8.8[c]	2006~2013	–0.6	2013~2016	17.8[c]	2016~2019	4.3[c]	—	—	6.9[c]	（5.7~8.0）
60-64 years	1990~2013	5.7[c]	2013~2016	20.1[c]	2016~2019	–4.0	—	—	—	—	6.1[c]	（4.9~7.2）
65-69 years	1990~2012	4.5[c]	2012~2016	23.6[c]	2016~2019	0.1	—	—	—	—	6.5[c]	（5.4~7.7）
70-74 years	1990~2001	12.2[c]	2001~2013	1.6[c]	2013~2016	30.3[c]	2016~2019	2.0	—	—	8.3[c]	（6.9~9.8）
75-79 years	1990~2010	8.3[c]	2010~2013	–3.3	2013~2016	18.1[c]	2016~2019	–3.3	—	—	6.7[c]	（5.1~8.4）
Female												
adjusted	1990~2004	10.4[c]	2004~2013	–3.4[c]	2013~2016	25.9[c]	2016~2019	1.5	—	—	6.4[c]	(5.2~7.6)
50-54 years	1990~2004	10.8[c]	2004~2012	–4.9[c]	2012~2016	18.9[c]	2016~2019	2.3	—	—	6.4[c]	（5.2~7.5）
55-59 years	1990~2005	10.4[c]	2005~2013	–6.0[c]	2013~2016	19.4[c]	2016~2019	2.9	—	—	5.7[c]	（4.0~7.4）
60-64 years	1990~2005	9.0[c]	2005~2012	–4.2[c]	2012~2016	19.5[c]	2016~2019	–2.9	—	—	5.8[c]	（5.0~6.6）
65-69 years	1990~2002	9.8[c]	2002~2013	–1.0	2013~2016	35.7[c]	2016~2019	2.2	—	—	7.1[c]	（5.5~8.7）
70-74 years	1990~2001	16.9[c]	2001~2013	–2.8[c]	2013~2016	36.1[c]	2016~2019	7.4[c]	—	—	9.0[c]	（7.0~11.2）
75-79 years	1990~1996	1.6	1996~2005	15.6[c]	2005~2013	–4.1[c]	2013~2016	18.6[c]	2016~2019	2.7	5.9[c]	（3.7~8.1）

[a]AAPC: average annual percentage change.

[b]APC: annual percentage change.

[c]Indicates that the APC and AAPC are significantly different from zero at the alpha=.05 level.

[d]Not applicable.

APC Model Analysis Results of HIV or AIDS Mortality

In this study, by fitting the age-period-cohort model, the IE algorithm was used for quantitative analysis of China's 1990-2019 elderly HIV or AIDS deaths among different age groups and periods. The result of the analysis of HIV or AIDS mortality is shown in Table 2 and Figure 5.

The age effect of male HIV or AIDS mortality showed a net increase of 0.59 (–0.21 to 0.38), from the age of 50-79 years; using the lowest value (55-59 years age group) of the male age effect as a reference, the highest value (75-79 years age group) is 1.81 times higher. The female population's effect was more complex than that of males, with the maximum occurring in the 65-69 years age group, and the minimum in the 70-74 years age group with less fluctuation.

According to period effects, there is a significant upward trend in the risk of death from HIV or AIDS among older people of both genders. Female groups had a slight decline after 2005 and then an increase. The risk of death is lowest at ages 50-54 years (male: –0.80; female: –0.78) and highest at ages 75-79 years (male: 0.86; female: 0.69). If the 1990 male period group is used as a reference, the period effect of HIV or AIDS mortality in 2015 increased by 5.23. This shows that the risk of HIV or AIDS deaths among older Chinese males increased by 522.61% over 30 years. Meanwhile, using the 1990 female population as the reference group, the period risk of HIV or AIDS deaths among older Chinese female individuals increased by 441.83% over 30 years.

According to the analysis of cohort effects, the mortality rates of male and female individuals living with HIV or AIDS have almost identical trends with complex and fluctuating variations similar to waves. The 1920-1924 period had the lowest cohort effect (male: –0.29; female: –0.38) on mortality risk from HIV or AIDS. Using the male population's lowest cohort effect (1920-1924) as a reference, the highest cohort effect risk (1950-1954) of death was 1.35. The female cohort effect has 2 peaks, occurring in cohorts 1950-1954 and 1965-1969. However, the cohort effects were not statistically different (P>.05; Table 2).

Table 2. The age-period-cohort model analysis results of HIV mortality.

Cohort and age range	Coefficient	SE	*P* value	Period	Coefficient	SE	*P* value	Cohort	Coefficient	SE	*P* value
Male[a]											
50-54[b] years	–0.0798	0.0226	.003	1990-1994[b]	–0.7965	0.0203	<.001	1915-1919[b]	–0.1228	0.0466	.02
55-59[b] years	–0.2136	0.0211	<.001	1995-1999[b]	–0.3831	0.0218	<.001	1920-1924[b]	–0.2915	0.0366	<.001
60-64[b] years	–0.1692	0.0218	<.001	2000-2004[b]	–0.1217	0.022	<.001	1925-1929	–0.0307	0.0328	.36
65-69 years	–0.0264	0.0219	.25	2005-2009[b]	0.1332	0.0217	<.001	1930-1934[b]	0.0644	0.0302	.049
70-74[b] years	0.1067	0.0215	<.001	2010-2014[b]	0.3109	0.0208	<.001	1935-1939	–0.0146	0.0275	.60
75-79[b] years	0.3824	0.0213	<.001	2015-2019[b]	0.8572	0.0235	<.001	1940-1944[b]	–0.076	0.0238	.006
—[c]	—	—	—	—	—	—	—	1945-1949[b]	0.0779	0.0261	.009
—	—	—	—	—	—	—	—	1950-1954[b]	0.1764	0.0278	<.001
—	—	—	—	—	—	—	—	1955-1959	0.056	0.0299	.08
—	—	—	—	—	—	—	—	1960-1964	0.013	0.0338	.71
—	—	—	—	—	—	—	—	1965-1969[b]	0.148	0.0595	.02
Female[d]											
50-54[b] years	0.1007	0.0316	.006	1990-1994[b]	–0.776	0.0284	<.001	1915-1919[b]	–0.17	0.0651	.02
55-59 years	–0.0206	0.0295	.496	1995-1999[b]	–0.2952	0.0304	<.001	1920-1924[b]	–0.3801	0.0512	<.001
60-64 years	0.0561	0.0304	.08	2000-2004[b]	0.0971	0.0308	.006	1925-1929	–0.0048	0.0458	.92
65-69[b] years	0.0799	0.0306	.019	2005-2009[b]	0.1991	0.0303	<.001	1930-1934[b]	0.0929	0.0422	.04
70-74[b] years	–0.1227	0.0300	<.001	2010-2014[b]	0.0857	0.0291	.01	1935-1939	0.0125	0.0384	.75
75-79[b] years	–0.0934	0.0297	.006	2015-2019[b]	0.6893	0.0328	<.001	1940-1944	–0.0124	0.0333	.72
—[c]	—	—	—	—	—	—	—	1945-1949[b]	0.1271	0.0365	.003
—	—	—	—	—	—	—	—	1950-1954[b]	0.2084	0.0389	<.001
—	—	—	—	—	—	—	—	1955-1959	0.0722	0.0417	.10
—	—	—	—	—	—	—	—	1960-1964	0.0022	0.0473	.96
—	—	—	—	—	—	—	—	1965-1969	0.052	0.0832	.54

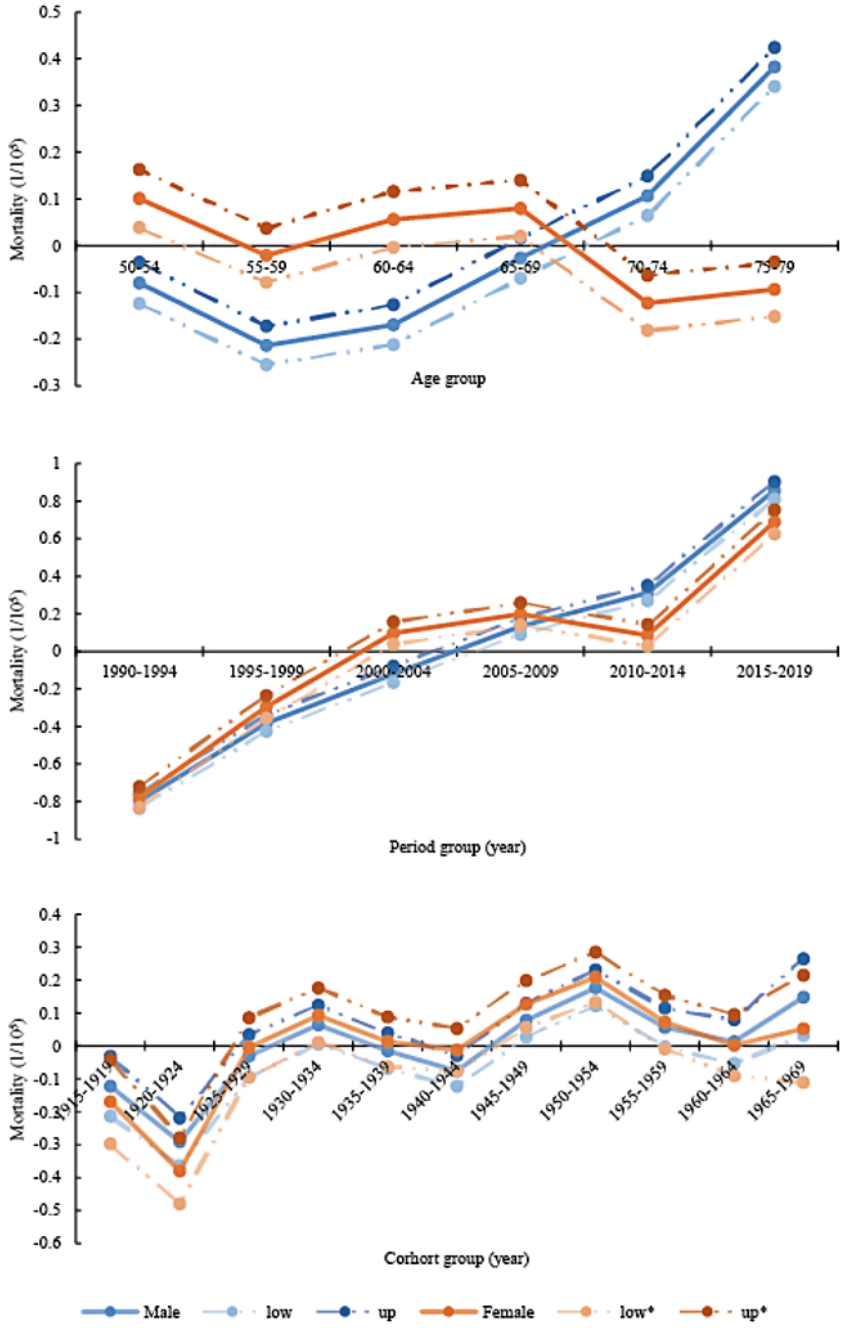
[a]R^2=0.9981; adjusted R^2=0.9957.

[b]Indicates that age, period, and cohort effects are significantly different from zero at the alpha=.05 level.

[c]Not applicable.

[d]R^2=0.9941; adjusted R^2=0.9867.

Figure 5. The age-period-cohort effect and 95% CI.



Discussion

Principal Findings

Due to the increased effectiveness of ART, life expectancy has increased for people with HIV. Although disparities in life expectancy among people with HIV continue to persist, there is an increasing prevalence of people with HIV at 50 years of age and older [28]. However, this study showed that elderly HIV or AIDS mortality rates in China increased from 1990 to 2019, with ASMR ranging from $0.50/10^5$ to $4.54/10^5$ for male and $0.19/10^5$ to $1.43/10^5$ for female individuals. In addition to the aging population, a proportion of HIV infections occurs in older persons [4,29,30], exacerbating the severity of the HIV epidemic in the older people. Our results indicated that elderly HIV mortality in China increased rapidly, especially in male individuals (average annual percentage change=7.0). Furthermore, the ASMR showed that the mortality rate was more pronounced for male individuals as they get older, especially at 75-79 years old, but for female individuals, it peaked at 65-69 years old. It may be because with the increasing efficiency of antiretroviral therapy, the age of survival of patients who have AIDS can reach 77.3 years, which is the average life expectancy of the Chinese population [31] regardless of whether they die of diseases or natural causes.

The mortality rate of older male individuals was 2-4 times that of the older female individuals [32], both in crude rates and in ASMR, which indicated a significant gender difference in the mortality of the older people with AIDS. Possible reasons for this are that the physiological functions of people older than 50

years of age have not declined, and the physical condition and sexual needs of older male individuals are still at a high level; the standard of living of mainland Chinese residents has improved in the early 21st century [9], whereas in older female individuals, incidences are mainly due to spousal transmission [33]. However, owing to the lack of sex education, these people did not have the most basic reproductive health education and had a low perception of risk, leading to the frequent occurrence of high-risk sexual behaviors [8]. Unprotected commercial sex is the main route of HIV transmission among older males [33]. Therefore, long-term, in-depth, comprehensive HIV or AIDS health education for older male individuals is essential for critical groups.

The joinpoint regression analysis showed that the mortality rates for older male individuals have continued to increase over time (at different rates per period), while for female individuals, there was a downward trend compared with male individuals from 2003 to 2013. However, HIV or AIDS mortality rates also increased more slowly during this period compared with other periods. In 2004, the Chinese government announced its "Four Frees and One Care" policy [1], which may reduce HIV- or AIDS-related mortality or a reduction in the rate of increase. The policy has increased ART facilities from 671 in 2004 to 3733 in 2013, facilitating access to standardized ART for the HIV or AIDS population. It also strengthens the cooperation between medical institutions and the Centers for Disease Control and Prevention, continuously adjusts the types of antiviral drugs and treatment standards according to the actual ART needs of each region, and operates and establishes a system for the procurement, supply, and funding of relevant drugs.

To present more realistic results, the APC model was used to divide the influencing factors into age, period, and cohort. Age is one of the most important demographic factors affecting HIV mortality, and many surveys have shown that ages older than 40 years are strongly associated with mortality from AIDS-related diseases [34,35]. The age effect in the change of HIV or AIDS mortality among Chinese older male individuals reflects a quantitative relationship that the higher the age, the larger the effect coefficient, with the most significant age effect coefficient of 0.38 for the 75-79 years group, indicating that the high-risk group for death among Chinese older men with HIV is still people in the higher age group. In contrast to male individuals, the risk of death among older female individuals with HIV is generally decreasing. However, there is a slight increase between 55 and 69 years. Therefore, prevention and control for female populations should focus on the 50-69 years age group.

According to the analysis of the period effect on HIV mortality, there was a net increase of 1.653 from 1990-1994 to 2015-2019. Such rapid growth may suggest that the period effect is an essential factor influencing HIV- or AIDS-related deaths in older people. The continuous improvement of the quality of life and the neglect of the sexual needs of older people by their families will lead to unsafe sexual behaviors [6,36]. At the same time, due to the lack of sexual knowledge, older males often have the mentality of not being afraid or not caring. More unprotected commercial behaviors [37,38] increase HIV mortality risk during these periods. Hence, in this era of increasing material abundance, the trend will continue to affect older people living with HIV. Therefore, at a time of continuous economic and social progress, openness to sexuality, and significance of aging [39,40], we should use multidisciplinary approaches to curb the growing severity of HIV- or AIDS-related problems.

The cohort effect is a comprehensive indicator, and it is impacted by age and period effects. Only by fundamentally solving the above problems can we effectively reduce the mortality rate of HIV in older people. Community organizations should be focused on carrying out more sex education, especially among older male populations, enriching the cultural life of older people and promoting healthy and safe sexual attitudes [41]. More attention must be paid to HIV or AIDS education in low-income and rural areas, raising awareness about its health risks and impact on families and society in an acceptable manner [42-44]. In particular, maintaining a single sexual partner and the correct use of condoms must be the focus of education. Moreover, continuously carrying out voluntary counseling and testing, actively mobilizing the older population for HIV testing for early detection of elderly HIV, and providing timely care and effective treatment [45] are required.

Limitations

This study also has some limitations. First, this paper only provides a descriptive analysis of the GBD 2019 database without etiological and attribution analyses. Second, we could not discuss China's provinces and the differences between regions due to data inadequacy. Third, the results of GBD 2019 are mainly estimates obtained from calculations by combining a system dynamics model with a statistical model, which may differ from the actual observed data and cannot avoid distortion of the results. Finally, our study has ecological fallacies and unique limitations associated with the APC model (including identifiability issues and the uncertainty principle). Therefore, future large-scale cohort studies are needed to confirm the relevant hypotheses in this study.

Conclusion

In conclusion, our study shows a marked increase in HIV mortality for both sexes in China from 1990 to 2019. These trends may be due to changes in socioeconomic growth and lifestyle in the population. The aging trend of the population is still a significant problem for HIV prevention and treatment in older people. It is essential to carry out early HIV screening and health education for people aged 50 years and older, as is urging infected individuals to receive ART as soon as possible to prevent HIV infection and reduce mortality rate. These findings may help predict future changes in HIV mortality and identify priority populations.

Acknowledgments

This paper was supported by Sichuan Research Centre for Sociology of Sexuality and Sexuality Education, Sichuan Key Research Base of Philosophy and Social Sciences Project (grant number: SXJYB2004), and the Luzhou Social Science Union (grant LZ21A079). The funders had no role in study design, data collection, data analysis, data interpretation, writing the manuscript, and decision to publish.

Data Availability

The data set supporting the conclusions of this article are available in the GBD Data Tool repository [46].

Demographic data were collected from the Statistical Yearbook of Population and Employment of China by the National Bureau of Statistics for the calendar years 1990 to 2019 [47].

Authors' Contributions

SF, NR, and YL participated in the study conception and design, literature search, and statistical analysis. NR, YL, RW, and WZ participated in gathering data, tabulating the table, and plotting the graphic. NR drafted and wrote the report. SF, AL, ZW, and EED participated in language polishing and provided comments on the manuscript. All authors participated in interpreting data and study findings as well as critically reviewing and substantively revising the manuscript. All authors have approved the final version of the manuscript to be published. All authors agreed to be accountable for all aspects of the work and ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

References

1. Wu Z, Chen J, Scott SR, McGoogan JM. History of the HIV epidemic in China. Curr HIV/AIDS Rep 2019 Dec 26;16(6):458-466. [doi: 10.1007/s11904-019-00471-4] [Medline: 31773405]
2. National epidemiological profile of statutory infectious diseases in 2020. Chinese Centers for Disease Control and Prevention. 2021. URL: http://www.nhc.gov.cn/jkj/s3578/202103/f1a448b7df7d4760976fea6d55834966.shtml [accessed 2022-11-07]
3. Shiau S, Bender AA, O'Halloran JA, Sundermann E, Aggarwal J, Althoff KN, et al. The current state of HIV and aging: Findings presented at the 10th international workshop on HIV and aging. AIDS Res Hum Retroviruses 2020 Dec;36(12):973-981 [FREE Full text] [doi: 10.1089/AID.2020.0128] [Medline: 32847368]
4. Erlandson KM, Karris MY. HIV and aging: Reconsidering the approach to management of comorbidities. Infect Dis Clin North Am 2019 Sep;33(3):769-786 [FREE Full text] [doi: 10.1016/j.idc.2019.04.005] [Medline: 31395144]
5. Aging with HIV. Centers for Disease Control and Prevention. 2018 Nov 14. URL: https://www.cdc.gov/hiv/group/age/olderamericans/index.html [accessed 2022-11-07]
6. Smit M, Brinkman K, Geerlings S, Smit C, Thyagarajan K, Sighem AV, ATHENA observational cohort. Future challenges for clinical care of an ageing population infected with HIV: a modelling study. Lancet Infect Dis 2015 Jul;15(7):810-818 [FREE Full text] [doi: 10.1016/S1473-3099(15)00056-0] [Medline: 26070969]
7. Wing EJ. HIV and aging. Int J Infect Dis 2016 Dec;53:61-68 [FREE Full text] [doi: 10.1016/j.ijid.2016.10.004] [Medline: 27756678]
8. Autenrieth CS, Beck EJ, Stelzle D, Mallouris C, Mahy M, Ghys P. Global and regional trends of people living with HIV aged 50 and over: Estimates and projections for 2000-2020. PLoS One 2018;13(11):e0207005 [FREE Full text] [doi: 10.1371/journal.pone.0207005] [Medline: 30496302]
9. Abel T, Werner M. HIV risk behaviour of older persons. Eur J Public Health 2003 Dec;13(4):350-352. [doi: 10.1093/eurpub/13.4.350] [Medline: 14703323]
10. Wang LY, Qin QQ, Ge L, Ding ZW, Cai C, Guo W, et al. [Characteristics of HIV infections among over 50-year-olds population in China]. Zhonghua Liu Xing Bing Xue Za Zhi 2016 Feb;37(2):222-226. [doi: 10.3760/cma.j.issn.0254-6450.2016.02.015] [Medline: 26917520]
11. Zhang HX, Han MJ, Zhou Y, Xiu XF, Xu F, Wang L. [Interrupted time series analysis for influence on HIV related fatality of implementation of 'Four Free Services One Care' policy in China]. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Mar 10;41(3):406-411. [doi: 10.3760/cma.j.issn.0254-6450.2020.03.024] [Medline: 32294844]
12. Tarone RE, Chu KC, Gaudette LA. Birth cohort and calendar period trends in breast cancer mortality in the United States and Canada. J Natl Cancer Inst 1997 Feb 05;89(3):251-256. [doi: 10.1093/jnci/89.3.251] [Medline: 9017006]
13. Ma Y, Cui Y, Hu Q, Mubarik S, Yang D, Jiang Y, et al. Long-term changes of HIV/AIDS incidence rate in China and the U.S. population from 1994 to 2019: A join-point and age-period-cohort analysis. Front Public Health 2021;9:652868 [FREE Full text] [doi: 10.3389/fpubh.2021.652868] [Medline: 34869132]

14. Gao D, Zou Z, Zhang W, Chen T, Cui W, Ma Y. Age-period-cohort analysis of HIV mortality in China: Data from the Global Burden of Disease Study 2016. Sci Rep 2020 Apr 27;10(1):7065 [FREE Full text] [doi: 10.1038/s41598-020-63141-1] [Medline: 32341364]
15. Fu W. Preliminary analysis of age-period-cohort data—Basic models. In: The Identification Problem and Beyond. Boca Raton, Florida, US: CRC Press; 2018.
16. GBD 2017 DiseaseInjury Incidence Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
17. Jia Z, Ruan Y, Lu Z. HIV incidence and mortality in China. Lancet 2015 Apr 18;385(9977):1510. [doi: 10.1016/S0140-6736(15)60753-X] [Medline: 25933281]
18. Adler WH, Baskar PV, Chrest FJ, Dorsey-Cooper B, Winchurch RA, Nagel JE. HIV infection and aging: mechanisms to explain the accelerated rate of progression in the older patient. Mech Ageing Dev 1997 Jun;96(1-3):137-155 [FREE Full text] [doi: 10.1016/s0047-6374(97)01888-5] [Medline: 9223117]
19. UNAID report on global AIDS epidemic. UNAIDS. URL: https://aids2020.unaids.org/report/ [accessed 2021-11-09]
20. Glenn ND, Mason WM, Fienberg SE. Cohort analysis in social research: Beyond the identification problem. Journal of the American Statistical Association 1986 Sep;81(395):870. [doi: 10.2307/2289041]
21. Mason KO, Mason WM, Winsborough HH, Poole WK. Some methodological issues in cohort analysis of archival data. American Sociological Review 1973 Apr;38(2):242. [doi: 10.2307/2094398]
22. O'Brien R. Age-Period-Cohort Models: Approaches and Analyses with Aggregate Data. New York, US: Chapman and Hall/CRC; 2014.
23. Osmond C, Gardner MJ. Age, period and cohort models applied to cancer mortality rates. Stat Med 1982;1(3):245-259. [doi: 10.1002/sim.4780010306] [Medline: 7187097]
24. James IR, Segal MR. On a method of mortality analysis incorporating age--year interaction, with application to prostate cancer mortality. Biometrics 1982 Jun;38(2):433-443. [Medline: 7115872]
25. Holford TR. The estimation of age, period and cohort effects for vital rates. Biometrics 1983 Jun;39(2):311-324. [Medline: 6626659]
26. Yang Y, Land KC. Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications. New York, US: Chapman and Hall/CRC; 2013:9781466507524.
27. Fu W. A Practical Guide to Age-Period Cohort Analysis Using R: The Identification Problem and Beyond. New York, US: Chapman and Hall/CRC; 2017.
28. Althoff KN, Chandran A, Zhang J, Arevalo WM, Gange SJ, Sterling TR, North American AIDS Cohort Collaboration on ResearchDesign (NA-ACCORD) of IeDEA. Life-expectancy disparities among adults with HIV in the United States and Canada: The impact of a reduction in drug- and alcohol-related deaths using the Lives Saved simulation model. Am J Epidemiol 2019 Dec 31;188(12):2097-2109 [FREE Full text] [doi: 10.1093/aje/kwz232] [Medline: 31602475]
29. Sundermann EE, Erlandson KM, Pope CN, Rubtsova A, Montoya J, Moore AA, et al. Current challenges and solutions in research and clinical care of older persons living with HIV: Findings presented at the 9th international workshop on HIV and aging. AIDS Res Hum Retroviruses 2019;35(11-12):985-998 [FREE Full text] [doi: 10.1089/AID.2019.0100] [Medline: 31373216]
30. 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. [doi: 10.1016/S2352-3018(17)30066-8] [Medline: 28501495]
31. Statistical bulletin on the development of health care in China in 2019. Committee Health and Wellness. 2020. URL: http://www.gov.cn/guoqing/2021-04/09/content_5598657.htm [accessed 2022-11-07]
32. Wu Z, Sullivan SG, Wang Y, Rotheram-Borus MJ, Detels R. Evolution of China's response to HIV/AIDS. Lancet 2007 Feb 24;369(9562):679-690 [FREE Full text] [doi: 10.1016/S0140-6736(07)60315-8] [Medline: 17321313]
33. Chen X, Li X, Qin B, Zheng J, He J, Wang L, et al. Older HIV-positive adults in Xiangxi, China: infection modes and associated risk factors. Sex Transm Dis 2012 Sep;39(9):716-719. [doi: 10.1097/OLQ.0b013e31825af361] [Medline: 22902669]
34. Zhang F, Dou Z, Ma Y, Zhang Y, Zhao Y, Zhao D, et al. Effect of earlier initiation of antiretroviral treatment and increased treatment coverage on HIV-related mortality in China: a national observational cohort study. Lancet Infect Dis 2011 Jul;11(7):516-524. [doi: 10.1016/S1473-3099(11)70097-4] [Medline: 21600849]
35. Chen L, Pan X, Ma Q, Yang J, Xu Y, Zheng J, et al. HIV cause-specific deaths, mortality, risk factors, and the combined influence of HAART and late diagnosis in Zhejiang, China, 2006-2013. Sci Rep 2017 Feb 15;7:42366. [doi: 10.1038/srep42366] [Medline: 28198390]
36. Lindau ST, Schumm LP, Laumann EO, Levinson W, O'Muircheartaigh CA, Waite LJ. A study of sexuality and health among older adults in the United States. N Engl J Med 2007 Aug 23;357(8):762-774 [FREE Full text] [doi: 10.1056/NEJMoa067423] [Medline: 17715410]

37. Xie T, Wu N. Epidemiological and mortality analysis of older adults with HIV in eastern China. Clin Interv Aging 2013;8:1519-1525 [FREE Full text] [doi: 10.2147/CIA.S53657] [Medline: 24277983]
38. Cheng S, Siankam B. The impacts of the HIV/AIDS pandemic and socioeconomic development on the living arrangements of older persons in sub-Saharan Africa: a country-level analysis. Am J Community Psychol 2009 Sep;44(1-2):136-147. [doi: 10.1007/s10464-009-9243-y] [Medline: 19543825]
39. Zeng Y, Fan J, Zhang Q, Wang PC, Tang DJ, Zhon SC, et al. Detection of antibody to LAV/HTLV-III in sera from hemophiliacs in China. AIDS Res 1986 Dec;2 Suppl 1:S147-S149. [Medline: 3103638]
40. Lu H, Liu Y, Dahiya K, Qian H, Fan W, Zhang L, et al. Effectiveness of HIV risk reduction interventions among men who have sex with men in China: a systematic review and meta-analysis. PLoS One 2013;8(8):e72747 [FREE Full text] [doi: 10.1371/journal.pone.0072747] [Medline: 24137497]
41. Shippy RA, Karpiak SE. The aging HIV/AIDS population: fragile social networks. Aging Ment Health 2005 May;9(3):246-254. [doi: 10.1080/13607860412331336850] [Medline: 16019278]
42. Nogueira JDA, Silva AO, de Sá LR, Almeida SAD, Monroe AA, Villa TCS. AIDS in adults 50 years of age and over: characteristics, trends and spatial distribution of the risk. Rev Lat Am Enfermagem 2014;22(3):355-363 [FREE Full text] [doi: 10.1590/0104-1169.3327.2424] [Medline: 25029044]
43. Sabin CA, Reiss P. Epidemiology of ageing with HIV: what can we learn from cohorts? AIDS 2017 Jun 01;31 Suppl 2:S121-S128. [doi: 10.1097/QAD.0000000000001374] [Medline: 28471942]
44. Puhr R, Kumarasamy N, Ly PS, Ng OT, Van Nguyen K, Merati TP, et al. HIV and aging: Demographic change in the Asia-Pacific region. J Acquir Immune Defic Syndr 2017 Apr 15;74(5):e146-e148 [FREE Full text] [doi: 10.1097/QAI.0000000000001258] [Medline: 28267699]
45. Ren N, Li Y, Zhang W, Wang R, Fan S, Li A. Disease burden and trend of HIV/AIDs among the elderly in China during 1990-2019. Chin J Dis Prev. 2022 2022;26(06):639-644. [doi: 10.16462/j.cnki.zhjbkz.2022.06.004]
46. GBD Results. Institute for Health Metrics and Evaluation. URL: http://ghdx.healthdata.org/gbd-results-tool [accessed 2022-11-07]
47. Statistical Yearbook of Population and Employment of China, 1990-2019. National Bureau of Statistics. URL: https://data.cnki.net/Trade/yearbook/single/N2020020031?z=Z001 [accessed 2022-11-07]

Abbreviations

APC: age-period-cohort
ART: antiretroviral therapy
ASMR: age-standardized mortality rates
GBD: Global Burden of Disease
IE: Intrinsic Estimator

Edited by G Eysenbach; submitted 21.12.21; peer-reviewed by Q Qin, J Park; comments to author 10.05.22; revised version received 29.05.22; accepted 23.10.22; published 17.11.22.

Please cite as:
Ren N, Li Y, Wan Z, Wang R, Zhang W, Dzakah EE, Zhang J, Li A, Fan S
Patterns of HIV or AIDS Mortality Among Older People From 1990 to 2019 in China: Age-Period-Cohort Analysis
JMIR Public Health Surveill 2022;8(11):e35785
URL: https://publichealth.jmir.org/2022/11/e35785
doi:10.2196/35785
PMID:36394944

Original Paper

The Relationship Between Population-Level SARS-CoV-2 Cycle Threshold Values and Trend of COVID-19 Infection: Longitudinal Study

Paria Dehesh[1], DVM, PhD; Hamid Reza Baradaran[1,2*], MD, PhD; Babak Eshrati[1,3*], MD, PhD; Seyed Abbas Motevalian[1], MD, PhD; Masoud Salehi[4], PhD; Tahereh Donyavi[5], MD, PhD

[1]Department of Epidemiology, School of Public Health, Iran University of Medical Sciences, Tehran, Iran

[2]Ageing Clinical and Experimental Research Team, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, United Kingdom

[3]Preventive Medicine and Public Health Research Center, Tehran, Iran

[4]Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran

[5]Department of Biotechnology, School of Allied Medical Sciences, Iran University of Medical Sciences, Tehran, Iran

[*]these authors contributed equally

Corresponding Author:
Babak Eshrati, MD, PhD
Department of Epidemiology
School of Public Health
Iran University of Medical Sciences
Hemmat Highway
Tehran, 1449614535
Iran
Phone: 98 9183616737
Email: babak.eshrati@gmail.com

Abstract

Background: The distribution of population-level real-time reverse transcription-polymerase chain reaction (RT-PCR) cycle threshold (Ct) values as a proxy of viral load may be a useful indicator for predicting COVID-19 dynamics.

Objective: The aim of this study was to determine the relationship between the daily trend of average Ct values and COVID-19 dynamics, calculated as the daily number of hospitalized patients with COVID-19, daily number of new positive tests, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19 by age. We further sought to determine the lag between these data series.

Methods: The samples included in this study were collected from March 21, 2021, to December 1, 2021. Daily Ct values of all patients who were referred to the Molecular Diagnostic Laboratory of Iran University of Medical Sciences in Tehran, Iran, for RT-PCR tests were recorded. The daily number of positive tests and the number of hospitalized patients by age group were extracted from the COVID-19 patient information registration system in Tehran province, Iran. An autoregressive integrated moving average (ARIMA) model was constructed for the time series of variables. Cross-correlation analysis was then performed to determine the best lag and correlations between the average daily Ct value and other COVID-19 dynamics–related variables. Finally, the best-selected lag of Ct identified through cross-correlation was incorporated as a covariate into the autoregressive integrated moving average with exogenous variables (ARIMAX) model to calculate the coefficients.

Results: Daily average Ct values showed a significant negative correlation (23-day time delay) with the daily number of newly hospitalized patients (P=.02), 30-day time delay with the daily number of new positive tests (P=.02), and daily number of COVID-19 deaths (P=.02). The daily average Ct value with a 30-day delay could impact the daily number of positive tests for COVID-19 (β=–16.87, P<.001) and the daily number of deaths from COVID-19 (β=–1.52, P=.03). There was a significant association between Ct lag (23 days) and the number of COVID-19 hospitalizations (β=–24.12, P=.005). Cross-correlation analysis showed significant time delays in the average Ct values and daily hospitalized patients between 18-59 years (23-day time delay, P=.02) and in patients over 60 years old (23-day time delay, P<.001). No statistically significant relation was detected in the number of daily hospitalized patients under 5 years old (9-day time delay, P=.27) and aged 5-17 years (13-day time delay, P=.39).

Conclusions: It is important for surveillance of COVID-19 to find a good indicator that can predict epidemic surges in the community. Our results suggest that the average daily Ct value with a 30-day delay can predict increases in the number of positive confirmed COVID-19 cases, which may be a useful indicator for the health system.

(JMIR Public Health Surveill 2022;8(11):e36424) doi:10.2196/36424

KEYWORDS

cycle threshold value; COVID-19; trend; surveillance; epidemiology; disease surveillance; surveillance; digital surveillance; prediction model; epidemic modeling; health system; infectious disease

Introduction

Coronaviruses are zoonotic pathogens that can be transmitted to humans after acquiring particular mutations [1]. SARS-CoV-2, which causes COVID-19, is mainly transmitted via airborne respiratory droplets. Although ocular secretions and oral-fecal transmission have also been indicated, these transmission methods remain uncertain [2,3].

A real-time reverse transcription-polymerase chain reaction (RT-PCR) test is used for detecting SARS-CoV-2 in respiratory samples as routine surveillance worldwide. The RT-PCR test has high sensitivity and specificity for diagnosing COVID-19 and offers faster turnaround times than the viral culture method; thus, this test has become the main method for diagnosing COVID-19. RT-PCR presents both qualitative and quantitative results with respect to the viral load [4]. The RT-PCR cycle threshold (Ct) value is identified as the number of amplification cycles needed to detect the target gene in samples [5]. The Ct value is a semiquantitative result of RT-PCR that reflects the amount of viral nucleic acids in a sample, and can thus be used as a proxy for viral load and may help decision-making in epidemic control. The Ct value has a reverse relationship with viral load so that each 3.3 increase in Ct value causes a 10-fold decrease in viral load [6]; the highest viral burden is on the first day of disease symptoms onset [7]. The positive result of COVID-19 RT-PCR tests has a lower Ct value than the recommended cutoff. In the United States, the Food and Drug Administration considers a Ct value <37 as the cutoff for a positive result of COVID-19 [8]. In more than 70% of samples with a Ct value <25, SARS-CoV-2 may be cultured, whereas only 3% of samples with a Ct value >35 can be cultured [9]. Several studies have reported that the Ct value also has an association with disease severity and mortality, and that the Ct values in patients who have more severe symptoms are low [5,10-12]. In addition, hospitalized patients who died from COVID-19 had lower Ct values [13]. A systematic review showed a significant correlation between Ct value and disease severity in hospitalized patients but not in nonhospitalized COVID-19 patients [5]. There is controversy among studies on the use of Ct values at an individual level for the prognosis of the disease or treatment planning. The Ct value may vary due to the collection method among laboratories [14] or the target gene selected for RT-PCR [15]. Moreover, the RT-PCR test can detect any viral material and does not distinguish between live viruses and viral debris, which may persist for a long time beyond the point of infectiousness [12].

To the best of our knowledge, few studies have examined the use of population-level Ct values as a measure of COVID-19 dynamics in communities. As Ct values have a significant relationship with disease severity and infectivity, a higher average Ct value in daily testing samples from a population may predict epidemic growth in a community. Hay et al [16] analyzed simulation and surveillance data and found that decreases in the proportion of Ct values in a population may cause a local increase in transmission or a new number of patients [16]. In addition, the median Ct value may be an effective measure for forecasting a pandemic surge.

To resolve these issues, the aims of this study were to determine the relationships between the daily trend of average Ct value and COVID-19 dynamics, including the daily number of hospitalized patients with COVID-19, daily number of new positive tests, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19 by age. We further aimed to determine the lag between these series.

Methods

Samples and RT-PCR

The samples included in this study were collected from March 21, 2021, to December 1, 2021. Inclusion criteria were samples obtained from individuals suspected of having COVID-19 and were referred to a laboratory in Tehran, Iran, to confirm the diagnosis. Daily results of Ct values of all patients referred to the laboratory for RT-PCR tests were recorded. The daily number of positive cases and the number of hospitalized people by age group for 9 months were extracted from the COVID-19 patient information registration system in Tehran province, Iran.

This study included samples of the upper respiratory tract (both nasopharyngeal and anterior nares swab samples) taken using a sterile Dacron thin swab with a plastic or aluminum handle as the main test specimen. The samples were collected by a physician, nurse, laboratory expert, and other staff with sufficient training and experience. All biological samples were sent to the Molecular Diagnostic Laboratory of Iran University of Medical Sciences in Tehran, Iran. All samples were analyzed using the Pishtazteb One-step RT-PCR COVID-19 Kit (dual-target gene diagnosis), and RNA extraction was performed using a Zybio nucleic acid extraction kit (magnetic bead method). To confirm the diagnosis, the target genes were the SARS-CoV-2 nucleocapsid gene and RdRp gene [17]. For each sample, the Ct value was recorded. The samples that produced a positive result in the RT-PCR test and had a Ct value ≤37 were recorded to determine the daily average Ct values.

Statistical Analysis

Overview

The daily median Ct value among all patients referred to the laboratory and the daily number of hospitalized patients with COVID-19 by age group were plotted over time. The autoregressive integrated moving average (ARIMA) and autoregressive integrated moving average with exogenous variables (ARIMAX) models were used to determine significant associations between the daily average Ct value and the daily number of COVID-19 hospitalizations by age, daily number of COVID-19 deaths, and daily number of positive tests in Tehran province, Iran.

ARIMA Model

Time-series analyses are appropriate when dealing with a set of data that has a time trend [18]. The Box-Jenkins time-series approach, especially the ARIMA model, is one of the best methods in time-series analysis of autocorrelated data [19], such as the daily average Ct value. In autoregressive models, the outcome (Y_t) is a linear function of the previous values and a random component. Nonseasonal ARIMA model parameters are (*p*, *d*, *q*) overall, where *p* is the order of autoregression (AR), *d* is the degree of trend difference, and *q* is the order of moving average (MA). To perform time-series analysis, it is first necessary to check the stability of the mean and variance. For this purpose, the augmented Dickey-Fuller (ADF) test is used [20] for checking the stability of the mean and the Box-Cox test is used to check the stability of the variance. Logarithm transformation and differentiation were used to establish stability in the variance and mean, respectively. The first-time differences can be expressed as:

$$Y'_t=Y_t-Y_{t-1} \quad \textbf{(1)}$$

Where Y_t represents nonstationary time-series data and Y'_t is the time series after the first-time differences. If the time series has a seasonal trend, seasonal differences are used to stabilize the series. The AR parameter *p* represents the linear correlation of the current value of the time series Y_t with the previous values Y_{t-1}, Y_{t-2},... and current residuals ε_t [21]. The MA parameter *q* shows the linear correlation of the current value of the time series Y_t with the current and previous residuals of the time series ε_t, ε_{t-1},… [22]. The general formula of AR (*p*) and MA (*q*) models are represented in equations (2) and (3), respectively:

$$Yt=C+\beta_1 Y_{t-1}+\beta_2 Y_{t-2}+\ldots+\beta pYt-n+\varepsilon_t \quad \textbf{(2)}$$

$$Yt=C+\varepsilon_t-\varphi_1\varepsilon_{t-1}-\varphi_2\varepsilon_{t-2}\ldots-\varphi_q\varepsilon_{t-q} \quad \textbf{(3)}$$

where C is a constant; β_1, β_2,…, β_p are AR model terms; and φ_1, φ_2,…, φ_q are MA model terms. The number of AR and MA parameters was determined by the autocorrelation function and partial autocorrelation function.

The general form of the ARIMA model can be written as:

$$Y'_t=C+\beta_1 Y_{t-1}+\beta_2 Y_{t-2}+\ldots+\beta_p Y_{t-p}+\varphi_1\varepsilon_{t-1}+\varphi_2\varepsilon_{t-2}+\ldots+\varphi_q\varepsilon_{t-q}+\varepsilon_t\ldots \quad \textbf{(4)}$$

Four main steps for the development of the ARIMA model include checking mean and variance stability (see Table S1 in Multimedia Appendix 1), and identifying *p* and *q* terms (see Figure S1 in Multimedia Appendix 1).

Model Parameter Estimation

The maximum-likelihood approach was used for the model parameters. To determine the best ARIMA model, among the models that passed the residual test (normality and stability in the variance), the model with the lowest Bayesian information criterion (BIC) and Akaike information criterion (AIC) was selected as the final model. The BIC and AIC formulae are represented as follows:

$$BIC=-2.\ln(L)+k.\ln(m) \quad \textbf{(5)}$$

$$AIC=2k-2\ln(L) \quad \textbf{(6)}$$

Where m is the number of observations, k is the total number of parameters in the model, and ln(*L*) is the likelihood function.

The ARIMA model was developed to the time series of the daily average Ct value, daily number of hospitalized patients with COVID-19, new number of daily positive tests, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19. The detailed method for derivation of the ARIMA model is described in Multimedia Appendix 1.

Cross-correlation Function

To evaluate the time delay between the daily average Ct value and the daily number of hospitalized patients with COVID-19, daily number of new positive tests, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19 by age, the cross-correlation function was used. The independent (daily average Ct value) and dependent variables (daily number of hospitalized patients with COVID-19, new number of daily positive tests, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19 by age) were preprocessed by the previously fit ARIMA models. The cross-correlation coefficient is mathematically represented as follows:

$$r_{\alpha\beta(k)}=C\alpha\beta(k)/S_\alpha S_\beta \quad \textbf{(7)}$$

where $C\alpha\beta(k)$ is the value of covariance between the preprocessed input time series and preprocessed output time series at the lag *k*, is the value of the standard deviation of the preprocessing input time series, and is the value of the standard deviation of the preprocessing output time series [23]. Three indicators, Schwarz Bayesian information criterion (SBIC), Hannan-Quinn information criterion (HQIC), and AIC, were used to select the best lag.

$$SBIC=\log(n)k-2\log(L(\hat{\theta})) \quad \textbf{(8)}$$

$$HQIC=-2\ln(L(\hat{\theta}))+2k\log(\log n) \quad \textbf{(9)}$$

In equations (8) and (9), n is the sample size, k is the number of estimated parameters, θ is the set of all parameter values, and $L(\hat{\theta})$ is the likelihood of the model.

ARIMAX Model

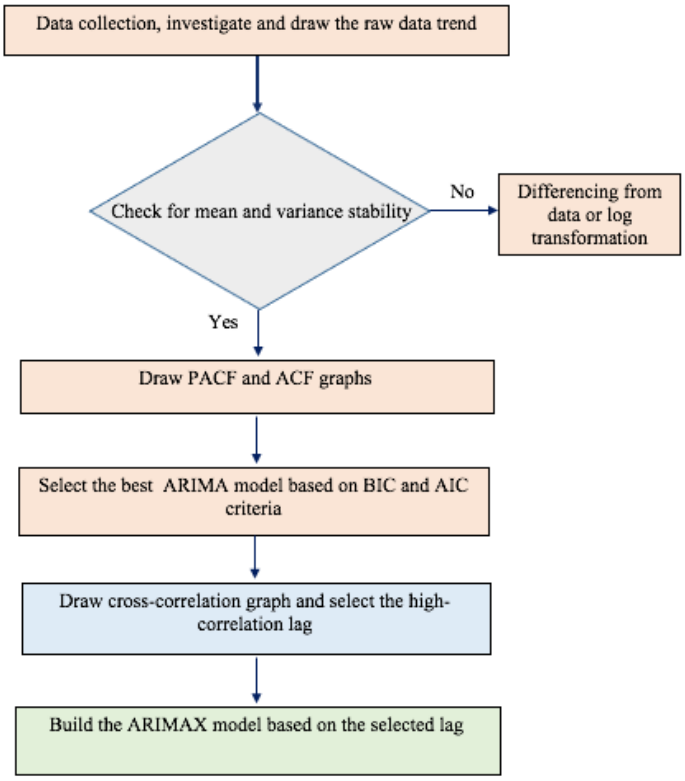
The ARIMAX model is an expansion of the ARIMA model by adding an explanatory independent variable. The ARIMAX model is the combination of multiple regression analysis and time-series analysis; therefore, it can determine the impact factor

of the relationship between different lags of Ct values and other study variables. The ARIMAX model formula is as follows:

$$Y_t = \beta x(t) + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_p Y_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \ldots + \varphi_q \varepsilon_{t-q} + \varepsilon_t \ldots \quad \textbf{(10)}$$

where x(t) is an independent variable at time t and β is its associated coefficient. $Y_{t-1)} \ldots Y_{t-p}$ is the previous value of a dependent variable, and $\varepsilon_t \ldots \varepsilon_{t-q}$ is the residual of the time series. To determine the association and coefficient of the association between the lags of the x_{t+m} time series and series Y_t, the ARIMAX model was used. The cross-correlation function was used to find the linear correlation between x_{t+m} and Y_t for different lags, which can help to find the best lags of the independent variable that might be used to predict the dependent variable [24]. The lags of Ct values that were selected through the correlation function were incorporated as covariates into the ARIMAX model with other dependent variables such as the daily number of hospitalized patients with COVID-19, number of new daily positive cases, daily number of COVID-19 deaths, and number of hospitalized patients with COVID-19 by age. The maximum-likelihood method was used for estimation of the parameters. The Ljung-Box Q test was applied to evaluate white noise for the residual series. Data were analyzed by Stata software version 14. Figure 1 shows the steps of building the best ARIMAX model.

Figure 1. Steps of building the best ARIMAX model. ACF: autocorrelation function; AIC: Akaike information criterion; ARIMA: autoregressive integrated moving average; ARIMAX: autoregressive integrated moving average with exogenous variables; BIC: Bayesian information criterion; PACF: partial autocorrelation function.



Ethics Considerations

Since individual data were not used in this study, no formal ethical assessment or informed consent was required. This study was approved by the Ethics Committee of Iran University of Medical Sciences (ethical code: IR.IUMS.REC.1400.799).

Results

Evaluation Outcomes

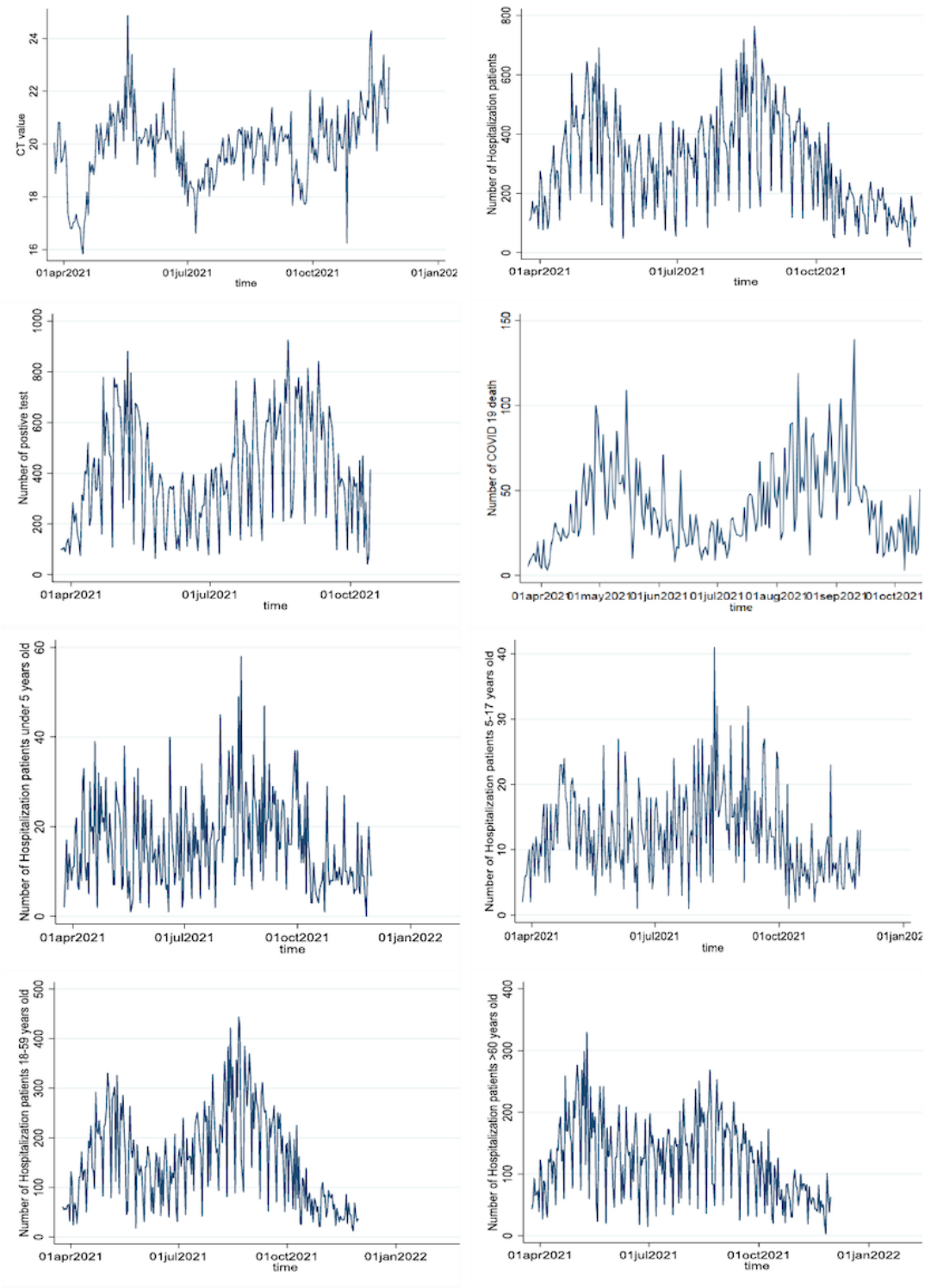
Table 1 shows descriptive statistics of the study variables that were included in the analysis. The minimum value of Ct was related to April 11, 2021, and the maximum frequency of hospitalized patients was related to August 23, 2021. Over 9 months, 80,882 positive COVID-19 tests were referred to the Molecular Diagnostic Laboratory of Iran University of Medical Sciences in Tehran, Iran.

Figure 2 shows the time trend of Ct values, along with the trends of the number of hospitalized patients, number of positive tests, number of COVID-19 deaths, number of hospitalized patients under 5 years old, number of hospitalized patients aged 5-17 years old, number of hospitalized patients aged 18-59 years old, and number of hospitalized patients over 60 years old over the 9 months. Similar to the a priori hypothesis, the daily average Ct value was negatively correlated with the daily number of hospitalized patients, daily count of positive COVID-19 tests (with a time delay), daily number of COVID-19 deaths, and daily number of hospitalized patients by age group. As shown in Figure 2, there was a time delay of approximately 28-32 days between the average daily Ct value and the daily number of hospitalized patients with COVID-19, daily count of positive COVID-19 tests, and daily number of COVID-19 deaths.

Table 1. Descriptive statistics of the study variables.

Variables	Maximum	Minimum	Mean (SD)
Dependent variable: cycle threshold value	24.87	15.83	19.89 (1.33)
Independent variables			
Number of hospitalized patients	763	47	310.65 (260.259)
Number of positive tests	925	42	396.48 (211.05)
Number of COVID-19 deaths	72	0	15.98 (24.57)
Number of hospitalized patients under 5 years old	58	0	16.514 (10.23)
Number of hospitalized patients aged 5-17 years	41	1	12.35 (6.78)
Number of hospitalized patients aged 18-59 years	444	12	155.94 (91.61)
Number of hospitalized patients over 60 years old	330	3	123.58 (63.37)

Figure 2. Trends of cycle threshold (Ct) values and other study variables over 9 months.



ARIMA Model for Study Variables

Table 2 shows the best ARIMA models for the study variables. The ARIMA (1,0,1) model was the best model for the daily average Ct value in comparison with other models, having the lowest BIC value, daily number of the hospitalized patients, and daily count of positive COVID-19 tests. The ARIMA (1,0,2) model was the best model for the daily number of COVID-19 deaths. All models had the lowest number of significant estimated parameters, and the residual analysis showed a good fit (normality and stability in the variance) for the selected ARIMA models using the AIC. There was no seasonal pattern in the study variables. The ADF test was used for evaluating stability in the mean and the Box-Cox test was used to test the time-series stability in the variance. The time series of the daily number of hospitalized patients by age did not show stability for the variance, and therefore log transformation was applied to this variable.

Table 2. The best selected autoregressive integrated moving average (ARIMA) models using the Bayesian information criterion (BIC) and Akaike information criterion (AIC).

Variable	ARIMA	Log likelihood	AIC	BIC
Cycle threshold value	(1,0,1)[a]	–355.99	702.38	716.42
Number of hospitalized patients	(1,0,1)	–1553.31	–1231.24	–1220.61
Number of positive COVID-19 tests	(1,0,1)	–1198.16	2494.99	2827.07
Number of COVID-19 deaths	(1,0,2)	–933.16	1876.33	1893.88
Number of hospitalized patients under 5 years old	(1,0,1)	–905.64	481.26	494.797
Number of hospitalized patients aged 5-17 years	(1,0,1)	–819.80	393.80	407.322
Number of hospitalized patients aged 18-59 years	(1,0,1)	–1401.00	374.58	384.72
Number of hospitalized patients over 60 years old	(1,0,1)	–919.60	397.41	407.55

[a]The numbers in parentheses represent the parameters (*p*, *d*, *q*) of the model, where *p* is the order of autoregression, *d* is the degree of trend difference, and *q* is the order of moving average.

Cross-correlation Analysis

Figure 3 shows the cross-correlations between the study variables and Ct value. In this figure, negative lags would not be considered because the negative lag indicates that the study variables could affect the average Ct value in a certain period at a later point in time; therefore, the positive lag was used to show the effect of the Ct value on the study variables in the future. A cross-correlation function was performed between the preprocessed input and output series. Table 3 shows the best lag difference between the Ct value and the study variables. Indicators such as AIC, SBIC, and HQIC were used to examine the selected lag. There was no statistically significant (all $P>.05$) lag (time delays) between the average Ct value and the daily number of hospitalized patients under 5 years old and the number of hospitalized patients aged 5-17 years. However, a significant 23-day lag was found between the average Ct value and number of hospitalized patients. The daily count of positive COVID-19 tests as well as the daily number of COVID-19 deaths had a significant 30-day lag with the average Ct value.

Figure 3. Cross-correlations (y-axes) between cycle threshold (Ct) values and other study variables.

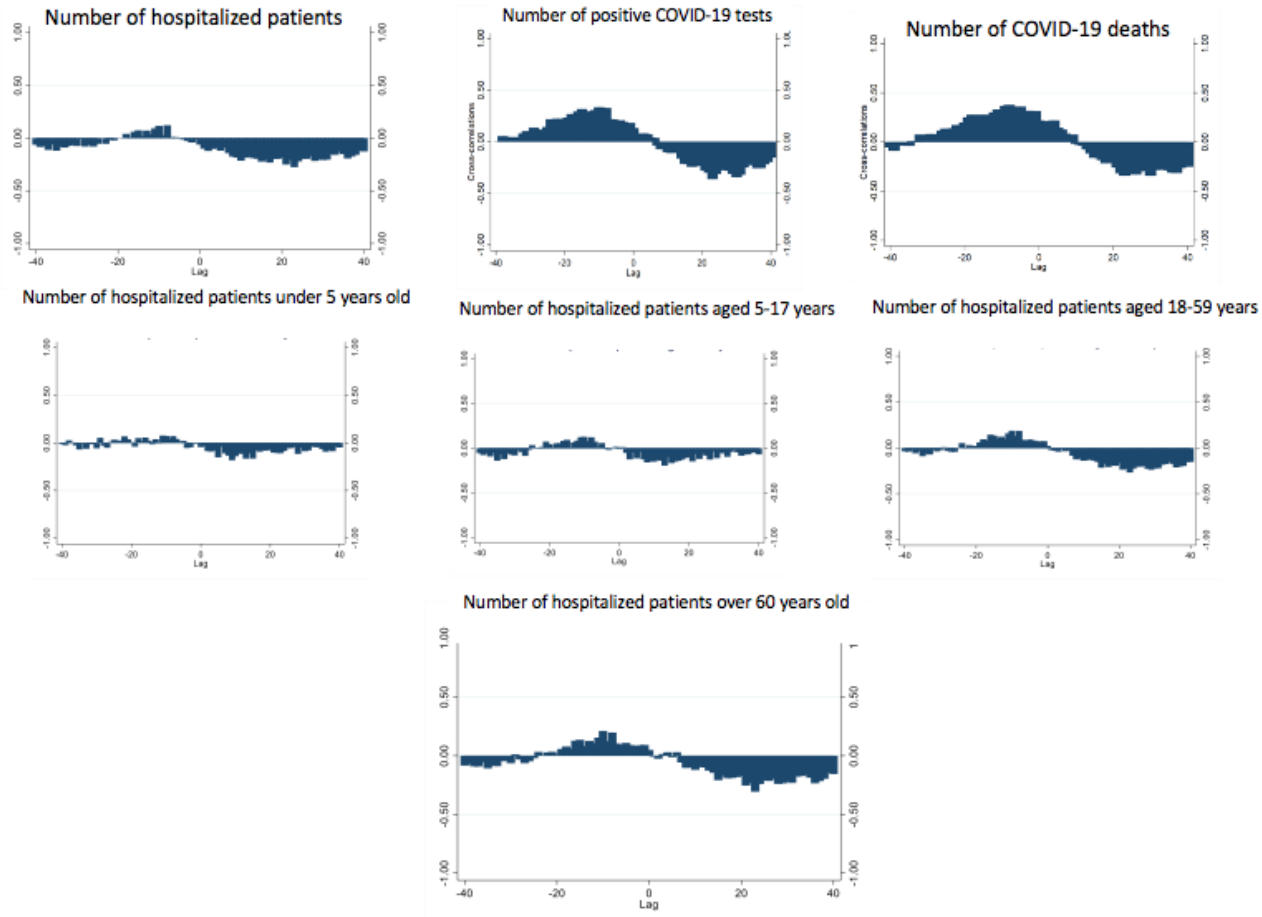


Table 3. High-correlation lags between the cycle threshold value and other study variables.

Variable	Lag	*r*	*P* value	AIC[a]	HQIC[b]	SBIC[c]
Number of hospitalized patients	23	–0.25	.02	15.34	15.95	16.85
Number of positive tests	30	–0.34	.02	15.83	16.83	8.24
Number of COVID-19 deaths	30	–0.26	.02	10.90	11.69	12.86
Number of hospitalized patients under 5 years old	9	–0.22	.76	10.38	10.73	11.25
Number of hospitalized patients aged 5-17 years	13	–0.23	.29	9.54	9.89	10.41
Number of hospitalized patients aged 18-59 years	23	–0.27	.04	14.24	14.85	15.75
Number of hospitalized patients over 60 years old	23	–0.30	.07	13.54	14.15	15.05

[a]AIC: Akaike information criterion.

[b]HQIC: Hannan-Quinn information criterion.

[c]SBIC: Schwarz Bayesian criterion.

Impact of the Ct Value on Study Variables (ARIMAX Model)

After obtaining the best lag between the daily Ct value and other variables using cross-correlation analysis (Table 3), ARIMAX was used to calculate the impact coefficients of the selected lags. Table 4 shows that a Ct value with a 30-day delay could affect the daily number of positive COVID-19 tests and the daily number of deaths from COVID-19. Specifically, a decrease in Ct value may cause an increase of approximately 16.87 times in the average number of new positive tests for COVID-19 after 30 days. In addition, the daily number of deaths from COVID-19 will increase by approximately 1.52 times after 30 days with a decrease in the Ct value. There was a significant coefficient between Ct lag (23 days) and the number of COVID-19 hospitalizations. There was also a significant association of the Ct value with a 23-day delay and the number of COVID-19 hospitalizations for patients aged 18-59 years and patients aged more than 60 years.

Table 4. Estimated coefficients obtained using autoregressive integrated moving average with exogenous variables models.

Variables and parameters	Coefficient (β)	95% CI	*P* value
Number of hospitalized patients; best model: (1,0,1)			
Ct[a] (23)[b]	–24.12	–41.08 to –7.16	.005
AR[c] (1)	.99	.95 to 1.02	<.001
MA[d] (1)	–.87	–.96 to –.78	<.001
Number of COVID-19 deaths; best model: (1,0,2)			
Ct (30)	–1.52	–2.86 to –.18	.03
AR (1)	.96	.89 to 1.03	<.001
MA (1)	–1.07	–1.22 to –.92	<.001
MA (2)	.21	.09 to .34	.001
Number of positive tests; best model: (1,0,1)			
Ct (30)	–16.87	–28.93 to –4.82	<.001
AR (1)	.96	.84 to 1.07	<.001
MA (1)	–.89	–1.06 to –.71	<.001
Number of hospitalized patients under 5 years old; best model: (1,0,1)			
Ct (9)	–.60	–1.68 to .47	.27
AR (1)	.96	.84 to 1.07	<.001
MA (1)	–.89	–1.06 to –.71	<.001
Number of hospitalized patients aged 5-17 years (1,0,1)			
Ct (13)	–.40	–1.30 to .50	.39
AR (1)	.97	.92 to 1.03	<.001
MA (1)	–.89	–.99 to –.79	<.001
Number of hospitalized patients aged 18-59 years; best model: (1,0,1)			
Ct (23)	–11.87	–21.81 to –1.94	.02
AR (1)	.99	.95 to 1.02	<.001
MA (1)	–.85	–.94 to –.76	<.001
Number of hospitalized patients over 60 years old; best model: (1,0,1)			
Ct (23)	–11.44	–17.82 to –5.07	<.001
AR (1)	.99	.96 to 1.02	<.001
MA (1)	–.90	–.98 to –.81	<.001

[a]Ct: cycle threshold.

[b]The numbers in parentheses indicate the lag in days.

[c]AR: autoregressive.

[d]MA: moving average.

Discussion

Principal Findings

The Ct value is a good proxy for viral load, which can offer the possibility of isolating people who have a higher viral load (lower Ct value) and those who have been in contact with these people for the past 5 days to reduce the transmission rate [11]. Therefore, the Ct value can be a good indicator for predicting the state of the disease process in the future. This study investigated the relationship between the population distribution of Ct values obtained from SARS-CoV-2–positive RT-PCR tests and COVID-19 dynamics. The results showed that the daily average Ct value has a significant negative relationship with three study variables of COVID-19 dynamics: daily number of hospitalized patients, daily count of positive COVID-19 tests, and daily COVID-19 deaths. The Ct value can predict the peak of the epidemic curve of the number of new positive COVID-19 patients with an interval of 30 days earlier.

Comparison With Prior Work

This result is consistent with the results of a study by Walker et al [21] showing that a declining population-level Ct value preceded increases in SARS-CoV-2 positivity tests. Another

study showed a negative association between individual Ct values and severity of symptoms of COVID-19 [25]. A few studies have focused on the effect of the population-level Ct value as an indicator for predicting pandemic surges. Consistent with this study, Tso et al [26] showed that daily median Ct values have a negative correlation with the daily count of positive tests, daily transmission rates, and daily number of COVID-19 hospitalizations in the greater El Paso area; they also showed a significant 33-day time delay between daily median Ct values and the daily number of COVID-19 hospitalizations. In this study, we found a significant 23-day time delay between the daily average Ct value and the number of hospitalized COVID-19 patients aged 18-59 years and aged more than 60 years. The former age group represents the major workforce, and are thus more likely to be exposed and become infected with the SARS-CoV-2 virus. Buchan et al [27] showed that the average Ct values were statistically similar among age groups, but patients in the age group of 80-89 years had slightly lower Ct values. According to an epidemiology study in Iran, the majority of hospitalized COVID-19 patients were in the age group of 50-60 years [28]. The relationship between the daily average Ct value and the number of COVID-19 patients aged under 5 years was not significant in this study.

Hay et al [16] estimated the epidemic trajectory in Massachusetts, United States, using a mathematical model for population-level Ct values, and also found that an increasing epidemic wave will be accompanied by a high frequency of recently infected patients with high viral loads (lower Ct values), whereas a declining epidemic wave occurs when the number of patients with older infections is high. Therefore, Ct values obtained from the disease care system during the epidemic of SARS-CoV-2 can determine the course of the epidemic process at short intervals [16]. In this study, the ARIMAX model was used to find the effect of Ct value delay time on the number of positive COVID-19 tests, and a 30-day delay was found between the average population-level Ct value and the number of positive COVID-19 cases.

Limitations

Differences in how measurements of Ct value or assurance about the quality of the data sets that are used to measure population-level Ct values in different geographical areas may affect the power of the Ct value for predicting local COVID-19 epidemic waves. Previous studies have indicated that changes in the population-level Ct values of surveillance samples may lead to a disease outbreak [16,29]. There is a hypothesis that if only patients with clinical symptoms who had positive tests were used to calculate the daily average Ct value, the association between the daily Ct value and COVID-19 cases would be more readily detected; thus, a decrease in Ct values may be more closely associated with the increasing number of COVID-19 patients. To investigate this hypothesis, only the Ct value of patients with symptoms was used to calculate the daily average Ct value in this study.

Conclusions

The daily average population-level Ct value has a relationship with the number of positive SARS-CoV-2 tests and time delay. Thirty days after reducing the daily average Ct value, the number of new COVID-19 cases is expected to increase. It is important to find a good indicator that can predict epidemic surges in the community for improved COVID-19 surveillance. Faster prediction of a new wave of disease will help health policymakers to initiate appropriate public health policies such as lockdowns for decreasing an anticipated pandemic surge, and will provide health systems an opportunity to meet the needs of medicine and facilities to support additional patients.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional data related to the statistical analysis steps and detailed description of the ARIMA model.
[DOCX File , 722 KB - publichealth_v8i11e36424_app1.docx]

References

1. Woo PCY, Lau SKP, Huang Y, Yuen K. Coronavirus diversity, phylogeny and interspecies jumping. Exp Biol Med 2009 Oct;234(10):1117-1127. [doi: 10.3181/0903-MR-94] [Medline: 19546349]
2. Seah IYJ, Anderson DE, Kang AEZ, Wang L, Rao P, Young BE, et al. Assessing viral shedding and infectivity of tears in coronavirus disease 2019 (COVID-19) patients. Ophthalmology 2020 Jul;127(7):977-979 [FREE Full text] [doi: 10.1016/j.ophtha.2020.03.026] [Medline: 32291098]
3. Liang L, Wu P. There may be virus in conjunctival secretion of patients with COVID-19. Acta Ophthalmol 2020 May 18;98(3):223-223 [FREE Full text] [doi: 10.1111/aos.14413] [Medline: 32189460]
4. Cheng H, Jian S, Liu D, Ng T, Huang W, Lin H, Taiwan COVID-19 Outbreak Investigation Team. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. JAMA Intern Med 2020 Sep 01;180(9):1156-1163 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2020] [Medline: 32356867]
5. Rao S, Manissero D, Steele V, Pareja J. A systematic review of the clinical utility of cycle threshold values in the context of COVID-19. Infect Dis Ther 2020 Sep;9(3):573-586 [FREE Full text] [doi: 10.1007/s40121-020-00324-3] [Medline: 32725536]

6. Tom M, Mina MJ. To interpret the SARS-CoV-2 test, consider the cycle threshold value. Clin Infect Dis 2020 Nov 19;71(16):2252-2254 [FREE Full text] [doi: 10.1093/cid/ciaa619] [Medline: 32435816]
7. Walsh KA, Jordan K, Clyne B, Rohde D, Drummond L, Byrne P, et al. SARS-CoV-2 detection, viral load and infectivity over the course of an infection. J Infect 2020 Sep;81(3):357-371 [FREE Full text] [doi: 10.1016/j.jinf.2020.06.067] [Medline: 32615199]
8. Emergency use authorization (EUA) summary COVID-19 RT-PCR test (Laboratory Corporation of America). US Food and Drug Administration. 2020. URL: https://www.fda.gov/media/136151/download [accessed 2022-10-23]
9. Jaafar R, Aherfi S, Wurtz N, Grimaldier C, Van Hoang T, Colson P, et al. Correlation between 3790 quantitative polymerase chain reaction-positives samples and positive cell cultures, including 1941 severe acute respiratory syndrome coronavirus 2 isolates. Clin Infect Dis 2021 Jun 01;72(11):e921-e921 [FREE Full text] [doi: 10.1093/cid/ciaa1491] [Medline: 32986798]
10. La Scola B, Le Bideau M, Andreani J, Hoang VT, Grimaldier C, Colson P, et al. Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards. Eur J Clin Microbiol Infect Dis 2020 Jun 27;39(6):1059-1061 [FREE Full text] [doi: 10.1007/s10096-020-03913-9] [Medline: 32342252]
11. Sarkar B, Sinha R, Sarkar K. Initial viral load of a COVID-19-infected case indicated by its cycle threshold value of polymerase chain reaction could be used as a predictor of its transmissibility - An experience from Gujarat, India. Indian J Community Med 2020;45(3):278. [doi: 10.4103/ijcm.ijcm_593_20]
12. Zhang X, Lu S, Li H, Wang Y, Lu Z, Liu Z, et al. Viral and antibody kinetics of COVID-19 patients with different disease severities in acute and convalescent phases: a 6-month follow-up study. Virol Sin 2020 Dec 22;35(6):820-829 [FREE Full text] [doi: 10.1007/s12250-020-00329-9] [Medline: 33351168]
13. Choudhuri J, Carter J, Nelson R, Skalina K, Osterbur-Badhey M, Johnston A, et al. SARS-CoV-2 PCR cycle threshold at hospital admission associated with patient mortality. PLoS One 2020 Dec 31;15(12):e0244777 [FREE Full text] [doi: 10.1371/journal.pone.0244777] [Medline: 33382805]
14. Miranda R, Guterres A, de Azeredo Lima CH, Filho PN, Gadelha MR. Misinterpretation of viral load in COVID-19 clinical outcomes. Virus Res 2021 Apr 15;296:198340 [FREE Full text] [doi: 10.1016/j.virusres.2021.198340] [Medline: 33592214]
15. Binnicker MJ. Challenges and Controversies to Testing for COVID-19. J Clin Microbiol 2020 Oct 21;58(11):e01695-20 [FREE Full text] [doi: 10.1128/JCM.01695-20] [Medline: 32817231]
16. Hay J, Kennedy-Shaffer L, Kanjilal S, Lennon N, Gabriel S, Lipsitch M, et al. Estimating epidemiologic dynamics from cross-sectional viral load distributions. Science 2021 Jul 16;373(6552):2021 [FREE Full text] [doi: 10.1126/science.abh0635] [Medline: 34083451]
17. Sheikhzadeh E, Eissa S, Ismail A, Zourob M. Diagnostic techniques for COVID-19 and new developments. Talanta 2020 Dec 01;220:121392 [FREE Full text] [doi: 10.1016/j.talanta.2020.121392] [Medline: 32928412]
18. Brockwell P, Brockwell P, Davis R. Introduction to time series and forecasting. New York, NY: Springer; 2016.
19. Malki Z, Atlam E, Ewis A, Dagnew G, Alzighaibi AR, ELmarhomy G, et al. ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound. Neural Comput Appl 2021 Oct 23;33(7):2929-2948 [FREE Full text] [doi: 10.1007/s00521-020-05434-0] [Medline: 33132535]
20. Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Solitons Fractals 2020 Jun;135:109864 [FREE Full text] [doi: 10.1016/j.chaos.2020.109864] [Medline: 32390691]
21. Diop ML, Kengne W. Piecewise autoregression for general integer-valued time series. J Stat Plan Inference 2021 Mar;211:271-286. [doi: 10.1016/j.jspi.2020.07.003]
22. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. Sci Total Environ 2020 Aug 10;729:138817 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138817] [Medline: 32360907]
23. Agustina D, Yosmar S, Rizal J. The identification of tsunami height correlation model with earthquake parameters. 2017 Presented at: IOP Conference Series: Earth and Environmental Science, The 4th International Seminar on Sciences; October 19-20, 2017; Bogor, Indonesia. [doi: 10.1088/1755-1315/187/1/012076]
24. Abolmaali S, Shirzaei S. A comparative study of SIR model, linear regression, logistic function and ARIMA model for forecasting COVID-19 cases. AIMS Public Health 2021;8(4):598-613 [FREE Full text] [doi: 10.3934/publichealth.2021048] [Medline: 34786422]
25. Rabaan AA, Tirupathi R, Sule AA, Aldali J, Mutair AA, Alhumaid S, Muzaheed, et al. Viral dynamics and real-time RT-PCR Ct values correlation with disease severity in COVID-19. Diagnostics 2021 Jun 15;11(6):1091 [FREE Full text] [doi: 10.3390/diagnostics11061091] [Medline: 34203738]
26. Tso CF, Garikipati A, Green-Saxena A, Mao Q, Das R. Correlation of population SARS-CoV-2 cycle threshold values to local disease dynamics: exploratory observational study. JMIR Public Health Surveill 2021 Jun 03;7(6):e28265 [FREE Full text] [doi: 10.2196/28265] [Medline: 33999831]
27. Buchan B, Hoff J, Gmehlin C, Perez A, Faron M, Munoz-Price L, et al. Distribution of SARS-CoV-2 PCR cycle threshold values provide practical insight into overall and target-specific sensitivity among symptomatic patients. Am J Clin Pathol 2020 Sep 08;154(4):479-485 [FREE Full text] [doi: 10.1093/ajcp/aqaa133] [Medline: 32687186]
28. Nikpouraghdam M, Jalali Farahani A, Alishiri G, Heydari S, Ebrahimnia M, Samadinia H, et al. Epidemiological characteristics of coronavirus disease 2019 (COVID-19) patients in IRAN: a single center study. J Clin Virol 2020 Jun;127:104378 [FREE Full text] [doi: 10.1016/j.jcv.2020.104378] [Medline: 32353762]

29. Walker A, Pritchard E, House T, Robotham J, Birrell P, Bell I, COVID-19 Infection Survey team. Ct threshold values, a proxy for viral load in community SARS-CoV-2 cases, demonstrate wide variation across populations and over time. Elife 2021 Jul 12;10:e64683. [doi: 10.7554/eLife.64683] [Medline: 34250907]

Abbreviations

ADF: augmented Dickey-Fuller
AIC: Akaike information criterion
AR: autoregressive
ARIMA: autoregressive integrated moving average
ARIMAX: autoregressive integrated moving average with exogenous variables
BIC: Bayesian information criterion
Ct: cycle threshold
HQIC: Hannan-Quinn information criterion
MA: moving average
RT-PCR: reverse transcription-polymerase chain reaction
SBIC: Schwarz Bayesian information criterion

Edited by A Mavragani, G Eysenbach; submitted 14.01.22; peer-reviewed by M Yousefi, M Mahmoodi, M Mahmoodi; comments to author 18.05.22; revised version received 30.05.22; accepted 11.10.22; published 08.11.22.

Please cite as:
Dehesh P, Baradaran HR, Eshrati B, Motevalian SA, Salehi M, Donyavi T
The Relationship Between Population-Level SARS-CoV-2 Cycle Threshold Values and Trend of COVID-19 Infection: Longitudinal Study
JMIR Public Health Surveill 2022;8(11):e36424
URL: https://publichealth.jmir.org/2022/11/e36424
doi:10.2196/36424
PMID:36240022

Original Paper

A Key Comprehensive System for Biobehavioral Surveillance of Populations Disproportionately Affected by HIV (National HIV Behavioral Surveillance): Cross-sectional Survey Study

Dafna Kanny[1], PhD; Dita Broz[1], PhD; Teresa Finlayson[1], PhD; Kathryn Lee[1], MPH; Catlainn Sionean[1], PhD; Cyprian Wejnert[1], PhD; NHBS Study Group[2]

[1]Behavioral and Clinical Surveillance Branch, Division of HIV Prevention, Centers for Disease Control and Prevention, Atlanta, GA, United States
[2]see Acknowledgments

Corresponding Author:
Dafna Kanny, PhD
Behavioral and Clinical Surveillance Branch
Division of HIV Prevention
Centers for Disease Control and Prevention
1600 Clifton Rd NE
Mailstop US8-4
Atlanta, GA, 30329
United States
Phone: 1 770 488 5411
Email: dkk3@cdc.gov

Abstract

Background: The National HIV Behavioral Surveillance (NHBS) is a comprehensive system for biobehavioral surveillance conducted since 2003 in 3 populations disproportionately affected by HIV: gay, bisexual, and other men who have sex with men (MSM); people who inject drugs; and heterosexually active persons at increased risk for HIV infection (HET). This ongoing and systematic collection and analysis of data is needed to identify baseline prevalence of behavioral risk factors and prevention service use, as well as to measure progress toward meeting HIV prevention goals among key populations disproportionately affected by HIV.

Objective: This manuscript provides an overview of NHBS from 2003 to 2019.

Methods: NHBS is conducted in rotating, annual cycles; these 3 annual cycles are considered a round. Venue-based, time-space sampling is used for the MSM population. Respondent-driven sampling is used for people who inject drugs and HET populations. A standardized, anonymous questionnaire collects information on HIV-related behavioral risk factors, HIV testing, and use of prevention services. In each cycle, approximately 500 eligible persons from each participating area are interviewed and offered anonymous HIV testing.

Results: From 2003 to 2019, 168,600 persons were interviewed and 143,570 agreed to HIV testing across 17 to 25 cities in the United States. In the fifth round (2017 to 2019), over 10,000 (10,760-12,284) persons were interviewed each of the 3 population cycles in 23 cities. Of those, most (92%-99%) agreed to HIV testing. Several cities also conducted sexually transmitted infection or hepatitis C testing.

Conclusions: NHBS is critical for monitoring the impact of the Ending the HIV Epidemic in the United States initiative. Data collected from NHBS are key to describe trends in key populations and tailor new prevention activities to ensure high prevention impact. NHBS data provide valuable information for monitoring and evaluating national HIV prevention goals and guiding national and local HIV prevention efforts. Furthermore, NHBS data can be used by public health officials and researchers to identify HIV prevention needs, allocate prevention resources, and develop and improve prevention programs directed to the populations of interest and their communities.

(JMIR Public Health Surveill 2022;8(11):e39053) doi:10.2196/39053

KEYWORDS

HIV; biobehavioral surveillance; men who have sex with men; persons who inject drugs, heterosexually active persons at increased risk for HIV infection; HIV risk; public health; surveillance; HIV prevention; HIV epidemic

Introduction

More than 40 years into the public health response to HIV, tremendous progress to prevent HIV transmission and save lives has been made globally and in the United States. Today, the tools to eliminate HIV exist, yet effective health interventions are not reaching populations that have been marginalized and are experiencing disproportionate impact of HIV [1,2]. Key members of the population and their partners, including gay, bisexual, and other men who have sex with men (MSM) and people who inject drugs, remain disproportionately affected by HIV [3]. Furthermore, social deprivation and poverty continue to be associated with high rates of HIV [4,5]. Biobehavioral surveillance of populations disproportionately affected by HIV has been critical to monitoring HIV prevention efforts and identifying areas of need, and it will continue to inform HIV prevention efforts, including those of Ending the HIV Epidemic in the United States by 2030 [6].

In 2003, the US Centers for Disease Control and Prevention (CDC), in collaboration with state and local partners and other surveillance and methodology experts, developed the National HIV Behavioral Surveillance (NHBS) as a comprehensive system for conducting biobehavioral surveillance among populations disproportionately affected by HIV [7]. To assure successful implementation, NHBS is focused on building relationships with community members, the intended populations, and prevention providers who work with these populations. NHBS has been funded through a series of cooperative agreements with collaborating state and local health departments. Health departments eligible to participate in NHBS are among those whose jurisdictions include a metropolitan statistical area (MSA) or a metropolitan division with high prevalence of HIV. Funded health departments conduct project activities within specified MSAs or metropolitan divisions. The key objectives of NHBS are to describe and monitor HIV behavioral risk factors, HIV testing, use of prevention services, and prevalence and trends in HIV infection in 3 populations disproportionately affected by HIV: MSM, people who inject drugs, and heterosexually active persons at increased risk for HIV infection (HET).

Male-to-male sexual contact is the most commonly reported route of HIV transmission in the United States, accounting for more than two-thirds of new diagnoses of HIV infection [3]. People who inject drugs are at high risk for HIV through sharing needles, syringes, or other drug injection equipment and through sexual contact. In the United States, about 1 in 10 HIV infections diagnosed are attributed to unsafe injection drug use or male-to-male sexual contact among people who inject drugs [8]. Among people who inject drugs, three-quarters of those who received a diagnosis of HIV infection live in urban areas [9]. About a quarter of new HIV diagnoses in the United States are associated with heterosexual sex [3]. Low-income HET in urban areas have highest HIV burden [10,11]. Stigma and discrimination related to male-male sex and drug use and overall health disparities linked with social and economic disadvantages make the populations surveyed in NHBS susceptible to multiple physical and health problems and can affect whether they seek HIV testing, treatment, and other health services [12-16]. Active community recruitment in NHBS ensures that impactful data are collected to inform prevention efforts for these populations and monitor progress. This manuscript provides an overview of NHBS from 2003-2019 focusing on the MSM, people who inject drugs, and HET populations.

Methods

Participants

HIV behavioral surveillance has been conducted in rotating, annual cycles since 2003 in populations disproportionately affected by HIV: MSM cycle [17], people who inject drugs cycle [18], and HET cycle [11]. For the HET cycle, NHBS considers poverty a qualifying risk factor for HIV infection. Specifically, participants are considered to have met the HET definition if they have income at or below 150% of the federal poverty level, adjusted for geographic cost of living differences. Participants in the HET cycle are asked about their combined monthly or yearly household income (in US $) from all sources for the calendar year before interview. Poverty is determined by using the US Department of Health and Human Services poverty guidelines. Because the poverty guidelines are not defined for the territory of Puerto Rico, the guidelines for the contiguous states and Washington, DC, are used for this jurisdiction. These 3 annual cycles are considered a round. In addition to the core cycles, a limited number of project areas had the option of conducting surveys in other key populations affected by HIV. In 2015, NHBS sampled young MSM aged 13 to 18 years in 3 project areas (NHBS-YMSM) [19]. In 2019-2020, NHBS received funding from the Secretary's Minority AIDS Initiative Fund to conduct a pilot program to collect data among transgender women (NHBS-Trans) in 7 project areas [20]. All participants provide their informed consent to take part in the interview, HIV testing, specimen storage (eg, dry blood spots), and if applicable, other testing (eg, hepatitis, sexually transmitted infection [STI]). Participants must consent to the survey to be eligible for the other components; however, if participants do not consent to the survey but still wish to receive HIV testing or other testing, project staff in each NHBS project area will provide referrals and information for the person to access these resources.

Ethics Approval

Activities for NHBS are approved by the CDC; NHBS is reviewed annually and determined to be a routine disease surveillance activity and thus exempt from ongoing CDC institutional review board (IRB) review (45 CFR § 46.102(l)(2)). Copy of this determination is provided in the NHBS protocol [21]. This project determination also covers secondary analyses of collected data and evaluation of NHBS, which is conducted on an ad hoc basis. These evaluations may include surveillance

evaluations, program evaluations, and evaluation activities such as inclusion of different populations (eg, transgender persons, sex partners of MSM, people who inject drugs, or HET) or different cities (eg, Southern MSAs, which are not eligible for NHBS but have a high prevalence of HIV among heterosexuals). NHBS is also reviewed by applicable local IRBs in each participating project area.

NHBS is covered under the Assurance of Confidentiality for HIV data. NHBS data are anonymous. Participants are not required to provide their names or other personal identifiers as a condition for participation. To prevent inadvertent linkage, consent forms that must be signed (due to local IRB requirement) are not labeled with a survey ID number and are maintained separately from other documents. Blood specimens, lab slips, coupons, and questionnaires are linked by survey ID numbers only. As a component of CDC HIV surveillance, NHBS data are protected by the Assurance of Confidentiality (Section 308[d] of the Public Health Service Act, 42 US Code § 242 m[d]). This assurance prohibits the disclosure of any information by the CDC that could be used to identify individuals directly or indirectly. Data collection, management, and analysis for this project are conducted in compliance with the CDC's Data Security and Confidentiality Guidelines for HIV, Viral Hepatitis, Sexually Transmitted Disease, and Tuberculosis Programs: Standards to Facilitate Sharing and Use of Surveillance Data for Public Health Action [22].

It is the responsibility of the CDC NHBS Publications Workgroup to facilitate the analysis and dissemination of NHBS data. NHBS data sets that contain aggregated data for all participating MSAs for a given cycle are maintained by the CDC. The NHBS Publications Workgroup has developed guidance to establish the methods for proposing and evaluating NHBS data analyses so that investigators can fairly participate in the process of publishing findings. All analyses of these multisite data sets must occur on the CDC premises in Atlanta, GA, or on the premises of a currently funded NHBS health department where they are housed.

Study Design

NHBS cycles are repeated cross-sectional surveys of persons disproportionately affected by HIV. The survey methods used to recruit participants are venue-based sampling (VBS) and respondent-driven sampling (RDS). VBS and RDS have been found effective for recruiting populations that are hidden. Hidden populations are those for which no sampling frame exists or whose members engage in stigmatized or illegal activities, making them reticent to divulge information that may compromise their privacy. VBS recruits attendees of MSM-focused venues (eg, clubs, organizations, street locations) within the project area to obtain the desired sample and is used in the MSM cycles [23]. RDS is a chain recruitment method that begins with a set of seeds who recruit members of their social networks to participate in project activities, who in turn recruit other members of their social networks. RDS is used in the people who inject drugs, HET, and Trans cycles [24]. YMSM used 3 sampling methods: VBS, RDS, and Facebook sampling, which used targeted banner ads to identify and recruit YMSM [19].

Procedures and Data Collection

NHBS activities are described in annual HIV surveillance reports and model protocols [21,25-27]. Trained interviewers use a standardized, anonymous questionnaire to collect information on HIV-related behavioral risk factors, HIV testing, and the use of HIV prevention services [28]. In each cycle, approximately 500 eligible persons from each participating project area are interviewed and offered optional, anonymous HIV testing. For each cycle, general NHBS eligibility criteria include age of 18 years or older, residence in participating MSA, no previous participation during the current survey cycle, ability to complete the survey in either English or Spanish, and ability to provide informed consent. In the past 5 rounds, for the MSM cycles, additional eligibility criteria included male sex at birth, male gender identity, and ever had oral or anal sex with a man. For the people who inject drugs cycles, additional eligibility criteria included injected drugs in the past 12 months and physical signs of recent injection or knowledge of injection. For the HET cycles, additional eligibility criteria included identify as male or female, had one or more opposite sex partner in the past 12 months, and aged 18 and 60 years.

There are 3 phases for NHBS implementation repeating annually. Every cycle starts with about 5 months (January to May) of formative assessment that includes interviews with people with lived experience and others closely knowledgeable about the populations [29,30]. Formative assessment helps project areas refine and develop their methods and operations for recruitment and data collection. Project areas often use formative assessments to answer key implementation questions, such as the appropriate incentive for participation, a safe, conveniently located field site location for data collection in RDS cycles, or identification of venues in the MSM cycle. The formative assessment also helps build community support for the survey. Formative assessment methods include a review of existing data, reports, and publications; qualitative interviews with key community partners, including service providers and community key informants; and ethnographic observations. From June to November, project areas collect biobehavioral data using different strategies to implement recruitment and data collection [31]. For MSM cycles, each project area conducts recruitment events at or near venues frequented by MSM. For people who inject drugs and HET cycles, project areas conduct recruitment and data collection at established field sites (eg, rented storefront, mobile van parked in an established location). In December, project areas begin closing out their projects.

Questionnaire

The NHBS interview uses a standardized, anonymous questionnaire that takes 30 to 40 minutes to complete on average [28]. Eligible individuals who consent complete an interviewer-administered, standardized, in-person anonymous questionnaire using portable computers, such as laptops or tablets. NHBS uses a single instrument for each cycle in a round. With few exceptions (eg, cycle-specific eligibility criteria), the NHBS questionnaire uses the same standardized items for all 3 cycles to assess demographics and key indicators in the following domains: sexual behaviors, alcohol use, injection and noninjection drug use, HIV testing experiences, history of

sexually transmitted diseases and hepatitis, social determinants or social conditions, and prevention activities, including pre-exposure prophylaxis. In accordance with the Paperwork Reduction Act, the Office of Management and Budget has approved the NHBS questionnaire [32]. For each round, the NHBS questionnaire is updated as needed based on feedback from interviewers, partners, and input from subject matter experts and experts in survey design. Project areas have an option to ask locally relevant questions for up to 10 additional minutes after the NHBS interview.

HIV Testing

All participants are offered HIV testing regardless of their self-reported HIV status. Testing methods include conducting a rapid test to screen for infection. If this rapid test is positive, a follow up lab-based test or a different type of rapid test to confirm infection is used. Participants are given the option of receiving their rapid test result after completing the questionnaire. Appropriate risk-reduction counseling is provided to all participants who elect testing for HIV. Counselors tailor prevention messages to specific risks identified during the behavioral surveillance interview. Counselors provide referrals for treatment and other health and social services identified during the counseling session. All laboratory tests conducted in the United States used to diagnose infection are regulated by Clinical Laboratory Improvement Amendments (CLIA). Project areas select tests from a list of CLIA-waived HIV rapid tests, which are diagnostic tests approved for use in field settings by nonlaboratory staff.

Additional Biological Testing

CDC's Division of HIV Prevention has established collaborations with other divisions and agencies to fund additional biological testing as part of NHBS in select project areas [33]. These include collaboration with the CDC's Division of STD Prevention on sexually transmitted infection (STI) testing [34] for (1) gonorrhea (Neisseria gonorrhoeae) and chlamydia (Chlamydia trachomatis) at the pharynx and rectum offered to MSM in 5 project areas in 2017 [35], (2) gonorrhea and chlamydia testing at the pharynx and vagina offered to young heterosexually active females aged 18 to 30 years in 5 project areas in 2019, and (3) pharyngeal, rectal, and urogenital gonorrhea and chlamydia testing offered to transgender women in 5 project areas in 2019-2020. All specimens were self-collected via swabs or urine in nonclinical settings. Additionally, in 2018, Division of HIV Prevention collaborated with the National Institutes of Health's National Institute on Drug Abuse [36] and CDC's Division of Viral Hepatitis [37] to provide hepatitis C virus (HCV) testing to people who inject drugs in 10 NHBS project areas [38]. Blood-based rapid HCV testing in the field and laboratory HCV RNA testing was offered to all people who inject drugs participants in the 10 project areas, results were provided to participants within 2 weeks of testing, and participant were referred to applicable care and treatment. In addition to the HIV testing offered as part of NHBS, project areas could conduct other testing with local funds if local regulations permit anonymous testing. Results of all biological testing conducted as part of NHBS are paired with the interview.

Incentives

Participants are offered incentives in exchange for their participation, both for the interview and for HIV testing. If additional testing are offered, such as STI and HCV, participants are also offered incentives. Participants may receive incentive payments in person (eg, cash, gift card). Participant compensation for incomplete surveys may be offered in accordance with local policies. Incentives are given to those interviewed and tested for HIV (approximately $25 for each). In cycles using RDS, additional rewards (approximately $10) are paid to those who successfully recruit others. Additional incentives are generally provided for any additional testing (eg, HBV, HCV, STI). Local project areas determine the exact amount and type of incentives deemed appropriate for the local populations being interviewed and tested.

Results

From 2003 to 2019, 5 rounds of NHBS were conducted (Figure 1). The number of completed interviews, HIV testing, STI testing, and HCV testing for each cycle between 2003 to 2019 are presented in Table 1. Overall, from 2003 to 2019, 168,600 persons were interviewed and 143,570 agreed to HIV testing. The fifth round was conducted from 2017 to 2019 in 23 MSAs (Table 2) [39], which represented 59% of all persons living with diagnosed HIV in urban areas with a population of at least 500,000 at the start of the funding cycle (year's end 2016). In each cycle of the last round over 10,000 persons were interviewed (range 10,760-12,284), and of those interviewed, 33,433 HIV testing were completed (92%-99%). Several NHBS project areas conducted STI or HCV testing.

Additional rounds of NHBS are ongoing. The sixth round of NHBS was planned to start in 2020; however, due to the COVID-19 pandemic, NHBS data collection in 2020 was disrupted. Thus, the MSM cycle was repeated in 2021. The people who inject drugs cycle is conducted in 2022. Round 7 is scheduled to resume with routine cycle implementation in 2023. Since 2003, NHBS data have been used in over 400 peer-reviewed manuscripts authored by CDC, local project areas, and collaborators [40]. Local and aggregate level NHBS data have also been disseminated through surveillance reports and infographics [41] and scientific, community, and internal presentations.

Figure 1. National HIV Behavioral Surveillance core and additional populations, 2003-2019. MSM: gay, bisexual, and other men who have sex with men; PWID: persons who inject drugs; HET: heterosexually active persons at increased risk for HIV infection; YMSM: young men who have sex with men; Trans: transgender women.

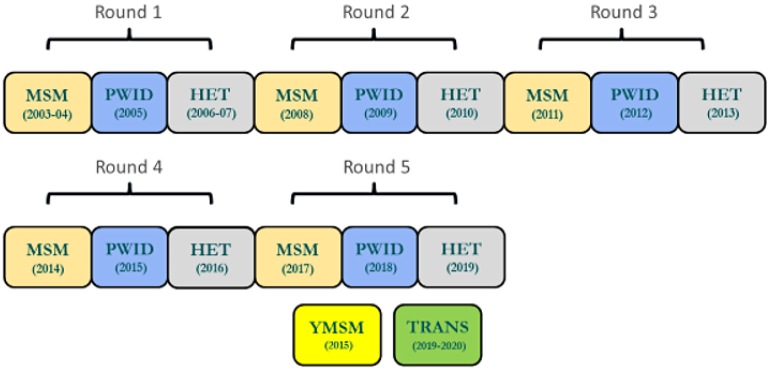


Table 1. Project areas, records, HIV testing, sexually transmitted infection testing, and hepatitis C virus testing by population/cycle, National HIV Behavioral Surveillance, 2003-2019.

Year	Population/cycle	Number of project areas	Total number of records[a]	HIV testing[b], n (%)	STI[c] testing, n (%)	HCV[d] testing, n (%)
2003-05	MSM[e]	15	10,030	2150 in 5 cities (36.7)	—[f]	—
2005	PWID[g]	22	11,613	—	—	—
2006-07	HET[h,i]	24	18,278	17,553 (96.0)	—	—
2008	MSM	21	9874	8654 (87.6)	—	—
2009	PWID	20	10,256	10,144 (98.9)	—	—
2010	HET	21	10,933	10,851 (99.2)	—	—
2011	MSM	20	9819	8922 (90.9)	—	—
2012	PWID	20	10,171	10,056 (98.9)		1461 in 4 cities (81.0)
2013	HET	20	10,535	10,479 (99.5)	—	—
2014	MSM	20	10,369	9384 (90.5)	—	—
2015	PWID	20	10,487	10,402 (99.2)	—	—
2015	YMSM[j]	3	569	508 (89.3)	—	—
2016	HET	17	9541	9445 (99.0)	—	—
2017	MSM	23	10,760	9888 (91.9)	2075[k] in 5 cities (83.1)	—
2018	PWID	23	11,444	11,355 (99.2)	—	5190 in 10 cities (99.5)
2019	HET	23	12,284	12,190 (99.2)	456[k] in 5 cities among women aged 18-30 (93.1)	—
2019-20	Trans[l]	7	1637	1589 (97.1)	824[k] in 5 cities (90.4)	—
Total	—	—	168,600	143,570	3355	6651

[a]Total number of records in each cycle's analysis is harmonized across the years within a cycle and includes the number of records that were eligible, consented to the survey, completed the interview, and provided valid answers.

[b]Valid rapid or enzyme immunoassay test for HIV antibodies.

[c]STI: sexually transmitted infection.

[d]HCV: hepatitis C virus.

[e]MSM: gay, bisexual, and other men who have sex with men.

[f]Not collected.

[g]PWID: persons who inject drugs.

[h]HET: heterosexually active persons at increased risk for HIV infection.

[i]The first HET cycle was a pilot of the optimal operational definition of HET at increased risk for HIV as well as the optimal sampling strategy (venue-based sampling vs respondent driven sampling) to reach them. The first HET cycle was also the first population and cycle that HIV testing was offered in all project areas.

[j]YMSM: young men who have sex with men.

[k]At least one valid test for gonorrhea or chlamydia from pharyngeal swabs (all cycles), rectal swab (MSM and Trans cycles only), vaginal swab (HET cycle only), or urine specimen (Trans cycle only).

[l]Trans: transgender women.

Table 2. Participating project area (funded health department), by round and cycle, National HIV Behavioral Surveillance, 2003-2019.

	Round 1, 2003-2007			Round 2, 2008-2010			Round 3, 2011-2013			Round 4, 2014-2016			Round 5, 2017-2019		
	MSM1[a]	PWID1[b]	HET1[c]	MSM2	PWID2	HET2	MSM3	PWID3	HET3	MSM4	PWID4	HET4	MSM5	PWID5	HET5
Atlanta (Georgia Dept of Human Resources)	x[d]	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Baltimore (Maryland Dept of Health and Mental Hygiene)	x	x	x	x	x	x	x	x	x	x	x		x	x	x
Boston (Massachusetts Dept of Public Health)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Chicago (Chicago Dept of Public Health)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Dallas (Texas Dept of Health)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Denver (Colorado Dept of Public Health)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Detroit (Michigan Dept of Community Health)		x	x	x	x	x	x	x	x	x	x	x	x	x	x
Fort Lauderdale (Florida Dept of Health)	x	x	x												
Houston (Houston Dept of Health and Human Services)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Indianapolis (Indiana State Dept of Health)															
Las Vegas (Nevada Dept of Health)		x	x												
Los Angeles (Los Angeles County Health Dept)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Memphis (Tennessee Dept of Health)												x	x	x	x
Miami (Florida Dept of Health)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Nassau (New York State Dept of Health)		x	x	x	x	x	x	x	x	x	x	x	x	x	x
New Haven (Connecticut Dept of Public Health)		x	x												
New Orleans (Louisiana Dept of Human Services)		x	x	x	x	x	x	x	x	x	x	x	x	x	x
New York City (NYC Dept of Health and Mental Hygiene)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

	Round 1, 2003-2007			Round 2, 2008-2010			Round 3, 2011-2013			Round 4, 2014-2016			Round 5, 2017-2019		
	MSM1[a]	PWID1[b]	HET1[c]	MSM2	PWID2	HET2	MSM3	PWID3	HET3	MSM4	PWID4	HET4	MSM5	PWID5	HET5
Newark (New Jersey Dept of Health and Senior Services)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Norfolk (Virginia Dept of Health)		x	x									x	x	x	x
Philadelphia (Philadelphia Dept of Public Health)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Portland (Oregon Health Authority)												x	x	x	x
San Diego (California Dept of Health Services)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
San Francisco (San Francisco Dept of Public Health)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
San Juan (Puerto Rico Health Dept)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Seattle (Washington Dept of Health)		x	x	x	x	x	x	x	x	x	x	x	x	x	x
St Louis (Missouri Dept of Health and Senior Services)		x	x	x	x	x									
Washington DC (DC Dept of Health)	x		x	x	x	x	x	x	x	x	x	x	x	x	x
Total project areas	17	24	25	21	21	21	20	20	20	20	20	22	23	23	23

[a]MSM: gay, bisexual, and other men who have sex with men.

[b]PWID: people who inject drugs.

[c]HET: heterosexually active persons at increased risk for HIV infection.

[d]x: indicates the specific round and cycle in which the project areas participated.

Discussion

Principal Findings

NHBS data have been used to provide behavioral and community context for trends seen in HIV diagnoses reported to the CDC's National HIV Surveillance System [42]. NHBS data have also described populations with high burden of HIV and thus have provided indications for intervention to prevent HIV transmission. Given the high levels of stigma, discrimination, and health inequity experienced by populations included in NHBS, this system provides data to address systemic and structural factors of HIV disparities. Through systematic, ongoing surveillance in groups disproportionately affected by HIV, NHBS has provided important information for planning and assessing efforts to prevent HIV in key populations.

NHBS populations often experience myriad comorbidities beyond HIV. To better serve these populations and assure successful implementation, NHBS seeks and maintains extensive collaborations. These collaborations include building relationships with community members, the intended populations, and prevention providers that work with these populations throughout the life cycle of the surveillance system. CDC and collaborators meet annually following data collection to debrief on methodological lessons learned in the preceding year and incorporate these into future iterations of NHBS. Collaborations to conduct additional biological testing that expands its public health mission beyond HIV have provided testing and referral to care for chlamydia, gonorrhea, and hepatitis C virus. These data have enhanced our knowledge of STIs and hepatitis C among NHBS populations, especially persons who may not access medical care [35]. Further, data gathered during these activities have addressed gaps in information about the prevalence of acute and chronic HCV infection among people who inject drugs in the United States [38].

Although HIV behavioral surveillance data cannot be used to evaluate the efficacy of specific interventions, these are important for monitoring whether HIV prevention efforts are reaching populations disproportionately affected by HIV within a community and whether these efforts meet local and national prevention goals. At the individual level, NHBS participants

have benefited directly from HIV prevention counseling, knowledge of their HIV status, and referrals for additional HIV prevention information and linkage to care. Participants who have preliminary HIV-positive or confirmed HIV-positive test results were counseled and referred for treatment and case management services.

Limitations

NHBS is not nationally representative and might not be generalizable to all US urban areas, nonurban areas, or all MSM, people who inject drugs, or HET populations. However, the hidden and hard-to-reach nature of these populations prevents collection of nationally representative samples. NHBS data represent the gold standard of national level data used to inform HIV prevention among these population in the United States. There are several sources of bias in RDS: (1) groups that are more insular (ie, more likely to recruit only within their own group) are more likely to be overrepresented (if recruitment chains become trapped inside the group) or underrepresented (if recruitment chains cannot access the group) in the sample than less insular groups; (2) groups with larger networks may be overrepresented in the sample because more recruitment paths lead to their members; and (3) some groups may be less willing or able to participate in the survey and would be underrepresented in the sample. There are several ways to assess this bias and compensate for it. Some of the potential sources of bias were controlled by NHBS project area staff; for instance, staff are encouraged to ensure that their initial peer recruits, or seeds, are diverse by race/ethnicity, gender, age, geographic location, and other important factors that would have the effect of increasing the insularity of recruitment and of homophily (ie, groups that recruit only within their own group). Project areas also implement lessons learned during formative assessment to mitigate potential participation bias. For example, information from formative assessment is used to optimize location and setup of field sites to ensure all population members have safe, convenient access to participants [43,44]. If necessary, multiple field sites are used.

Other sources of bias are considered during data analysis using information obtained during the survey. To calculate the population estimates and sample variances derived from RDS, participants' network size and information on who recruited whom (made possible through the coupon tracking system) are factored in to arrive at population estimates that reflect the underlying population. If these sources of bias cannot be satisfactorily controlled and measured, or if there are unknown barriers to peer recruitment, some assumptions on which RDS is based may not be met and the resulting estimates may not reflect the true population parameters of the NHBS population. Formative assessment and monitoring the sample throughout data collection is critical to minimize the effect of these sources of bias.

Findings from venue-based sampling methods can only be generalized to venue-attending MSM [17,45]. Some persons who are otherwise eligible (eg, by age, sexual behavior, and residence) may not attend the venues eligible for NHBS operations during the data collection cycle or not attend venues at all. To minimize the effect of this bias, formative assessment is conducted throughout the data collection period to update venue and daytime periods. If new venues or daytime periods are identified or become accessible, they should be added to the sampling frames. Similarly, if a venue becomes inaccessible (eg, lost owner approval for NHBS operations) or ineligible (eg, venue closure), it should be removed from the venue frame. Despite these limitations, venue-based sampling has obtained large and diverse samples in other studies, including earlier cycles of NHBS.

Biases in enrollment and agreement to HIV testing may result in over- or underestimation of HIV prevalence or incidence. If those who agree to be tested differ from those who decline in terms of age, race/ethnicity, or sex, findings may be less generalizable.

Conclusion

NHBS contributes to the nation's program of HIV surveillance by being the only multisite system that provides estimates on key HIV prevention measures among populations disproportionately affected by HIV, including HIV-negative individuals. NHBS data provide valuable information for monitoring and evaluating national HIV prevention goals and for guiding national and local HIV prevention efforts. Furthermore, NHBS data can be used by public health officials and researchers to identify HIV prevention needs, allocate prevention resources, and develop and improve prevention programs directed to the populations of interest and their communities.

Acknowledgments

We acknowledge the National HIV Behavioral Surveillance (NHBS) study group in the 23 participating cities and participants. The authors received no financial support for the research, authorship, or publication of this article. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. The members of the NHBS Study Group are as follows: Pascale Wortley, Jeff Todd, David Melton, Colin Flynn, Danielle German, Monina Klevens, Shauna Onofrey, Conall O'Cleirigh, Antonio D Jimenez, Irina Tabidze, David Kern, Margaret Vaaler, Jie Deng, Alia Al-Tayyib, Daniel Shodell, Emily Higgins, Vivian Griffin, Corrine Sanger, Salma Khuwaja, Zaida Lopez, Paige Padgett, Ekow Kwa Sey, Yingbo Ma, Hugo Santacruz, Monica Kent, Jack Marr, Meredith Brantley, Emma Spencer, David Forrest, Monica Faraldo, Bridget J Anderson, Ashley Tate, Meaghan Abrego, William T Robinson, Narquis Barak, Jacob Chavez, Sarah Braunstein, Alexis Rivera, Sidney Carrillo, Abdel R Ibrahim, Afework Wogayehu, Corey Rosmarin-DeStafano, Kathleen A Brady, Jennifer Shinefeld, Tanner Nassau, Timothy W Menza, E Roberto Orellana, Lauren Lipira, Sheryl Williams, Anna Flynn, Adam Bente, Willi McFarland, Desmond Miller, Danielle Veloso, Sandra Miranda De León, Yadira Rolón-Colón, María

Pabón Martínez, Tom Jaenicke, Sara Glick, Jennifer Reuer, Jennifer Kienzle, Brandie Smith, Toyah Reid, Jenevieve Opoku, and Irene Kuo.

Authors' Contributions

DK, DB, TF, KL, CS, and CW were responsible for study concept and design. TF and DK performed the analysis and interpreted the data. DK drafted the manuscript. DK, DB, TF, KL, CS, and CW were responsible for critical revision of the manuscript for important intellectual content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Lancet. 40 years of HIV/AIDS: a painful anniversary. Lancet 2021 Jun 05;397(10290):2125. [doi: 10.1016/S0140-6736(21)01213-7] [Medline: 34087107]
2. De Cock KM, Jaffe HW, Curran JW. Reflections on 40 years of AIDS. Emerg Infect Dis 2021 Jun;27(6):1553-1560 [FREE Full text] [doi: 10.3201/eid2706.210284] [Medline: 34013858]
3. HIV Surveillance Report 2019, volume 32: diagnoses of HIV infection in the United States and dependent areas. Atlanta: Centers for Disease Control and Prevention; 2021 May. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2022-08-30]
4. Pellowski JA, Kalichman SC, Matthews KA, Adler N. A pandemic of the poor: social disadvantage and the U.S. HIV epidemic. Am Psychol 2013;68(4):197-209 [FREE Full text] [doi: 10.1037/a0032694] [Medline: 23688088]
5. HIV Surveillance Supplemental Report, Volume 25, No 3: social determinants of health among adults with diagnosed HIV infection, 2018. Atlanta: Centers for Disease Control and Prevention; 2018. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-supplemental-report-2020-vol25-no3.pdf [accessed 2022-09-02]
6. 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]
7. Gallagher KM, Sullivan PS, Lansky A, Onorato IM. Behavioral surveillance among people at risk for HIV infection in the U.S.: the National HIV Behavioral Surveillance System. Public Health Rep 2007;122 Suppl 1:32-38 [FREE Full text] [doi: 10.1177/00333549071220S106] [Medline: 17354525]
8. HIV Surveillance Report, 2018, Volume 31: diagnoses of HIV infection in the United States and dependent areas, 2018 (Updated). Atlanta: Centers for Disease Control and Prevention; 2020 May. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2018-updated-vol-31.pdf [accessed 2022-08-30]
9. HIV Surveillance Supplemental Report, Volume 24, No 2: diagnoses of HIV infection among adults and adolescents in Metropolitan Statistical Areas—United States and Puerto Rico, 2017. Atlanta: Centers for Disease Control and Prevention; 2019. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-24-2.pdf [accessed 2022-08-30]
10. Centers for Disease Control and Prevention. Characteristics associated with HIV infection among heterosexuals in urban areas with high AIDS prevalence—24 cities, United States, 2006-2007. MMWR Morb Mortal Wkly Rep 2011 Aug 12;60(31):1045-1049 [FREE Full text] [Medline: 21832975]
11. Dinenno EA, Oster AM, Sionean C, Denning P, Lansky A. Piloting a system for behavioral surveillance among heterosexuals at increased risk of HIV in the United States. Open AIDS J 2012;6:169-176 [FREE Full text] [doi: 10.2174/1874613601206010169] [Medline: 23049666]
12. Silvestri F, Tilchin C, Wagner J, Hamill MM, Rompalo A, Ghanem KG, et al. Enacted sexual minority stigma, psychological distress, and sexual and drug risk behaviors among urban men who have sex with men (MSM). AIDS Behav 2022 Jul 13:1. [doi: 10.1007/s10461-022-03784-5] [Medline: 35831493]
13. Harrison SE, Muessig K, Poteat T, Koester K, Vecchio A, Paton M, et al. Addressing racism's role in the US HIV epidemic: qualitative findings from three ending the HIV epidemic prevention projects. J Acquir Immune Defic Syndr 2022 Jul 01;90(S1):S46-S55. [doi: 10.1097/QAI.0000000000002965] [Medline: 35703755]
14. Algarin AB, Ibañez GE, Forrest DW, Faraldo M, Spencer EC, Maddox L. Examining the psychometrics of the national hiv behavioral surveillance measure for community HIV-related stigma. AIDS Behav 2022 Jan;26(1):252-260. [doi: 10.1007/s10461-021-03378-7] [Medline: 34283342]
15. Baugher AR, Whiteman A, Jeffries WL, Finlayson T, Lewis R, Wejnert C, NHBS Study Group. Black men who have sex with men living in states with HIV criminalization laws report high stigma, 23 U.S. cities, 2017. AIDS 2021 Aug 01;35(10):1637-1645 [FREE Full text] [doi: 10.1097/QAD.0000000000002917] [Medline: 34270489]
16. Balaji AB, Bowles KE, Hess KL, Smith JC, Paz-Bailey G, NHBS study group. Association between enacted stigma and HIV-related risk behavior among MSM, National HIV Behavioral Surveillance System, 2011. AIDS Behav 2017 Jan;21(1):227-237. [doi: 10.1007/s10461-016-1599-z] [Medline: 27830344]

17. MacKellar DA, Gallagher KM, Finlayson T, Sanchez T, Lansky A, Sullivan PS. Surveillance of HIV risk and prevention behaviors of men who have sex with men—a national application of venue-based, time-space sampling. Public Health Rep 2007;122 Suppl 1:39-47 [FREE Full text] [doi: 10.1177/00333549071220S107] [Medline: 17354526]
18. Lansky A, Abdul-Quader AS, Cribbin M, Hall T, Finlayson TJ, Garfein RS, et al. Developing an HIV behavioral surveillance system for injecting drug users: the National HIV Behavioral Surveillance System. Public Health Rep 2007;122 Suppl 1:48-55 [FREE Full text] [doi: 10.1177/00333549071220S108] [Medline: 17354527]
19. National HIV Behavioral Surveillance—Populations/Projects: young men who have sex with men (NHBS-YMSM). Atlanta: Centers for Disease Control and Prevention; 2021. URL: https://www.cdc.gov/hiv/statistics/systems/nhbs/populations-projects/ymsm.html [accessed 2022-09-02]
20. National HIV Behavioral Surveillance—Populations/Projects: transgender women (NHBS-Trans). Atlanta: Centers for Disease Control and Prevention; 2021. URL: https://www.cdc.gov/hiv/statistics/systems/nhbs/populations-projects/trans.html [accessed 2022-09-02]
21. Behavioral Surveillance Team NCHHSTP/DHAP-SE/BCSB. Model Surveillance Protocol: National HIV Behavioral Surveillance—Round 6 COVID-19 Pandemic Update for 2021. Atlanta: Centers for Disease Control and Prevention; 2020 Dec 18. URL: https://www.cdc.gov/hiv/statistics/systems/nhbs/methods-questionnaires.html#msm-pwid-het-protocols [accessed 2022-09-02]
22. Data security and confidentiality guidelines for HIV, viral hepatitis, sexually transmitted disease, and tuberculosis programs: standards to facilitate sharing and use of surveillance data for public health action. Atlanta: Centers for Disease Control and Prevention; 2011. URL: http://www.cdc.gov/nchhstp/programintegration/docs/PCSIDataSecurityGuidelines.pdf [accessed 2022-09-02]
23. MacKellar D, Valleroy L, Karon J, Lemp G, Janssen R. The Young Men's Survey: methods for estimating HIV seroprevalence and risk factors among young men who have sex with men. Public Health Rep 1996;111 Suppl 1:138-144 [FREE Full text] [Medline: 8862170]
24. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. Soc Problems 1997 May;44(2):174-199. [doi: 10.1525/sp.1997.44.2.03x0221m]
25. HIV Surveillance Special Report, No 22—HIV infection risk, prevention, and testing behaviors among men who have sex with men: National HIV Behavioral Surveillance, 23 U.S. cities, 2017. Atlanta: Centers for Disease Control and Prevention; 2017. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-special-report-number-22.pdf [accessed 2022-09-02]
26. HIV Surveillance Special Report No 24—HIV infection risk, prevention, and testing behaviors among persons who inject drugs: National HIV Behavioral Surveillance: Injection Drug Use, 23 U.S. Cities, 2018. Atlanta: Centers for Disease Control and Prevention; 2018. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-special-report-number-24.pdf [accessed 2022-09-02]
27. HIV Surveillance Special Report No 26—HIV Infection, Risk, Prevention, and Testing Behaviors Among Heterosexually Active Adults at Increased Risk for HIV Infection: National HIV Behavioral Surveillance, 23 U.S. Cities, 2019. 2019. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-special-report-number-26.pdf [accessed 2022-09-02]
28. National HIV Behavioral Surveillance Questionnaires. Atlanta: Centers for Disease Control and Prevention; 2021. URL: https://www.cdc.gov/hiv/pdf/statistics/systems/nhbs/cdc-hiv-nhbs-round6-crq.pdf [accessed 2022-09-02]
29. Allen DR, Finlayson T, Abdul-Quader A, Lansky A. The role of formative research in the National HIV Behavioral Surveillance System. Public Health Rep 2009 Feb;124(1):26-33 [FREE Full text] [doi: 10.1177/003335490912400106] [Medline: 19413025]
30. National HIV Behavioral Surveillance MSM2021 Formative Assessment Guide. Atlanta: Centers for Disease Control and Prevention; 2021. URL: https://www.cdc.gov/hiv/pdf/statistics/systems/nhbs/cdc-hiv-nhbs-msm6-formative-assessment-guide.pdf [accessed 2022-09-02]
31. National HIV Behavioral Surveillance MSM2021 Operations Manual. Atlanta: Centers for Disease Control and Prevention; 2021. URL: https://www.cdc.gov/hiv/pdf/statistics/systems/nhbs/cdc-hiv-nhbs-msm6-operations-manual.pdf [accessed 2022-09-02]
32. Office of Information and Regulatory Affairs. Washington: US General Services Administration, Office of Management and Budget; 2020. URL: https://www.reginfo.gov/public/do/PRAOMBHistory?ombControlNumber=0920-0770 [accessed 2022-09-02]
33. National HIV Behavioral Surveillance: lab collaborations. Atlanta: Centers for Disease Control and Prevention; 2020. URL: https://www.cdc.gov/hiv/statistics/systems/nhbs/collaborations.html [accessed 2022-09-02]
34. Sexually Transmitted Disease Surveillance. Atlanta: Centers for Disease Control and Prevention; 2022. URL: https://www.cdc.gov/std/default.htm [accessed 2022-09-02]
35. Johnson Jones ML, Chapin-Bardales J, Bizune D, Papp JR, Phillips C, Kirkcaldy RD, National HIV Behavioral Surveillance Sexually Transmitted Infection Study Group. Extragenital chlamydia and gonorrhea among community venue-attending men who have sex with men—five cities, United States, 2017. MMWR Morb Mortal Wkly Rep 2019 Apr 12;68(14):321-325 [FREE Full text] [doi: 10.15585/mmwr.mm6814a1] [Medline: 30973847]

36. National Institute on Drug Use. URL: https://www.drugabuse.gov/ [accessed 2022-09-02]
37. Viral Hepatitis. Atlanta: Centers for Disease Control and Prevention; 2021. URL: https://www.cdc.gov/hepatitis/ [accessed 2022-09-02]
38. Chapin-Bardales J, Asher A, Broz D, Teshale E, Hayden T, Blanco C, et al. Hepatitis C virus infection and coinfection with HIV among PWID in 10 US cities. 2020 Presented at: Conference on Retroviruses and Opportunistic Infections; 2020; Boston URL: https://www.croiconference.org/abstract/hepatitis-c-virus-infection-and-coinfection-with-hiv-among-pwid-in-10-us-cities/
39. National HIV Behavioral Surveillance project areas. Atlanta: Centers for Disease Control and Prevention; 2021. URL: https://www.cdc.gov/hiv/statistics/systems/nhbs/projectareas.html [accessed 2022-09-02]
40. National HIV Behavioral Surveillance (NHBS) full NHBS bibliography. Atlanta: Centers for Disease Control and Prevention; 2021. URL: https://www.cdc.gov/hiv/statistics/systems/nhbs/bibliography.html [accessed 2022-09-02]
41. National HIV Behavioral Surveillance. Atlanta: Centers for Disease Control and Prevention; 2021. URL: https://www.cdc.gov/hiv/statistics/systems/nhbs/index.html [accessed 2022-09-02]
42. HIV Surveillance Overview. Atlanta: Centers for Disease Control and Prevention; 2020. URL: https://www.cdc.gov/hiv/statistics/surveillance/index.html [accessed 2022-09-02]
43. Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. AIDS 2005 May;19 Suppl 2:S67-S72. [doi: 10.1097/01.aids.0000172879.20628.e1] [Medline: 15930843]
44. McKnight C, Des Jarlais D, Bramson H, Tower L, Abdul-Quader AS, Nemeth C, et al. Respondent-driven sampling in a study of drug users in New York City: notes from the field. J Urban Health 2006 Nov;83(6 Suppl):i54-i59 [FREE Full text] [doi: 10.1007/s11524-006-9102-1] [Medline: 16977493]
45. Iachan R, Finlayson T, Kyle T, Le B, Wejnert C, Paz-Bailey G. Weighting for venue-based sampling: the MSM3 study. 2013 Presented at: Joint Statistical Meetings 2013; 2013; Montreal.

Abbreviations

CDC: Centers for Disease Control and Prevention
CLIA: Clinical Laboratory Improvement Amendments
HCV: hepatitis C virus
HET: heterosexually active persons at increased risk for HIV infection
IRB: institutional review board
MSA: metropolitan statistical area
MSM: gay, bisexual, and other men who have sex with men
NHBS: National HIV Behavioral Surveillance
RDS: respondent-driven sampling
STI: sexually transmitted infection
Trans: transgender women
VBS: venue-based sampling
YMSM: young men who have sex with men

Edited by T Sanchez, A Mavragani; submitted 27.04.22; peer-reviewed by C Folch, K Card, S Goodreau; comments to author 12.08.22; revised version received 02.09.22; accepted 29.09.22; published 15.11.22.

Please cite as:
Kanny D, Broz D, Finlayson T, Lee K, Sionean C, Wejnert C, NHBS Study Group
A Key Comprehensive System for Biobehavioral Surveillance of Populations Disproportionately Affected by HIV (National HIV Behavioral Surveillance): Cross-sectional Survey Study
JMIR Public Health Surveill 2022;8(11):e39053
URL: https://publichealth.jmir.org/2022/11/e39053
doi:10.2196/39053
PMID:36378503

Original Paper

New Surveillance Metrics for Alerting Community-Acquired Outbreaks of Emerging SARS-CoV-2 Variants Using Imported Case Data: Bayesian Markov Chain Monte Carlo Approach

Amy Ming-Fang Yen[1], PhD; Tony Hsiu-Hsi Chen[2], PhD; Wei-Jung Chang[2], PhD; Ting-Yu Lin[2], PhD; Grace Hsiao-Hsuan Jen[1], PhD; Chen-Yang Hsu[2,3], MD, PhD; Sen-Te Wang[4,5], MD, PhD; Huong Dang[6], PhD; Sam Li-Sheng Chen[1,7], PhD

[1]School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan

[2]Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan

[3]Daichung Hospital, Miaoli, Taiwan

[4]Department of Family Medicine, Taipei Medical University Hospital, Taipei, Taiwan

[5]Department of Family Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[6]Department of Economics and Finance, University of Canterbury, Christchurch, New Zealand

[7]Research Center of Cancer Translational Medicine, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Sam Li-Sheng Chen, PhD
School of Oral Hygiene
College of Oral Medicine
Taipei Medical University
No 250, Wuxing St
Taipei, 110
Taiwan
Phone: 886 27361661 ext 5211
Email: samchen@tmu.edu.tw

Abstract

Background: Global transmission from imported cases to domestic cluster infections is often the origin of local community-acquired outbreaks when facing emerging SARS-CoV-2 variants.

Objective: We aimed to develop new surveillance metrics for alerting emerging community-acquired outbreaks arising from new strains by monitoring the risk of small domestic cluster infections originating from few imported cases of emerging variants.

Methods: We used Taiwanese COVID-19 weekly data on imported cases, domestic cluster infections, and community-acquired outbreaks. The study period included the D614G strain in February 2020, the Alpha and Delta variants of concern (VOCs) in 2021, and the Omicron BA.1 and BA.2 VOCs in April 2022. The number of cases arising from domestic cluster infection caused by imported cases (Dci/Imc) per week was used as the SARS-CoV-2 strain-dependent surveillance metric for alerting local community-acquired outbreaks. Its upper 95% credible interval was used as the alert threshold for guiding the rapid preparedness of containment measures, including nonpharmaceutical interventions (NPIs), testing, and vaccination. The 2 metrics were estimated by using the Bayesian Monte Carlo Markov Chain method underpinning the directed acyclic graphic diagram constructed by the extra-Poisson (random-effect) regression model. The proposed model was also used to assess the most likely week lag of imported cases prior to the current week of domestic cluster infections.

Results: A 1-week lag of imported cases prior to the current week of domestic cluster infections was considered optimal. Both metrics of Dci/Imc and the alert threshold varied with SARS-CoV-2 variants and available containment measures. The estimates were 9.54% and 12.59%, respectively, for D614G and increased to 14.14% and 25.10%, respectively, for the Alpha VOC when only NPIs and testing were available. The corresponding figures were 10.01% and 13.32% for the Delta VOC, but reduced to 4.29% and 5.19% for the Omicron VOC when NPIs, testing, and vaccination were available. The rapid preparedness of containment measures guided by the estimated metrics accounted for the lack of community-acquired outbreaks during the D614G period, the early Alpha VOC period, the Delta VOC period, and the Omicron VOC period between BA.1 and BA.2. In contrast,

community-acquired outbreaks of the Alpha VOC in mid-May 2021, Omicron BA.1 VOC in January 2022, and Omicron BA.2 VOC from April 2022 onwards, were indicative of the failure to prepare containment measures guided by the alert threshold.

Conclusions: We developed new surveillance metrics for estimating the risk of domestic cluster infections with increasing imported cases and its alert threshold for community-acquired infections varying with emerging SARS-CoV-2 strains and the availability of containment measures. The use of new surveillance metrics is important in the rapid preparedness of containment measures for averting large-scale community-acquired outbreaks arising from emerging imported SARS-CoV-2 variants.

(JMIR Public Health Surveill 2022;8(11):e40866) doi:10.2196/40866

KEYWORDS

COVID-19; imported case; surveillance metric; early detection; community-acquired outbreak

Introduction

During the COVID-19 pandemic lasting for over 2.5 years, countries around the world have experienced cyclical COVID-19 changes alternating between lifting and operating nonpharmaceutical interventions (NPIs) and between the protective and waning effects of vaccines when facing the incessant epidemics of the COVID-19 pandemic [1-4]. The cyclical resurgence of COVID-19 at the country, continental, and global levels is mainly caused by emerging SARS-CoV-2 variants, particularly variants of concern (VOCs). In response to the resurgence of community-acquired outbreaks, 2 containment measures have become important, including the timely adjustment of NPIs (strengthened border control strategies and restricted social activities) combined with testing and the launch of mass primary and booster vaccinations [5-7].

It should be noted that the typical pattern of transmission from an imported case to domestic cluster infection is often the root of local community-acquired outbreaks caused by emerging SARS-CoV-2 variants from any region or country across the globe [2,7-9]. Such an importation-cluster transmission mode has been clearly demonstrated by the resurgence of global epidemic waves following the emergence of dominant strains of Alpha, Beta, Gamma, Delta, and Omicron VOCs. To avert local community-acquired outbreaks of emerging SARS-CoV-2 variants, rapid preparedness of containment measures and effective contact tracing are mandatory when domestic cluster infections are identified after the introduction of emerging imported cases. In addition, the risk of domestic cluster infection on the introduction of imported cases varies with each emerging SARS-CoV-2 strain owing to the evolutionary characteristics of invading VOCs, including an increase in transmissibility and a higher likelihood of escaping immune response after vaccination [4,7-12].

It is therefore important to have new surveillance metrics for monitoring the odds of having domestic cluster infection transmitted from few imported cases and setting up the alert threshold for forestalling community-acquired outbreaks, as traditional surveillance metrics, like effective reproductive number (R_t), are tailored for assessing the spread and control of community-acquired outbreaks at the population level, which may only involve a single country and a specific SARS-CoV-2 strain in a short period rather than the country of the imported case across the world and the full spectrum of SARS-CoV-2 strains with a long period [13-16]. Such a traditional epidemic surveillance model (eg, the SEIR [Susceptible-Exposed-Infected-Recovery] model) is not only limited to model the relationship of sparse cases of domestic cluster infection and small samples of imported cases from each original country, but also inflexible to make allowance for the heterogeneity of the imported-domestic transmission mode across countries and SARS-CoV-2 strains across time, as well as the variation across local regions in question. To consider these issues of heterogeneity, it is therefore necessary to develop new surveillance models and their corresponding metrics with a new statistical approach, such as a sampling method of machine learning, particularly the Bayesian Markov Chain Monte Carlo (MCMC) method, in conjunction with a sparse event history regression model, such as the extra-Poisson (random-effect) regression model with relevant parameters and random variables parameterized under the directed acyclic graphic (DAG) diagram.

Developing these new surveillance metrics for quantifying the effect size of the transmission from importation to domestic cluster infection would not only be helpful for alerting emerging community-acquired outbreaks, but also aid health professionals having rapid preparedness of SARS-CoV-2 strain–dependent containment measures, including effective and efficient contact tracing. Using a series of chronological epidemic data on COVID-19 divided into 2 phases (non-VOC phase [wild type and D614G] and VOC phase) in Taiwan, this study aimed to develop new surveillance metrices across the periods of various SARS-CoV-2 strains for alerting emerging community-acquired outbreaks by monitoring the risk of small domestic cluster infections originating from the transmission of few imported cases of emerging variants in order to forestall community-acquired outbreaks when facing emerging SARS-CoV-2 variants. The Bayesian MCMC sampling method was therefore used to estimate and predict the new surveillance metrics underpinning the DAG diagram of the Poisson or negative binomial random-effect regression model.

Methods

Data Sources

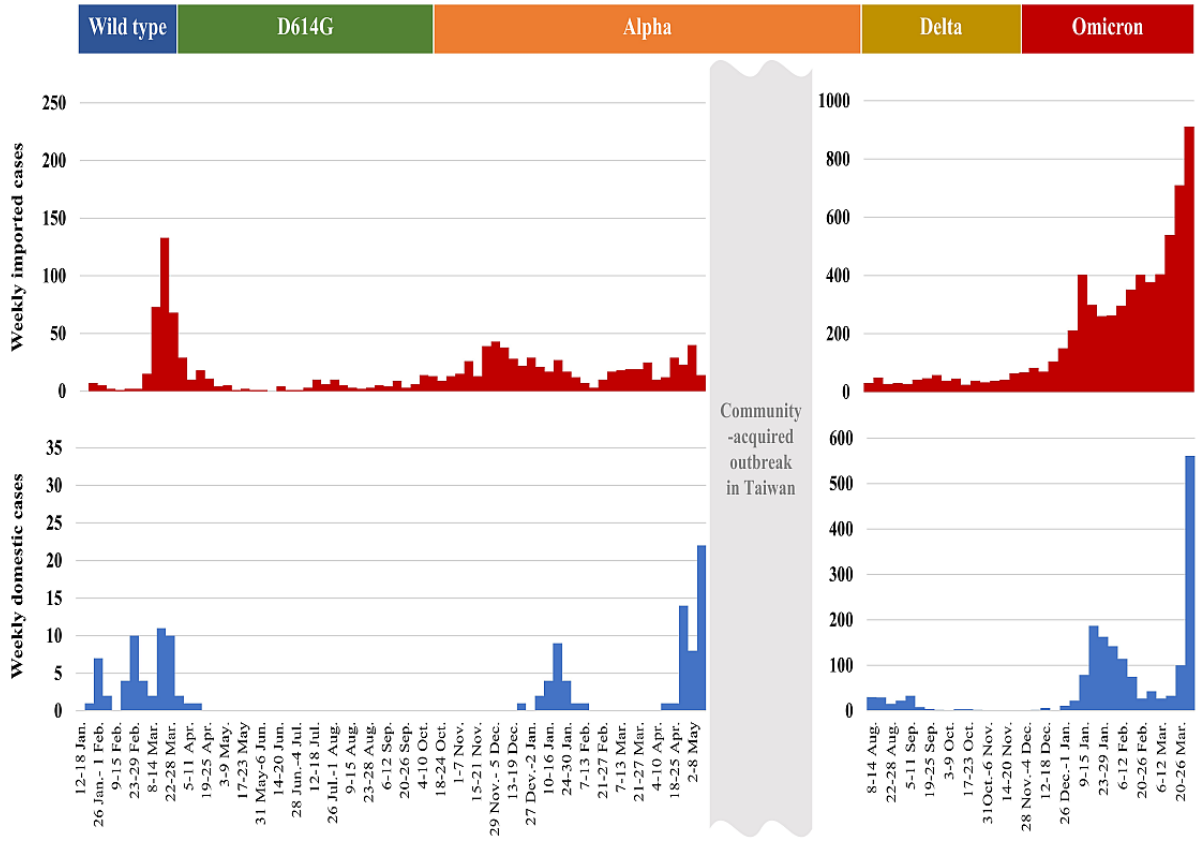
Publicly available information on COVID-19, including the daily number of cases, recovered patients, and deaths from January 1, 2020, to April 2, 2022, in Taiwan, was extracted from the report of the Central Epidemic Command Centre and the Taiwan National Infectious Disease Statistics System maintained by the Taiwan Centre for Disease Control [17].

During the period between January 11 and June 20, 2020, tabular data with epidemiologic information on COVID-19 mentioned above by county and origin of cases (domestic versus imported) were obtained. After October 2020, only aggregated numbers of imported and domestic COVID-19 cases without detailed information at the county and city levels were provided. The population sizes of 23 counties and cities in Taiwan were extracted from the official website of the Department of Household Registration [18].

Containment Measures for the Non-VOC Phase in Taiwan

The containment measures for the non-VOC phase in Taiwan centered on 2 strategies, namely border control with quarantine and isolation, and NPIs without various strategies. Multimedia Appendix 1 shows the timelines of the evolution of border control measures for this non-VOC phase. The number of imported and domestic cases of COVID-19 by the date of onset on a weekly basis is presented in Figure 1.

Figure 1. COVID-19 epidemics in Taiwan for periods without outbreaks by the origin of cases (imported vs domestic).



The COVID-19 cases in Taiwan in 2020 mainly included imported cases (Figure 1). The risk of an outbreak following the transmission of COVID-19 to the community from these imported cases was largely reduced by the very strict border control strategies with quarantine and isolation in conjunction with NPIs, such as wearing masks and social distancing [19]. Multimedia Appendix 2 provides details on the criteria and guidelines for the implementation of 4 COVID-19 alert levels to target outbreaks in Taiwan.

Containment Measures for the VOC Phase in Taiwan

The Alpha VOC became the predominant strain of the global pandemic by the end of 2020. Several cluster infections occurred in hospitals and households since January 2021, but were still under control until mid-May 2021, when a large-scale outbreak of the Alpha VOC occurred. On the top of border control measures with quarantine and isolation implemented since the non-VOC phase in Taiwan, the focus of containment measures for averting community-acquired outbreaks turned to community-based active surveillance with rapid test stations for the hotspots of outbreaks and enhanced NPIs, including strict regulation for wearing masks, restriction of public gathering, setting up of check points for high-risk areas such as public transportation sites and markets, and restriction of nonessential services such as restaurants and pubs. Multimedia Appendix 1 summarizes the timeline of the implementation of a series of containment measures for the VOC phase starting from the enhancement of NPIs from level 1 to level 2 alert until high restriction border control for travelers. The level 3 alert was rapidly extended to a nationwide level 3 alert on May 19, 2021 [20,21]. During the Delta VOC period (from August to December 2021) and Omicron VOC period (December 2021 onward), transmission in the community has been threatened by imported cases. In addition to containment measures, high coverage of vaccination has been an effective prevention strategy during these 2 periods. In response to the rapid spread of the Omicron VOC, inbound passengers have to follow updated regulations with more frequent reverse transcription-polymerase chain reaction (RT-PCR) testing plus rapid testing, and a possible mandatory 14-day quarantine based on the vaccination status. Additionally, inbound passengers have to provide negative COVID-19 RT-PCR test reports within

2 days and have to take a government-funded rapid RT-PCR test on arrival starting January 11, 2022. Owing to waning of the effects of vaccines, booster shots have been allowed for all adults who have received 2 vaccine doses for 12 weeks (84 days), since January 7, 2022.

Statistical Analysis

New Surveillance Metrics for Quantifying Imported-Domestic Transmission

We used an extra-Poisson regression model with a Bayesian DAG approach [22] to calculate the expected weekly domestic cluster infections associated with imported cases of COVID-19, as shown in the right panel of Multimedia Appendix 3. For the *j*th county or city with the Y_{tj} domestic case at week *t*, the extra-Poisson regression model can be specified by

$$Y_{jt} \sim \text{Poisson}(\mu_{jt}),$$
$$\log(\mu_{jt}) = \text{offest}_j + \alpha_j + \beta X_{j,t-1}$$
$$\alpha_j \sim \text{Normal}(\alpha_0, \sigma^2_\alpha) \quad \textbf{(1)}$$

where offest_j is the population of log scale, and the heterogeneity of imported-domestic transmission across counties and cities in Taiwan is captured by a normal distributed random intercept parameter, α_j. While the common intercept parameter, α_0, represents the common risk of transmission in Taiwan, the heterogeneity is captured by the variance parameter, σ^2_α. With this framework, the number of cases arising from domestic cluster infection caused by imported cases per week before ($X_{j,t-1}$) can be assessed by using the regression coefficient β, which becomes the first surveillance metric and is denoted as Dci/Imc per week for estimating the effect size of domestic cluster infection. The larger the value of this metric estimated, the larger the domestic cluster infection. The extra variation across cities and counties regarding the transmission of COVID-19 associated with imported cases was captured by a random effect (α_j) incorporated into the Poisson regression model, which is also called the random-effect Poisson regression model. The predicted distribution of the number of expected domestic cases in the next week (μ.pred[t+1]; Multimedia Appendix 3) can be generated by using the number of imported cases in the current week (X[t]; Multimedia Appendix 3) in conjunction with the posterior distribution of the force of transmission (β), standing for the metric of Dci/Imc per week, and the common intercept (α_0) taking into account the county-level heterogeneity of COVID-19 transmission (σ^2_α). The Poisson model has been widely applied to sparse counts of domestic infection, which occur independently if there is a lack of larger cluster infections, with a high potential of developing into a large-scale community-acquired outbreak. If the observed value of our model is beyond the upper limit of the 95% credible interval (CrI), it means that sparse and independent assumptions based on the Poisson distribution are violated and implies a high potential of yielding a large-scale community-acquired outbreak. Accordingly, the second surveillance metric is to build up the alert threshold of emerging community-acquired outbreaks and to provide guidance for the rapid preparedness of containment measures (including effective and efficient contact tracing) for forestalling community-acquired outbreaks.

As mentioned above, data were divided into the non-VOC phase and VOC phase. The former period used for estimating the parameters of the following extra-Poisson regression model was based on imported and domestic cases between January 11 and June 20, 2020, covering the wild-type and D614G period in Taiwan. Because imported cases require an incubation time to generate secondary cases, we tested the lag time of imported cases by 0 weeks (concurrent, X_{jt}), 1 week ($X_{j,t-1}$), and 2 weeks ($X_{j,t-2}$), and further selected the optimal lag time interval with the smallest deviance information criterion (DIC).

Regarding the impact of imported cases on the occurrence of domestic cases for the early Alpha (October 11, 2020, to May 12, 2021), Delta (August 8 to December 9, 2021), and Omicron (December 12, 2021, to April 2, 2022) VOC periods without outbreaks in Taiwan, a Bayesian negative binomial regression model was applied to take into account the heterogeneity across counties and cities associated with the imported-domestic transmission of COVID-19 owing to the lack of detailed information on the cases in counties and cities. Multimedia Appendix 4 shows the DAG model for assessing the force of imported-domestic transmission by using a Bayesian negative binomial regression model. Following the approach applied for the wild-type and D614G period, the 1-week lag model was adopted. For week *t*, the number of cases Y_t resulting from imported cases 1 week prior, X_{t-1}, can be modeled by using the negative binomial regression model as follows:

$$Y_t \sim \text{Negative Binomial}(\mu_t, k),$$
$$\log(\mu_t) = \beta X_{t-1} \quad \textbf{(2)}$$

where the heterogeneity is captured by the dispersion parameter 1/*k*. Similar to the extra-Poisson regression model as above, the risk of imported-domestic transmission can thus be assessed by using the regression coefficient β for estimating the effect size of Dci/Imc. Following the extra-Poisson approach, the predicted distribution of the number of expected domestic cases in the next week (μ.pred[t+1]; Multimedia Appendix 4) for the Bayesian negative binomial model can be generated from the current number of imported cases (X[t]; Multimedia Appendix 4) by using the posterior distribution of imported-domestic transmission (β) and the dispersion parameter (1/*k*).

Estimation With the Bayesian MCMC Method

The Bayesian MCMC method was used to generate the samples derived from the posterior distributions of parameters for estimating 2 surveillance metrics. With the Markov chain underpinning, a stationary distribution for parameters can be reached in the long run under regular conditions. Independent samples can thus be generated from such a stationary posterior distribution on the basis of which inferences can be made [23]. The DAG models depicted in Multimedia Appendix 3 and Multimedia Appendix 4 were applied to facilitate the decomposition of joint distribution into full conditional density distribution by using the relationship between parent and child nodes [24]. Taking the extra-Poisson regression model as an example, the joint distribution,

$$P(Y, \mu, \alpha, \alpha_0, \sigma^2_\alpha, \beta) \quad (3)$$

is proportional to the product of the kernel distribution written by

$$L(Y, \mu, \alpha, \alpha_0, \sigma^2_\alpha, \beta)=P(Y \mid \mu)P(\mu \mid \alpha, \beta)P(\alpha \mid \alpha_0, \sigma^2_\alpha)P(\beta)P(\alpha_0)P(\sigma^2_\alpha) \quad (4)$$

In our application, noninformative priors were used to derive the samples from the stationary posterior distribution of parameters, including the risk of imported-domestic transmission (β), the common intercept (α_0), and the county-specific random effect (σ_α).

A block-wise Metropolic-Hasting sampler was applied to generate samples from the stationary posterior distribution. The sampling algorithm is detailed as follows:

1. Start with an initial value ($\beta^{(0)}$, $\alpha^{(0)}$, $\alpha_0{}^{(0)}$, $\sigma^2{}_\alpha{}^{(0)}$) selected from the support of each parameter.

2. Draw the candidate value for the first parameter, say $\beta^{(1)}$, from a normal proposal distribution, $q(\beta)$.

3. Compute the acceptance probability

4. Draw u from uniform (0,1) and update $\beta^{(0)}$ with $\beta^{(1)}$ if $u < r(\beta^{(1)}, \beta^{(0)} \mid \alpha^{(0)}, \alpha_0{}^{(0)}, \sigma^2{}_\alpha{}^{(0)})$; otherwise, repeat steps 2 and 3.

5. Draw the candidate value for the next parameter, $\alpha^{(1)}$, to update the parameter sample with ($\alpha^{(0)} \mid \alpha_0{}^{(0)}, \sigma^2{}_\alpha{}^{(0)}, \beta^{(1)}$) by using steps 2 to 4.

6. Repeat steps 2 to 5 for the rest of the parameters, (α_0, σ^2_α), to derive ($\beta^{(1)}$, $\alpha^{(1)}$, $\alpha_0{}^{(1)}$, $\sigma^2{}_\alpha{}^{(1)}$) to complete an iteration of the update for parameter samples.

Thinning intervals of 10 and 100,000 iterations were used to generate the 10,000 posterior samples after 250,000 burn-in iterations by using the Bayesian MCMC methods mentioned above.

We estimated the effect size of Dci/Imc per week for each period corresponding to the type of SARS-CoV-2 variant. We built up the alert threshold by using the upper limit of the 95% CrI of the predicted number of domestic cases (μ.pred[t+1]; Multimedia Appendix 3 and Multimedia Appendix 4) generated by the parameters after updating the data on the non-VOC phase in Taiwan for alerting the possibility of yielding a large-scale community-acquired outbreak through imported-domestic transmission in the subsequent epochs. The possibility of a community-acquired outbreak was deemed low if observed domestic cases were not more than the alert threshold, namely the upper limit of the 95% CrI. Otherwise, an outbreak was likely to occur, and therefore, the rapid preparedness of containment measures, including effective and efficient contact tracing, would be flagged to forestall the ensuing community-acquired outbreak.

To validate the proposed surveillance model for the transmission from imported to domestic cases during the non-VOC period in Taiwan, the publicly available COVID-19 data provided by the Ministry of Health in New Zealand were used [25]. The chronological order of the incidence of COVID-19 for the hotspots was compared to validate the epidemic surveillance model for an outbreak.

All statistical analyses were performed using SAS 9.4 software (SAS Institute Inc).

Results

Evaluation of the Optimal Time Lag Model

After the application of the Bayesian MCMC method for the identification of the optimal time (in weeks) lag of imported cases prior to cases arising from domestic cluster infection, the 1-week lag of imported cases yielded a DIC of 255.8, which was smaller than the DICs of the model with concurrent week imported cases (260.3) and the model with a 2-week lag of imported cases (279.5).

Surveillance Metrics for the Imported-Domestic Transmission Mode

The weekly observed number (red dot) and expected number (green circle) of domestic cases are shown in Figure 2 (wild-type and D614G period, January to September 2020), Figure 3 (Alpha VOC period, October 2020 to May 2021), and Figure 4 (Delta VOC period, mid-August to mid-December 2021; and Omicron VOC period, mid-December 2021 to early-April 2022). Table 1 shows the details of the estimated results of the parameters encoded in the Bayesian extra-Poisson regression model with a 1-week lag of imported cases regarding the 3 periods without outbreaks in Taiwan, namely the wild-type and D614G period, early Alpha VOC period, and Delta VOC period.

The upper bound of the 95% CrI of expected cases (dotted line, Figures 2-4) has been plotted to provide the alert threshold of domestic cluster infection in the community caused by transmission from imported cases 1 week before. This 1-week prior alert on the risk of elevated Dci/Imc per week guided the vigilance on NPIs for averting further community-acquired outbreaks.

Figure 2. Number of observed (dotted point) and expected (green circle) domestic cases with the upper limit of the 95% credible interval (CrI) (dotted line) by week in the wild-type/D614G period.

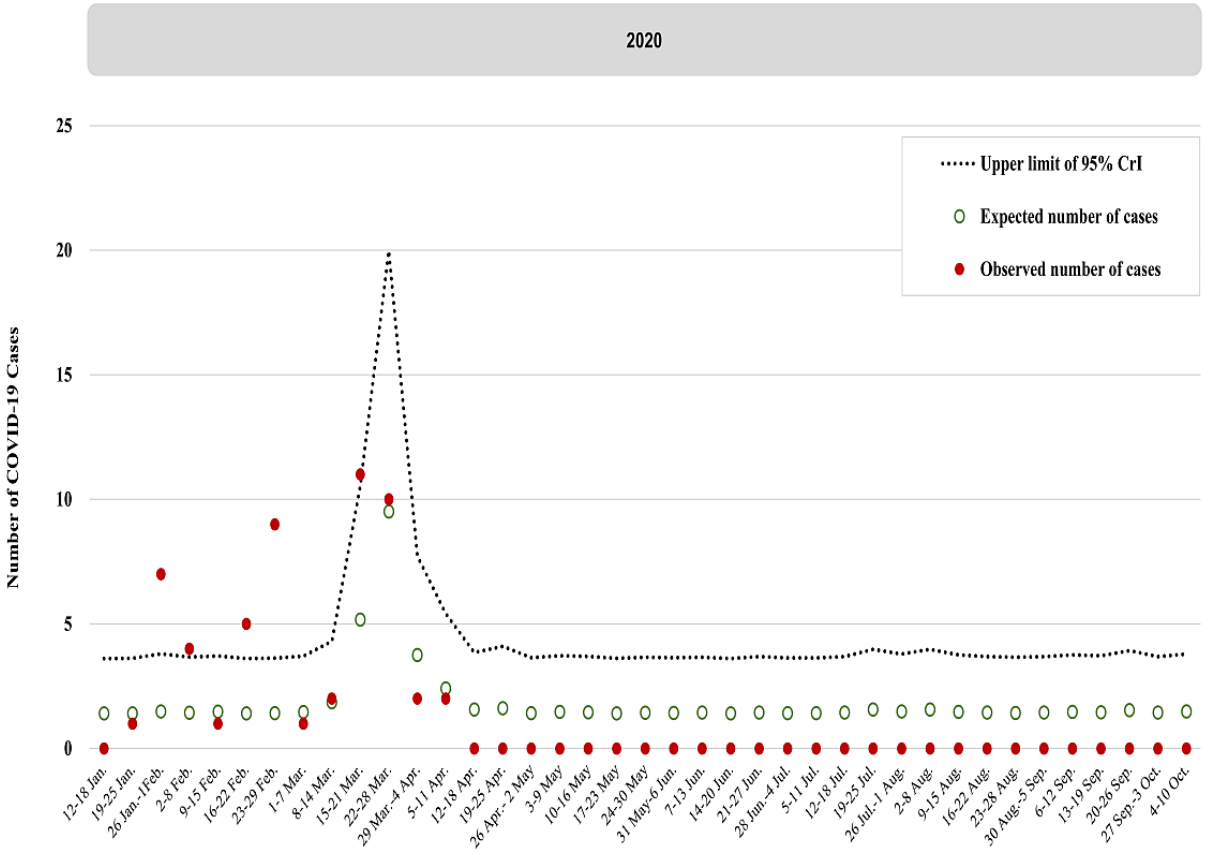


Figure 3. Number of observed (dotted point) and expected (green circle) domestic cases with the upper limit of the 95% credible interval (CrI) (dotted line) by week in the Alpha variant of concern period.

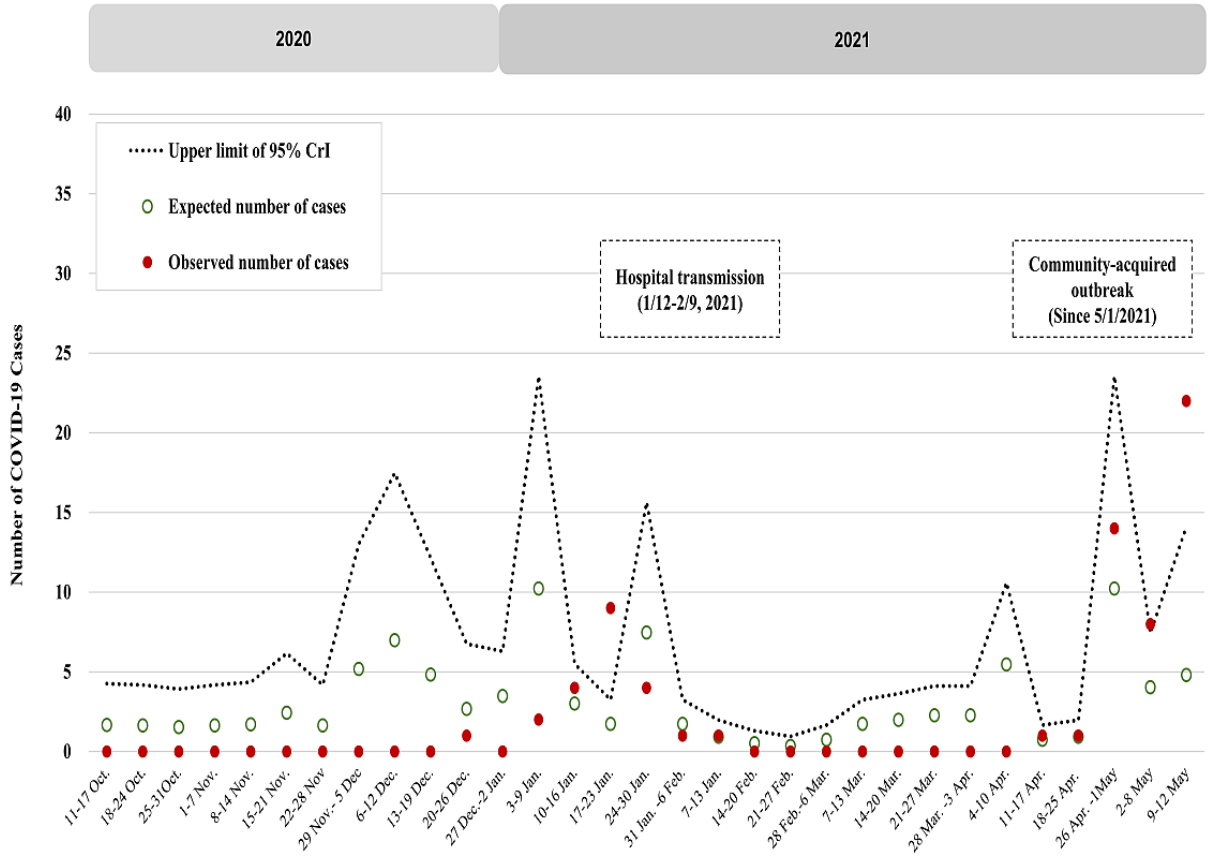


Figure 4. Number of observed (dotted point) and expected (green circle) domestic cases with the upper limit of the 95% credible interval (CrI) (dotted line) by week in the Delta (August 8, 2021, to December 9, 2021) and Omicron (December 12, 2021, to April 2, 2022) variant of concern periods.

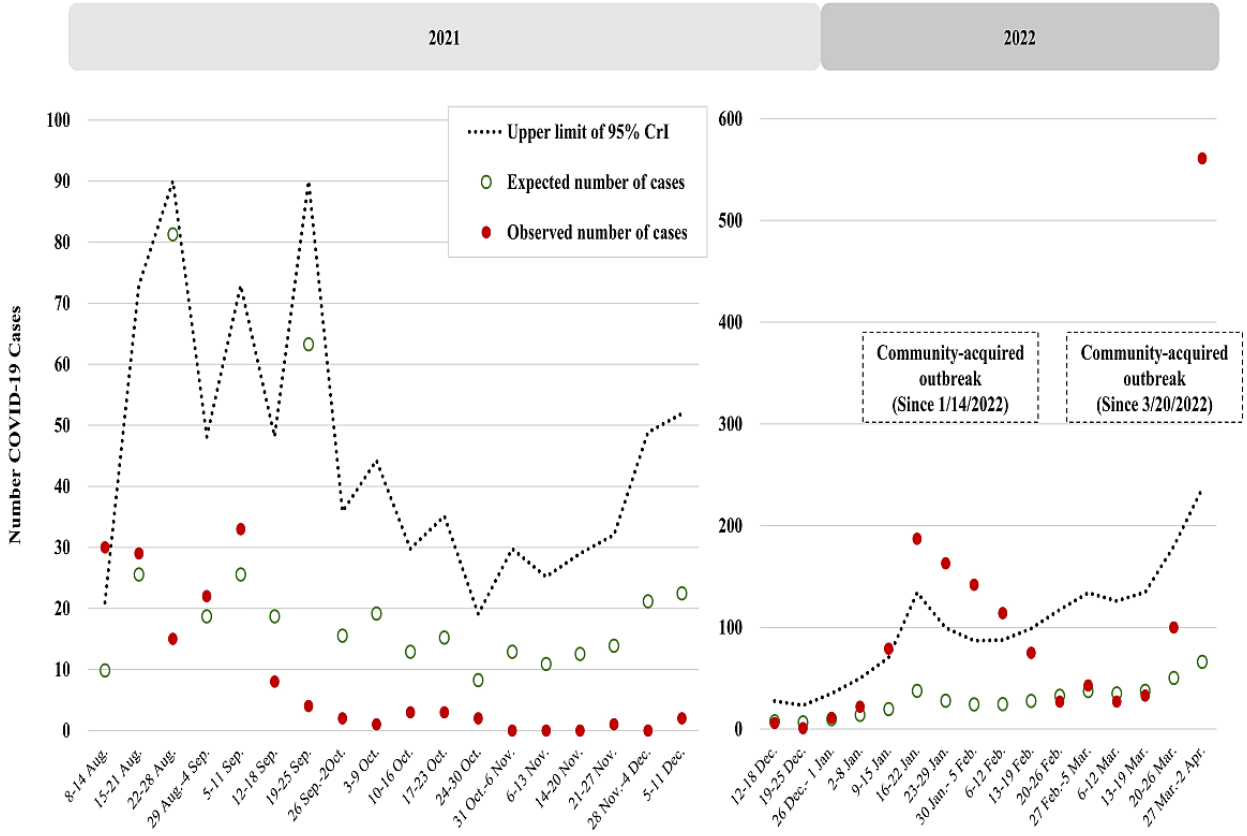


Table 1. Estimated results for the risk of imported-domestic transmission of COVID-19 for 3 periods in Taiwan.

Parameter	Estimate	95% CI
Wild-type/D614G period (January to September 2020)		
Common intercept	−3.5457	−5.1978 to −2.4413
Risk of imported-domestic transmission	0.0954	0.0644 to 0.1259
Standard error of random intercept, σ_v	1.8116	0.9861 to 3.7359
Alpha VOC[a] period (October 2020 to May 2021)		
Intercept	−1.9448	−4.1238 to 0.0712
Risk of imported-domestic transmission	0.1414	0.0541 to 0.2510
Dispersion parameter	1.7438	0.2734 to 3.8218
Delta VOC period (August to December 2021)		
Risk of imported-domestic transmission	0.1005	0.0685 to 0.1332
Dispersion parameter	1.5454	0.3954 to 3.3562
Omicron VOC period		
December 2021 to January 2022		
Risk of imported-domestic transmission	0.0459	0.0366 to 0.0575
Dispersion parameter	1.0054	0.4192 to 2.2245
February to April 2022		
Risk of imported-domestic transmission	0.0429	0.0352 to 0.0519
Dispersion parameter	1.5969	0.8484 to 2.8928

[a]VOC: variant of concern.

Wild-Type and D614G Period

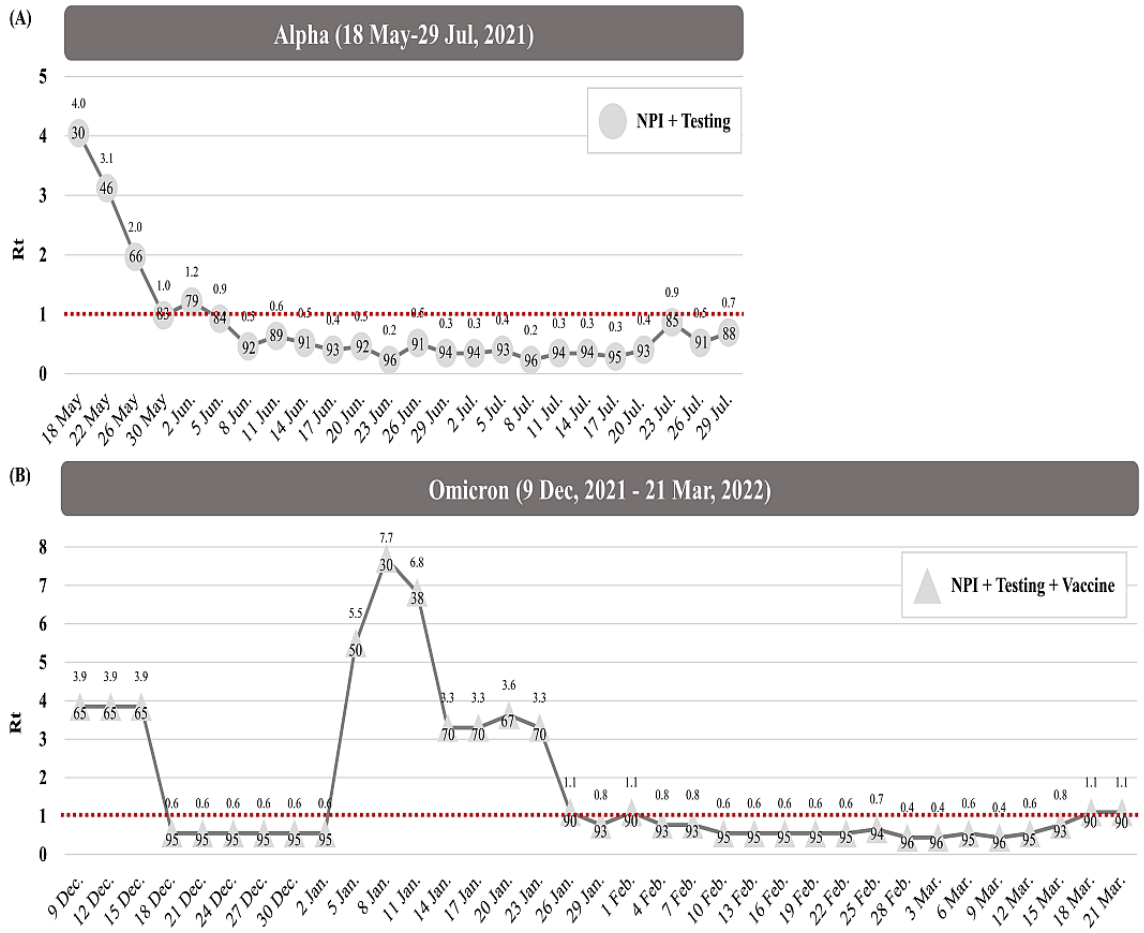
During the wild-type/D614G period, the estimated Dci/Imc per week was 9.54% (95% CrI 6.44%-12.59%; Table 1). Figure 2 shows that there were 5 weeks (January 26 to February 1, February 2 to February 8, February 16 to February 22, February 23 to February 29, and March 15 to March 21, 2020) in which the observed numbers of domestic cases exceeded the alert thresholds. This period yielded 81% of clustered cases (22 of 27 community-acquired cases) in 5 clusters in Taiwan, including 3 household clusters (with 5, 3, and 6 COVID-19 cases, respectively), 1 medical institute cluster (with 9 COVID-19 cases), and 1 academic institute cluster (with 4 COVID-19 cases). Guidance of the alert thresholds from this early period of the wild-type COVID-19 strain provided a strong rationale for being on alert for the ensuing cluster infections from the preceding 1 week when imported cases were introduced. This accounted for why none of these 5 cluster events led to any large-scale community-acquired outbreaks in Taiwan. There was rapid preparedness of containment measures with strict NPIs together with effective and efficient contact tracing of all possible susceptible individuals.

Figure 2 shows that this surveillance metric was very useful, particularly when there was a substantial surge in imported cases resulting from the large-scale COVID-19 pandemic worldwide. This could be seen in our cases between March and April 2020, as shown in Figure 1. Again, the surveillance metric was used for alerting about possible cluster infections to forestall further community-acquired outbreaks. Alerted by the threshold (20 domestic cases per week), the observed domestic cases were kept lower than the alert threshold to avoid large-scale outbreaks in April. Since then, there had not been any domestic case until December 2020.

Alpha VOC Period

There had been no outbreak during the early Alpha VOC phase of the COVID-19 pandemic from October to December 2020. The second surge of imported cases occurred from January 2021 onwards (Figure 1). Again, there was 1 week (January 17 to January 23, 2021) in which the observed number of domestic cases was beyond the alert threshold, resulting from hospital-based cluster infections and 3 subsequent household clusters (11 family members) (Figure 3). The source of this cluster infection was later identified as an imported case infected with the Alpha VOC. After being alerted by the proposed surveillance metric and following timely contact tracing and containment measures, including quarantine and isolation for all staff members in the hospital and their close contacts for 14 days, there was no further outbreak until early May 2021, when the number of observed domestic cases reached beyond 20, which was higher than the expected surveillance curve (5 cases) and the corresponding upper bound of the 95% CrI (14 cases). The estimated results on the basis of the empirical data during this Alpha VOC period showed that the Dci/Imc increased to 14.14% (95% CrI 5.41%-25.10%; Table 1). The peak of domestic cases far beyond the threshold value in early May 2021 not only presaged the ensuing outbreaks, but also revealed the loose NPIs at that time without an available vaccine in Taiwan. If effective contact tracing and timely containment measures had been deployed in advance on the basis of the increased risk and the alert threshold in the week between May 2 and May 8, the ensuing community-acquired outbreak involving the Alpha VOC could have been prevented (bottom panel of Figure 1). This outbreak lasted for 2.5 months and subsided in July 2021. The corresponding periods of the estimated results for the effectiveness of NPIs and testing implemented in Taiwan during the Alpha VOC phase are shown in Figure 5A.

Figure 5. Effective reproductive number (Rt) and effectiveness of nonpharmaceutical interventions (NPIs) and testing in Taiwan. (A) Alpha variant. (B) Omicron variant.



Delta and Omicron VOC Period

After controlling the outbreak in the Alpha VOC period, there was monitoring of domestic cluster infections for the Delta VOC since August 8, 2021, and the Omicron VOC from December 10, 2021. After mid-August, there had been a cluster infection during the Delta VOC phase of the COVID-19 pandemic. The estimated results (Table 1, Delta VOC period) showed that the Dci/Imc remained at 10.01% (95% CrI 6.85%-13.32%), although the Delta VOC had higher transmissibility and escaped vaccine-induced immunity. The left panel of Figure 4 shows that for 2 weeks (August 22 to August 28, 2021, and September 19 to September 25, 2021) when a surge in the number of domestic cases as a result of Dci/Imc was expected (green circle), the observed number of domestic cases was far below the threshold of outbreak. This can be attributed to the implementation of enhanced containment measures, including the strengthening of border control strategies with multiple tests on arrival and during quarantine, the collective quarantine strategy, and the elevated alerts of NPIs to levels 2 and 3 since the outbreak in May 2021 in Taiwan. Since then, the weekly observed cases were below the alert threshold owing to NPIs with a level 2 alert and the rapid administration of vaccines.

The right panel of Figure 4 shows the observed and predicted numbers of cases along with the alert threshold during the Omicron VOC period. Given the increasing coverage rate of vaccination, the risk of domestic cluster infection per imported case for the Omicron BA.1 VOC reduced to 4.59% (95% CrI 3.66%-5.75%), but an upsurge in domestic cases was still observed because the Omicron BA.1 VOC was considered to have a high transmission probability. In the week from January 9 to January 15, 2022, the observed number exceeded the alert threshold, indicating a high potential for a community-acquired outbreak. There was indeed a small-scale community-acquired outbreak of the Omicron BA.1 VOC. After a series of containment measures, including rapid RT-PCR tests for inbound passengers on arrival coupled with stringent quarantine and isolation, rapid booster vaccination, and enhanced NPIs with a level 2 alert in the community, the community-acquired outbreak subsided by the end of February 2022 (Figure 5).

There was a return to the imported-domestic transmission model, with the surveillance metric of Dci/Imc estimated at 4.29% (95% CrI 3.52%-5.19%) from February until March 20, 2022. After that, a similar circumstance beyond the alert threshold was noted for the invasion of the imported Omicron BA.2 VOC, and a community-acquired outbreak started from March 20 to 26, 2022, resulting in a large-scale community-acquired outbreak from early April until July 2022.

External Validation of the Surveillance Metrics for Domestic Cluster Infections Using Imported Cases in New Zealand

To validate the proposed model and extend its application to different periods of SARS-CoV-2 variants, the proposed extra-Poisson regression model was applied to data on the New Zealand COVID-19 outbreak in 2020. Multimedia Appendix 5 shows the estimated results obtained. Notably, in New Zealand, the risk of Dci/Imc per week increased to 9.38% (95% CrI 8.88%-9.86%), which was close to the estimated results based on Taiwan data (9.54%, 95% CrI 6.44%-12.59%) in the same period. Details regarding the spatial temporal distribution of COVID-19 outbreaks by types of cases in New Zealand are provided in Multimedia Appendix 6. Multimedia Appendix 7 shows the predicted number of domestic cases by using the parameters trained from the empirical data of New Zealand (Multimedia Appendix 5). Similar to the application in Taiwan, the risk of an outbreak associated with imported cases could be assessed by comparing the observed cases (red dot in Multimedia Appendix 7) with the alert threshold (dotted line in Multimedia Appendix 7). The detailed interpretation of the results of this external validation is elaborated in Multimedia Appendix 8.

Discussion

Many cyclical community-acquired outbreaks in each country or region during the COVID-19 pandemic have been noted from 2020 to 2022, and these epidemics have occurred in parallel with the evolution of various emerging SARS-CoV-2 variants, including the wild-type/D614G strain during the non-VOC phase and the Alpha, Beta, Gamma, Delta, and Omicron strains during the VOC phase. More importantly, when facing emerging variants, there were corresponding chains of containment measures with the following 3 serial steps: (1) border control of imported cases together with quarantine and isolation; (2) contact tracing and epidemic investigation of domestic cluster infection together with testing for detecting the foci of infection earlier (small households to large institutions); and (3) control of large-scale community-acquired infection with population-based approaches, mainly involving NPIs and mass vaccination. Although most countries focus on steps 1 and 2 for averting community-acquired outbreaks in the beginning, they end up having no choice but to adopt step 3 involving population-based approaches. Accordingly, most epidemic surveillance models still follow traditional surveillance metrics like the effective basic reproductive number (R_t) for containing community-acquired outbreaks. However, it is still important to develop a new surveillance model with new surveillance metrics commensurate with steps 1 and 2 for forestalling community-acquired outbreaks when facing emerging SARS-CoV-2 variants like the updated Omicron subvariants BA.4/BA.5. Most importantly, such a new surveillance model with useful metrics can be robustly applied across countries and time, covering various emerging SARS-CoV-2 variants detected from imported cases.

Using Taiwan empirical data, the proposed new surveillance model for monitoring cluster infections in the wake of imported cases was assessed in 5 periods (wild-type, D614G, Alpha VOC, Delta VOC, and Omicron VOC). The first metric of Dci/Imc per week was used to estimate the effect size of the risk of infection through imported-domestic transmission. The second metric involving the upper bound of the 95% CrI for predicted domestic cases of cluster infections derived from imported cases 1 week before provided the alert threshold for guiding the preparedness of containment measures for preventing community-acquired outbreaks in each country or region. Such an alert threshold would be affected by the characteristics of each emerging SARS-CoV-2 variant, as well as the underlying coverage rate of vaccination and the extent of NPIs. By using empirical data on imported and domestic COVID-19 cases in Taiwan, the effect size of Dci/Imc and the alert threshold were estimated as 9.54% and 12.59% for the wild-type/D614G strain, 14.14% and 25.10% for the Alpha VOC, 10.05% and 13.32% for the Delta VOC, 4.59% and 5.75% for the Omicron BA.1 VOC, and 4.29% and 5.19% for the Omicron BA.2 VOC, respectively, in 2 periods. It should be noted that the interpretation of the absolute effect size of Dci/Imc across various emerging SARS-CoV-2 variants should be taken with great caution.

When a similar logic is applied to the alert threshold, the threshold value for weekly domestic cases during 3 SARS-CoV-2 variant periods would not go beyond 20 cases under the low coverage rate of vaccination and good performance of NPIs before the Omicron VOC period. After the Omicron VOC period, the alert threshold for weekly domestic cases would be 180 cases under the high vaccination rate and minimal NPIs. Empirical evidence on whether and how community-acquired outbreaks can be averted through different periods of SARS-CoV-2 variants with the proposed new surveillance model has been demonstrated by Taiwan data. Outbreaks were averted during the wild-type and D614G periods. In contrast, large-scale outbreaks could not be averted during the Alpha VOC period when the expected number of domestic cases was far beyond the alert threshold for an outbreak between May 9 and May 12, 2021, because of the increased transmissibility of the Alpha VOC that was supported by the increased risk of imported-domestic transmission in comparison with the wild-type/D614G variant (14.14% vs 9.54%). However, a low level of NPIs might also have contributed to such an outbreak around mid-May 2021 (30%; Figure 5). In the Delta VOC period after excluding the outbreak related to the Alpha VOC in Taiwan, the level of the NPI alert and the strict border control strategies implemented since the outbreak period reduced the risk of imported-domestic transmission to 10.05% (Table 1) and averted a community-acquired outbreak of the Delta VOC.

Given the high coverage of full vaccination, lower estimates of Dci/Imc for the Omicron VOC were seen, ranging from 4.3% to 4.6%. As the protective effect of the Oxford/AstraZeneca vaccine in particular started to wane in the community, the observed number of domestic cases was beyond the threshold of outbreak during the early Omicron BA.1 VOC period. Guided by the alert threshold, several containment strategies, including more restricted border control and rapid RT-PCR testing on arrival for travelers, and rapid booster shots for eligible adults,

were implemented to avert a large-scale outbreak. However, another large-scale community-acquired outbreak could not be averted in late March 2022 owing to the observed cases going beyond the alert threshold partially due to waning of the protective effect of the mRNA-1273 vaccine (Moderna) or BNT162b2 vaccine (Pfizer–BioNTech).

Although our metrics of the risk for domestic cluster infection and the alert threshold are pivotal in imported cases 1 week before applying the surveillance model for monitoring imported cases 1 week prior to the formation of cluster infections of domestic COVID-19 cases, they are very useful for alerting the surrounding community in proximity to imported cases beyond the threshold of the upper bound of the 95% CrI to enhance NPIs and active rapid testing with effective contact tracing and epidemic investigation for the observed cases. This accounted for the lack of community-acquired outbreaks before mid-May 2021 in Taiwan. Such good control over the COVID-19 epidemic has been reported in previous studies by evaluating NPIs at the individual and population levels [5,19,26], and using the traditional surveillance model for assessing the duration from R_t larger than 1 to R_t smaller than 1 and the case load following the machining learning model [27].

Several previous studies have proposed an early warning model in relation to contact tracing and epidemic investigation before community-acquired outbreaks. However, our study differs from 2 recent previous studies [15,16] that developed an early warning model of COVID-19 outbreaks, in 2 main aspects. First, both studies covered a short period that reflected 1 or 2 SARS-CoV-2 strains and used data based on a single country. They were therefore unable to test the robustness of their models for a series of SARS-CoV-2 variants and samples across countries. Guan et al used human mobility data in Israel over the period from February 1, 2020, to January 7, 2021. Specifically, they trained their model over the period from April 6 to October 24, 2020, and evaluated the model's predictive ability over 2 very short periods (November 1-30, 2020, and December 1-31, 2021) [15]. Kogan et al employed data in the United States that had been obtained from multiple digital traces over 2 short periods (March 1-May 31, 2020, and June 1-September 30, 2020) [16]. In contrast, the proposed new surveillance model made use of the full chronological empirical data in Taiwan from January 1, 2020, to April 2, 2022, covering various emerging SARS-CoV-2 variants. Our model is robust across a long period involving various variants and across countries covering different geographical and cultural conditions when using imported cases. Second, the data derived from digital traces, for example, Google Trends, used in the studies by Guan et al and Kogan et al may be affected by media activities. Furthermore, it is not easy to obtain accurate real-time data in countries with unavailable technological infrastructure, strong information censoring, and a lack of transparency. Instead of focusing on digital traces, we made use of imported cases that may be less likely to be affected by confounding factors. The use of imported data in our proposed surveillance model can be applied to countries with different political and social conditions and at different technological development stages. Moreover, our study is highly relevant to health regulators and public health policy makers, particularly in countries that have opened their borders and eased/removed NPI measures.

In addition to the illustration of the Taiwan experience, the external validation involving New Zealand further adds credibility to the application of this surveillance model in a scenario without an outbreak. This model can also be applied to those countries with controllable community-acquired outbreaks, such as Israel [28] and Qatar [29], after the mass vaccination program since early 2021, to monitor the impact of imported cases on the risk of domestic cluster infection. This is especially important for outbreaks resulting from vaccine breakthrough in countries or regions with high vaccine coverage, such as Singapore [30] and Israel [31], or the rapid waning of booster effectiveness worldwide, possibly affecting the community-level spread of SARS-CoV-2 VOCs [32].

One of the major limitations of the current model pertains to the generalization of the proposed new surveillance model. There are 2 major circumstances that require the refinement of the current proposed epidemic surveillance model. The proposed model has not incorporated health care capacity for accommodating the threshold of tolerable COVID-19 cases responsible for each episode of the outbreak. In consideration of resuming prepandemic activity, making allowance for this factor is of paramount importance for the implementation of NPIs and testing given the vaccine coverage rate. Different countries and regions may require different outbreak thresholds based on this global surveillance model. With increasing cases of vaccine breakthrough; the rapid emergence of VOCs with a wide spectrum of immunogenicity, high transmissibility, and resistance to antibodies associated with natural infection or vaccination; and the waning of immunity in the population, there is a high likelihood for the continued spread of SARS-CoV-2 in the population [33]. Given the possibility of a long-term association between SARS-CoV-2 and the human population, the goal of epidemic surveillance may shift from the elimination of this pathogen to a balance among health care capacity, socioeconomic activity, and population immunity. If this occurs, the proposed surveillance model should take this factor into account and should be used as a guide to inform the containment measures required to mitigate large-scale outbreaks according to health care capacity. Moreover, as the border control policy on quarantine and isolation of imported cases gets altered with the advent of high-performance rapid testing and the gradual expansion of vaccine coverage worldwide, the surveillance model for monitoring imported-domestic transmission to avert outbreaks may vary from country to country, depending on the extent of NPIs, administration of tests, coverage rate of vaccines, and administration of vaccine boosters. Such a heterogeneity should be taken into account to refine the surveillance model on imported-domestic transmission when it is applied to avert a large-scale outbreak. More importantly, our new surveillance model and metrics are not meant to replace conventional surveillance and corresponding metrics like R_t for assessing how to eliminate the spread of large-scale community-acquired outbreaks. When a community-acquired outbreak occurs, the conventional surveillance SEIR model is needed to assess the effectiveness of containment measures, as shown in Figure 5. Based on the

SEIR model, the R_t decreased from 4.0 to 0.7 from May 18, 2021, to July 31, 2021. The effectiveness of NPIs and testing, which reflects the strategies implemented 2 weeks ago, was over 60% after May 26, 2021, and increased to over 90% after June 14, 2021. A similar finding was noted for a community-acquired outbreak of Omicron BA.2. Again, R_t reduced from 7.7 (value of R_t in early January) to less than 1 (around the end of January; Figure 5B). Our proposed new surveillance model has a supplementary role as a global vigilance method for forestalling large-scale local community-acquired outbreaks of emerging SARS-CoV-2 VOCs in each country and region worldwide.

In conclusion, a global new surveillance model and metrics have been proposed for monitoring imported cases of SARS-CoV-2 variants from the non-VOC phase to the VOC phase, using the Taiwan scenario. The new surveillance model and metrics are very useful for forestalling a new large-scale community-acquired outbreak through monitoring of the imported-domestic transmission mode associated with emerging infectious diseases in the future.

Acknowledgments

The authors received financial support from the Ministry of Science and Technology, Taiwan (MOST 111-2118-M-002-004-MY2; MOST 111-2118-M-038-001-MY2; MOST 111-2118-M-038-002-MY2; and MOST 111-2321-B-002-017).

Authors' Contributions

AMY, THC, and SLC conceptualized and designed the study. AMY was responsible for data analysis and drafting of the manuscript. WJC, TYL, GHJ, and CYH were responsible for statistical analysis. WJC, TYL, and GHJ were in charge of data collection and management. THC, CYH, STW, HD, and SLC interpreted the results and revised the manuscript. All authors agreed with the findings and provided input on the revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Timelines of the SARS-CoV-2 variant of concern outbreak and the implementation of key containment measures in Taiwan.
[DOCX File , 432 KB - publichealth_v8i11e40866_app1.docx]

Multimedia Appendix 2
Criteria and guidelines for the 4 COVID-19 alert levels in Taiwan.
[DOCX File , 18 KB - publichealth_v8i11e40866_app2.docx]

Multimedia Appendix 3
Directed acyclic graphic model of the Bayesian extra-Poisson regression model for assessing the force of imported-domestic transmission.
[DOCX File , 75 KB - publichealth_v8i11e40866_app3.docx]

Multimedia Appendix 4
Directed acyclic graphic model of the Bayesian negative binomial model for assessing the force of imported-domestic transmission.
[DOCX File , 53 KB - publichealth_v8i11e40866_app4.docx]

Multimedia Appendix 5
Estimated results for the risk of the imported-domestic transmission of COVID-19 in New Zealand with the consideration of heterogeneity across counties using the Poisson model.
[DOCX File , 17 KB - publichealth_v8i11e40866_app5.docx]

Multimedia Appendix 6
Epidemic curve of the COVID-19 outbreak in New Zealand by types of cases.
[DOCX File , 43 KB - publichealth_v8i11e40866_app6.docx]

Multimedia Appendix 7
Number of observed (blue dot) and expected (green circle) domestic cases with the upper limit of the 95% credible interval (dotted line) by week.
[DOCX File , 32 KB - publichealth_v8i11e40866_app7.docx]

Multimedia Appendix 8
Surveillance metrics for domestic cluster infections using imported cases in New Zealand.
[DOCX File , 17 KB - publichealth_v8i11e40866_app8.docx]

References

1. Chen SL, Yen AM, Lai C, Hsu C, Chan C, Chen TH. An index for lifting social distancing during the COVID-19 pandemic: Algorithm recommendation for lifting social distancing. J Med Internet Res 2020 Sep 17;22(9):e22469 [FREE Full text] [doi: 10.2196/22469] [Medline: 32886622]
2. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, CMMID COVID-19 Working Group, COVID-19 Genomics UK (COG-UK) Consortium, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science 2021 Apr 09;372(6538):eabg3055 [FREE Full text] [doi: 10.1126/science.abg3055] [Medline: 33658326]
3. Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. The Lancet Infectious Diseases 2021 Feb;21(2):193-202. [doi: 10.1016/s1473-3099(20)30785-4]
4. Riley S, Ainslie KEC, Eales O, Walters CE, Wang H, Atchison C, et al. Resurgence of SARS-CoV-2: Detection by community viral surveillance. Science 2021 May 28;372(6545):990-995 [FREE Full text] [doi: 10.1126/science.abf0874] [Medline: 33893241]
5. Chen C, Lai C, Luh D, Chuang S, Yang K, Yeh Y, et al. Review of epidemic, containment strategies, clinical management, and economic evaluation of COVID-19 pandemic. J Formos Med Assoc 2021 Jun;120 Suppl 1:S6-S18 [FREE Full text] [doi: 10.1016/j.jfma.2021.05.022] [Medline: 34116896]
6. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
7. Koelle K, Martin MA, Antia R, Lopman B, Dean NE. The changing epidemiology of SARS-CoV-2. Science 2022 Mar 11;375(6585):1116-1121 [FREE Full text] [doi: 10.1126/science.abm4915] [Medline: 35271324]
8. Brown CM, Vostok J, Johnson H, Burns M, Gharpure R, Sami S, et al. Outbreak of SARS-CoV-2 infections, including COVID-19 vaccine breakthrough infections, associated with large public gatherings - Barnstable County, Massachusetts, July 2021. MMWR Morb Mortal Wkly Rep 2021 Aug 06;70(31):1059-1062 [FREE Full text] [doi: 10.15585/mmwr.mm7031e2] [Medline: 34351882]
9. Ku M, Huang L, Chiu SY, Wang W, Jeng Y, Yen M, et al. Continental transmission of emerging COVID-19 on the 38° north latitude. J Formos Med Assoc 2021 Jun;120 Suppl 1:S19-S25 [FREE Full text] [doi: 10.1016/j.jfma.2021.05.008] [Medline: 34112588]
10. Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Euro Surveill 2021 Jun;26(24):2100509 [FREE Full text] [doi: 10.2807/1560-7917.ES.2021.26.24.2100509] [Medline: 34142653]
11. Krutikov M, Hayward A, Shallcross L. Spread of a variant SARS-CoV-2 in long-term care facilities in England. N Engl J Med 2021 Apr 29;384(17):1671-1673. [doi: 10.1056/nejmc2035906]
12. Scudellari M. How the coronavirus infects cells - and why Delta is so dangerous. Nature 2021 Jul;595(7869):640-644. [doi: 10.1038/d41586-021-02039-y] [Medline: 34321669]
13. Huisman J, Scire J, Angst D, Li J, Neher R, Maathuis M, et al. Estimation and worldwide monitoring of the effective reproductive number of SARS-CoV-2. eLife 2022;11:e71345. [doi: 10.7554/elife.71345]
14. Hodcroft E, Zuber M, Nadeau S, Vaughan T, Crawford K, Althaus C, SeqCOVID-SPAIN consortium, et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. Nature 2021 Jul 24;595(7869):707-712 [FREE Full text] [doi: 10.1038/s41586-021-03677-y] [Medline: 34098568]
15. Guan G, Dery Y, Yechezkel M, Ben-Gal I, Yamin D, Brandeau ML. Early detection of COVID-19 outbreaks using human mobility data. PLoS One 2021 Jul 20;16(7):e0253865 [FREE Full text] [doi: 10.1371/journal.pone.0253865] [Medline: 34283839]
16. Kogan NE, Clemente L, Liautaud P, Kaashoek J, Link NB, Nguyen AT, et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. Sci Adv 2021 Mar 05;7(10):eabd6989 [FREE Full text] [doi: 10.1126/sciadv.abd6989] [Medline: 33674304]
17. Severe Pneumonia with Novel Pathogens (COVID-19). Taiwan National Infectious Disease Statistics System. URL: https://nidss.cdc.gov.tw/en/nndss/disease?id=19CoV [accessed 2021-10-01]
18. Population and 5-year age group for counties and cities. Dept. of Household Registration, Ministry of the Interior, Republic of China (Taiwan). URL: https://www.ris.gov.tw/app/en/3910 [accessed 2021-10-01]
19. Hsu C, Wang J, Huang K, Fan AC, Yeh Y, Chen SL. Household transmission but without the community-acquired outbreak of COVID-19 in Taiwan. J Formos Med Assoc 2021 Jun;120 Suppl 1:S38-S45 [FREE Full text] [doi: 10.1016/j.jfma.2021.04.021] [Medline: 33994234]
20. Press Releases. Taiwan Centers for Disease Control. URL: https://www.cdc.gov.tw/En/Bulletin/List/7tUXjTBf6paRvrhEl-mrPg [accessed 2022-11-05]

21. Crucial policies for combating COVID-19. Ministry of Health and Welfare, Taiwan. URL: https://covid19.mohw.gov.tw/ch/sp-timeline0-205.html?fbclid=IwAR1eeqICUPFQdPuQkG7z3hIKYAwDaaLnigzqHR6xFXOVlZn0BcXT0dNhfkI [accessed 2021-10-01]
22. Congdon P. Bayesian Statistical Modelling. Hoboken, New Jersey: John Wiley & Sons, Ltd; 2006.
23. Cox DR. The Theory of Stochastic Processes. New York, New York: Routledge; 1965.
24. Lauritzen SL, Spiegelhalter DJ. Local computations with probabilities on graphical structurestheir application to expert systems. Journal of the Royal Statistical Society. Series B (Methodological) 1988;50(2):157-224 [FREE Full text]
25. COVID-19: Case demographics. Ministry of Health, New Zealand. URL: https://www.health.govt.nz/our-work/diseases-and-conditions/covid-19-novel-coronavirus/covid-19-current-situation/covid-19-current-cases/covid-19-current-cases-details#download [accessed 2021-09-05]
26. Lin T, Liao S, Lai C, Paci E, Chuang S. Effectiveness of non-pharmaceutical interventions and vaccine for containing the spread of COVID-19: Three illustrations before and after vaccination periods. J Formos Med Assoc 2021 Jun;120 Suppl 1:S46-S56 [FREE Full text] [doi: 10.1016/j.jfma.2021.05.015] [Medline: 34112587]
27. Wang W, Lin T, Chiu SY, Chen C, Sarakarn P, Ibrahim M, et al. Classification of community-acquired outbreaks for the global transmission of COVID-19: Machine learning and statistical model analysis. J Formos Med Assoc 2021 Jun;120 Suppl 1:S26-S37 [FREE Full text] [doi: 10.1016/j.jfma.2021.05.010] [Medline: 34083090]
28. Bar-On YM, Goldberg Y, Mandel M, Bodenheimer O, Freedman L, Kalkstein N, et al. Protection of BNT162b2 vaccine booster against Covid-19 in Israel. N Engl J Med 2021 Oct 07;385(15):1393-1400. [doi: 10.1056/nejmoa2114255]
29. Chemaitelly H, Tang P, Hasan MR, AlMukdad S, Yassine HM, Benslimane FM, et al. Waning of BNT162b2 vaccine protection against SARS-CoV-2 infection in Qatar. N Engl J Med 2021 Dec 09;385(24):e83 [FREE Full text] [doi: 10.1056/NEJMoa2114114] [Medline: 34614327]
30. Singapore COVID-19 situation reports. World Health Organization. URL: https://www.who.int/singapore/emergencies/covid-19-in-singapore/situation-reports [accessed 2021-10-13]
31. Wadman M. A grim warning from Israel: Vaccination blunts, but does not defeat Delta. Science. 2021. URL: https://www.science.org/content/article/grim-warning-israel-vaccination-blunts-does-not-defeat-delt [accessed 2022-11-05]
32. Levine-Tiefenbrun M, Yelin I, Alapi H, Herzel E, Kuint J, Chodick G, et al. Waning of SARS-CoV-2 booster viral-load reduction effectiveness. Nat Commun 2022 Mar 04;13(1):1237 [FREE Full text] [doi: 10.1038/s41467-022-28936-y] [Medline: 35246560]
33. Telenti A, Arvin A, Corey L, Corti D, Diamond MS, García-Sastre A, et al. After the pandemic: perspectives on the future trajectory of COVID-19. Nature 2021 Aug 08;596(7873):495-504. [doi: 10.1038/s41586-021-03792-w] [Medline: 34237771]

Abbreviations

CrI: credible interval
DAG: directed acyclic graphic
Dci/Imc: number of cases arising from domestic cluster infection caused by imported cases
DIC: deviance information criterion
MCMC: Markov Chain Monte Carlo
NPI: nonpharmaceutical intervention
Rt: effective reproductive number
RT-PCR: reverse transcription-polymerase chain reaction
SEIR: Susceptible-Exposed-Infected-Recovery
VOC: variant of concern

Edited by A Mavragani, G Eysenbach; submitted 08.07.22; peer-reviewed by K Nagar, J Khuntia; comments to author 25.07.22; revised version received 15.08.22; accepted 18.10.22; published 25.11.22.

Please cite as:
Yen AMF, Chen THH, Chang WJ, Lin TY, Jen GHH, Hsu CY, Wang ST, Dang H, Chen SLS
New Surveillance Metrics for Alerting Community-Acquired Outbreaks of Emerging SARS-CoV-2 Variants Using Imported Case Data: Bayesian Markov Chain Monte Carlo Approach
JMIR Public Health Surveill 2022;8(11):e40866
URL: https://publichealth.jmir.org/2022/11/e40866
doi:10.2196/40866
PMID:36265134

Original Paper

Factors Associated With the Intention to Receive the COVID-19 Vaccine: Cross-sectional National Study

Monica L Kasting[1,2], PhD; Jonathan T Macy[3], PhD; Shaun J Grannis[4,5], MS, MD; Ashley J Wiensch[5], MPH; Juan M Lavista Ferres[6], MSc; Brian E Dixon[5,7,8], MPA, PhD

[1]Department of Public Health, Purdue University, West Lafayette, IN, United States

[2]Cancer Prevention and Control Program, Simon Comprehensive Cancer Center, Indiana University, Indianapolis, IN, United States

[3]Department of Applied Health Science, School of Public Health, Indiana University, Bloomington, IN, United States

[4]Department of Family Medicine, School of Medicine, Indiana University, Indianapolis, IN, United States

[5]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, United States

[6]AI for Good Research Lab, Microsoft Corporation, Redmond, WA, United States

[7]Richard M Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States

[8]Center for Health Information and Communication, Health Services Research & Development Service, Richard L Roudebush VA Medical Center, Veterans Health Administration, Indianapolis, IN, United States

Corresponding Author:
Monica L Kasting, PhD
Department of Public Health
Purdue University
812 W. State Street, Room 216
West Lafayette, IN, 47907
United States
Phone: 1 765 496 9483
Email: mlkastin@purdue.edu

Abstract

Background: The COVID-19 pandemic is an unprecedented public health crisis, and vaccines are the most effective means of preventing severe consequences of this disease. Hesitancy regarding vaccines persists among adults in the United States, despite overwhelming scientific evidence of safety and efficacy.

Objective: The purpose of this study was to use the Health Belief Model (HBM) and reasoned action approach (RAA) to examine COVID-19 vaccine hesitancy by comparing those who had already received 1 vaccine to those who had received none.

Methods: This study examined demographic and theory-based factors associated with vaccine uptake and intention among 1643 adults in the United States who completed an online survey during February and March 2021. Survey items included demographic variables (eg, age, sex, political ideology), attitudes, and health belief variables (eg, perceived self-efficacy, perceived susceptibility). Hierarchical logistic regression analyses were used for vaccine uptake/intent. The first model included demographic variables. The second model added theory-based factors to examine the association of health beliefs and vaccine uptake above and beyond the associations explained by demographic characteristics alone.

Results: The majority of participants were male (n=974, 59.3%), White (n=1347, 82.0%), and non-Hispanic (n=1518, 92.4%) and reported they had already received a COVID-19 vaccine or definitely would when it was available to them (n=1306, 79.5%). Demographic variables significantly associated with vaccine uptake/intent included age (adjusted odds ratio [AOR] 1.05, 95% CI 1.04-1.06), other race (AOR 0.47, 95% CI 0.27-0.83 vs White), and political ideology (AOR 15.77, 95% CI 7.03-35.35 very liberal vs very conservative). The theory-based factors most strongly associated with uptake/intention were attitudes (AOR 3.72, 95% CI 2.42-5.73), self-efficacy (AOR 1.75, 95% CI 1.34-2.29), and concerns about side effects (AOR 0.59, 95% CI 0.46-0.76). Although race and political ideology were significant in the model of demographic characteristics, they were not significant when controlling for attitudes and beliefs.

Conclusions: Vaccination represents one of the best tools to combat the COVID-19 pandemic, as well as other possible pandemics in the future. This study showed that older age, attitudes, injunctive norms, descriptive norms, and self-efficacy are positively associated with vaccine uptake and intent, whereas perceived side effects and lack of trust in the vaccine are associated with lower uptake and intent. Race and political ideology were not significant predictors when attitudes and beliefs were considered.

Before vaccine hesitancy can be addressed, researchers and clinicians must understand the basis of vaccine hesitancy and which populations may show higher hesitancy to the vaccination so that interventions can be adequately targeted.

(JMIR Public Health Surveill 2022;8(11):e37203) doi:10.2196/37203

KEYWORDS

SARS-CoV-2; COVID-19 vaccines; vaccination intention; vaccine hesitancy; Health Belief Model; reasoned action approach; COVID-19; vaccination; public health; online survey; health intervention; logistic regression; demographic

Introduction

The COVID-19 pandemic, caused by the novel SARS-CoV-2 virus [1], represents an unprecedented public health crisis. On March 11, 2020, the World Health Organization officially declared COVID-19 a pandemic [1]. In less than 2 years, over 67 million cases and 850,000 deaths from COVID-19 occurred in the United States alone [2]. In December 2020, the US Food and Drug Administration (FDA) granted emergency use authorization for the first vaccine to protect against COVID-19. By April 2021, the FDA has issued emergency use authorization for vaccines by 3 different companies: Pfizer-BioNTech, Moderna, and Johnson & Johnson [3]. As of August 23, 2021, the FDA had granted full approval to the Pfizer-BioNTech vaccine [4]. In addition to data provided by the manufacturers to the FDA, multiple independent research studies demonstrate the vaccines are safe, effective, and widely available for individuals 5 years and older in the United States [5-7].

Hesitancy regarding COVID-19 vaccines persists among adults in the United States [8-12], despite overwhelming scientific evidence of their safety and efficacy. Vaccine hesitancy refers to delay in acceptance or refusal of vaccination, despite the availability of the vaccine and vaccine services [13]. This belief results in lower uptake of prophylactic vaccines and unnecessary morbidity and mortality from vaccine-preventable diseases [11,14,15]. Before vaccine hesitancy can be addressed through population-level intervention, researchers must better understand the basis of vaccine hesitancy and which populations may show higher hesitancy to the COVID-19 vaccine so that interventions can be adequately targeted.

Some of the strongest predictors of vaccine hesitancy and vaccine uptake are attitudes and beliefs derived from the Health Belief Model (HBM) and the reasoned action approach (RAA). Specifically, the HBM proposes that people will take action to prevent a disease if they believe that (1) they are susceptible, (2) the consequences are serious, (3) they can reduce susceptibility or severity through some action, (4) the benefits of taking action outweigh the barriers, and (5) they can engage in a specific behavior (self-efficacy) [16-19]. Previous research guided by this model shows vaccine intent and uptake across multiple diseases are associated with higher perceived benefits, lower perceived barriers, higher perceived severity of the disease, and higher perceived susceptibility/threat of disease [20-23]. However, because the current vaccines against COVID-19 were only approved for emergency use in December 2020, it is unknown whether these health beliefs will translate to how individuals perceive the new vaccine.

In addition to the HBM, this study is also informed by the RAA, which is the newest formulation of the theory of planned behavior and the theory of reasoned action [24]. The RAA extends the theory of planned behavior by differentiating between the attitude, subjective norms, and perceived behavioral control constructs that were integral in the original model [24]. RAA constructs, including experiential attitude, instrumental attitude, and injunctive norm, are significantly associated with the intent to engage in health behaviors [25]. Specifically, research shows that RAA constructs, in particular attitudes toward vaccination and perceived norms, are significantly associated with vaccine intent [26,27].

Therefore, the purpose of this study was to use the HBM and the RAA to examine COVID-19 vaccine hesitancy by exploring vaccine uptake and intent among a national convenience sample of adults in the United States. Specifically, we examined those who already received at least 1 dose of the COVID-19 vaccine or reported a strong intent to be vaccinated compared to those who did not report a strong likelihood of getting vaccinated, as well as demographic, attitudinal, and health belief variables associated with vaccination. Examining factors associated with vaccine uptake and intent provides valuable insight to inform future interventions to combat vaccine hesitancy, not only during the ongoing COVID-19 pandemic, but also during possible future pandemics.

Methods

Participants and Recruitment

We used recruitment methods developed during a pilot study by our team and previously published elsewhere [28]. Briefly, we partnered with Microsoft News to recruit participants to complete a 1-time online survey between February 25 and March 22, 2021. The survey questionnaire was developed for this study. The Microsoft News team created a banner advertisement, shown in Figure 1, which appeared across the top of a news page that a user was viewing. Microsoft News consumers with US browser settings were shown the survey twice in total if they did not click on it and never again after they clicked the link, regardless of whether they completed the survey. The link to the survey was additionally placed in an informational section of the Bing COVID-19 Tracker. Interested participants clicked on the banner and were directed to a survey developed using Qualtrics, a cloud-based survey tool licensed by Indiana University. Eligibility criteria included age 18 years or older, residing in the United States, and able to read English. The survey consisted of 35 individual questions and took approximately 5-10 minutes to complete, and participants were not provided with an incentive.

Figure 1. Study recruitment banner advertisement on Microsoft News.

Ethical Considerations

The study was given exempt status by the Indiana University Institutional Review Board. Because this was exempt research and no identifiable data were collected, this study received a waiver and did not collect written informed consent.

Measures

The primary outcome for this study was vaccine uptake or intent (among the unvaccinated). Vaccine uptake was measured with the question "Have you received at least one dose of any COVID-19 vaccine?" Response options included "Yes, I have received one dose of a vaccine," "Yes, I have received two doses of a vaccine," or "No, I have not received a dose of any vaccine." The people who had not received any doses of a COVID-19 vaccine were asked their vaccine intent with the question "If the vaccines were available where you live and offered to you at no cost, which of the following statements best describes your intention to get either of the vaccines?" Responses were scored on a 4-point Likert scale from "I would definitely get one of the vaccines" to "I would definitely not get either of these vaccines." Responses to these 2 questions were dichotomized such that the sample was divided into those who had already received at least 1 dose or indicated they definitely would get the vaccine (vaccinated/intenders) and those who had not received the vaccine and indicated they did not intend to get vaccinated (unvaccinated/nonintenders).

Covariates fell into 2 categories: demographic characteristics and theory-based attitudes and beliefs. Demographic characteristics included age, gender (female, male, nonbinary, no response), race (White, Asian, Black/African American, or other), ethnicity (yes/no Latinx ethnicity), and political ideology (on a 5-point scale from very conservative to very liberal).

Theory-based attitudes and belief variables were measured on a 5-point Likert scale from strongly agree to strongly disagree. Attitudes about getting the vaccine were assessed with the statement "Getting vaccinated is a good thing to do." To assess injunctive norms, we used the statement "Most people important to me think I should get vaccinated." The descriptive norms construct was measured with the statement "Most people like me will get vaccinated." To assess self-efficacy, participants responded to the statement "I am confident that I can get vaccinated." To assess perceived susceptibility to COVID-19, participants responded to the statement "I am worried about the likelihood of getting COVID-19 in the near future." We examined 3 separate barriers to vaccination: side effects ("The side effects of getting vaccinated interfere with my usual activities"), fear of needles ("I am scared of needles"), and trust in the vaccine ("I do not trust the vaccine"). All 3 items used the same 5-point Likert scale from strongly agree to strongly disagree and were analyzed as separate items.

Data Analysis

First, we described the study sample using n (%) or means and SDs. We then compared the vaccinated/intenders group (already received at least 1 vaccine dose or reported they definitely will get vaccinated) and the unvaccinated/nonintenders group using chi-square or *t* tests, as appropriate. We then conducted a hierarchical logistic regression analysis. We first added the demographic covariates age, gender, race, ethnicity, and political ideology. We next added the theory-based factors to test their unique contributions independent from demographic influences. Analyses were conducted using IBM SPSS Statistics version 28.

Results

Participant Details

A total of 1643 people participated in the survey between February 25 and March 22, 2021, and reported their vaccine status. The sample was 59.3% (n=974) male, 82.0% (n=1347) White, and 92.4% (n=1518) non-Hispanic, and the mean age was 59.4 (SD 14.6, range 18-105) years. There was representation in the sample from all 50 states as well as Washington, DC, and Puerto Rico. For political ideology, 5.5% (n=90) of the participants reported being very conservative, 16.3% (n=268) were conservative, 37.3% (n=613) were moderate, 19.2% (n=316) were liberal, 9.4% (n=154) were very liberal, and 12.3% (n=202) did not respond. Overall, the majority (n=920, 56.0%) were unvaccinated, with 345 (21.0%) receiving 1 dose of any vaccine and 378 (23.0%) receiving 2 doses. Of the unvaccinated, 583 (63.4%) reported they definitely will get the vaccine, 104 (11.3%) reported they probably will get the vaccine, 65 (7.1%) reported they probably will not get the vaccine, and 168 (18.3%) reported they definitely will not get the vaccine. Therefore, for the purposes of this analysis, the majority (n=1306, 79.5%) reported already being vaccinated or said they definitely will get vaccinated when it is available to them. The mean age for the vaccine-hesitant group was slightly less compared to the vaccinated group (53.4 vs 60.9 years, $P<.001$). Vaccine uptake/intent differed by political ideology, with 37.4% (n=126) of the vaccine-hesitant group reporting being either very conservative or conservative. In contrast, only 17.7% (n=232) of the vaccinated/intenders group reported being very conservative (n=50, 21.6%) or conservative (n=182, 78.4%; $P<.001$). For a sample description and bivariate comparisons of the 2 groups, see Table 1.

Table 1. Sample characteristics by vaccine hesitancy.

Characteristics	Total (N=1643)	Vaccinated/intenders (n=1306, 79.5%)	Unvaccinated/nonintenders (n=337, 20.5%)
Age (years), mean (SD); *t* (*df*)=7.81 (1642), *P*<.001	59.4 (14.6)	60.9 (13.7)	53.4 (16.2)
Gender, n (%); χ^2=40.57, *P*<.001			
Female	618 (37.6)	486 (37.2)	132 (39.2)
Male	974 (59.3)	797 (61.0)	177 (52.5)
Nonbinary	25 (1.5)	11 (0.8)	14 (4.2)
No response	26 (1.6)	12 (0.9)	14 (4.2)
Race, n (%); χ^2=41.21, *P*<.001			
Asian	55 (3.3)	42 (3.2)	13 (3.9)
Black/African American	102 (6.2)	73 (5.6)	29 (8.6)
White	1347 (82.0)	1107 (84.8)	240 (71.2)
Other	139 (8.5)	84 (6.4)	55 (16.3)
Ethnicity, n (%); χ^2=3.71, *P*=.05			
Latinx	125 (7.6)	91 (7.0)	34 (10.1)
Not Latinx	1518 (92.4)	1215 (93.0)	303 (89.9)
Political ideology, n (%); χ^2=103.31, *P*<.001			
Very conservative	90 (5.5)	50 (3.8)	40 (11.9)
Conservative	268 (16.3)	182 (13.9)	86 (25.5)
Moderate	613 (37.3)	517 (39.6)	96 (28.5)
Liberal	316 (19.2)	290 (22.2)	26 (7.7)
Very liberal	154 (9.4)	136 (10.4)	18 (5.3)
No response	202 (12.3)	131 (10.0)	71 (21.1)

For the logistic regression analysis that tested factors associated with vaccine uptake and intent, we included those who reported their gender as male or female, reported their political ideology, and answered all theory-based vaccine items, resulting in a full case analysis (n=1370, 83%). We present results from the adjusted logistic regression models in Table 2. In the model with demographic covariates, only age, race, and political ideology were significantly associated with vaccine uptake/intent (all *P*<.01). Specifically, as age increased, the odds of being in the vaccinated/intenders group increased (adjusted odds ratio [AOR] 1.05, 95% CI 1.04-1.06). The "other" race category had lower odds of being in the vaccinated/intenders group than White participants (AOR 0.47, 95% CI 0.27-0.83). The odds of being in the vaccinated/intenders group increased across the political spectrum from a very conservative to a very liberal political ideology, such that those who reported being very liberal had more than 15 times the odds of being in the vaccinated/intenders group compared to those who reported being very conservative (AOR 15.77, 95% CI 7.03-35.35).

However, when theory-based attitudes and belief variables were added to the model, the only demographic variable that remained significant was age. Race and political ideology were no longer significant when controlling for attitudes and beliefs. The attitudes and beliefs variables associated with an increased odds of being in the vaccinated/intenders group included attitudes (AOR 3.72, 95% CI 2.42-5.73), injunctive norms (AOR 1.60, 95% CI 1.18-2.17), descriptive norms (AOR 1.59, 95% CI 1.14-2.22), self-efficacy (AOR 1.75, 95% CI 1.34-2.29), and perceived susceptibility to COVID-19 (AOR 1.30, 95% CI 1.04-1.64). Attitudes and beliefs associated with a decreased odds of being in the vaccinated/intenders group included a concern about side effects (AOR 0.59, 95% CI 0.46-0.76) and lack of trust in the vaccine (AOR 0.73, 95% CI 0.56-0.95). The only attitudes and beliefs variable that was not significantly associated with vaccine uptake/intent was a fear of needles.

Table 2. Results of logistic regression complete case analysis (N=1370).

Characteristics	Model 1: demographic covariates only, AOR[a] (95% CI)	Model 2: demographic covariates plus theory-based factors, AOR (95% CI)
Age (years)	1.05[b] (1.04-1.06)	1.03[c] (1.01-1.05)
Gender		
Female (reference)	N/A[d]	N/A
Male	1.05 (0.77-1.45)	0.91 (0.52-1.59)
Race		
White (reference)	N/A	N/A
Asian	0.90 (0.37-2.22)	1.06 (0.23-4.88)
Black/African American	0.77 (0.41-1.46)	1.15 (0.40-3.30)
Other	0.47[c] (0.27-0.83)	1.08 (0.40-2.94)
Latinx ethnicity		
No (reference)	N/A	N/A
Yes	1.20 (0.64-2.26)	1.32 (0.45-3.89)
Political ideology		
Very conservative (reference)	N/A	N/A
Conservative	1.75[e] (1.03-2.96)	0.66 (0.22-1.95)
Moderate	5.19[b] (3.11-8.67)	0.85 (0.30-2.43)
Liberal	13.80[b] (7.20-26.43)	1.07 (0.33-3.54)
Very liberal	15.77[b] (7.03-35.35)	0.93 (0.22-3.92)
Attitudes	N/A	3.72[b] (2.42-5.73)
Injunctive norms	N/A	1.60[c] (1.18-2.17)
Descriptive norms	N/A	1.59[c] (1.14-2.22)
Self-efficacy	N/A	1.75[b] (1.34-2.29)
Susceptibility to COVID-19	N/A	1.30[e] (1.04-1.64)
Side-effects barrier	N/A	0.59[b] (0.46-0.76)
Fear-of-needles barrier	N/A	1.12 (0.91-1.36)
Do-not-trust-vaccine barrier	N/A	0.73[e] (0.56-0.95)

[a]AOR: adjusted odds ratio.
[b]*P*<.001.
[c]*P*<.01.
[d]N/A: not applicable.
[e]*P*<.05.

Discussion

Principal Findings

This study examined hesitancy in COVID-19 vaccine uptake and intent using a national sample from the United States. Overall, vaccine uptake and intent were high in this sample, with almost 80% of the participants indicating they either received a COVID-19 vaccine already or intended to receive one when it was available to them. However, approximately 1 in 5 participants indicated they had not received the vaccine and did not report they definitely would receive it, when available, indicating vaccine hesitancy. With highly contagious viral variants quickly spreading across the nation, public health officials perceive a new phase of the ongoing COVID-19 pandemic being dubbed a “pandemic of the unvaccinated” [29]. There is an urgent need to understand the beliefs and attitudes associated with vaccine hesitancy so that interventions to improve the vaccination rate worldwide can be developed and implemented.

Comparison With Prior Work

Our study found 3 demographic variables associated with vaccine uptake and intent in the model that included demographic characteristics only: age, race, and political affiliation. However, only age remained significant when accounting for the theory-based factors. Specifically, older age was associated with increased odds of being vaccinated or intending to be vaccinated. This is not surprising, given the vaccine rollout in the United States occurred largely by age group and is consistent with early research prior to vaccine availability that noted increasing age was associated with increasing vaccine intent [12]. All adults in the United States were eligible for vaccination by April 19, 2021 [30]. It is possible some of the adults who responded were not eligible for vaccination yet, because these data were collected in February and March. However, because we included people who reported they definitely would get the vaccine when it was available in with the vaccinated sample, this should not have affected our results. The association between age and vaccine uptake and intent may be due to the fact that older adults, if infected, are more likely to have severe disease [31]. However, this association persisted even when controlling for perceived susceptibility to COVID-19, indicating the association may not be explained by either availability or perceived susceptibility. Our study did not examine issues of access or logistics, particularly transportation barriers, time off work, and childcare, which likely affect younger adults more than older adults. Access and logistical barriers are important issues to examine in future research.

Political affiliation was also associated with vaccine uptake and intent, with the odds of uptake increasing across the sample, with very conservatives reporting the lowest uptake/intent and very liberals reporting the highest uptake/intent. However, this association was no longer significant when accounting for attitudes and beliefs. Another recent research study found increased vaccine hesitancy among moderates and conservatives (compared to liberals) when accounting for respondent characteristics and behaviors [32]. However, this research did not include beliefs in the model, which our data indicate may be an important predictor to analyze. An additional study examined COVID-19 vaccine intent while controlling for political affiliation and media exposure [33]. This study did find a difference in intent between Republicans and Democrats, with Democrats indicating a higher intent to be vaccinated. Although they controlled for preferred media for virus-related news (including social media Fox News, and CNN/MSNBC), and belief in conspiracy theories, they did not control for other attitudinal or belief variables, including injunctive and descriptive norms. It is essential to understand that this lack of association once we control for attitudes and beliefs does not imply political affiliation's lack of causal effect on vaccine hesitancy. Other political science research has found that instead of people's moral foundations predicting their political affiliations, it is in fact people's political affiliations that predict their moral foundations [34]. That is, people tend to switch their moral values, depending on how they fit with their political beliefs, as opposed to switching their political beliefs, depending on how they fit their moral values. Based on these findings, it is important for future research to examine the interplay between political affiliation, attitudes, and beliefs to better understand which is actually the driver of the association with vaccine hesitancy. Having a better understanding of the association between political affiliation, attitudes, beliefs, and vaccine hesitancy will enable researchers to develop community-based interventions that address these challenges.

Like political affiliation, race was significantly associated with uptake/intent in the model that included demographic characteristics only. Specifically, people who reported they were a race other than White, Asian, or Black/African American were approximately half as likely to be vaccinated or intend to be vaccinated compared to the White participants. However, this association was no longer significant when theoretical covariates were entered into the model. As was discussed earlier in regard to age, our study did not examine issues of access or logistics, particularly transportation barriers, time off work, and childcare, which may affect non-White respondents more than White respondents. Although research does indicate there is mistrust among non-White patients, there are also issues with health equity and access to care that seem to be driving the disparity [35]. A recent publication noted that the racial disparity in COVID-19 mortality is due more to structural racism than to race itself [36]. It is also important to note that although the association we found in our study was significant for the "other" race category, it was not significant for Black/African American participants or those who reported Latinx ethnicity. Future research should examine these associations to better understand the interplay between race, attitudes, beliefs, and vaccine hesitancy so that culturally appropriate community-based interventions can be developed.

The primary aim of this study was to identify the beliefs underlying the US adults' decision to get vaccinated against COVID-19. Of note, when we added the theory-based constructs to the regression model for vaccine uptake and intention, age remained the only statistically significant demographic variable. This points to the important contributions of the theoretical constructs in explaining the variation in the decision to get vaccinated, beyond the influence of several demographic factors. The theoretical construct most strongly associated with vaccine uptake and intention in this sample was attitude. This finding suggests that attitude could be an important focal point for interventions aimed at increasing COVID-19 vaccine uptake. Attitudes can be addressed through communication and education campaigns that present the advantages of getting vaccinated and address any potential negative consequences. One method some hospital systems have used is publishing infographics that demonstrate that the hospitalized patients are overwhelmingly unvaccinated [37]. Furthermore, a multilevel intervention that included a component addressing patient and provider attitudes toward human papillomavirus vaccination saw increased uptake of the vaccine in the intervention group compared to the control group [38]. However, the authors stated the increase was lower than expected. Future research should examine effective ways to improve attitudes and increase uptake of vaccines.

Self-efficacy was also significantly associated with vaccine uptake and intention in this sample. This suggests that public

health interventions should address adults' confidence that they can get vaccinated. There are 2 approaches to improving self-efficacy or capacity. One approach aims to address people's beliefs directly. Communication and educational campaigns can potentially help people see and come to believe that they have the capacity to get vaccinated. Modeling is 1 effective way to improve self-efficacy [39]. According to past research, modeling interventions should resemble the target group, start with small steps, look to succeed but not immediately, and be reinforced for the behavior of getting vaccinated [40]. Thus, these campaigns could include examples of how people successfully overcame their hesitancy to get the vaccine. The second approach is to address the actual environment by removing barriers to getting vaccinated or adding facilitators at local, organizational, and governmental levels. This could include removing the request for health insurance information and providing paid time off from work to get the vaccine and recover from any short-term side effects.

Both types of normative beliefs (injunctive and descriptive) were significantly associated with vaccine uptake and intention, albeit less strongly so than attitude and self-efficacy. Injunctive norms represent people's perceptions about what people who are important to them think they should do, and descriptive norms represent people's beliefs about how people like them are behaving. This suggests that, in this sample of US adults, the influence of important people in their lives and people like them might be key determinants of their intention to get vaccinated. Therefore, health communication messages tailored for US adults should emphasize that people important to them want them to get vaccinated and people like them are getting vaccinated.

Two of the barriers examined were associated with decreased odds of being in the vaccinated/intenders group. Specifically agreeing that the vaccine would cause side effects that would interfere with their usual activities and reporting they do not trust the vaccine were both associated with decreased odds of being vaccinated/intending to get vaccinated. This is consistent with other recent surveys examining people who have not yet been vaccinated and found that almost 1 in 5 of them reported not being vaccinated due to concerns over adverse effects or the vaccines' newness [41]. Many of these concerns among the population stem from misinformation encountered on social media. Indeed, 1 recent research study found that COVID-19 vaccine intent is significantly associated with not relying on social media for virus information [33]. Misinformation can shape people's decision-making and perceptions, particularly if left unchallenged [41]. Specifically, 1 study found that negative television news coverage of a medication can increase reporting of adverse events for that medication [42]. Furthermore, research shows that viewing a website critical of vaccines for just 5-10 minutes decreases the intention to vaccinate [43]. However, it is important to be transparent about the potential side effects of any medication or vaccine. Research in the HIV literature found that a failure to acknowledge potential negative effects of receiving an HIV test results in a "boomerang effect," where people who already perceive obstacles to testing are less likely to get tested if the negative effects aren't acknowledged [44]. However, to foster trust in these vaccines and combat the misinformation that people encounter regarding safety and efficacy, it is important to challenge their misperceptions and provide scientifically accurate information that is understandable to the layperson and delivered by a person they trust. This information should include that the vaccine side effects are mild, the risks of the vaccine are much lower than the risks of COVID-19 infection, and the vaccines are effective in preventing severe COVID-19. A key partner in this conversation is the person's health care provider, and providers should communicate to their patients that they strongly recommend vaccination. Research shows the intent to be vaccinated increases if the person's health care provider recommends they receive the vaccine [12].

Limitations

Although this study had numerous strengths, including using a national sample and examining relevant and understudied attitudes and beliefs, the results should be interpreted in the light of some limitations. First, these data are cross-sectional and causal associations cannot be determined. Second, the data were collected in February and March 2021. It is possible attitudes, intent, and uptake may have shifted in the intervening months. This period was slightly before all US adults could be vaccinated against COVID-19 and was also prior to widespread infection with the more contagious delta and subsequent omicron variants. Ongoing research on these topics is warranted. Third, although we did recruit nationally for this study, compared to the overall US population, our sample was a lower proportion of females (37.6% vs 50.8% nationwide) and Hispanic (7.6% vs 18.5% nationwide) and was older (mean age 59.4 years vs median age 37.7 years nationwide) [45]. Although we controlled for demographic variables in the regression analyses, our findings may not be generalizable to the broader US population. In addition, our recruitment strategy using Microsoft News limited our sample to only those who use Microsoft products and have this feature turned on, further limiting generalizability.

Conclusion

Vaccination represents one of the best tools to combat the ongoing COVID-19 pandemic [46]. Hesitancy regarding vaccines persists among adults in the United States, despite overwhelming scientific evidence of safety and efficacy. These beliefs result in lower uptake of vaccines and unnecessary morbidity and mortality from vaccine-preventable diseases. This research provides novel insight into the association between attitudes and beliefs with vaccine hesitancy. Specifically, older age, attitudes, injunctive norms, descriptive norms, and self-efficacy are positively associated with vaccine uptake and intent, whereas perceived side effects and lack of trust in the vaccine are associated with lower uptake and intent. Before vaccine hesitancy can be addressed, researchers need to understand the basis of vaccine hesitancy and intent as well as which populations may show higher hesitancy to the COVID-19 vaccine so that interventions can be adequately targeted.

Acknowledgments

The authors would like to acknowledge the Regenstrief Institute, Inc, for its support, especially the public relations team and the Center for Biomedical Informatics, of this nonfunded research project during the pandemic. The authors would additionally like to thank the Microsoft News team for its support in developing and implementing the banner advertisements (Dr Mary L Gray, Vera Chan, Matt Lindenburg, and Erin Van Noy).

MLK is supported by grant numbers KL2TR002530 (B Tucker Edmonds, principal investigator [PI]) and UL1TR002529 (S Moe and S Wiehe, co-PIs) from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award. BED is supported, in part, by funding from the Centers for Disease Control and Prevention (CDC), the Indiana Department of Health, and the Marion County Public Health Department (MCPHD) to track the spread of COVID-19. The funding agencies had no role in designing the study, interpreting the data, writing the manuscript, or making the decision to submit the manuscript for publication.

Conflicts of Interest

JMLF is employed at Microsoft Corporation. The other authors have no conflicts of interest to declare.

References

1. World Health Organization. Archived: WHO Timeline - COVID-19. 2020 Apr 27. URL: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 [accessed 2022-10-25]
2. Johns Hopkins University. COVID-19 Dashboard. URL: https://coronavirus.jhu.edu/map.html [accessed 2022-10-25]
3. Food and Drug Administration. COVID-19 Vaccines. URL: https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines [accessed 2022-10-25]
4. Food and Drug Administration. FDA Approves First COVID-19 Vaccine. 2021 Aug 23. URL: https://www.fda.gov/news-events/press-announcements/fda-approves-first-covid-19-vaccine [accessed 2022-10-25]
5. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, C4591001 Clinical Trial Group. Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. N Engl J Med 2020 Dec 31;383(27):2603-2615 [FREE Full text] [doi: 10.1056/NEJMoa2034577] [Medline: 33301246]
6. Centers for Disease Control and Prevention (CDC). COVID-19: Safety of COVID-19 Vaccines. URL: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/safety-of-vaccines.html [accessed 2022-10-25]
7. Hause AM, Baggs J, Marquez P, Myers TR, Gee J, Su JR, et al. COVID-19 vaccine safety in children aged 5-11 years - United States, November 3-December 19, 2021. MMWR Morb Mortal Wkly Rep 2021 Dec 31;70(5152):1755-1760 [FREE Full text] [doi: 10.15585/mmwr.mm705152a1] [Medline: 34968370]
8. Khubchandani J, Sharma S, Price JH, Wiblishauser MJ, Sharma M, Webb FJ. COVID-19 vaccination hesitancy in the United States: a rapid national assessment. J Community Health 2021 Apr;46(2):270-277 [FREE Full text] [doi: 10.1007/s10900-020-00958-x] [Medline: 33389421]
9. Quinn SC, Jamison AM, An J, Hancock GR, Freimuth VS. Measuring vaccine hesitancy, confidence, trust and flu vaccine uptake: results of a national survey of White and African American adults. Vaccine 2019 Feb 21;37(9):1168-1173. [doi: 10.1016/j.vaccine.2019.01.033] [Medline: 30709722]
10. Santibanez TA, Nguyen KH, Greby SM, Fisher A, Scanlon P, Bhatt A, et al. Parental vaccine hesitancy and childhood influenza vaccination. Pediatrics 2020 Dec;146(6):e2020007609 [FREE Full text] [doi: 10.1542/peds.2020-007609] [Medline: 33168671]
11. Thomas TL, DiClemente R, Snell S. Overcoming the triad of rural health disparities: how local culture, lack of economic opportunity, and geographic location instigate health disparities. Health Educ J 2014 May;73(3):285-294 [FREE Full text] [doi: 10.1177/0017896912471049] [Medline: 25242822]
12. Head KJ, Kasting ML, Sturm LA, Hartsock JA, Zimet GD. A national survey assessing SARS-CoV-2 vaccination intentions: implications for future public health communication efforts. Sci Commun 2020 Sep 23;42(5):698-723. [doi: 10.1177/1075547020960463]
13. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: definition, scope and determinants. Vaccine 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.036] [Medline: 25896383]
14. Zhai Y, Santibanez TA, Kahn KE, Srivastav A, Walker TY, Singleton JA. Rural, urban, and suburban differences in influenza vaccination coverage among children. Vaccine 2020 Nov 10;38(48):7596-7602. [doi: 10.1016/j.vaccine.2020.10.030] [Medline: 33071004]
15. Swiecki-Sikora AL, Henry KA, Kepka D. HPV vaccination coverage among US teens across the rural-urban continuum. J Rural Health 2019 Sep;35(4):506-517 [FREE Full text] [doi: 10.1111/jrh.12353] [Medline: 30703854]
16. Skinner C, Tiro J, Champion V. The Health Belief Model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass; 2015.
17. Bandura A. Self-efficacy mechanism in human agency. Am Psychol 1982 Feb;37(2):122-147. [doi: 10.1037/0003-066x.37.2.122]

18. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
19. Rosenstock IM. Historical origins of the Health Belief Model. Health Educ Monogr 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]
20. Donadiki E, Jiménez-García R, Hernández-Barrera V, Sourtzi P, Carrasco-Garrido P, López de Andrés A, et al. Health Belief Model applied to non-compliance with HPV vaccine among female university students. Public Health 2014 Mar;128(3):268-273. [doi: 10.1016/j.puhe.2013.12.004] [Medline: 24529635]
21. Schaefer Ziemer K, Hoffman MA. Beliefs and attitudes regarding human papillomavirus vaccination among college-age women. J Health Psychol 2013 Oct;18(10):1360-1370. [doi: 10.1177/1359105312462432] [Medline: 23188917]
22. Mehta P, Sharma M, Lee RC. Designing and evaluating a Health Belief Model-based intervention to increase intent of HPV vaccination among college males. Int Q Community Health Educ 2013;34(1):101-117. [doi: 10.2190/IQ.34.1.h] [Medline: 24366025]
23. Coe AB, Gatewood SBS, Moczygemba LR, Goode JKR, Beckner JO. The use of the Health Belief Model to assess predictors of intent to receive the novel (2009) H1N1 influenza vaccine. Innov Pharm 2012;3(2):1-11 [FREE Full text] [doi: 10.24926/iip.v3i2.257] [Medline: 22844651]
24. Fishbein M, Ajzen I. Predicting and Changing Behavior: The Reasoned Action Approach. Oxfordshire, UK: Taylor & Francis; 2011.
25. McEachan R, Taylor N, Harrison R, Lawton R, Gardner P, Conner M. Meta-analysis of the reasoned action approach (RAA) to understanding health behaviors. Ann Behav Med 2016 Aug 11;50(4):592-612 [FREE Full text] [doi: 10.1007/s12160-016-9798-4] [Medline: 27169555]
26. Jozkowski KN, Geshnizjani A. Using a reasoned action approach to examine US college women's intention to get the HPV vaccine. Health Educ J 2014 Dec 04;75(1):14-26. [doi: 10.1177/0017896914561100]
27. Lueck JA, Spiers A. Which beliefs predict intention to get vaccinated against COVID-19? A mixed-methods reasoned action approach applied to health communication. J Health Commun 2020 Oct 02;25(10):790-798. [doi: 10.1080/10810730.2020.1865488] [Medline: 33719876]
28. Dixon BE, Mukherjee S, Wiensch A, Gray ML, Ferres JML, Grannis SJ. Capturing COVID-19-like symptoms at scale using banner ads on an online news platform: pilot survey study. J Med Internet Res 2021 May 20;23(5):e24742 [FREE Full text] [doi: 10.2196/24742] [Medline: 33872190]
29. Anthes E, Petri A. C.D.C. Director Warns of a 'Pandemic of the Unvaccinated'. URL: https://www.nytimes.com/2021/07/16/health/covid-delta-cdc-walensky.html [accessed 2022-10-25]
30. Superville D, Jaffe A. Biden Makes All Adults Eligible for a Vaccine on April 19. URL: https://apnews.com/article/biden-move-vaccine-eligibility-date-april-19-021157c7bdf964181e3b63f51b89601e [accessed 2022-10-25]
31. Centers for Disease Control and Prevention (CDC). COVID-19: Older Adults. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html [accessed 2022-10-25]
32. Gatwood J, McKnight M, Fiscus M, Hohmeier KC, Chisholm-Burns M. Factors influencing likelihood of COVID-19 vaccination: a survey of Tennessee adults. Am J Health Syst Pharm 2021 May 06;78(10):879-889 [FREE Full text] [doi: 10.1093/ajhp/zxab099] [Medline: 33954426]
33. Ruiz JB, Bell RA. Predictors of intention to vaccinate against COVID-19: results of a nationwide survey. Vaccine 2021 Feb 12;39(7):1080-1086 [FREE Full text] [doi: 10.1016/j.vaccine.2021.01.010] [Medline: 33461833]
34. Hatemi P, Crabtree C, Smith K. Ideology justifies morality: political beliefs predict moral foundations. Am J Pol Sci 2019 Jul 30;63(4):788-806 [FREE Full text] [doi: 10.1111/ajps.12448]
35. Reverby SM. Racism, disease, and vaccine refusal: people of color are dying for access to COVID-19 vaccines. PLoS Biol 2021 Mar;19(3):e3001167 [FREE Full text] [doi: 10.1371/journal.pbio.3001167] [Medline: 33684102]
36. Nephew LD. Systemic racism and overcoming my COVID-19 vaccine hesitancy. EClinicalMedicine 2021 Feb;32:100713 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100713] [Medline: 33495751]
37. Zara C. Delta Variant Devastation: Florida Hospitals Show Vaccinated vs Unvaccinated ICU Patient Data. 2021 Aug 18. URL: https://www.fastcompany.com/90667213/delta-variant-devastation-florida-hospitals-show-vaccinated-vs-unvaccinated-icu-patient-data [accessed 2022-10-25]
38. Paskett ED, Krok-Schoen JL, Pennell ML, Tatum CM, Reiter PL, Peng J, et al. Results of a multilevel intervention trial to increase human papillomavirus (HPV) vaccine uptake among adolescent girls. Cancer Epidemiol Biomarkers Prev 2016 Apr;25(4):593-602 [FREE Full text] [doi: 10.1158/1055-9965.EPI-15-1243] [Medline: 27196093]
39. Bandura A. Self-Efficacy: The Exercise of Control. New York, NY: W H Freeman/Times Books/Henry Holt & Co; 1997.
40. Lee L, Arthur A, Avis M. Using self-efficacy theory to develop interventions that help older people overcome psychological barriers to physical activity: a discussion paper. Int J Nurs Stud 2008 Nov;45(11):1690-1699. [doi: 10.1016/j.ijnurstu.2008.02.012] [Medline: 18501359]
41. Rief W. Fear of adverse effects and COVID-19 vaccine hesitancy: recommendations of the treatment expectation expert group. JAMA Health Forum 2021 Apr 01;2(4):e210804 [FREE Full text] [doi: 10.1001/jamahealthforum.2021.0804] [Medline: 36218819]

42. Faasse K, Gamble G, Cundy T, Petrie KJ. Impact of television coverage on the number and type of symptoms reported during a health scare: a retrospective pre-post observational study. BMJ Open 2012 Aug 17;2(4):e001607 [FREE Full text] [doi: 10.1136/bmjopen-2012-001607] [Medline: 22904334]
43. Betsch C, Renkewitz F, Betsch T, Ulshöfer C. The influence of vaccine-critical websites on perceiving vaccination risks. J Health Psychol 2010 Apr;15(3):446-455. [doi: 10.1177/1359105309353647] [Medline: 20348365]
44. Kasting ML, Cox AD, Cox D, Fife KH, Katz BP, Zimet GD. The effects of HIV testing advocacy messages on test acceptance: a randomized clinical trial. BMC Med 2014 Nov 06;12:204 [FREE Full text] [doi: 10.1186/s12916-014-0204-4] [Medline: 25374047]
45. U.S. Census Bureau. Quick Facts: United States. URL: https://www.census.gov/quickfacts/fact/table/US/PST045219 [accessed 2022-10-25]
46. Centers for Disease Control and Prevention (CDC). COVID-19 Vaccines. URL: https://www.cdc.gov/coronavirus/2019-ncov/downloads/vaccines/facts-covid-vaccines-english-508.pdf [accessed 2022-10-25]

Abbreviations

AOR: adjusted odds ratio
FDA: Food and Drug Administration
HBM: Health Belief Model
RAA: reasoned action approach

Edited by G Eysenbach; submitted 11.02.22; peer-reviewed by G Shakerinejad, A Tannoubi; comments to author 02.05.22; revised version received 20.05.22; accepted 09.10.22; published 14.11.22.

Please cite as:
Kasting ML, Macy JT, Grannis SJ, Wiensch AJ, Lavista Ferres JM, Dixon BE
Factors Associated With the Intention to Receive the COVID-19 Vaccine: Cross-sectional National Study
JMIR Public Health Surveill 2022;8(11):e37203
URL: https://publichealth.jmir.org/2022/11/e37203
doi:10.2196/37203
PMID:36219842

Original Paper

Outcomes of a Community Engagement and Information Gathering Program to Support Telephone-Based COVID-19 Contact Tracing: Descriptive Analysis

Chi-Chi N Udeagu[1*], MPH; Masha Pitiranggon[1*], MPH; Kavita Misra[1*], MPH, PhD; Jamie Huang[1*], MPH; Thomas Terilli[1*]; Yasmin Ramos[1*]; Martha Alexander[1*], MHS; Christine Kim[1*], MPH; David Lee[1*], MPH, MBA; Kathleen Blaney[1*], MPH, RN; Chris Keeley[2*], MUP; Theodore Long[2*], MD; Neil M Vora[1*], MD

[1]New York City Test & Trace Corps, New York City Department of Health & Mental Hygiene, Queens, NY, United States

[2]New York City Test & Trace Corps, New York City Health + Hospitals, New York City, NY, United States

[*]all authors contributed equally

Corresponding Author:
Chi-Chi N Udeagu, MPH
New York City Test & Trace Corps
New York City Department of Health & Mental Hygiene
42-09 28th Street
Queens, NY, 11101
United States
Phone: 1 3473864909
Email: cudeagu@health.nyc.gov

Abstract

Background: Contact tracing is an important public health tool for curbing the spread of infectious diseases. Effective and efficient contact tracing involves the rapid identification of individuals with infection and their exposed contacts and ensuring their isolation or quarantine, respectively. Manual contact tracing via telephone call and digital proximity app technology have been key strategies in mitigating the spread of COVID-19. However, many people are not reached for COVID-19 contact tracing due to missing telephone numbers or nonresponse to telephone calls. The New York City COVID-19 Trace program augmented the efforts of telephone-based contact tracers with information gatherers (IGs) to search and obtain telephone numbers or residential addresses, and community engagement specialists (CESs) made home visits to individuals that were not contacted via telephone calls.

Objective: The aim of this study was to assess the contribution of information gathering and home visits to the yields of COVID-19 contact tracing in New York City.

Methods: IGs looked for phone numbers or addresses when records were missing phone numbers to locate case-patients or contacts. CESs made home visits to case-patients and contacts with no phone numbers or those who were not reached by telephone-based tracers. Contact tracing management software was used to triage and queue assignments for the telephone-based tracers, IGs, and CESs. We measured the outcomes of contact tracing–related tasks performed by the IGs and CESs from July 2020 to June 2021.

Results: Of 659,484 cases and 861,566 contact records in the Trace system, 28% (185,485) of cases and 35% (303,550) of contacts were referred to IGs. IGs obtained new phone numbers for 33% (61,804) of case-patients and 11% (31,951) of contacts; 50% (31,019) of the case-patients and 46% (14,604) of the contacts with new phone numbers completed interviews; 25% (167,815) of case-patients and 8% (72,437) of contacts were referred to CESs. CESs attempted 80% (132,781) of case and 69% (49,846) of contact investigations, of which 47% (62,733) and 50% (25,015) respectively, completed interviews. An additional 12,192 contacts were identified following IG investigations and 13,507 following CES interventions.

Conclusions: Gathering new or missing locating information and making home visits increased the number of case-patients and contacts interviewed for contact tracing and resulted in additional contacts. When possible, contact tracing programs should add information gathering and home visiting strategies to increase COVID-19 contact tracing coverage and yields as well as promote equity in the delivery of this public health intervention.

(JMIR Public Health Surveill 2022;8(11):e40977) doi:10.2196/40977

KEYWORDS

COVID-19; contact tracing; home visits; community health workers; health equity

Introduction

Worldwide, the emergence of COVID-19 as a public health crisis prompted a range of measures to curb the spread of SARS-CoV-2, the virus that causes COVID-19. The mitigation measures included nonpharmaceutical interventions, such as handwashing, stay-at-home order, self-masking, social distancing, and limits on the type and number of people at social gatherings [1-4]. In addition, public health jurisdictions applied contact tracing strategies to identify and notify exposed contacts of people with COVID-19 to stem ongoing disease transmission [5-10].

Contact tracing is a resource-intensive, multistep process [7]. The core feature of an efficient and effective contact tracing program is timely case identification and investigation to elicit exposed contacts and ensure self-isolating of case-patients as well as the notification and quarantining of their contacts [11,12]. Studies have found that early identification of cases through testing and contact tracing and quarantining of the exposed contacts can result in about 80% reduction in the transmission SARS-CoV-2, including transmissions by presymptomatic or asymptomatic individuals with COVID-19 [13-15].

Nonetheless, the high burden of COVID-19 cases presented overwhelming challenges to contact tracing programs in reaching every case-patient or contact and conducting timely manual contact tracing via telephone calls [16-19]. Therefore, many public health jurisdictions added digital contact tracing, which involves the use of smartphones to optimize the breadth of tracing and minimize delays in contact notifications [8-10]. Digital proximity contact tracing aims to rapidly identify people who may have been in contact with individuals subsequently diagnosed with COVID-19 for a certain amount of time, using electronic techniques including Bluetooth, Global Positioning System, or Wi-Fi.

Manual telephone calls and digital contact tracing rely on the ownership and use of smartphones, electronic tracking systems, and accurate telephone numbers. People with COVID-19 or their contacts may lack access to telephone or mobile technology or the skill and ability to operate them [20-22]. Furthermore, people with COVID-19 may be reluctant to respond to telephone calls from public health officials or to name their contacts, fearing stigma or quarantine, or they may be unwilling to opt into digital tracking due to privacy concerns [23-25]. A cornerstone of comprehensive contact tracing for infectious diseases is a community-based effort, including door-to-door visits to reach people who are unable or unwilling to engage via phone calls or digital platforms [26,27]. Face-to-face interactions between contact tracers with individuals with COVID-19 or their contacts may offer the opportunity to establish rapport and build trust needed to obtain personal information from reluctant individuals.

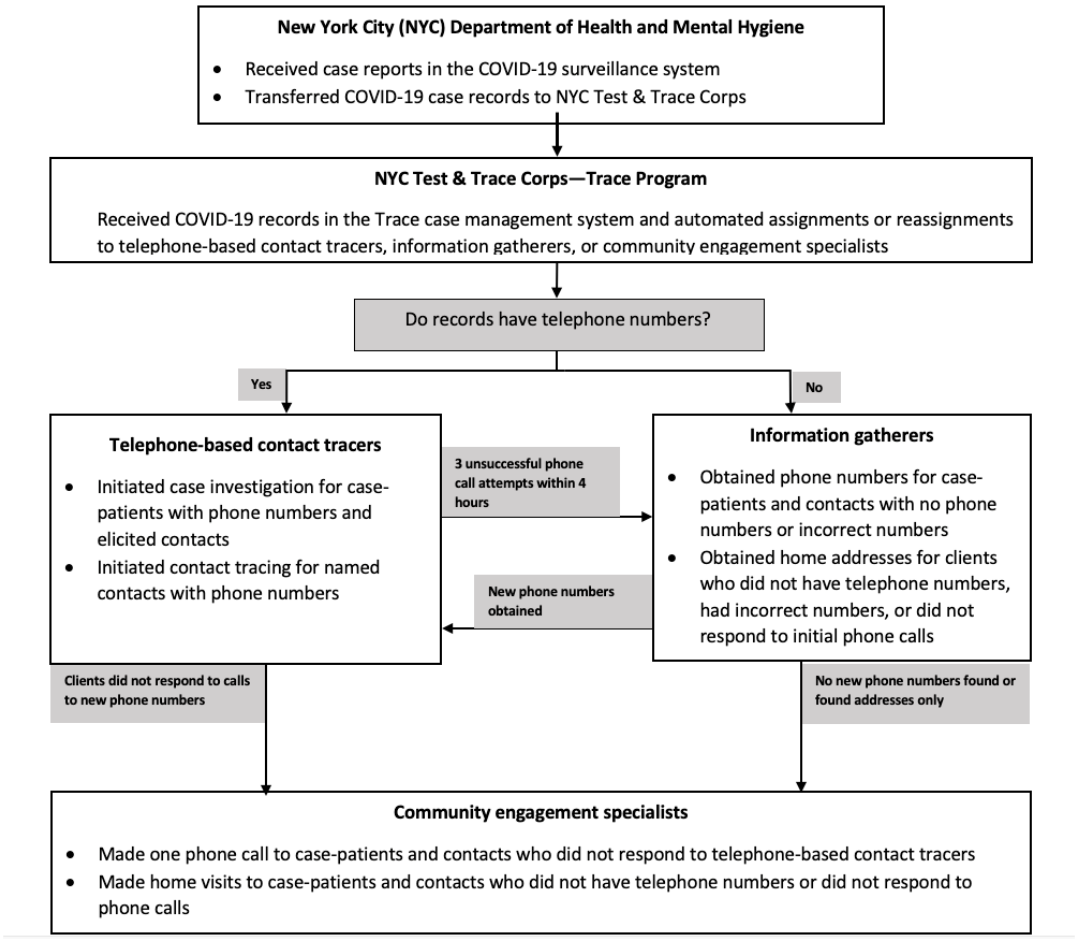
In June 2020, New York City (NYC) established the NYC COVID-19 Test & Trace Corps to develop and implement interventions to suppress COVID-19 transmission in NYC [28]. Beginning in June 2020, the contact tracing component of the Test & Trace Corps—Trace—attempted to reach people with COVID-19 and their contacts through telephone-based contact tracers. Between June and July 2020, Trace implemented 2 additional workflows with specialized staff to complement the efforts of the telephone-based tracers. These were efforts to (1) look for locating information of case-patients and contacts using Information Gatherers (IGs) when records lacked working telephone numbers; and (2) conduct home-based contact tracing using Community Engagement Specialists (CESs) when phone numbers were lacking or after unsuccessful telephone-based efforts. In this paper, we assess the contributions of the IGs and CESs to the NYC COVID-19 contact tracing efforts from July 2020-June 2021.

Methods

Study Population and Data Sources

All COVID-19 positive and negative results of tests performed by NYC laboratories and point-of-care testing sites were reported to the NYC Department of Health and Mental Hygiene's (DOHMH's) COVID-19 surveillance system. Daily, the DOHMH exported case records of confirmed or probable COVID-19 cases to the Trace case management system (Figure 1). To minimize the records with missing locating information, the DOHMH matched case records against available electronic medical record data systems of NYC medical institutions prior to data transfer to Trace. Data for our analysis were comprised of records forwarded to the Trace program from July 2020-June 2021. These records included the name and contact information of the ordering provider, demographic information of the case-patients (ie, name, phone number, address, and date of birth), date of specimen collection, and test type. We also analyzed records of the contacts named by case-patients. For each contact, contact tracers attempted to obtain name, phone number, address, date of birth, and the date of last exposure.

Figure 1. New York City Test & Trace Corps COVID-19 case investigation and contact tracing workflow.



Definitions

Case-patients were comprised of persons with probable or confirmed COVID-19 results or contacts with COVID-19 symptoms (or 'symptomatic contacts'), even if the contacts had no reported COVID-19 test results [29]. Contacts were persons who came within 6 feet of people with COVID-19 during their infectious period for a cumulative total of ≥10 minutes over a 24-hour period [30]. The infectious period began 2 days before the onset of symptoms for case-patients or, if asymptomatic, 2 days before the specimen collection date of their COVID-19 positive test. We referred to case-patients or contacts as clients [31].

Contact Tracing Workflow

Contact tracing encompassed case and contact investigations. Case investigation included the interview by a telephone-based tracer or CES to elicit contacts from case-patients, give isolation instructions, and make referrals for supportive services (eg, housing, groceries, and pet care). Contact investigation involved attempts to reach and interview the named contacts, inform them of their potential exposure to SARS-CoV-2, give recommendations for quarantine (ie, isolate, if symptomatic), and make referrals for supportive services.

Contact tracing workflow and the coordination of activities among telephone-based tracers, IGs, and CESs were managed within a software configured for Trace data management. Upon data transfer from the DOHMH, an automated algorithm assigned client records with telephone numbers to telephone-based tracers or to IGs if phone numbers were lacking (Figure 1). Furthermore, records were assigned to telephone-based tracers, IGs, or CESs based on the outcomes of the previous attempts. For example, if the telephone-based tracers could not reach clients at available phone numbers after 3 attempts within 4 hours of assignment, the records for those clients were then assigned to IGs to attempt to find new numbers. If IGs obtained new numbers, those records were reassigned to telephone-based tracers. If IGs obtained residential addresses only or clients did not respond to repeated outreach attempts by telephone-based tracers, those records were assigned to the CESs. Telephone-based tracers, IGs, and CESs recorded all interim and final outcomes they obtained in the Trace case management system in real time.

Overview of IG Activities

For cases, IGs (n=74 at peak) called the reporting laboratories or ordering medical providers to obtain any available locating information (eg, telephone number and address) in their medical records. In addition, IGs manually searched CLEAR, a subscription service that collects public record information, including phone numbers and addresses, for locating information. For contacts, IGs did not contact the persons with COVID-19 who had named the contacts; rather, IGs used CLEAR to look for phone numbers or addresses. During the searches, IGs used clients' first and last names, and full date of birth (ie, month, date, and 4-digit year) to confirm that

information was being obtained for the referenced client. IGs did not perform provider or record searches if clients' records were missing complete date of birth. During periods of high workloads relative to the number of IGs, IGs first prioritized information gathering for case-patients over contacts and same-day referrals over referrals from previous days.

IGs entered new phone numbers in the appropriate data fields in the Trace case management system, and the system queued the case or contact records for the telephone-based tracers. If only addresses were found, IGs updated the address field, and the records were then queued for the CESs. If neither phone numbers nor addresses were obtained, IGs made notes in the text field of the Trace case management system (July-November 2020) or assigned a final disposition of "unable to locate" new information (December 2020-June 2021).

Overview of CES Activities

CESs' COVID-19 prevention activities have been previously described [32]. In brief, CESs (n=540 at peak) performed in-person contact tracing and other COVID-19 prevention activities, such as the dissemination of COVID-19 information and sanitary supplies (eg, masks and hand sanitizers) at NYC schools, business establishments, and community settings. From July 2020-June 2021, the number of CESs assigned to perform contact tracing fluctuated daily (range: 192-492), depending on the need for them to engage in these other prioritized community-based COVID-19 prevention activities.

CESs' contact tracing activities entailed making telephone calls and home visits to clients who did not have phone numbers or did not respond to telephone-based tracers. First, CES supervisors (n=50 at peak) manually assessed the records assigned to the CESs in the Trace case management system and made individual CES assignments, prioritizing case investigations over contact investigations. Supervisors also grouped clients by zip code, address, and telephone number to improve efficiency; for example, clients residing at the same address or with the same phone numbers were assigned to the same CES. At the beginning of their workday, CESs logged into the Trace case management system on their iPads and sequentially planned their outreach to clients. CESs first attempted phone calls to clients for whom telephone numbers were available, then made home visits to the addresses of clients who either did not respond to those phone calls or who had no telephone numbers on their record.

CESs received training on universal infection control practices and the proper use of personal protective equipment (eg, mask and face shield) and were instructed to conduct interviews outside clients' front doors, standing at least 6 feet from the clients [32]. If clients were reached but could not complete phone calls or were located during a home visit but lacked privacy or space for physical distancing, CESs arranged for follow-up phone calls, encouraged clients to call Trace telephone-based tracers, or made another visit at a convenient time, within 24 hours. If clients' addresses were confirmed but they were not found during home visits, CESs left letters asking them to call the Trace call line and if needed, arranged follow-up visits within 24 hours.

If CESs reached clients via phone calls or home visits, potential outcomes were as follows: (1) "completed interview," (2) "declined to complete interview," or (3) "unable to complete interview" (eg, assigned for call-back, unable to respond, currently outside NYC, or residing in congregate facility). If CESs did not reach clients via phone calls or home visits, the outcome was recorded as "unable to locate" (eg, wrong or nonexisting address, address not confirmed, or not home). The "unable to locate" disposition option was not available for CES use from July-November 2020.

Data Analysis

We generated descriptive frequencies and proportions of the records referred to IGs and CESs from July 2020-June 2021 and summarized IG and CES workload and outcomes. Our analyses included only the records of clients who were referred to IGs or CESs for initial case and contact investigations. We deduplicated the records with multiple interviews and retained the first assignment and last outcome. We presented the proportions of select sociodemographic characteristics of clients by whether their records were ever referred to the IGs or CESs. Furthermore, we assessed the timeliness of the IG and CES activities by examining the median number of days and IQRs from the dates of referral (of cases) or identification (of contacts) to IGs or CESs to the date of initial attempt or final outcome (ie, interview or final disposition) and the date from initial attempt to final outcome. Data analysis was performed using R (version 3.5.2; R Foundation for Statistical Computing).

Ethics Approval

Contact tracing data collection is part of routine public health surveillance and intervention and was determined to be nonresearch. Contact tracing, as a public health activity, was determined not to be research, in accordance with the federal human subject's protection regulations at 45 Code of Federal Regulations 46.101c and 46.102d [33] and Centers for Disease Control and Prevention's Guidelines for Defining Public Health Research and Public Health Non-Research (protection of human subjects, US Federal Code Title 45 Part 46) [34]. Participants voluntarily participated in the activities. Informed consent from participants was not required for contact tracing interview.

Results

Characteristics of Clients Referred to IGs or CESs

Case and contact demographics stratified by referral status to IGs and CESs are described in Table 1. Overall, 266,156 of 659,484 (40%) cases and 331,483 of 861,566 (38%) contacts were ever referred to the IGs and CESs over the period of July 2020-June 2021. Most of the referred case records (155,356/266,156, 59%) were from just 2 of the 5 NYC boroughs (ie, Brooklyn and Queens).

Table 1. Select characteristics of cases and contacts ever referred or not referred to information gatherers or community engagement specialists for case or contact investigation interview from July 2020-June 2021.

Characteristics	Cases (n=659,484)		Contacts (n=861,566)	
	Referred	Not referred	Referred	Not referred
Total, n (%)	266,156 (40.36)	393,328 (59.64)	331,483 (38.47)	530,083 (61.53)
Borough, n (%)				
Bronx	45,793 (17.21)	69,025 (17.55)	35,536 (10.72)	77,882 (14.69)
Brooklyn	81,809 (30.74)	111,167 (28.26)	49,464 (14.92)	118,907 (22.43)
Manhattan	35,969 (13.51)	59,088 (15.02)	20,845 (6.29)	54,571 (10.29)
Queens	73,547 (27.63)	111,091 (28.24)	51,207 (15.45)	121,420 (22.91)
Staten Island	22,546 (8.47)	30,896 (7.86)	16,012 (4.83)	35,492 (6.70)
Unknown	6492 (2.44)	12,061 (3.07)	158,419 (47.79)	121,811 (22.98)
Race or ethnicity, n (%)				
Black (not Hispanic or Latino)	14,788 (5.56)	40,688 (10.34)	13,154 (3.97)	49,938 (9.42)
White (not Hispanic or Latino)	23,823 (8.95)	61,369 (16)	17,319 (5.22)	69,809 (13.17)
Hispanic or Latino	35,254 (13.25)	92,371 (23.48)	36,728 (11.08)	117,037 (22.08)
Asian (not Hispanic or Latino)	11,847 (4.45)	28,285 (7.19)	9131 (2.75)	32,034 (6.04)
Multiracial (not Hispanic or Latino)	775 (0.29)	2638 (0.67)	932 (0.28)	4174 (0.79)
Native Hawaiian or Pacific Islander, Native American or Alaskan Native (not Hispanic or Latino)	262 (0.10)	675 (0.17)	249 (0.08)	766 (0.14)
Did not identify with any race or ethnicity provided	2252 (0.85)	4926 (1.25)	2089 (0.63)	5725 (1.08)
Unknown	177,155 (66.56)	162,376 (41.28)	251,881 (75.99)	250,600 (47.28)
Age (years), median (IQR), range	38 (24-56), 0-117	36 (25-52), 0-111	27 (13-45), 0-109	28 (12-46), 0-109
Age group, n (%)				
0-12	24,621 (9.25)	35,504 (9.03)	48,197 (14.54)	120,006 (22.64)
13-24	43,866 (16.48)	61,191 (15.55)	42,395 (12.79)	80,119 (15.11)
25-44	90,463 (33.99)	153,299 (38.98)	57,748 (17.42)	130,253 (24.57)
45-64	69,565 (26.14)	104,014 (26.45)	39,862 (12.03)	95,072 (17.93)
≥65	37,180 (13.97)	39,255 (9.98)	12,874 (3.88)	27,023 (5.10)
Unknown	461 (0.17)	65 (0.02)	130,407 (39.34)	77,610 (14.64)
Gender identity, n (%)				
Male	126,417 (47.50)	171,135 (43.51)	40,590 (12.24)	119,129 (2247)
Female	131,747 (49.50)	203,973 (51.86)	48,442 (14.61)	158,043 (29.81)
Transgender, nonbinary, or queer	245 (0.09)	727 (0.18)	220 (0.07)	722 (0.14)
Unknown	7747 (2.91)	17,493 (4.45)	242,231 (73.07)	252,189 (47.58)
Preferred language, n (%)				
English	143,505 (53.92)	312,867 (79.54)	128,677 (38.82)	385,513 (72.73)
Spanish	25,670 (9.64)	52,958 (13.46)	32,288 (9.74)	70,264 (13.26)
Other	12,545 (4.71)	18,082 (4.60)	9048 (2.73)	15,510 (2.93)
Unknown	84,436 (31.72)	9421 (2.40)	161,470 (48.71)	58,796 (11.09)
Disability, n (%)				
Difficulty concentrating, remembering, or deciding	1260 (0.47)	4278 (1.09)	1516 (0.46)	5925 (1.12)
Difficulty doing errands	372 (0.14)	951 (0.24)	333 (0.10)	983 (0.19)
Difficulty dressing or bathing	112 (0.04)	266 (0.07)	100 (0.03)	270 (0.05)
Difficulty hearing	922 (0.35)	1269 (0.32)	654 (0.20)	1226 (0.23)

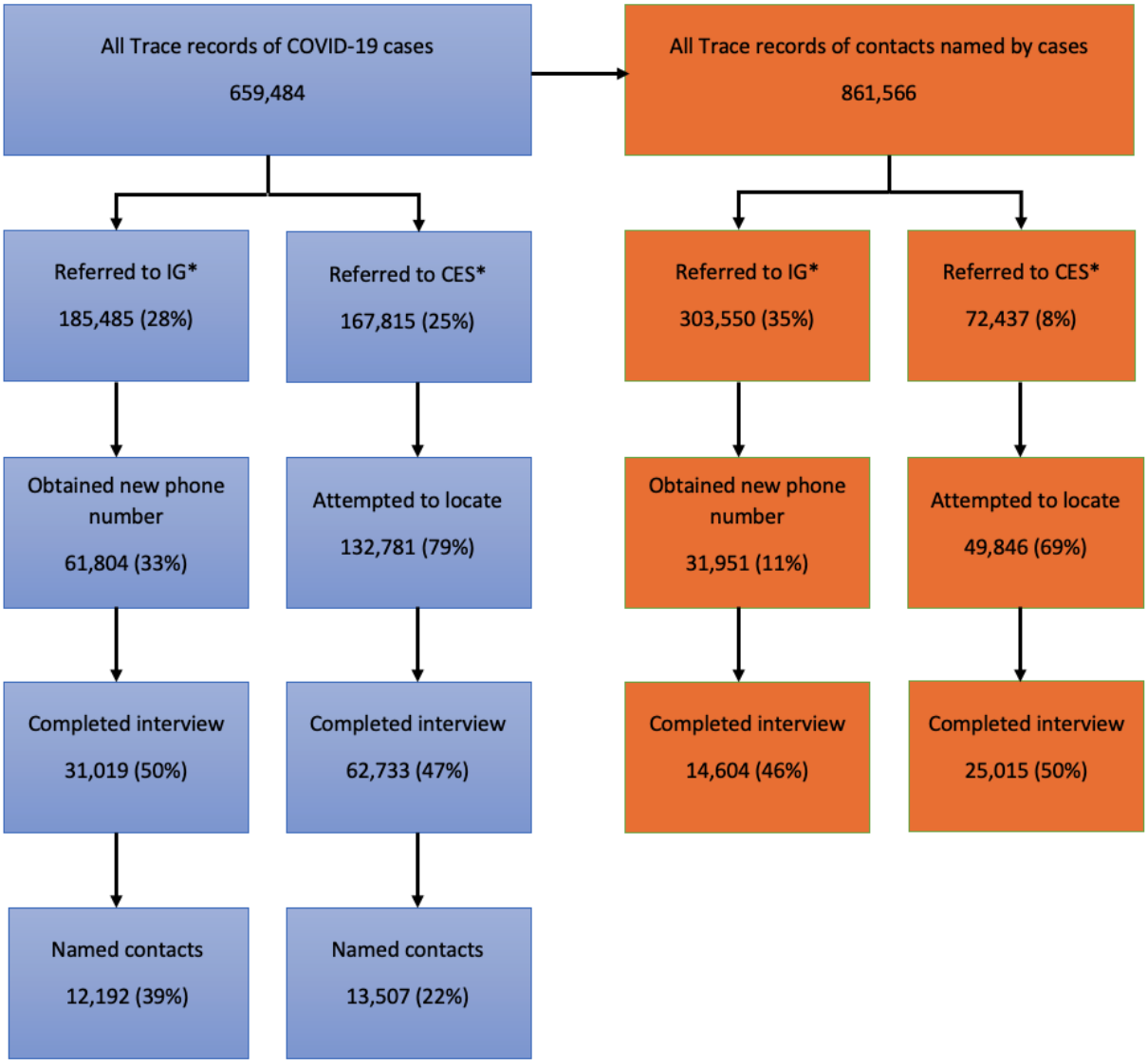
Characteristics	Cases (n=659,484)		Contacts (n=861,566)	
	Referred	Not referred	Referred	Not referred
Difficulty seeing	2336 (0.88)	5008 (1.27)	2060 (0.62)	5589 (1.05)
Difficulty walking or climbing stairs	1885 (0.71)	5577 (1.42)	137 (0.42)	4294 (0.81)
Multiple disabilities	3916 (1.47)	9031 (2.30)	3037 (0.92)	8491 (1.60)
No disability	80,968 (30.42)	199,514 (50.72)	70,946 (21.40)	223,498 (42.16)
Unknown	174,385 (66.52)	167,434 (42.57)	251,458 (75.86)	279,807 (52.79)

Workload and Outcomes of Referrals to IGs and CESs

Figure 2 depicts the numbers and proportions of clients' records referred to IGs and CESs from July 2020-June 2021 and the outcomes of those investigations.

Of the 659,484 Trace case records during this period, 185,485 (28%) were referred to IGs, and new phone numbers were obtained for 61,804 (33%) of the referred case-patient records. Subsequently, 31,019 (50%) of the case-patients with new phone numbers completed interviews, of whom 12,192 (39%) named contacts. During the same period, 303,550/861,566 (35%) contacts were referred to IGs. IGs obtained new phone numbers for 31,951 (11%) of the referred contact, of whom 14,604 (46%) completed interviews.

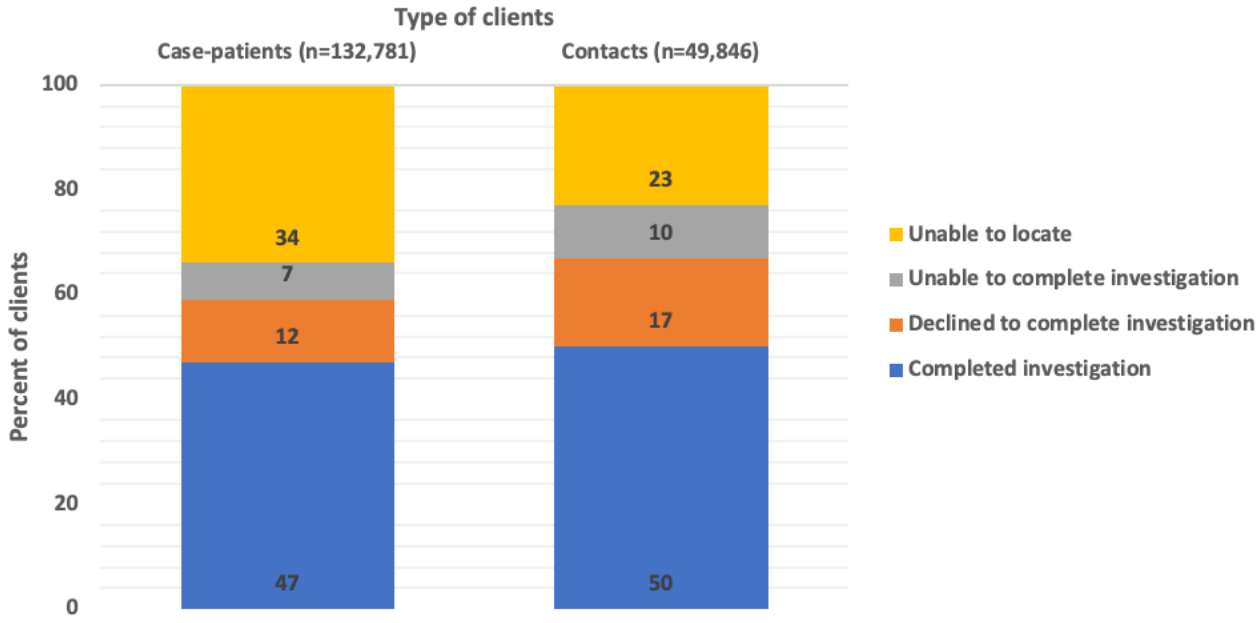
Figure 2. Workload and outcomes of referrals to information gatherers (IGs) and community engagement specialists (CESs), July 1, 2020-June 30, 2021. subsequent outcomes and proportions of subsequent steps were calculated based on the previous steps (eg, obtained new numbers were the proportions of records referred to IGs). The number of persons referred to IGs or CESs were not mutually exclusive. Some records may have been referred to both IGs and CESs work groups during the analysis period.



From July 2020-June 2021, 167,815/685,717 (24%) of Trace case records were referred to CESs. CESs attempted case investigation on 132,781 (79%) of the referrals; interviews were completed for 62,733 (47%) of the attempted referrals; and 13,507 (22%) of case-patients interviewed named contacts. Of the 861,566 contact records, 72,437 (8%) were referred to CESs. CESs investigated 49,846 (69%) of the referred contacts, and 25,015 (50%) of the contacts completed interviews.

Among the 132,781 case investigations attempted by CESs, 44,448 (34%) of case-patients sought were never located through phone calls or home visits, 9,310 (7%) and 16,290 (12%) were located but were unable or declined to complete interviews, respectively (Figure 3). Among the 49,846 contact investigations attempted, CESs did not locate 11,243 (23%); 5,104 (10%) and 8,484 (17%) of persons located were unable or declined to complete interviews, respectively.

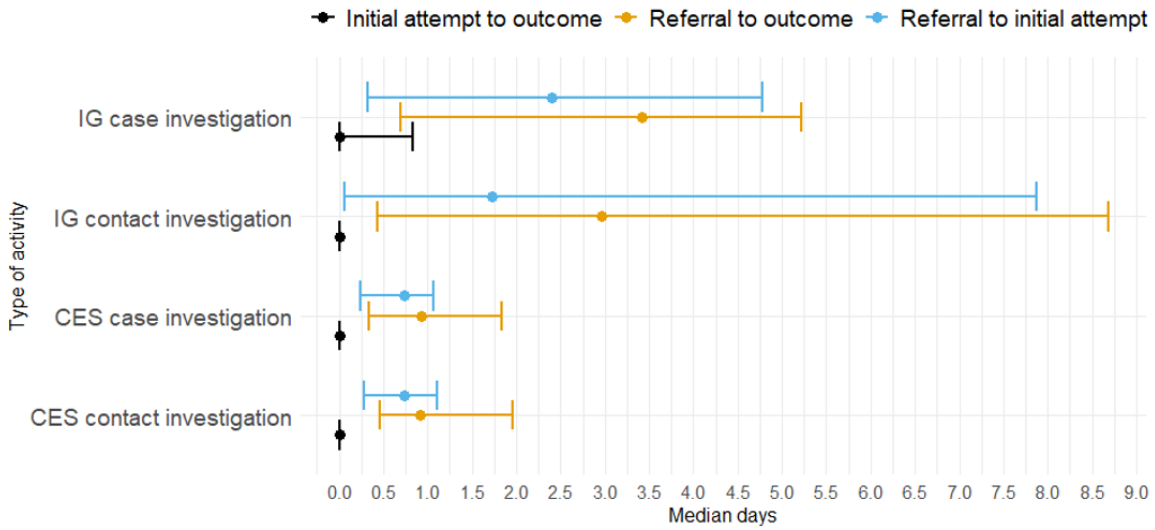
Figure 3. Outcomes of referrals attempted to locate by community engagement specialists for case and contact investigation, July 2020-June 2021. Number of cases: completed case investigation (62,733); declined to complete investigation (16,290); unable to complete investigation (9,310); unable to locate (44,448). Number of contacts: completed case investigation (25,015); declined to complete investigation (8,484); unable to complete investigation (5,104); unable to locate (11,243).



Timeliness of IG and CES activities

Among cases referred to IGs, the median interval was 2.4 (IQR 0.32-4.78) days from referral to the first attempt, 3.41 (IQR 0.7-5.22) days from referral to final outcome (eg, new phone number or declined to complete interviews), and 0 (IQR 0-0.83) days from first attempt to final outcome (Figure 4). Among contacts, the median interval was 1.72 (IQR 0.06-7.87) days from referral to IGs to first attempt, 2.96 (IQR 0.43-8.68) days from referral to final outcome, and 0 (IQR 0-0) days from first attempt to final outcome.

Figure 4. Timeliness measure of case and contact investigations referred and attempted by information gatherers (IGs) and community engagement specialists (CESs), July 2020-June 2021. Median days are from the dates of referral or initial attempts for case or contact investigations to outcomes (eg, found new number or interviewed clients). Error bars indicate IQRs of timeliness measures.



Among cases referred to CESs, the median interval was 0.74 (IQR 0.24-1.07) days from referral to first attempt to locate clients, 0.93 (IQR 0.33-1.83) days from referrals to final outcome, and 0 (IQR 0-0) days from first attempt to final outcome. Regarding contacts referred to CESs, the median intervals from referral to first attempt and to final outcome were 0.74 (IQR 0.28-1.11) days and 0.91 (IQR 0.46-1.96) days, respectively, and 0 (IQR 0-0) days from first attempts to final outcome.

Discussion

We assessed the value added by information gathering and home visit workforces to manual telephone-based contact tracing. From July 2020-June 2021, despite the NYC DOHMH's efforts to enrich the COVID-19 reports to the surveillance system with locating information from available electronic medical record sources, about 266,156/659,484 (40%) of case records and 331,483/861,566 (38%) contact records transferred to the Trace case management system lacked working telephone numbers or required home visit attempts to initiate contact tracing. This finding shows that missing locating information in reports from diagnostic providers and laboratories to public health disease surveillance systems delays or limits the already complex and multistepped manual contact tracing and supports the integration of digital proximity app-based contact tracing technique. Digital contact tracing using automated electronic information to identify individuals with new COVID-19 diagnosis and notify their exposed contacts has the potential to mitigate the lack of locating information on surveillance reports and shorten the time required for manual telephone contact notifications [8-10].

During the 1-year period of this study, the new phone numbers obtained by the IGs yielded interviews with an additional 31,019 case-patients and 14,604 contacts. The investigations attempted by the CESs added 62,733 completed interviews with case-patients and 25,015 with contacts. Furthermore, 12,192/31,019 (39%) and 13,507/63,733 (22%) of interviews with case-patients following the IG and CES interventions resulted in the identification of 12,192 and 13,507 contacts, respectively. Importantly, the median days for the completion of case and contact investigations was within 1 day of the IGs and CESs' initial attempts to find new phone numbers or locate clients. Our results support the findings of a study of multiple US jurisdictions showing the important role of case investigation and contact tracing in reaching COVID-19 case-patients and contacts to implement COVID-19 prevention measures and curb ongoing disease transmission [6].

Information gathering [35] and face-to-face interactions [26,27,32] are core features of contact tracing for other infectious diseases, such as tuberculosis, HIV, and sexually transmitted infections. An effective contact tracing program aims to reach as many case-patients as possible to identify all potentially exposed contacts and then locate, evaluate, and educate those contacts on infection control. The unprecedented high volumes of COVID-19 incident cases required mass outreach and time-sensitive contact tracing strategies accomplished with telephone calls and digital platforms. However, the populations living in dense urban conditions, such as in NYC, are often most susceptible to SARS-Cov-2 acquisition [36,37], and among them are people with limited or unreliable access to telephone or digital communication services. Furthermore, mental and physical disabilities [38,39] or reluctance to share personal confidential information with strangers over phone calls could impede contact tracing on electronic platforms alone [40-42].

Our program used a 3-pronged approach, prioritizing phone calls when possible while simultaneously searching for locating information, or as a last recourse, making home visits. This strategy offers a contact tracing model that enhances the reach and yields of a contact tracing program and promotes equitable delivery of COVID-19 interventions [20-22]. We strived to minimize mistrust and communication gaps with our clients by recruiting CESs from NYC communities heavily impacted by COVID-19 and with language skills beyond English [32]. Our approaches can be adapted to jurisdictions' resource levels and priorities. A jurisdiction could employ IGs alone to focus on obtaining missing locating information to increase case investigation and contact identification or use a small team of CESs to prioritize home visits for communities with the highest case counts or lowest response rates to telephone calls.

In addition to reaching the most people, another key factor to the success of contact tracing is the ability to reach people as quickly as possible following COVID-19 diagnosis or exposure [11,42]. Our results show that once our IGs and CESs initiated attempts to find new information or to locate clients, the median time to clients' interviews was within 1 day. Therefore, the addition of the IG and CES workflows while increasing the breadth and yield of contact tracing outcomes did not markedly delay case investigations and contact notifications. For our program, this efficiency was enabled by the integrated Trace case management system, which allowed for real-time data sharing and automated algorithms for assignments and reassignments of investigations among the telephone-based tracers, IGs, and CESs.

Although the median times from referrals to CESs to their initial attempts or final outcomes were all within 1 day, we observed longer time intervals for the IGs (2->3 days) from referral to initial attempts or final outcomes. The reason for these delays were twofold. First CESs were required to complete all investigation within 24 hours. Second, from July-November 2020, the CESs and IGs lacked the ability to assign a final disposition code of "unable to locate" to clients, and these records remained on the IG queue for further investigation. Until the final disposition code was introduced, the IGs and CESs were instructed to sort and attempt assignments based on the most current date of referral.

Face-to-face interactions between contact tracers or health care practitioners with clients can help establish rapport and build trust, thus facilitating the sharing of confidential information [42-44] Although CESs reached the vast majority of the case-patients (88,333/132,781, 66%) and contacts (38,603/49,846, 77%) sought, fairly sizable proportions of each (25,600/132,781, 19%) and 13,588/49,846, 27%, respectively) either declined to be interviewed or postponed but never completed interviews. To address these refusals, our CESs'

standard operation procedures included the routine provision of brief COVID-19 prevention education materials and information on how to receive free services (eg, testing, vaccine when it became available, and social services) and instructions on safe isolation and quarantine.

About one-third of case-patients and one-fourth of contacts sought by CESs were never reached at their available telephone numbers or addresses. Prior reports on the outcomes of home-based contact tracing for COVID-19 are lacking. The rate of nonresponse among our study population highlights the importance of augmenting manual telephone contact tracing with digital contact tracing [8-10]; promoting mass testing and vaccination [45-47]; and widespread dissemination of COVID-19 prevention education through mass media campaigns, social network sites, and community settings [32,44,48-51]. In fact, during the study period, more than half of our CES workforce were regularly mobilized to participate in the dissemination of these COVID-19 prevention information and resources in community settings [32].

Our study is subject to several limitations. First, IGs and CESs could not attempt all the referrals due to the mounting caseload and with no increase in staffing. In particular, the number of CESs available for contact tracing was the lowest during the periods of COVID-19 resurgences in NYC when many CESs were reassigned to conduct community outreach to distribute COVID-19 sanitary supplies and COVID-19 information flyers to promote community COVID-19 testing sites. Second, despite the provision of an official contact tracing letter, some laboratory staff and medical providers did not give the IGs clients' locating information, often citing the Health Insurance Portability and Accountability Act. This deficiency in public health case reporting requirement of full patient contact information and during IG follow-up impeded the completeness and timeliness of contact tracing. Third, there may have been some overlap between the number of interviews or additional contacts identified among the IG and CES outcomes. Some clients with new telephone numbers may have been forwarded to CESs for home visits. Fourth, missing data on clients' sociodemographic characteristics prevented us from assessing the potential differences between the clients who were reached by telephone-based tracers and those referred to the IGs or CESs.

Manual telephone contact tracing even when augmented with information gathering and home visits faces limitations, including being labor and time-intensive and insufficient staffing. Although digital contact tracing has the potential to rapidly notify exposed contacts and provide risk reduction information and resources, it relies on mass ownership and adoption of the digital platforms and minimal concerns of individuals for their privacy. These limitations underscore the importance of generalized COVID-19 prevention measures, such as universal self-masking, sanitary supplies, vaccination, and antiviral treatment for severe illness.

Our program's approaches demonstrate that the efforts of manual telephone-based tracers can be complemented by information gathering and in-person contact tracing to achieve increases in the number of people reached for case investigation and contact identification, and therefore, in contact notification. Missing or incomplete telephone numbers and locating information on surveillance reports initially sent to the NYC DOHMH from diagnostic providers and laboratories show the need for improvements in data collections at the time of diagnosis or the completeness of data reported by providers to health departments. In settings with limited resources for information gathering and home visits, targeted applications of these strategies could focus on geographic areas or demographics with the highest incidence of COVID-19 or low contact tracing participation rates.

Acknowledgments

The authors are grateful to the teams of information gatherers and community engagement specialists for their efforts to improve the coverage of COVID-19 contact tracing in New York City (NYC).

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. COVID-19 prevention actions: how to protect yourself and others. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html [accessed 2022-09-26]
2. Parodi SM, Liu VX. From containment to mitigation of COVID-19 in the US. JAMA 2020 Apr 21;323(15):1441-1442. [doi: 10.1001/jama.2020.3882] [Medline: 32167525]
3. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Imperial College COVID-19 Response Team, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020 Aug 08;584(7820):257-261. [doi: 10.1038/s41586-020-2405-7] [Medline: 32512579]

4. Kucharski AJ, Klepac P, Conlan AJK, Kissler SM, Tang ML, Fry H, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. Lancet Infect Dis 2020 Oct;20(10):1151-1160. [doi: 10.1016/s1473-3099(20)30457-6]
5. Rainisch G, Jeon S, Pappas D, Spencer KD, Fischer LS, Adhikari BB, et al. Estimated COVID-19 cases and hospitalizations averted by case investigation and contact tracing in the US. JAMA Netw Open 2022 Mar 01;5(3):e224042 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.4042] [Medline: 35333362]
6. Lash RR, Moonan PK, Byers BL, Bonacci RA, Bonner KE, Donahue M, COVID-19 Contact Tracing Assessment Team. COVID-19 case investigation and contact tracing in the US, 2020. JAMA Netw Open 2021 Jun 01;4(6):e2115850 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.15850] [Medline: 34081135]
7. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/overview.html [accessed 2022-05-03]
8. Bonnell TJ, Revere D, Baseman J, Hills R, Karras BT. Equity andaccessibility of Washington state's COVID-19 digital exposure notification tool (WA Notify): survey and listening sessions among community leaders. JMIR Form Res 2022 Aug 03;6(8):e38193 [FREE Full text] [doi: 10.2196/38193] [Medline: 35787520]
9. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
10. Zeng K, Bernardo SN, Havins WE. The use of digital tools to mitigate the COVID-19 pandemic: comparative retrospective study of six countries. JMIR Public Health Surveill 2020 Dec 23;6(4):e24598 [FREE Full text] [doi: 10.2196/24598] [Medline: 33302255]
11. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. The Lancet Public Health 2020 Aug;5(8):e452-e459. [doi: 10.1016/s2468-2667(20)30157-2]
12. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect Dis 2020 Aug;20(8):911-919 [FREE Full text] [doi: 10.1016/s1473-3099(20)30287-5]
13. Al-Sadeq DW, Nasrallah GK. The incidence of the novel coronavirus SARS-CoV-2 among asymptomatic patients: a systematic review. Int J Infect Dis 2020 Sep;98:372-380 [FREE Full text] [doi: 10.1016/j.ijid.2020.06.098] [Medline: 32623083]
14. Jefferson T, Spencer EA, Brassey J, Onakpoya IJ, Rosca EC, Plüddemann A, et al. Transmission of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) from pre and asymptomatic infected individuals: a systematic review. Clin Microbiol Infect 2022 Feb;28(2):178-189 [FREE Full text] [doi: 10.1016/j.cmi.2021.10.015] [Medline: 34757116]
15. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, Hossmann S, Imeri H, Ipekci AM, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis. PLoS Med 2020 Sep 22;17(9):e1003346 [FREE Full text] [doi: 10.1371/journal.pmed.1003346] [Medline: 32960881]
16. Ruebush E, Fraser M, Poulin A, Allen M, Lane J, Blumenstock J. COVID-19 case investigation and contact tracing: early lessons learned and future opportunities. J Public Health Manag Pract 2021;27 Suppl 1, COVID-19 and Public Health: Looking Back, Moving Forward:S87-S97. [doi: 10.1097/PHH.0000000000001290] [Medline: 33239569]
17. Groves R, Travis-Bassett M, Hout M. Encouraging Participation and Cooperation in Contact Tracing: Lessons From Survey Research. Washington, DC: The National Academies Press; 2020.
18. Kalyanaraman N, Fraser MR. Containing COVID-19 through contact tracing : a local health agency approach. Public Health Rep 2021 Nov 10;136(1):32-38 [FREE Full text] [doi: 10.1177/0033354920967910] [Medline: 33170094]
19. Leser K, Hay M, Henebry B, Virden J, Patel M, Luttrell-Freeman J, et al. An academic-health department community partnership to expand disease investigation and contact tracing capacity and efficiency during the COVID-19 pandemic. J Public Health Manag Pract 2022;28(1):E16-E22. [doi: 10.1097/PHH.0000000000001379] [Medline: 34016907]
20. Mathevet I, Ost K, Traverson L, Zinszer K, Ridde V. Accounting for health inequities in the design of contact tracing interventions: a rapid review. Int J Infect Dis 2021 May;106:65-70 [FREE Full text] [doi: 10.1016/j.ijid.2021.03.010] [Medline: 33716194]
21. Benjamin GC. Ensuring health equity during the COVID-19 pandemic: the role of public health infrastructure. Rev Panam Salud Publica 2020:1-4. [doi: 10.26633/rpsp.2020.70]
22. Perry BL. Contact tracing could exacerbate COVID-19 health disparities: the role of economic precarity and stigma. Am J Public Health 2021 May;111(5):778-781. [doi: 10.2105/ajph.2021.306244]
23. Sun W, Zhou Y, Chen W, Huang F, Sun M, Shen L, et al. Disclosure experience among COVID-19-confirmed patients in China: a qualitative study. J Clin Nurs 2021 Mar 15;30(5-6):783-792 [FREE Full text] [doi: 10.1111/jocn.15616] [Medline: 33349988]
24. Zhang B, Kreps S, McMurry N, McCain RM. Americans' perceptions of privacy and surveillance in the COVID-19 pandemic. PLoS One 2020 Dec 23;15(12):e0242652 [FREE Full text] [doi: 10.1371/journal.pone.0242652] [Medline: 33362218]

25. Shelby T, Hennein R, Schenck C, Clark K, Meyer AJ, Goodwin J, et al. Implementation of a volunteer contact tracing program for COVID-19 in the United States: a qualitative focus group study. PLoS One 2021 May 5;16(5):e0251033 [FREE Full text] [doi: 10.1371/journal.pone.0251033] [Medline: 33951107]
26. Thorpe L. Local acts. Public Health Rep 2012 Jan 01;127(1):107-114 [FREE Full text] [doi: 10.1177/003335491212700112] [Medline: 22298929]
27. Khatana GH, Haq I, Khan SS. Effectiveness, acceptance and feasibility of home-based intervention model for tuberculosis contact tracing in Kashmir. J Clin Tuberc Other Mycobact Dis 2019 Feb;14:19-25 [FREE Full text] [doi: 10.1016/j.jctube.2019.01.001] [Medline: 31720414]
28. New York City Health + Hospitals. URL: https://www.nychealthandhospitals.org/test-and-trace [accessed 2021-06-24]
29. Coronavirus disease 2019 (COVID-19) case definition. Centers for Disease Control and Prevention. URL: https://ndc.services.cdc.gov/case-definitions/coronavirus-disease-2019-2021/ [accessed 2022-09-22]
30. Centers for Disease Control. Contact tracing for COVID-19: close contact through proximity and duration of exposure. Accessed September 22, 2022. URL: https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/appendix.html#contact [accessed 2022-09-22]
31. Scientific brief: SARS-CoV-2 transmission. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/sars-cov-2-transmission.html [accessed 2022-09-22]
32. Udeagu CN, Huang J, Misra K, Terilli T, Ramos Y, Alexander M, et al. Community-based workforce for COVID-19 contact tracing and prevention activities in New York City, July-December 2020. Public Health Rep 2022 Jul 21:333549221110833 [FREE Full text] [doi: 10.1177/00333549221110833] [Medline: 35861302]
33. 45 code of federal regulations 46. US Department of Health and Human Services. 2021. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html [accessed 2022-11-09]
34. Code of federal regulations. Part 46 - protection of human subjects. National Archives. URL: https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46 [accessed 2022-11-09]
35. Padilla M, Mattson CL, Scheer S, Udeagu CN, Buskin SE, Hughes AJ, et al. Locating people diagnosed with HIV for public health action: utility of HIV case surveillance and other data sources. Public Health Rep 2018;133(2):147-154. [doi: 10.1177/0033354918754541] [Medline: 29486143]
36. Sy KTL, White LF, Nichols BE. Population density and basic reproductive number of COVID-19 across United States counties. PLoS One 2021 Apr 21;16(4):e0249271 [FREE Full text] [doi: 10.1371/journal.pone.0249271] [Medline: 33882054]
37. Tsori Y, Granek R. Epidemiological model for the inhomogeneous spatial spreading of COVID-19 and other diseases. PLoS One 2021 Feb 19;16(2):e0246056 [FREE Full text] [doi: 10.1371/journal.pone.0246056] [Medline: 33606684]
38. Agaronnik N, Campbell EG, Ressalam J, Iezzoni LI. Communicating with patients with disability: perspectives of practicing physicians. J Gen Intern Med 2019 Jul 18;34(7):1139-1145 [FREE Full text] [doi: 10.1007/s11606-019-04911-0] [Medline: 30887435]
39. Hossain MM, Tasnim S, Sultana A, Faizah F, Mazumder H, Zou L, et al. Epidemiology of mental health problems in COVID-19: a review. F1000Res 2020 Jun 23;9:636 [FREE Full text] [doi: 10.12688/f1000research.24457.1] [Medline: 33093946]
40. Otterman S. N.Y.C. hired 3,000 workers for contact tracing. It's off to a slow start. The New York Times. 2020 Jun 21. URL: https://www.nytimes.com/2020/06/21/nyregion/nyc-contact-tracing.html [accessed 2021-06-24]
41. von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, et al. Drivers of acceptance of COVID-19 proximity tracing apps in Switzerland: panel survey analysis. JMIR Public Health Surveill 2021 Jan 06;7(1):e25701 [FREE Full text] [doi: 10.2196/25701] [Medline: 33326411]
42. Dowthwaite L, Fischer J, Perez Vallejos E, Portillo V, Nichele E, Goulden M, et al. Public adoption of and trust in the NHS COVID-19 contact tracing app in the United Kingdom: quantitative online survey study. J Med Internet Res 2021 Sep 17;23(9):e29085 [FREE Full text] [doi: 10.2196/29085] [Medline: 34406960]
43. Sousa S, Kalju T. Modeling trust in COVID-19 contact-tracing apps using the human-computer Trust Scale: online survey study. JMIR Hum Factors 2022 Jun 13;9(2):e33951 [FREE Full text] [doi: 10.2196/33951] [Medline: 35699973]
44. Bologna L, Stamidis K, Paige S, Solomon R, Bisrat F, Kisanga A, et al. Why communities should be the focus to reduce stigma attached to COVID-19. Am J Trop Med Hyg 2021 Jan;104(1):39-44. [doi: 10.4269/ajtmh.20-1329] [Medline: 33258438]
45. Massetti GM, Jackson BR, Brooks JT, Perrine CG, Reott E, Hall AJ, et al. Summary of guidance for minimizing the impact of COVID-19 on individual persons, communities, and health care systems - United States, August 2022. MMWR Morb Mortal Wkly Rep 2022 Aug 19;71(33):1057-1064 [FREE Full text] [doi: 10.15585/mmwr.mm7133e1] [Medline: 35980866]
46. James A, Plank MJ, Hendy S, Binny R, Lustig A, Steyn N, et al. Successful contact tracing systems for COVID-19 rely on effective quarantine and isolation. PLoS One 2021 Jun 3;16(6):e0252499 [FREE Full text] [doi: 10.1371/journal.pone.0252499] [Medline: 34081709]
47. Peeling RW, Heymann DL, Teo Y, Garcia PJ. Diagnostics for COVID-19: moving from pandemic response to control. The Lancet 2022 Feb;399(10326):757-768. [doi: 10.1016/s0140-6736(21)02346-1]

48. Gilmore B, Ndejjo R, Tchetchia A, de Claro V, Mago E, Diallo AA, et al. Community engagement for COVID-19 prevention and control: a rapid evidence synthesis. BMJ Glob Health 2020 Oct;5(10):e003188 [FREE Full text] [doi: 10.1136/bmjgh-2020-003188] [Medline: 33051285]
49. Zhang Y, Tambo E, Djuikoue IC, Tazemda GK, Fotsing MF, Zhou X. Early stage risk communication and community engagement (RCCE) strategies and measures against the coronavirus disease 2019 (COVID-19) pandemic crisis. Glob Health J 2021 Mar;5(1):44-50 [FREE Full text] [doi: 10.1016/j.glohj.2021.02.009] [Medline: 33850632]
50. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593 [FREE Full text] [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
51. Anwar A, Malik M, Raees V, Anwar A. Role of mass media and public health communications in the COVID-19 pandemic. Cureus 2020 Sep 14;12(9):e10453 [FREE Full text] [doi: 10.7759/cureus.10453] [Medline: 33072461]

Abbreviations

CES: community engagement specialist
DOHMH: Department of Health and Mental Hygiene
IG: information gatherer
NYC: New York City

Edited by A Mavragani; submitted 11.07.22; peer-reviewed by M Fantini, V von Wyl; comments to author 07.09.22; revised version received 27.09.22; accepted 13.10.22; published 15.11.22.

Please cite as:
Udeagu CCN, Pitiranggon M, Misra K, Huang J, Terilli T, Ramos Y, Alexander M, Kim C, Lee D, Blaney K, Keeley C, Long T, Vora NM
Outcomes of a Community Engagement and Information Gathering Program to Support Telephone-Based COVID-19 Contact Tracing: Descriptive Analysis
JMIR Public Health Surveill 2022;8(11):e40977
URL: https://publichealth.jmir.org/2022/11/e40977
doi:10.2196/40977
PMID:36240019

Original Paper

Incidence and Prevalence of Peripheral Arterial Disease in South Korea: Retrospective Analysis of National Claims Data

Gi Wook Ryu[1,2,3], RN, PhD; Young Shin Park[2], RN, PhD; Jeewuan Kim[4], MA; Yong Sook Yang[3], RN, PhD; Young-Guk Ko[5*], MD, PhD; Mona Choi[2*], RN, PhD

[1]Department of Nursing, Hansei University, Gunpo-si, Republic of Korea

[2]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea

[3]College of Nursing, Yonsei University, Seoul, Republic of Korea

[4]Department of Statistics and Data Science, Yonsei University, Seoul, Republic of Korea

[5]Division of Cardiology, Severance Cardiovascular Hospital, College of Medicine, Yonsei University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Mona Choi, RN, PhD
Mo-Im Kim Nursing Research Institute
College of Nursing
Yonsei University
50 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3341
Fax: 82 2 2227 8303
Email: monachoi@yuhs.ac

Abstract

Background: Peripheral arterial disease (PAD) causes blood vessel narrowing that decreases blood flow to the lower extremities, with symptoms such as leg pain, discomfort, and intermittent claudication. PAD increases risks for amputation, poor health-related quality of life, and mortality. It is estimated that more than 200 million people worldwide have PAD, although the paucity of PAD research in the East detracts from knowledge on global PAD epidemiology. There are few national data–based analyses or health care utilization investigations. Thus, a national data analysis of PAD incidence and prevalence would provide baseline data to enable health promotion strategies for patients with PAD.

Objective: This study aims to identify South Korean trends in the incidence and prevalence of PAD and PAD treatment, in-hospital deaths, and health care utilization.

Methods: This was a retrospective analysis of South Korean national claims data from 2009 to 2018. The incidence of PAD was determined by setting the years 2010 and 2011 as a washout period to exclude previously diagnosed patients with PAD. The study included adults aged ≥20 and <90 years who received a primary diagnosis of PAD between 2011 and 2018; patients were stratified according to age, sex, and insurance status for the incidence and prevalence analyses. Descriptive statistics were used to assess incidence, prevalence, endovascular revascularization (EVR) events, amputations, in-hospital deaths, and the health care utilization characteristics of patients with PAD.

Results: Based on data from 2011 to 2018, there were an average of 124,682 and 993,048 incident and prevalent PAD cases, respectively, in 2018. PAD incidence (per 1000 persons) ranged from 2.68 to 3.09 during the study period. From 2012 to 2018, the incidence rate in both sexes showed an increasing trend. PAD incidence continued to increase with age. PAD prevalence (per 1000 persons) increased steadily, from 3.93 in 2011 to 23.55 in 2018. The number of EVR events varied between 933 and 1422 during the study period, and both major and minor amputations showed a decreasing trend. Health care utilization characteristics showed that women visited clinics more frequently than men, whereas men used tertiary and general hospitals more often than women.

Conclusions: The number of incident and prevalent PAD cases generally showed an increasing trend. Visits to tertiary and general hospitals were higher among men than women. These results indicate the need for attention not only to Western and male

patients, but also to Eastern and female patients with PAD. The results are generalizable, as they are based on national claims data from the entire South Korean population, and they can promote preventive care and management strategies for patients with PAD in clinical and public health settings.

(JMIR Public Health Surveill 2022;8(11):e34908) doi:10.2196/34908

KEYWORDS

peripheral arterial disease; insurance claims; incidence; prevalence; endovascular revascularization; amputation; population-based study; blood flow; intermittent claudication; age; sex

Introduction

Peripheral arterial disease (PAD) is a major vascular condition that decreases blood flow to the affected limbs; it is mostly caused by atherosclerosis, a progressive disease characterized by the intra-arterial accumulation of lipids and fibrous elements [1,2]. The PAD symptoms of claudication and critical limb ischemia (CLI) occur following the reduction of blood flow to affected limbs, with resultant resting pain and cramps [3-5]. Worldwide, more than 236 million people are affected by PAD, and the PAD burden could increase with population aging [6,7], as PAD prevalence consistently and globally increases with age, especially in older age groups [6,7]. In the United States, treatment of CLI symptoms in older patients (aged >65 years) with PAD incurs an estimated cost of US $1.2 billion yearly [8,9].

Vessel patency in the affected limb is essential for adequate blood flow, as vessel obstruction increases risk for amputation, mortality, and poor health-related quality of life [10-12]. Endovascular revascularization (EVR) by percutaneous transluminal angiography (PTA) is the preferred method to open affected vessels, thereby improving the clinical manifestations of claudication and CLI [13] and reducing major amputations of the affected limbs [14]. However, the prognosis of the surgical procedure is associated with procedural characteristics, such as method and target region, and patient characteristics, such as age, smoking, and comorbidities [15]. A systematic review revealed that major amputation events after surgical intervention were significantly related to comorbidities, such as cardiovascular disease, chronic kidney disease, diabetes, chronic occlusive pulmonary disease, dementia, and frailty [16].

Some patients with PAD have atypical presentation, without intermittent claudication or clear limb symptoms [4,9], and may attempt to alleviate limb symptoms by reducing physical activity, which may eventually cause a worse prognosis [4,17]. Thus, patients with asymptomatic PAD may not be properly diagnosed and may not receive adequate treatment [18,19]. Furthermore, chronic diseases such as PAD affect psychological well-being by inducing depression, anxiety, and low quality of life [20,21]. Pain and difficulty in walking distances and climbing stairs in patients with PAD are significantly related to quality of life [22] and well-being [23].

Understanding trends in the incidence, prevalence, and clinical manifestations of PAD and related procedures, treatments, and health outcomes are crucial for public health interventions. Thus, identifying the incidence and prevalence of PAD using recent national data may provide baseline data to facilitate the development of health promotion strategies and interventions for patients with PAD and public health promotion. However, most previous studies have examined the incidence and prevalence in Western countries [1,24], and PAD has been studied only as part of atherosclerotic disease [25].

Currently, studies on PAD in Eastern countries are scarce, which limits understanding of the global features of PAD. Moreover, few studies have investigated national data on health care utilization characteristics. This study used nationwide data obtained from the Health Insurance Review and Assessment (HIRA) Service of South Korea from 2011 to 2018 to investigate (1) trends in the incidence, prevalence, and treatment of PAD (eg, EVR events) and PAD-related amputations and in-hospital deaths and (2) health care utilization characteristics of patients with PAD.

Methods

Ethical Considerations

This study was reviewed by the Yonsei University health system institutional review board (Y-2019-0105) and was conducted using secondary data analysis with a descriptive study design. This study used South Korea–specific research data obtained by HIRA (M20190923977).

Data Source

We acquired data from the HIRA database for patients with PAD from January 1, 2009, to December 31, 2018. The National Health Insurance system in Korea is a single-payer system that covers 98% of the total population. More than 99% of medical institutions are mandatorily included in the system, and the HIRA collects claims data to reimburse health care providers [26]; these data cover all South Korean citizens and can be used as anonymized information on diagnoses, procedures, prescription records, demographic information, and direct medical costs [26,27].

Study Population

Patients with PAD were defined as those with the following Korean Standard Classification of Diseases, 7th revision (KCD-7) codes: I70.2, I73.9, I73.9, I74.3, I74.4, I74.5, I74.8, and I74.9; these are primary PAD diagnoses (Multimedia Appendix 1). The KCD-7 codes were developed in Korea based on the International Classification of Diseases, Tenth Revision (ICD-10) codes, and the KCD-7 codes for PAD are identical to the ICD-10 codes. We selected the codes by referring to published studies that analyzed PAD-related data with similar codes [24,25,28]. Adolescents (ie, those aged 19 years or younger) were excluded. Adult patients (aged ≥20 and <90

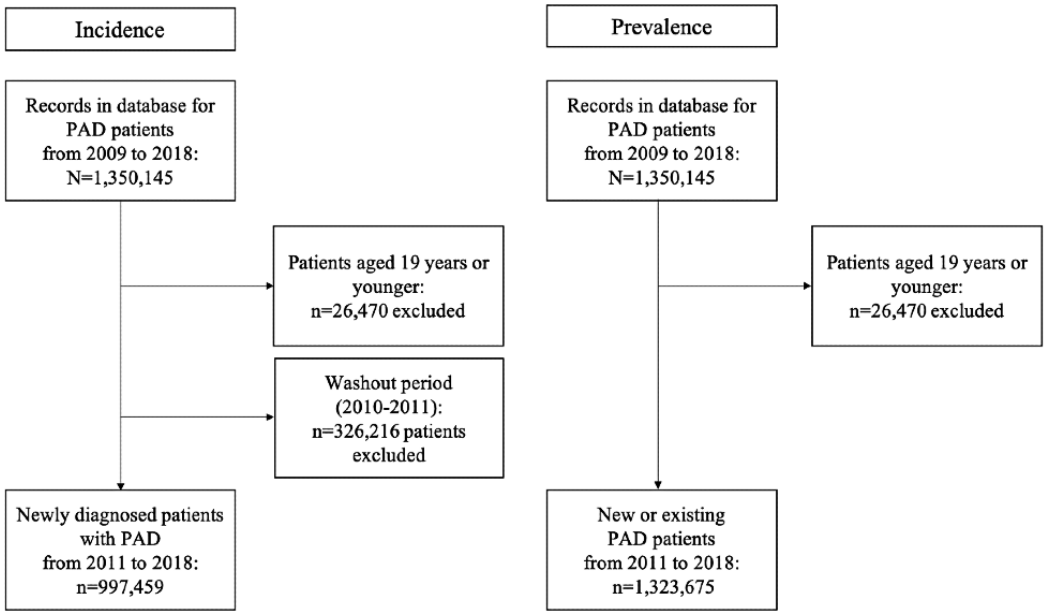
years) who were diagnosed with PAD by a physician between January 1, 2009, and December 31, 2018, as outpatients or inpatients at health facilities ranging from clinics to tertiary hospitals were enrolled.

Trends in Incidence and Prevalence

For PAD incidence, data from 2009 to 2018 were used. Data from patients treated for PAD from January 1, 2009, to December 31, 2010, were excluded to identify newly diagnosed patients. In general, the incidence of chronic diseases, such as diabetes and PAD, is calculated after excluding data from a 2-year washout period [14], and the same method was used in this study. The index date was defined as the date between 2011 and 2018 on which a patient was first diagnosed with PAD, and these dates were analyzed to determine PAD incidence. PAD prevalence was ascertained from data from 2011 to 2018 to identify patients treated for PAD every year. The index date was determined as the date on which a patient was first diagnosed for every year from 2011 to 2018 (Figure 1).

The annual number of incident and prevalent PAD cases and PAD incidence and cumulative prevalence were assessed. PAD incidence and prevalence are reported as the number of patients with PAD per 1000 individuals.

Figure 1. Selection of patients with peripheral arterial disease using Health Insurance Review and Assessment data. PAD: peripheral arterial disease.



Trends in Treatment and Deaths

To identify the annual number of EVR events, open surgical procedures, amputations, and in-hospital deaths, the numbers of cases from January 1, 2011, to December 31, 2018, were assessed based on the prevalence database used in this study. The codes for EVR events, surgical procedures, and amputations were selected based on a previous study [25] (Multimedia Appendix 2). EVR events included PTA, stent grafts, and atherectomies, whereas amputations included major and minor amputations. In-hospital death was assessed based on the results of medical treatment for patients diagnosed with PAD from January 1, 2011, to December 31, 2018.

Health Care Use Characteristics

To identify health care use characteristics, we assessed all claims from January 1, 2011, to December 31, 2018, and grouped them to identify the number of visits to tertiary hospitals, general hospitals, small hospitals, long-term care facilities, and clinics. Tertiary hospitals are defined as large hospitals with more than 20 medical departments, with each department having relevant specialists. General hospitals are defined as having 100 or more beds, and hospitals are defined as small hospitals with 30 or more beds. Long-term care facilities provide medical and nursing care for inpatients and outpatients. Clinics provide treatment and care to outpatients.

Statistical Analysis

The number of incident PAD cases was the number of patients who were newly diagnosed with PAD. The annual PAD incidence was calculated as the number of newly diagnosed patients with PAD in a year divided by the size of the population at risk. The "population at risk" for this calculation was defined by excluding preexisting patients with PAD from the midyear population [29,30]. As we had already excluded patients from 2010 from the analysis, the incidence could not be calculated for 2011, and this study therefore only analyzed incidence from 2012 to 2018.

The number of prevalent cases was the number of patients who were previously or newly diagnosed with PAD and underwent treatment, whereas PAD prevalence was the total accumulated number of patients with PAD every year. In this study, census data were used for the total population of South Korea.

Incident cases, incidence, prevalent cases, prevalence, EVR events, open surgical procedures, amputations, in-hospital deaths, and health care utilization characteristics were analyzed with descriptive statistics. The frequency of PAD incident cases was stratified according to age, sex, and insurance status. Incidence was adjusted by age and sex using a standardization method that calculates a weight based the study population for

the year 2011 [31]. The incidence trends were analyzed according to sex and age.

Changes in the frequency of EVR events, open surgical procedures, and amputations per year were assessed to determine treatment trends. For the analysis of health care utilization, all claims were grouped by sex to identify the number of visits to tertiary hospitals, general hospitals, small hospitals, long-term care facilities, and clinics.

SAS (version 9.3; SAS Institute, Inc) and R (2020 version, R Foundation for Statistical Computing) were used for statistical analysis.

Results

Demographics

A total of 997,459 new patients with PAD from 2011 to 2018 were identified. In 2011 and 2018, the numbers of new patients with PAD were 117,876 and 142,211, respectively. The total number of new female patients was 603,788 (60.5%), which was greater than the number of new male patients (n=393,671, 39.5%). Among patients who were newly diagnosed with PAD during the study period, those in their 50s were the most common by age at 242,425 (24.3%). In 2018, the number of prevalent PAD cases was 993,048. From 2011 to 2018, the number of prevalent PAD cases consistently increased (Table 1).

Table 1. Incident and prevalent cases of peripheral arterial disease from 2011 to 2018 in South Korea. Data represent the number of patients with peripheral arterial disease. Percentages are based on incident cases of peripheral arterial disease.

	2011	2012	2013	2014	2015	2016	2017	2018	Total (2011-2018)	Average
Incident cases, n	117,876	102,153	111,345	116,025	125,674	139,191	142,984	142,211	997,459	124,682
Incident cases by sex , n (%)										
Male	46,399 (39.4)	41,601 (40.7)	44,484 (40)	45,503 (39.2)	49,185 (39.1)	53,495 (38.4)	55,981 (39.2)	57,023 (40.1)	393,671 (39.5)	49,209
Female	71,477 (60.6)	60,552 (59.3)	66,861 (60)	70,522 (60.8)	76,489 (60.9)	85,696 (61.6)	87,003 (60.9)	85,188 (59.9)	603,788 (60.5)	75,474
Incident cases by age group (years), n (%)										
20s	3527 (3.0)	2967 (2.9)	3090 (2.7)	3044 (2.7)	3415 (2.7)	3863 (2.8)	3853 (2.7)	4086 (2.9)	27,845 (2.8)	3481
30s	7451 (6.3)	6125 (6.0)	6375 (5.7)	6172 (5.3)	6464 (5.1)	7215 (5.2)	7249 (5.1)	7100 (5.0)	54,151 (5.4)	6769
40s	17,343 (14.7)	14,295 (14.0)	14,982 (13.5)	15,087 (13.0)	15,896 (12.6)	17,069 (12.3)	17,101 (12.0)	16,177 (11.4)	127,950 (12.8)	15,994
50s	29,491 (25.0)	25,948 (25.4)	27,616 (24.8)	28,696 (24.7)	30,470 (24.2)	33,346 (24.0)	33,748 (23.6)	33,110 (23.3)	242,425 (24.3)	30,303
60s	27,907 (23.7)	23,870 (23.4)	25,765 (23.1)	26,873 (23.2)	29,617 (23.6)	34,356 (24.7)	35,952 (25.1)	36,541 (25.7)	240,881 (24.1)	30,110
70s	24,647 (20.9)	22,185 (21.7)	25,617 (23.0)	26,834 (23.1)	29,117 (23.2)	31,344 (22.5)	31,977 (22.4)	31,759 (22.3)	223,480 (22.4)	27,935
80s	7510 (6.4)	6763 (6.6)	7946 (7.1)	9273 (8.0)	10,695 (8.5)	11,998 (8.6)	13,104 (9.2)	13,438 (9.4)	80,727 (8.1)	10,091
Incident cases by insurance status, n (%)										
Health insurance	108,481 (92.0)	94,597 (92.6)	103,333 (92.8)	107,898 (93.0)	116,467 (92.7)	128,866 (92.6)	132,635 (92.8)	132,165 (92.9)	924,442 (92.7)	115,555
Medical aid	9250 (7.8)	7458 (7.3)	7911 (7.1)	8032 (6.9)	9133 (7.3)	10,248 (7.4)	10,219 (7.1)	9984 (7.0)	72,235 (7.2)	9029
Veteran	145 (0.1)	98 (0.1)	101 (0.1)	95 (0.1)	74 (0.1)	77 (0.1)	130 (0.1)	62 (0.0)	782 (0.1)	98
Prevalent cases[a], n	154,296	256,449	367,794	474,627	595,621	725,163	854,630	993,048	993,048	N/A[b]

[a]Prevalent cases refers to patients who were undergoing treatment after diagnosis of peripheral arterial disease; the values are accumulated values.

[b]N/A: not applicable.

Overall Trends in Incidence and Prevalence

In 2012, the total PAD incidence per 1000 patients was 2.92, and the absolute change was 0.07 between 2012 and 2018. The trend did not noticeably increase or decrease (Table 2).

The total incidence trend per year showed an increase from 2012 to 2018 (Figure 2).

The incidence in men increased from 2.12 per 1000 individuals in 2012 to 2.73 per 1000 individuals in 2018, for an absolute

increase of 0.61. In women, the incidences in 2012 and 2018 were 3.04 and 4.03 per 1000 individuals, respectively, for an absolute increase of 0.99. From 2012 to 2018, the incidence trend was consistently higher in women than men (Figure 3A).

PAD incidence continued to increase with age from 20 to 70 years, and the average incidence among those in their 80s or older was higher than among those in their 70s. Among individuals in their 80s, PAD incidence in 2012 and 2018 was 7.84 and 8.94, respectively, for an absolute increase of 1.10, which was the highest among all age groups. In 2012 and 2018, the prevalence was 6.46 and 23.55, respectively, representing a consistently increasing trend. As shown in Figure 3B, the slope showed an increasing trend without a plateau.

Table 2. Incidence and prevalence per 1000 individuals. Peripheral arterial disease incidence and prevalence are based on the number of incident cases and the overall South Korean population. Incidence was adjusted by age and sex.

Year	2011	2012	2013	2014	2015	2016	2017	2018	AC[a] (2012-2018)	Average (2012-2018)
Incidence	N/A[b]	2.92	2.68	2.71	2.86	3.09	3.09	3.00	+0.07	2.91
Incidence by sex										
Male	N/A	2.12	2.21	2.26	2.42	2.60	2.70	2.73	+0.61	2.43
Female	N/A	3.04	3.28	3.46	3.72	4.12	4.15	4.03	+0.99	3.69
Incidence by age (years)										
20s	N/A	0.53	0.46	0.47	0.51	0.57	0.57	0.60	+0.07	0.53
30s	N/A	0.91	0.80	0.79	0.84	0.96	0.98	0.98	+0.07	0.89
40s	N/A	1.97	1.68	1.69	1.80	1.94	1.97	1.91	–0.06	1.85
50s	N/A	3.80	3.45	3.50	3.67	3.97	3.99	3.86	+0.06	3.75
60s	N/A	6.49	5.79	5.74	5.87	6.43	6.39	6.18	–0.31	6.13
70s	N/A	8.37	8.40	8.60	9.24	9.76	9.55	9.16	+0.79	9.01
80s	N/A	7.84	7.75	8.31	8.80	9.13	9.31	8.94	+1.10	8.58
Prevalence	3.93	6.46	9.17	11.70	14.52	17.49	20.43	23.55	+19.62	13.41[c]

[a]AC: absolute change.

[b]N/A: not applicable.

[c]Average prevalence is the average from 2011 to 2018.

Figure 2. Trends in incidence and prevalence. Incidence was adjusted by age and sex.

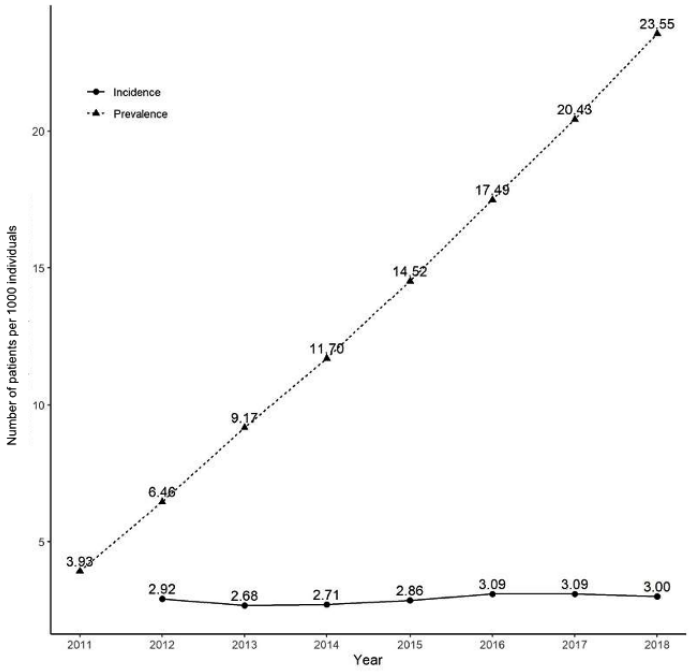
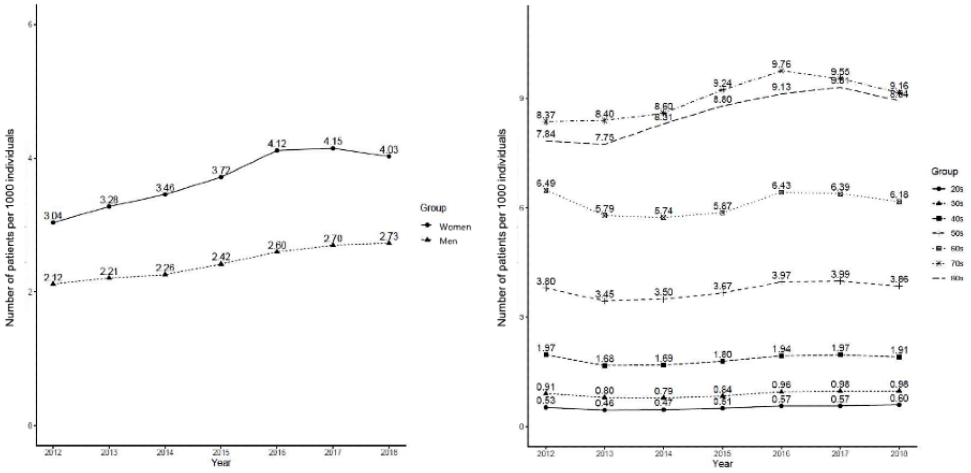


Figure 3. Trends in incidence by sex and age groups.



(A) Trend in incidence by sex

(B) Trend in incidence by age groups

Trends in Treatment and Death

From 2011 to 2018, the number of EVR events showed fluctuations (Figure 4).

In 2011, 933 EVR events were observed, increasing to 1206 cases in 2018, an absolute increase of 273 cases. In the same period, amputations decreased from 143 to 89, an absolute decrease of 54 cases. Major amputations decreased from 61 in 2011 to 35 in 2018, and minor amputations decreased from 82 in 2011 to 54 in 2018.

In-hospital deaths decreased from 95 in 2011 to 46 in 2018, an absolute decrease of 49. The number of in-hospital deaths was greater within 7 days than between 30 and 90 days. In 2011, 53 and 89 in-hospital deaths occurred within 7 and 30 days, respectively. In-hospital deaths within 30 days included deaths within 0 days, and in 2011, there were 36 deaths between 7 and 30 days. In 2018, 33 and 44 in-hospital deaths were observed within 7 days and 30 days, respectively (Table 3).

Figure 4. Trends in the annual numbers of endovascular revascularization events, open surgical procedures, amputations, and all-cause in-hospital deaths.

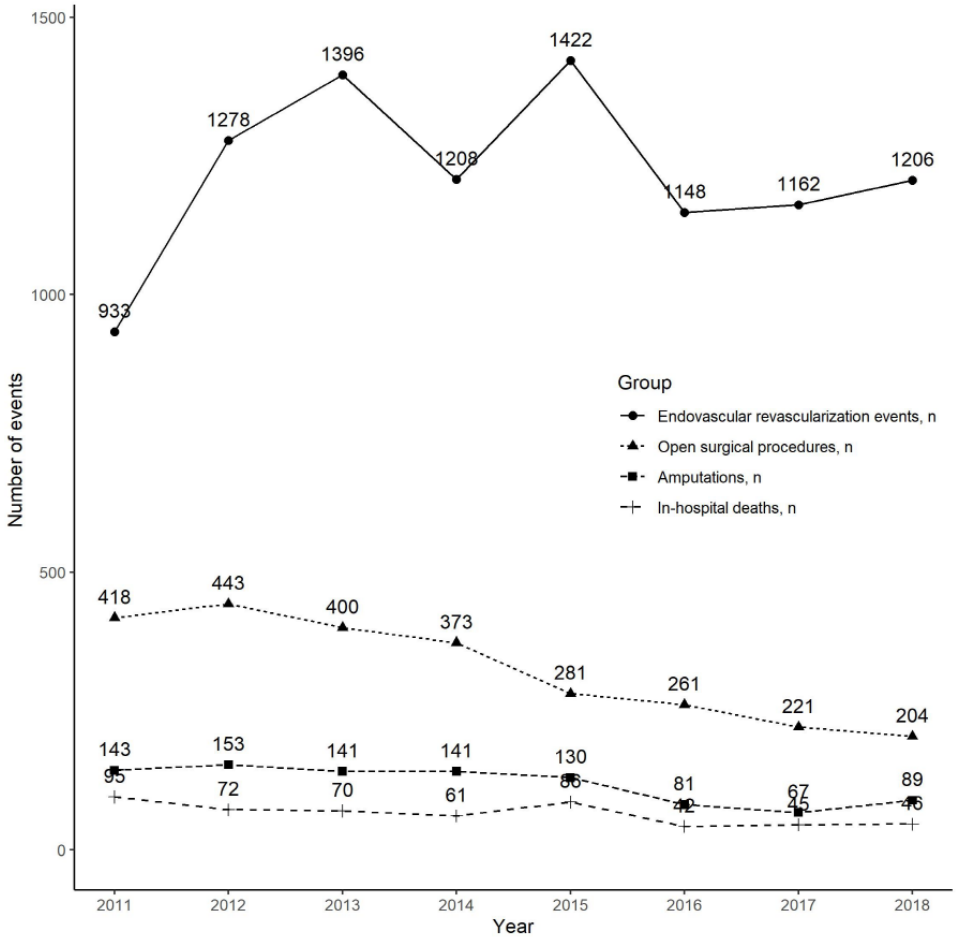


Table 3. Annual number of endovascular revascularization events, open surgical procedures, amputations, and all-cause in-hospital deaths.

	2011, n	2012, n	2013, n	2014, n	2015, n	2016, n	2017, n	2018, n	Total (2011-2018), n	AC[a] (2011-2018)	Average (2011-2018), n
EVR[b] events	933	1278	1396	1208	1422	1148	1162	1206	9753	+273	1219
Open surgical procedures	418	443	400	373	281	261	221	204	2601	–214	325
Amputations											
Total	143	153	141	141	130	81	67	89	945	–54	118
Major amputations	61	68	59	56	47	28	27	35	381	–26	48
Minor amputations	82	85	82	85	83	53	40	54	564	–28	71
In-hospital deaths											
Total	95	72	70	61	86	42	45	46	517	–49	65
Within 7 days	53	37	37	47	48	24	33	33	312	–20	39
Within 30 days	89	64	64	58	82	41	42	44	484	–45	61
Within 90 days	95	72	70	61	86	42	45	46	517	–49	57

[a]AC: absolute change.

[b]EVR: endovascular revascularization.

Health Care Utilization by Sex

Total claims from 2011 to 2018 were 4,222,726. Male patients used tertiary hospitals more than than female patients (n=177,274, 9.8% vs n=76,636, 3.2%, respectively). Female patients used clinics more than male patients (n=2,027,490, 83.8% vs n=1,222,519, 67.8%, respectively) (Table 4).

Table 4. Types of health care utilization based on claims by sex (N=4,222,726). Tertiary hospitals are large, with at least 20 medical departments and specialists for each department. General hospitals have at least 100 beds and hospitals at least 30. Long-term care facilities treat inpatients and outpatients. Clinics treat outpatients.

	Men			Women		
	Total (n=1,802,405), n (%)	Inpatients (n=30,402), n (%)	Outpatients (n=1,772,003), n (%)	Total (n=2,420,321), n (%)	Inpatients (n=11,198), n (%)	Outpatients (n=2,409,123), n (%)
Tertiary hospitals	177,274 (9.8)	14,173 (46.6)	163,101 (9.2)	76,636 (3.2)	3904 (34.9)	72,732 (3)
General hospitals	307,692 (17.1)	12,778 (42)	294,914 (16.6)	196,391 (8.1)	4118 (36.8)	192,273 (8)
Small hospitals	85,367 (4.7)	1922 (2.3)	83,445 (4.71)	106,292 (4.4)	1621 (14.5)	104,671 (4.3)
Long-term care facilities	9553 (0.5)	1294 (4.3)	8259 (0.47)	13,512 (0.6)	1226 (11)	12,286 (0.5)
Clinics	1,222,519 (67.8)	235 (0.8)	1,222,284 (69)	2,027,490 (83.8)	329 (2.9)	2,027,161 (84.2)

Discussion

This study identified the incidence and prevalence of PAD and PAD treatment trends, in-hospital deaths, and health care utilization in South Korea over the past 8 years through a retrospective analysis of national claims data.

The average PAD incidence was 2.91 per 1000 individuals from 2012 to 2018. Previously, a United States–based study used MarketScan data, which includes commercial, Medicare, and Medicaid health insurance data, to identify patients with a PAD or CLI diagnosis and found that the mean annual incidence of PAD was 2.34 [32]. A study conducted in the United Kingdom used a database of 11 million patients from 2000 to 2014 to search for symptomatic patients with PAD with at least 1 medical record in at least 2 years and found that the overall PAD incidence was 1.73 to 3.85 per 1000 individuals [33]. Our findings show that PAD incidence in South Korea was higher than in the United States and similar to the United Kingdom. Considering the characteristics of the participants, this study included claims with 1 PAD diagnosis in 8 years. However, if our analysis had used the same criteria as the study conducted in the United Kingdom, the PAD incidence in South Korea would have been lower.

The sex-stratified incidence and prevalence trends of PAD differed from those in previous studies and were higher in women than in men. The proportion of female patients with PAD ranged from 59.3% (60,552/102,153) to 61.6% (85,696/139,191), whereas for male patients, it ranged from 38.4% (53,495/139,191) to 40.7% (41,601/102,153). In a previous study, PAD incidence was 23.05 per 10,000 person-years in males, which was higher than the reported 12.37 per 10,000 person-years in females [33]. PAD has traditionally been reported to be a male-dominant disease [34]. However, PAD has recently been reported to affect women as much as men in the general population [34]. A systematic review reported that women had a slightly higher prevalence than men by the age of 75 years in high-income countries, measured by an arterial ankle brachial index (ABI) of 0.90 or less [6]. Classifying health care utilization by sex in this study revealed differences in claims between women and men. In terms of health care use, the number of tertiary hospital claims was high for men, whereas the number of clinic claims was high for women. In a Korean study, men accounted for a higher proportion than women of patients who received procedures at tertiary hospitals [35].

In our study, the PAD incidence trend among individuals in their 20s to 70s increased with age, which is similar to the findings of studies based in the United Kingdom [33] and United States [32]. Aging increases PAD-associated risk factors, such as hypertension, hyperlipidemia, and diabetes, and thereby increases the prevalence of PAD [2,36]. In terms of absolute change, patients with PAD in their 80s had the highest increase, at 1.10. Aging has increased the proportion of people in their 80s in the general population, and accordingly, the proportion of patients with PAD has also increased.

In our study, the prevalence of PAD was 3.93 and 23.55 per 1000 individuals in 2011 and 2018, respectively, indicating a steadily increasing trend, without decline. The prevalence of PAD has been reported to consistently increase [1,37]. In a study based on health insurance claims data in Germany, the number of prevalent cases of PAD consistently increased [24], which is similar to the findings of this study. In a meta-analysis, the prevalence of PAD was 5.56% in adults older than 25 years worldwide [6]. The results of this study are consistent with those of previous studies. Considering patients who do not visit hospitals due to having an asymptomatic condition, the incidence and prevalence of PAD may have been underestimated.

In this study, the number of EVR events fluctuated during the study period and increased from 933 in 2011 to 1206 in 2018. Similarly, an increasing trend in EVR events was reported for the US population from 1996 to 2011 among patients with PAD and diabetes [38]. PTA is recommended as the first-line revascularization intervention for PAD and is known to be effective, safe, and widely applicable, with few complications

[13]; therefore, PTA has a positive effect on preventing major amputations [14], and the trend of PTA use has increased.

In our study, the rates of both major and minor amputations decreased slightly. The finding of a decreasing trend in major amputations over time is similar to the results of a study of patients with PAD based on health insurance claims data in Germany between 2008 and 2016 [24].

In this study, the numbers of patients who died in the hospital within 7 days and 30 days of a PAD diagnosis were 24 to 53 and 41 to 89, respectively, suggesting that a high proportion of in-hospital deaths occurred within 7 days. Patients with PAD may develop complications, such as ischemic myocardial and cerebrovascular events and sepsis [39-41], and older adults have a higher risk of complications than younger age groups [40]. Interventions and intensive monitoring of complications in the first month are necessary for hospitalized patients with PAD.

This study had some limitations. The data were claims data for health insurance that were collected through administrative processes and were not intended for research purposes. Furthermore, the previously entered diagnosis codes might not have been changed, or the doctor might have only entered the diagnosis and treatment codes that were required for the health insurance claims. Thus, the number of patients, procedures, or surgeries may have been underestimated. Furthermore, PAD was defined using disease codes only. Therefore, individuals who were not diagnosed with PAD but had CLI-associated symptoms and those with an ABI of less than 0.9 may have been excluded from the study, leading to underestimated results. As this was a retrospective study that was based on claims data, many potential confounders were not adjusted for in the analyses. Moreover, the use of descriptive statistics limits statistical inferences across study groups. Therefore, it is necessary to carefully interpret our results on the incidence and prevalence of PAD; future studies that investigate this topic should adjust for confounders, such as risk factors, geographical heterogeneities, and medical disparities. Furthermore, we suggest that health outcomes, not only medical procedures and surgeries but also psychological well-being (eg, depression and anxiety) and quality of life, should be considered in association with PAD.

Our findings provide evidence for strategies for health promotion and intervention for patients with PAD and may help with strategies to manage risk factors, such as ceasing smoking, following a low-fat diet, and managing weight. The American Heart Association guidelines recommended walking as an exercise for controlling risk factors [15,42]. PAD causes pain when walking, which makes it difficult to carry out daily activities and can influence psychological health, such as by inducing depression and anxiety in patients with PAD. Therefore, the management of psychological health deserves attention in PAD care for aging populations.

In this study, increasing trends in incident cases and the prevalence of PAD in South Korea were observed between 2011 and 2018. PAD incidence was higher in women than men in this study. A strength of this study is that, methodologically, the epidemiological trends of the entire South Korean population and all patients with PAD in South Korea were ascertained through public data analysis. Furthermore, the health care utilization of patients with PAD was determined based on national data, which enables the generalization of results for the provision of information to undertake both prevention and treatment in the clinical setting and for further research.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea, funded by the Ministry of Education (2019R1A2C1007185, 2020R1A6A1A03041989) and by the Brain Korea 21 Four Project, funded by the National Research Foundation of Korea and Yonsei University College of Nursing. GWR and YSY received scholarships from the Brain Korea 21 Four Project. JK received a scholarship from the Brain Korea 21 Project (Big Data–Based Interdisciplinary Education and Research for Data Science). This study used Health Insurance Review and Assessment Service research data (M20190923977). The views expressed in this paper are those of the authors and not necessarily those of the Health Insurance Review and Assessment Service or the Ministry of Health and Welfare in South Korea.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Code for patients with peripheral arterial disease.
[XLSX File (Microsoft Excel File), 10 KB - publichealth_v8i11e34908_app1.xlsx]

Multimedia Appendix 2
Procedure codes and names for patients with peripheral arterial disease.
[XLSX File (Microsoft Excel File), 11 KB - publichealth_v8i11e34908_app2.xlsx]

References

1. Fowkes FGR, Aboyans V, Fowkes FJI, McDermott MM, Sampson UKA, Criqui MH. Peripheral artery disease: epidemiology and global perspectives. Nat Rev Cardiol 2017 Mar;14(3):156-170. [doi: 10.1038/nrcardio.2016.179] [Medline: 27853158]
2. Hiatt WR, Armstrong EJ, Larson CJ, Brass EP. Pathogenesis of the limb manifestations and exercise limitations in peripheral artery disease. Circ Res 2015 Apr 24;116(9):1527-1539. [doi: 10.1161/CIRCRESAHA.116.303566] [Medline: 25908726]
3. Aboyans V, Ricco J, Bartelink M, Björck M, Brodmann M, Cohner T, et al. 2017 ESC Guidelines on the Diagnosis and Treatment of Peripheral Arterial Diseases, in collaboration with the European Society for Vascular Surgery (ESVS). Kardiol Pol 2017 Nov 22;75(11):1065-1160. [doi: 10.5603/kp.2017.0216]
4. Aday AW, Matsushita K. Epidemiology of peripheral artery disease and polyvascular disease. Circ Res 2021 Jun 11;128(12):1818-1832 [FREE Full text] [doi: 10.1161/CIRCRESAHA.121.318535] [Medline: 34110907]
5. Kim M, Kim C, Kim E, Choi M. Effectiveness of mobile health-based exercise interventions for patients with peripheral artery disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Feb 15;9(2):e24080 [FREE Full text] [doi: 10.2196/24080] [Medline: 33587042]
6. Song P, Rudan D, Zhu Y, Fowkes FJI, Rahimi K, Fowkes FGR, et al. Global, regional, and national prevalence and risk factors for peripheral artery disease in 2015: an updated systematic review and analysis. Lancet Glob Health 2019 Aug;7(8):e1020-e1030 [FREE Full text] [doi: 10.1016/S2214-109X(19)30255-4] [Medline: 31303293]
7. Shu J, Santulli G. Update on peripheral artery disease: epidemiology and evidence-based facts. Atherosclerosis 2018 Aug;275:379-381 [FREE Full text] [doi: 10.1016/j.atherosclerosis.2018.05.033] [Medline: 29843915]
8. Mustapha JA, Katzen BT, Neville RF, Lookstein RA, Zeller T, Miller LE, et al. Determinants of long-term outcomes and costs in the management of critical limb ischemia: a population-based cohort study. J Am Heart Assoc 2018 Aug 21;7(16):e009724 [FREE Full text] [doi: 10.1161/JAHA.118.009724] [Medline: 30369325]
9. Levin SR, Arinze N, Siracuse JJ. Lower extremity critical limb ischemia: A review of clinical features and management. Trends Cardiovasc Med 2020 Apr;30(3):125-130. [doi: 10.1016/j.tcm.2019.04.002] [Medline: 31005554]
10. Bauersachs R, Debus S, Nehler M, Huelsebeck M, Balradj J, Bowrin K, et al. A targeted literature review of the disease burden in patients with symptomatic peripheral artery disease. Angiology 2020 Apr 30;71(4):303-314. [doi: 10.1177/0003319719896477] [Medline: 31884807]
11. Salisbury DL, Whipple MO, Burt M, Brown RJ, Hirsch A, Foley C, et al. Translation of an evidence-based therapeutic exercise program for patients with peripheral artery disease. J Vasc Nurs 2018 Mar;36(1):23-33 [FREE Full text] [doi: 10.1016/j.jvn.2017.09.003] [Medline: 29452626]
12. Reinecke H, Unrath M, Freisinger E, Bunzemeier H, Meyborg M, Lüders F, et al. Peripheral arterial disease and critical limb ischaemia: still poor outcomes and lack of guideline adherence. Eur Heart J 2015 Apr 14;36(15):932-938. [doi: 10.1093/eurheartj/ehv006] [Medline: 25650396]
13. Bisdas T, Borowski M, Torsello G, First-Line Treatments in Patients With Critical Limb Ischemia (CRITISCH) Collaborators. Current practice of first-line treatment strategies in patients with critical limb ischemia. J Vasc Surg 2015 Oct;62(4):965-973.e3 [FREE Full text] [doi: 10.1016/j.jvs.2015.04.441] [Medline: 26187290]
14. Kim J, Chun D, Kim S, Yang H, Kim JH, Cho J, et al. Trends in lower limb amputation in patients with diabetic foot based on vascular intervention of peripheral arterial disease in Korea: a population-based nationwide study. J Korean Med Sci 2019 Jul 08;34(26):e178 [FREE Full text] [doi: 10.3346/jkms.2019.34.e178] [Medline: 31269542]
15. Writing Committee Members, Gerhard-Herman MD, Gornik HL, Barrett C, Barshes NR, Corriere MA, ACC/AHA Task Force Members, et al. 2016 AHA/ACC Guideline on the Management of Patients with Lower Extremity Peripheral Artery Disease: executive summary. Vasc Med 2017 Jun;22(3):NP1-NP43. [doi: 10.1177/1358863X17701592] [Medline: 28494710]
16. Kim C, Yang Y, Ryu G, Choi M. Risk factors associated with amputation-free survival for patients with peripheral arterial disease: a systematic review. Eur J Cardiovasc Nurs 2021 May 22;20(4):295-304. [doi: 10.1093/eurjcn/zvaa022] [Medline: 33786599]
17. Abaraogu U, Ezenwankwo E, Dall P, Tew G, Stuart W, Brittenden J, et al. Barriers and enablers to walking in individuals with intermittent claudication: A systematic review to conceptualize a relevant and patient-centered program. PLoS One 2018 Jul 26;13(7):e0201095 [FREE Full text] [doi: 10.1371/journal.pone.0201095] [Medline: 30048501]
18. Firnhaber JM, Powell CS. Lower extremity peripheral artery disease: diagnosis and treatment. Am Fam Physician 2019 Mar 15;99(6):362-369 [FREE Full text] [Medline: 30874413]
19. Signorelli SS, Marino E, Scuto S, Di Raimondo D. Pathophysiology of Peripheral Arterial Disease (PAD): A Review on Oxidative Disorders. Int J Mol Sci 2020 Jun 20;21(12):4393 [FREE Full text] [doi: 10.3390/ijms21124393] [Medline: 32575692]
20. Chan ASW, Ho JMC, Li JSF, Tam HL, Tang PMK. Impacts of COVID-19 pandemic on psychological well-being of older chronic kidney disease patients. Front Med (Lausanne) 2021;8:666973 [FREE Full text] [doi: 10.3389/fmed.2021.666973] [Medline: 34124096]
21. Conversano C. Common psychological factors in chronic diseases. Front Psychol 2019;10:2727 [FREE Full text] [doi: 10.3389/fpsyg.2019.02727] [Medline: 31866912]
22. Kim M, Kim Y, Ryu GW, Choi M. Functional status and health-related quality of life in patients with peripheral artery disease: a cross-sectional study. Int J Environ Res Public Health 2021 Oct 18;18(20):10941 [FREE Full text] [doi: 10.3390/ijerph182010941] [Medline: 34682683]

23. Rafnsson SB, Fowkes G. Positive and negative well-being of older adults with symptomatic peripheral artery disease: A population-based investigation. JRSM Cardiovasc Dis 2020 Oct 20;9:2048004020961717 [FREE Full text] [doi: 10.1177/2048004020961717] [Medline: 33520199]
24. Kreutzburg T, Peters F, Rieß HC, Hischke S, Marschall U, Kriston L, et al. Editor's choice - comorbidity patterns among patients with peripheral arterial occlusive disease in Germany: a trend analysis of health insurance claims data. Eur J Vasc Endovasc Surg 2020 Jan;59(1):59-66 [FREE Full text] [doi: 10.1016/j.ejvs.2019.08.006] [Medline: 31744786]
25. Kim H, Kim S, Han S, Rane PP, Fox KM, Qian Y, et al. Prevalence and incidence of atherosclerotic cardiovascular disease and its risk factors in Korea: a nationwide population-based study. BMC Public Health 2019 Aug 14;19(1):1112 [FREE Full text] [doi: 10.1186/s12889-019-7439-0] [Medline: 31412823]
26. Kim J, Yoon S, Kim L, Kim D. Towards actualizing the value potential of Korea Health Insurance Review and Assessment (HIRA) data as a resource for health research: strengths, limitations, applications, and strategies for optimal use of HIRA data. J Korean Med Sci 2017 May;32(5):718-728 [FREE Full text] [doi: 10.3346/jkms.2017.32.5.718] [Medline: 28378543]
27. Kwon S. Thirty years of national health insurance in South Korea: lessons for achieving universal health care coverage. Health Policy Plan 2009 Jan 12;24(1):63-71. [doi: 10.1093/heapol/czn037] [Medline: 19004861]
28. Londero LS, Hoegh A, Houlind K, Lindholt J. Major amputation rates in patients with peripheral arterial disease aged 50 years and over in Denmark during the period 1997-2014 and their relationship with demographics, risk factors, and vascular services. Eur J Vasc Endovasc Surg 2019 Nov;58(5):729-737 [FREE Full text] [doi: 10.1016/j.ejvs.2019.06.007] [Medline: 31551135]
29. Gordis L. Epidemiology, 5th edition. Philadelphia, PA: Elsevier /Saunders; 2014.
30. Joo S, Goo Y, Ryu J, Lee S, Lee WK, Chung D, et al. Epidemiology of trichomoniasis in South Korea and increasing trend in incidence, health insurance review and assessment 2009-2014. PLoS One 2016 Dec 9;11(12):e0167938 [FREE Full text] [doi: 10.1371/journal.pone.0167938] [Medline: 27936227]
31. Lin C, Li C, Hsiao C, Liu C, Yang S, Lee C, et al. Time trend analysis of the prevalence and incidence of diagnosed type 2 diabetes among adults in Taiwan from 2000 to 2007: a population-based study. BMC Public Health 2013 Apr 09;13:318 [FREE Full text] [doi: 10.1186/1471-2458-13-318] [Medline: 23570503]
32. Nehler MR, Duval S, Diao L, Annex BH, Hiatt WR, Rogers K, et al. Epidemiology of peripheral arterial disease and critical limb ischemia in an insured national population. J Vasc Surg 2014 Sep;60(3):686-95.e2 [FREE Full text] [doi: 10.1016/j.jvs.2014.03.290] [Medline: 24820900]
33. Cea-Soriano L, Fowkes FGR, Johansson S, Allum AM, García Rodriguez LA. Time trends in peripheral artery disease incidence, prevalence and secondary preventive therapy: a cohort study in The Health Improvement Network in the UK. BMJ Open 2018 Jan 21;8(1):e018184 [FREE Full text] [doi: 10.1136/bmjopen-2017-018184] [Medline: 29358428]
34. Schramm K, Rochon P. Gender differences in peripheral vascular disease. Semin Intervent Radiol 2018 Mar 05;35(1):9-16 [FREE Full text] [doi: 10.1055/s-0038-1636515] [Medline: 29628610]
35. Ko Y, Ahn C, Min P, Lee J, Yoon C, Yu CW, K-VIS investigators. Baseline characteristics of a retrospective patient cohort in the Korean Vascular Intervention Society Endovascular Therapy in Lower Limb Artery Diseases (K-VIS ELLA) Registry. Korean Circ J 2017 Jul;47(4):469-476 [FREE Full text] [doi: 10.4070/kcj.2017.0020] [Medline: 28765738]
36. Hiramoto JS, Katz R, Ix JH, Wassel C, Rodondi N, Windham BG, Health ABC study. Sex differences in the prevalence and clinical outcomes of subclinical peripheral artery disease in the Health, Aging, and Body Composition (Health ABC) study. Vascular 2014 Apr 13;22(2):142-148 [FREE Full text] [doi: 10.1177/1708538113476023] [Medline: 23512905]
37. Fowkes FGR, Rudan D, Rudan I, Aboyans V, Denenberg JO, McDermott MM, et al. Comparison of global estimates of prevalence and risk factors for peripheral artery disease in 2000 and 2010: a systematic review and analysis. Lancet 2013 Oct 19;382(9901):1329-1340. [doi: 10.1016/S0140-6736(13)61249-0] [Medline: 23915883]
38. Goodney PP, Tarulli M, Faerber AE, Schanzer A, Zwolak RM. Fifteen-year trends in lower limb amputation, revascularization, and preventive measures among medicare patients. JAMA Surg 2015 Jan 01;150(1):84-86 [FREE Full text] [doi: 10.1001/jamasurg.2014.1007] [Medline: 25409197]
39. Pande RL, Perlstein TS, Beckman JA, Creager MA. Secondary prevention and mortality in peripheral artery disease: National Health and Nutrition Examination Study, 1999 to 2004. Circulation 2011 Jul 05;124(1):17-23 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.110.003954] [Medline: 21690489]
40. Skonetzki S, Lüders F, Engelbertz C, Malyar NM, Freisinger E, Meyborg M, et al. Aging and outcome in patients with peripheral artery disease and critical limb ischemia. J Am Med Dir Assoc 2016 Oct 01;17(10):927-932. [doi: 10.1016/j.jamda.2016.06.004] [Medline: 27427216]
41. Kim M, Yang YS, Ko Y, Choi M. Major adverse events in patients with peripheral artery disease after endovascular revascularization: a retrospective study. J Clin Med 2022 May 01;11(9):2547 [FREE Full text] [doi: 10.3390/jcm11092547] [Medline: 35566674]
42. Criqui MH, Aboyans V. Epidemiology of peripheral artery disease. Circ Res 2015 Apr 24;116(9):1509-1526. [doi: 10.1161/circresaha.116.303849]

Abbreviations

ABI: arterial ankle brachial index
CLI: clinical limb ischemia
EVR: endovascular revascularization
HIRA: Health Insurance Review and Assessment
ICD-10: International Classification of Diseases, Tenth Revision
KCD-7: Korean Standard Classification of Diseases, 7th revision
PAD: peripheral arterial disease
PTA: percutaneous transluminal angiography

Edited by T Sanchez, A Mavragani; submitted 12.11.21; peer-reviewed by ASW Chan , C Zeng; comments to author 18.08.22; revised version received 01.10.22; accepted 13.10.22; published 18.11.22.

Please cite as:
Ryu GW, Park YS, Kim J, Yang YS, Ko YG, Choi M
Incidence and Prevalence of Peripheral Arterial Disease in South Korea: Retrospective Analysis of National Claims Data
JMIR Public Health Surveill 2022;8(11):e34908
URL: https://publichealth.jmir.org/2022/11/e34908
doi:10.2196/34908
PMID:36399371

Original Paper

Associations Among Multimorbid Conditions in Hospitalized Middle-aged and Older Adults in China: Statistical Analysis of Medical Records

Yan Zhang[1,2*], MPH; Chao Chen[1,3*], MSc; Lingfeng Huang[1,3*], MM; Gang Liu[4*], MM; Tingyu Lian[1,3], MPH; Mingjuan Yin[1,3], MM; Zhiguang Zhao[5], PhD; Jian Xu[2], PhD; Ruoling Chen[6], PhD; Yingbin Fu[4], BM; Dongmei Liang[1,3], MM; Jinmei Zeng[1,3], MPH; Jindong Ni[1,3], PhD

[1]Precision Key Laboratory of Public Health, School of Public Health, Guangdong Medical University, Dongguan, China

[2]Department of Elderly Health Management, Shenzhen Center for Chronic Disease Control, Shenzhen, China

[3]Institute of Public Health and Wellness, Guangdong Medical University, Dongguan, China

[4]Department of Primary Public Health Promotion, Shenzhen Center for Disease Control and Prevention, Shenzhen, China

[5]Administration Office, Shenzhen Center for Chronic Disease Control, Shenzhen, China

[6]Faculty of Education, Health and Wellbeing, University of Wolverhampton, Wolverhampton, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Jindong Ni, PhD
Precision Key Laboratory of Public Health
School of Public Health
Guangdong Medical University
No.1 Xincheng Road, Songshan Lake
Dongguan, 523808
China
Phone: 86 15817668208
Email: nijd-gw@gdmu.edu.cn

Abstract

Background: Multimorbidity has become a new challenge for medical systems and public health policy. Understanding the patterns of and associations among multimorbid conditions should be given priority. It may assist with the early detection of multimorbidity and thus improve quality of life in older adults.

Objective: This study aims to comprehensively analyze and compare associations among multimorbid conditions by age and sex in a large number of middle-aged and older Chinese adults.

Methods: Data from the home pages of inpatient medical records in the Shenzhen National Health Information Platform were evaluated. From January 1, 2017, to December 31, 2018, inpatients aged 50 years and older who had been diagnosed with at least one of 40 conditions were included in this study. Their demographic characteristics (age and sex) and inpatient diagnoses were extracted. Association rule mining, Chi-square tests, and decision tree analyses were combined to identify associations between multiple chronic conditions.

Results: In total, 306,264 hospitalized cases with available information on related chronic conditions were included in this study. The prevalence of multimorbidity in the overall population was 76.46%. The combined results of the 3 analyses showed that, in patients aged 50 years to 64 years, lipoprotein metabolism disorder tended to be comorbid with multiple chronic conditions. Gout and lipoprotein metabolism disorder had the strongest association. Among patients aged 65 years or older, there were strong associations between cerebrovascular disease, heart disease, lipoprotein metabolism disorder, and peripheral vascular disease. The strongest associations were observed between senile cataract and glaucoma in men and women. In particular, the association between osteoporosis and malignant tumor was only observed in middle-aged and older men, while the association between anemia and chronic kidney disease was only observed in older women.

Conclusions: Multimorbidity was prevalent among middle-aged and older Chinese individuals. The results of this comprehensive analysis of 4 age-sex subgroups suggested that associations between particular conditions within the sex and age groups occurred more frequently than expected by random chance. This provides evidence for further research on disease clusters and for health

care providers to develop different strategies based on age and sex to improve the early identification and treatment of multimorbidity.

(JMIR Public Health Surveill 2022;8(11):e38182) doi:10.2196/38182

KEYWORDS

multimorbidity; chronic conditions; aging; association rule mining; decision tree analysis

Introduction

Background

China is the world's most populous country and has the largest aging population. The population aged 65 years and older has markedly increased in recent years, and there were approximately 190 million people aged 65 years and older in China in 2020 [1]. With such a large aging population, chronic conditions are a major contributor to health burden, inequalities in health outcomes, and economic burden in China [2]. Multimorbidity (defined as 2 or more coexisting chronic conditions) has become a new challenge for medical systems and public health policy [3-5]. Multimorbidity is often associated with functional limitations, reduced quality of life, higher mortality, higher rates of adverse drug events, and frequent use of health services [6,7]. Despite the growing number of studies suggesting that multimorbidity is normal for older adults, the majority of health care systems and public health policies is focused on the treatment of individual diseases rather than a complex network of diseases [3]. The incidence of multimorbidity is latent, and the progression is slow [8]. If early detection and diagnosis are not efficient and timely, this not only will delay treatment and prognosis and affect the development of the disease but also may lead to premature death [9]. Therefore, understanding the patterns and associations among multimorbid conditions should be given priority, which may assist the early diagnosis of multimorbidity and thus improve quality of life of older adults [10].

An increasing number of studies have reported on the frequent combinations of diseases and described the patterns of multimorbidity. These studies used various methods, such as generating all possible combinations of chronic diseases, estimating observed-to-expected ratios or relative risk among the most common combination of 2 or 3 chronic conditions [11], cluster analysis [12,13], latent class analysis [14,15], factor analyses [16,17], and network analysis [4,18]. These methods are similar and investigate combinations of conditions but do not elucidate associations and the prioritization of associations between individual conditions. Furthermore, these disease combinations are mainly based on a single algorithm and lack further methods to verify their stability.

Association rule mining (ARM) is now being used to explore associations between frequent diseases [6]. ARM, a data mining technique used extensively in health care, attempts to identify and predict rules by extracting simple structures from a set of items in a database [19]. However, extrapolation of the association results based on existing samples and the priority of the associated condition of the target conditions are not taken into account in traditional ARM. With the addition of the Chi-square test and decision tree analysis, these disadvantages can be avoided. The Chi-square test is a statistical method based on the difference in rate distribution, which can be used to test the statistical significance of the associations between the antecedent conditions and the consequent conditions in the association results, in order to extrapolate the sample results to the population situation. Decision tree analysis, a powerful statistical tool, has been successfully applied to recursively split independent variables into groups to predict an outcome [20,21]. In previous studies, it was also utilized to explore associated factors with survival in breast cancer patients [22], examine the interaction of shared variables to predict survival in patients with newly diagnosed malignant pleural mesothelioma [23], and investigate the prognostic importance of each factor for overall survival [24]. Unlike common methods, decision tree analysis can be used to classify factors to determine their importance to the target variables and decide which factor has the strongest association with the dependent variable at each point in the tree structure [25]. The combination of the 3 methods can obviously strengthen the evidence of the association between conditions, which enables accurate clinical decision support in practice. For more details on comparisons with currently used methods, please refer to Multimedia Appendix 1.

In addition, most studies on multimorbidity in China were conducted in community-dwelling populations, and self-reported questionnaires were used to define chronic diseases, which may have been affected by recall and reporting bias [11,26]. Hospital medical records describe the occurrence, development, diagnosis, and treatment of patients, and more objective clinical diagnoses are used to define multimorbidity. Obtaining the medical records of hospitalized patients to study multimorbidity could avoid recall or reporting bias. Furthermore, although multimorbidity is strongly associated with sociodemographic factors, few studies have focused on multimorbidity associations by age and sex.

Objectives

To better understand the multimorbidity patterns in middle-aged and older people, this study used the novel method of combining ARM with a traditional statistical significance test and decision tree analysis to examine and compare associations among multimorbid conditions by age and sex in a large number of middle-aged and older Chinese adults using the home pages of inpatient medical records in Shenzhen, China. It was hoped that the results would provide possible potential trajectories between multimorbid conditions and improve population-specific approaches to early detection and management of multimorbidity.

Methods

Data Source

This study used data from the home pages of inpatient medical records in the Shenzhen National Health Information Platform, a data center that collects medical information on cases from all medical institutions in Shenzhen. The home pages of inpatient medical records, including information on hospitalized patients' demographic characteristics (age and sex), inpatient diagnoses, International Classification of Diseases version 10 (ICD-10) codes, and personal identifiers, were removed. All clinical visits by patients were linked to their unique encrypted identification number.

Measurement of Multimorbidity and Study Population

In this study, the following 40 chronic conditions were selected based on the most frequently mentioned diseases in multimorbidity by previous studies that were considered to significantly impact long-term treatment and quality of life among middle-aged and older Chinese individuals [7,27]: hypertension (HT), diabetes mellitus (DM), lipoprotein metabolism disorder (LMD), chronic gastritis, chronic obstructive pulmonary disease, cerebrovascular disease (CBD), chronic kidney disease (CKD), spleen disease, peripheral vascular disease (PVD), varicose veins, schizophrenia, malignant tumor (MT), dementia, Alzheimer disease, bronchiectasis, glaucoma, senile cataract (SC), asthma, chronic nasopharyngitis, chronic viral hepatitis, thyroid disorders, hearing loss, dermatitis and eczema, anemia, migraine, chronic liver disease (CLD), depression, epilepsy, anxiety, Parkinson disease, sleep disorder, heart disease (HD), chronic gastric ulcer, gout, osteoporosis, transient cerebral ischemia, arthropathy, spondylosis, and dizziness/vertigo. Conditions were identified if they had been documented using inpatient ICD-10 codes in an individual's medical records. Multimedia Appendix 2 lists all chronic conditions included and their corresponding ICD-10 codes. For this study, multimorbidity was defined as having 2 or more concurrent chronic conditions.

In this study, patient inclusion criteria included (1) diagnoses with at least one of the aforementioned 40 conditions in all inpatient records from January 1, 2017, to December 31, 2018, and (2) aged 50 years or older on earlier records. Middle-aged patients with multimorbidity represent a large group, and the prevalence of multimorbidity ranges from 45% to 72% among middle-aged and older people older than 50 years [28]. Exclusion criteria were that none of the conditions were diagnosed in any inpatient records during the study period. A total of 306,264 patients were included.

Statistical Analyses

Descriptive Statistics

First, descriptive statistics were used in the study population, including number, proportion (%), median, and IQR of age for sex (female and male). The top 10 prevalent chronic conditions with the largest composition ratio, including the average number (mean [SD]) of coexisting conditions, were evaluated. Furthermore, age was categorized into 2 subgroups (50-64 years and ≥65 years) and cross-combined with sex into 4 age-sex subgroups. The number and proportion were used to describe the distribution of patients with or without multimorbidity, and the Chi-square test was performed to compare differences in the characteristics of patients with and without multimorbidity.

Association Rule Mining Based on Subgroups

To identify the associations between conditions by age and sex, 4 age-sex–based subgroup analyses were then performed. ARM was applied to determine common multimorbidity patterns that met a minimum requirement of measurement indicators. Association rules were relationships between sets of conditions from "antecedent" to "consequent" [29]. We used 3 common measurement indicators: (1) support (how frequently the condition combinations appear in the data set), (2) confidence (how frequently the consequent conditions occur, conditional on the antecedent conditions), and (3) lift (the ratio of the observed support to that expected if antecedent and consequent were independent) [30]. Lift was considered the main measure of significance in ARM. A lift of "1" means that the probabilities of occurrence of the antecedent and consequent are independent of each other. Hence, a higher lift indicates a higher chance of co-occurrence of the consequent with the antecedent and a more significant association [31]. Setting a higher threshold value would reduce the number of rules that might result in missing essential rules with low frequencies, and setting a lower threshold value could result in a large number of rules that might hinder the management from summarizing rules [29]. Thus, many rounds of testing and evaluation were carried out before defining final thresholds to mine reasonable rules and to ensure the robustness of the model performance. Considering the vast number of disease types in the data set, the rules satisfying support >1%, confidence >50%, and lift >1 were selected. All association rules were sorted by lifts, and the top 10 association rules with larger lifts in 4 subgroups were described.

Chi-square Tests

To evaluate the statistical significance of the aforementioned association rules, Chi-square tests were applied. Odds ratios (ORs) and 95% CIs between antecedent conditions and consequent conditions in the association rules of the 4 age-sex subcategories are shown.

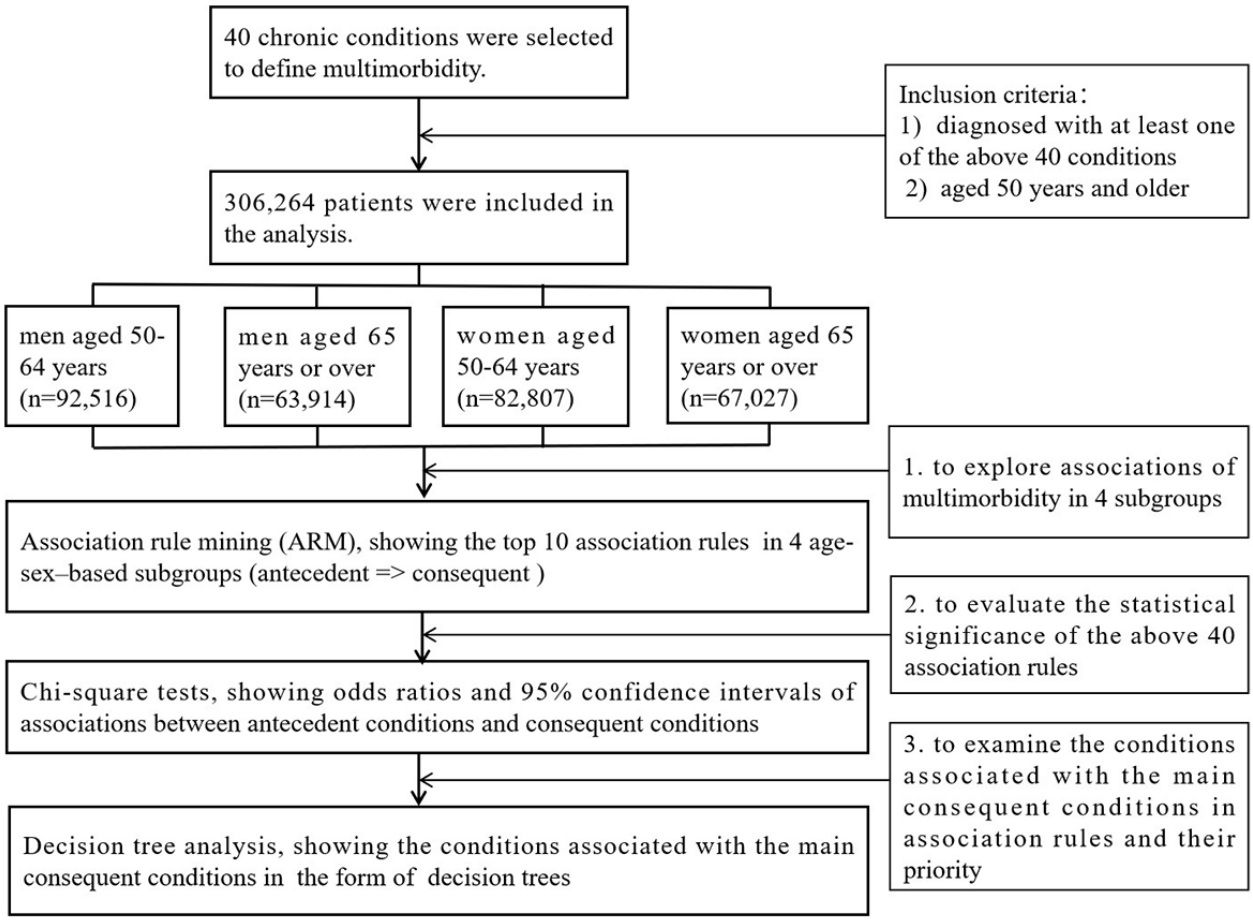
Decision Tree Analysis

Furthermore, decision tree analysis was conducted to examine the conditions associated with the main consequent conditions in the association rules. Decision tree analysis examines the relationship between influencing factors and target variables [32]. The decision tree process is a nonparametric method that creates a tree-based classification model [33]. A decision tree contains 3 main parts: decision nodes, branches, and leaves. The internal variables of the model represent a tree structure in which a decision is made in each branch according to the data features [25]. The tree starts with a node and extends to the leaf. The risky paths are identified and shown in nodes [34]. In this study, we used decision tree analysis to determine the relationship between the conditions and main consequent conditions in rule results. Thus, the consequent conditions in rule results were used as target variables, while the remaining conditions were used as the independent variables. Splitting

criteria provides a rate for each predictor variable. Variables that have the best rate of splitting criteria are selected to remain in the model [25], which have a greater impact on the target variables, and in this study, various conditions were screened based on this feature. In the decision tree, the first variable or root node is the most important factor, and other variables can be classified in order of importance [35]. The decision trees were drawn to show the associated conditions with main consequent conditions in association rules.

The flowchart of the analyses is shown in Figure 1. All descriptive analyses and Chi-square tests were performed using SPSS version 25.0 (IBM Corp, Armonk, NY), with a .05 level of significance. ARM and decision tree analysis were carried out using R 3.4.0 (The R Foundation for Statistics and Mathematics, Vienna, Austria) with the arules package and the tree package. To make the results more intuitive, GraphPad Prism version 8.0 (GraphPad Software, San Diego, CA) was used to show the ORs and 95% CIs, and PowerPoint software 2021 version (Microsoft Corp, Redmond, WA) was used to draw decision trees.

Figure 1. The flowchart of the main research steps.



Ethics Approval

This study was approved by the Institutional Review Ethics Committee of the Affiliated Hospital of Guangdong Medical University (YJYS202008). Informed consent was not required from participants as all data provided were deidentified.

Results

Characteristics of the Participants

In total, 306,264 hospitalized cases with available information on related chronic conditions were included in this study. The median age of the study population was 62 (IQR 55-71) years. There were more men than women, with men accounting for 51.08% (156,430/306,264) of the sample. The median age of the male and female participants was 62 (IQR 54-71) years and 63 (IQR 56-72) years, respectively (Table 1).

Table 1. Characteristics of the study population (N=306,264).

Characteristics	All respondents	Men	Women
Number of people, n (%)	306,264 (100)	156,430 (51.08)	149,834 (48.92)
Age (years), median (IQR)	62 (55-71)	62 (54-71)	63 (56-72)

Characteristics of the Chronic Conditions

As shown in Table 2, 44.72% (136,972/306,264) of the study population had HT, which was the most prevalent condition. This was followed by HD (74,535/306,264, 24.34%), DM (70,917/306,264, 23.16%), CBD (68,151/306,264, 22.25%), LMD (65,385/306,264, 21.35%), CKD (63,470/306,264, 20.72%), CLD (61,829/306,264, 20.19%), PVD (51,311/306,264, 16.75%), spondylosis (42,982/306,264, 14.03%), and gout (33,984/306,264, 11.10%). Patients with these chronic conditions had an average multimorbidity burden of ≥4 chronic conditions per patient.

Table 2. Top 10 conditions with the largest composition ratio in all cases (N=306,264).

Rank	Chronic conditions	Presence in all participants, n (%)	Number of co-occurring conditions, mean (SD)
1	Hypertension	136,972 (44.72)	4.79 (0.76)
2	Heart disease	74,535 (24.34)	4.97 (0.27)
3	Diabetes mellitus	70,917 (23.16)	4.84 (0.67)
4	Cerebrovascular disease	68,151 (22.25)	4.89 (0.53)
5	Lipoprotein metabolism disorder	65,385 (21.35)	4.91 (0.49)
6	Chronic kidney disease	63,470 (20.72)	4.81 (0.77)
7	Chronic liver disease	61,829 (20.19)	4.87 (0.59)
8	Peripheral vascular disease	51,311 (16.75)	4.95 (0.38)
9	Spondylosis	42,982 (14.03)	4.94 (0.39)
10	Gout	33,984 (11.10)	4.97 (0.29)

Differences in the Characteristics of Patients With and Without Multimorbidity

Of the 306,264 patients included, over 50% (175,323/306,264, 57.25%) were between 50 years and 64 years old (Table 3). The prevalence of multimorbidity in the overall population was 76.46% (234,156/306,264), with a higher prevalence in patients aged 65 years or older (108,937/306,264, 83.20%) than in those aged 50 years to 64 years (125,219/306,264, 71.42%). There were statistically significant sex differences in the prevalence of multimorbidity in the overall population, and patients aged 50 years to 64 years showed a higher prevalence in men than in women.

Table 3. Differences in the characteristics of patients with and without multimorbidity (N=306,264).

Age groups	Multimorbidity, n (%)	No multimorbidity, n (%)	*P* value
≥50 and ≤64 years			
Men	67,665 (73.14)	24,851 (26.86)	<.001
Women	57,554 (69.50)	25,253 (30.50)	
Total	125,219 (71.42)	50,104 (28.58)	—[a]
≥65 years			
Men	53,305 (83.40)	10,609 (16.60)	.052
Women	55,632 (83.00)	11,395 (17.00)	
Total	108,937 (83.20)	22,004 (16.80)	—
Overall sample			
Men	120,970 (77.33)	35,460 (22.67)	<.001
Women	113,186 (75.54)	36,648 (24.46)	
Total	234,156 (76.46)	72,108 (23.54)	—

[a]Not applicable.

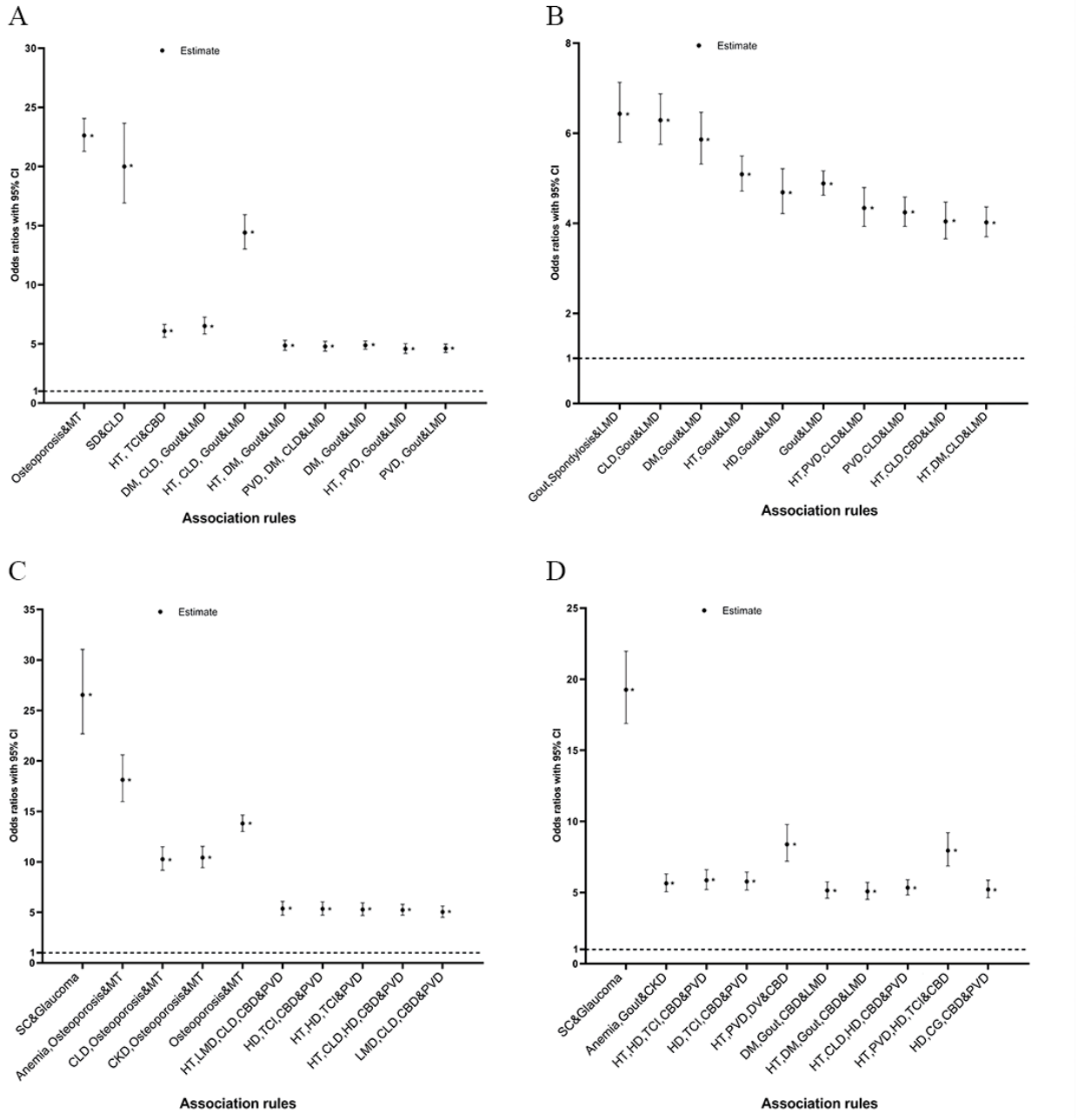
Association Rules and Statistical Analysis Results

The top 10 association rules in 4 age-sex–based subgroups according to lifts are shown in Multimedia Appendix 3. Among men and women aged 50 years to 64 years, LMD tended to be comorbid with DM, CLD, gout, HT, and PVD, which occurred in 7 association rules in men and in 10 rules in women. In addition, the combination of osteoporosis and MT was observed to have the strongest association in men, with a lift of 6.60, whereas this combination was not found in women. For patients aged 65 years or older, PVD tended to be present in combination with HT, LMD, CBD, and HD, which occurred in 5 association rules in men and 4 in women among the top 10 rules, indicating that these antecedent combinations positively correlated with

the occurrence of PVD. Furthermore, the strongest associations were observed between SC and glaucoma in men (lift=6.65) and in women (lift=4.93). In particular, the 4 association rules including osteoporosis and MT were only observed in men, and their lifts were all greater than 4, while the associations between anemia, gout, and CKD (lift=3.00) were only observed in women.

Statistical analysis (Chi-square tests) of the association rules in 4 age-sex–based subgroups was carried out, and the results are shown in Figure 2 and Multimedia Appendix 4. For all 40 rules, the ORs of the associations between antecedent conditions and consequent conditions were greater than "1," and the 95% CIs did not include "1," indicating that the latter conditions were more likely to be positive when the combinations of antecedent conditions were positive than negative.

Figure 2. Point estimates of the odds ratios and 95% CIs (1.96 SE) of the associations between antecedent conditions and consequent conditions in the association rules of 4 age-sex subgroups in (A) men aged 50-64 years, (B) women aged 50-64 years, (C) men aged 65 years or older, and (D) women aged 65 years or older. The * indicate significant findings. CBD: cerebrovascular disease; CG: chronic gastritis; CKD: chronic kidney disease; CLD: chronic liver disease; DM: diabetes mellitus; DV: dizziness/vertigo; HD: heart disease; HT: hypertension; LMD: lipoprotein metabolism disorder; MT: malignant tumor; PVD: peripheral vascular disease; SC: senile cataract; SD: spleen disease; TCI: transient cerebral ischemia.

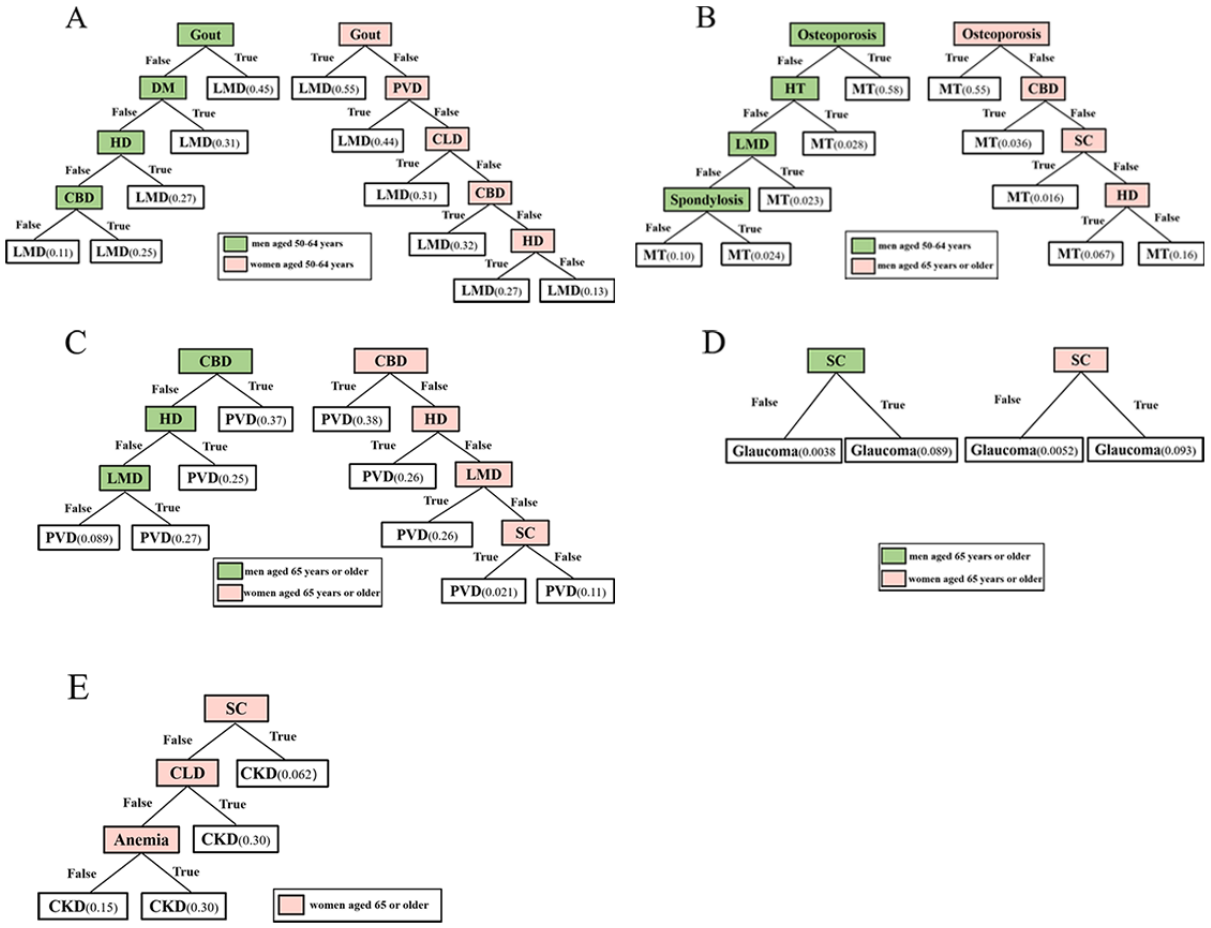


Decision Tree Analysis of the Main Association Rules

Decision tree analysis was used to examine the associated comorbidities of the main consequent conditions in the rule results. The main decision trees are shown in Figure 3. Figure 3A shows that, in patients aged 50 years to 64 years, the decision tree with LMD as the dependent variable included nodes of gout, DM, HD, and CBD in men, and gout was at the top of the tree, indicating that 45.05% (5830/12,940) of patients with gout had LMD. More importantly, gout, CBD, and HD remained in

the LMD decision tree for women. Gout was still at the top of the tree, and 55.37% (3270/5920) of patients with gout had LMD. Figure 3B shows that, in men aged 50 years to 64 years and 65 years or older, osteoporosis remained at the top of the decision tree of MT in men, indicating that more than 50% of patients with osteoporosis had a comorbidity of MT. Among patients aged 65 years or older, condition nodes reserved in the decision tree of PVD included CBD, HD, and LMD in men and CBD, HD, LMD, and SC in women (Figure 3C). SC was the only node in the glaucoma decision tree in both sexes (Figure 3D). Furthermore, in women, SC, CLD, and anemia were observed in the decision tree of CKD (Figure 3E).

Figure 3. Decision trees with main consequent conditions as the target variables (other conditions divided into 2 subcategories: positive = "True"; negative = "False) in the association rules of different age-sex subgroups: (A) lipoprotein metabolism disorder (LMD) as the target variable in men and women aged 50-64 years, (B) malignant tumor (MT) as the target variable in men aged 50-64 years and 65 years or older, (C) peripheral vascular disease (PVD) as the target variable in men and women aged 65 years or older, (D) glaucoma as the target variable in men and women aged 65 years or older, (E) chronic kidney disease (CKD) as the target variable in women aged 65 years or older. All decimal values represent the proportion of the target conditions that were positive when the associated conditions were in the corresponding subgroup. CBD: cerebrovascular disease; CLD: chronic liver disease; DM: diabetes mellitus; HD: heart disease; HT: hypertension; SC: senile cataract.



Discussion

Principal Findings

Understanding multimorbidity associations is an important public health priority for clinicians, academics, and funders alike [9]. This study was conducted to comprehensively evaluate the associations among multimorbid conditions based on the electronic hospitalized medical record home pages of a large sample of middle-aged and older Chinese people. To the best of our knowledge, this study is the first to evaluate multimorbidity associations using a comprehensive analysis with ARM, Chi-square tests, and decision tree analysis. Our analysis process not only revealed associations between particular conditions within different age-sex subgroups but also examined the importance of these associated chronic conditions for certain target conditions.

In this study of more 300,000 cases, more than 76% of patients were found to have 2 or more chronic conditions in the comprehensive list of 40 chronic conditions examined. The results showed that multiple conditions including HT, HD, DM, CBD, and LMD were the most common among hospitalized middle-aged and older Chinese adults, and they co-occurred with more than 3 other conditions. This is similar to the findings reported in previous studies [4,7]. The prevalence of multimorbidity varied across the 2 age groups (50-64 years old and ≥65 years old) and both sex groups, reflecting the strong associations between multimorbidity and both age and sex [6]. Therefore, our subsequent analysis was based on specific

age-sex subgroups to identify and compare the associations among multimorbid conditions within age and sex.

Association rules can reflect the interdependence and relevance between one condition and others. In our study, the ranked lift of most association rules indicated that LMD was the dominant condition among men and women aged 50 years to 64 years and was directly and indirectly associated with multiple conditions, including DM, CLD, gout, HT, and PVD or combinations of these conditions, which was also confirmed by statistical analysis. The potential mechanisms might include increased systemic inflammatory mediators and some adverse effects, such as physical inactivity, which are also risk factors for associated conditions [7]. Furthermore, in both men and women, gout appeared at the top of the decision tree with LMD as the dependent variable, which proved that the strong association between gout and LMD was not coincidental. Our findings are consistent with those from previous studies. In a review, the author concluded that complex interconnections between gout and metabolic syndromes including LMD existed, showing that gout may play an important role in the manifestation of metabolic syndromes [36]. Therefore, proper management of one disease may have implications for early detection and prevention of another.

Among patients aged 65 years or older, ARM, statistical analysis, and decision tree analysis consistently found that PVD was closely interlinked with CBD, HD, and LMD. It was previously reported that these diseases share risk, pathophysiological, and prognostic features and their coexistence would cause a cumulative burden [27]. People with PVD are at significantly higher risk of myocardial infarction and stroke than the general population [37]. Although PVD can lead to adverse health outcomes, it has received little attention [38]. As an important comorbidity, PVD needs to be emphasized, and patients diagnosed with associated conditions should be targeted for PVD screening. Similarly, a significant association between SC and glaucoma was confirmed by all 3 methods, indicating that the probability of glaucoma was higher than the probability of other conditions when SC was present. This finding was consistent with that in a study based on large medical claims data among a Chinese population of 2 million [7]. The incidence of glaucoma and comorbid SC will increase with age, and measurements targeting those shared specific factors may benefit 2 or more related diseases [39].

Men aged 50 years to 64 years or 65 years or older reported a high prevalence of MT, with a high probability of co-occurrence with osteoporosis in the association rules and statistical analysis, which was consistent with the results of the decision tree analysis. Osteoporosis was found to be the condition most related with MT. Certain biological links have indeed been found between osteoporosis and MT, including the presence of important cytokines, hormones, and oxidative stress [40]. However, the sex difference between the 2 conditions in our research was inconsistent with some previous studies, which showed that osteoporosis and some types of MT, including breast cancer, thyroid cancer, and colorectal cancer, were more closely linked in women than in men [40-43]. This may be affected by factors such as MT type and age of the population, which requires further investigation in cancer subgroups. However, recognizing the existence of this association may help to guide the early screening of MT in Chinese middle-aged and older men with osteoporosis, especially the type with a high incidence in men.

The strong association between anemia, gout, and CKD was only detected in women aged 65 years or older by ARM and statistical analysis. The lift of 3.00 indicated that these conditions were 3 times as likely to occur simultaneously as they were alone. In the decision tree analysis, SC, CLD, and anemia were observed to be CKD-associated conditions. The common results of these 3 methods seemed to imply that there was a special association between anemia and CKD in this subgroup. Anemia is a common complication and contributes to increased morbidity and mortality in CKD patients, which has been demonstrated previously [44,45]. A systematic review concluded that excess was a main contributor to the disordered iron homeostasis and anemia of CKD by impairing dietary iron absorption and iron mobilization from body stores [46]. Furthermore, possible explanations for this relationship only found in older women included shared risk factors of 2 conditions, such as aging and female sex [47,48]. Therefore, for older women, active improvement of anemia may be of great significance in preventing and delaying the development of CKD.

The main strength of this study is that a novel method was used, that is, the combination of ARM with a traditional statistical significance test and decision tree analysis, to examine the associations of multimorbidity. In particular, this was the first time that decision tree analysis was used in a multimorbidity study. Second, the disease diagnoses that defined multimorbidity in our analysis were based on a large sample of inpatient medical records, which avoided recall or reporting bias. Finally, our association analysis was based on age-sex subgroups, avoiding the confounding effects of age and sex. The present findings indicated that combinations of particular conditions within sex and age groups occur more frequently than expected by random chance. This provides evidence for further research on the potential mechanisms and risk factors for specific combinations and to encourage health care providers to develop population-specific approaches for early detection and management of multimorbidity according to sex and age.

Limitations

Several limitations of our study must be acknowledged. First, our sample consisted of hospitalized cases, and mild and early cases may not have been included. In view of the fact that the research on multimorbidity in China is still at an early stage, our findings based on more severe cases may provide ideas for research on the early prevention of combinations of specific conditions. Second, we could not draw conclusions about causality effects between multiple conditions due to the cross-sectional design of the study. Finally, patients' socioeconomic status, family history, and lifestyle factors were not incorporated into the model in this analysis due to data availability, and the data set anonymized participants to avoid possible misuse; therefore, some potential confounding factors were not taken into consideration. However, given the advantages of our large sample size, the findings do provide

support and a new perspective for future longitudinal or experimental studies to identify potential mechanisms and risk factors for specific combinations.

Conclusions

Multimorbidity was prevalent among middle-aged and older Chinese individuals. The results of this comprehensive analysis of 4 age-sex subgroups suggested that associations among particular conditions within sex and age groups occurred more frequently than expected by random chance. This provides evidence for further research on disease clusters and for health care providers to develop different strategies, according to age and sex, to improve the early identification and treatment of multimorbidity.

Acknowledgments

This work was funded by the Natural Science Foundation of Guangdong Province, China (grant number 2021A1515011038), Discipline construction project of Guangdong Medical University (grant number 4SG21001G), and Dongguan Science and Technology Commissioner Project (grant number 20201800500352).

Authors' Contributions

All authors contributed to this research. JN conceived and designed this research. YZ and CC extracted and analyzed the data and wrote the main part of the manuscript. LH and GL performed the main analysis and helped revise the manuscript. TL and MY searched the literature and designed the list of conditions. ZZ and JX provided great assistance with the whole process of research design and data analysis. RC and YF provided key ideas to write the paper. DL organized the tables, and JZ drew the figures. JN, YZ, CC, LH, and GL provided major suggestions for the revision of the paper. All authors edited and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of methods used in multimorbidity studies.
[DOCX File , 18 KB - publichealth_v8i11e38182_app1.docx]

Multimedia Appendix 2
ICD-10 numbers of 40 conditions included in the analysis.
[DOCX File , 15 KB - publichealth_v8i11e38182_app2.docx]

Multimedia Appendix 3
The top 10 association rules in the order of lifts.
[DOCX File , 33 KB - publichealth_v8i11e38182_app3.docx]

Multimedia Appendix 4
The results of the statistical analysis of association rules.
[DOCX File , 43 KB - publichealth_v8i11e38182_app4.docx]

References

1. China Statistical Yearbook 2021. National Bureau of Statistics of China. 2021. URL: http://www.stats.gov.cn/tjsj/ndsj/2021/indexeh.htm [accessed 2022-09-24]
2. Zhao Y, Atun R, Oldenburg B, McPake B, Tang S, Mercer SW, et al. Physical multimorbidity, health service use, and catastrophic health expenditure by socioeconomic groups in China: an analysis of population-based panel data. The Lancet Global Health 2020 Jun;8(6):e840-e849. [doi: 10.1016/s2214-109x(20)30127-3]
3. Park B, Ock M, Lee HA, Lee S, Han H, Jo M, et al. Multimorbidity and health-related quality of life in Koreans aged 50 or older using KNHANES 2013-2014. Health Qual Life Outcomes 2018 Sep 15;16(1):186 [FREE Full text] [doi: 10.1186/s12955-018-1016-6] [Medline: 30219061]
4. Hernández B, Reilly RB, Kenny RA. Investigation of multimorbidity and prevalent disease combinations in older Irish adults using network analysis and association rules. Sci Rep 2019 Oct 10;9(1):14567 [FREE Full text] [doi: 10.1038/s41598-019-51135-7] [Medline: 31601959]
5. Shi X, Lima SMDS, Mota CMDM, Lu Y, Stafford RS, Pereira CV. Prevalence of multimorbidity of chronic noncommunicable diseases in Brazil: population-based study. JMIR Public Health Surveill 2021 Nov 25;7(11):e29693 [FREE Full text] [doi: 10.2196/29693] [Medline: 34842558]

6. Lee Y, Kim H, Jeong H, Noh Y. Patterns of multimorbidity in adults: an association rules analysis using the Korea Health Panel. Int J Environ Res Public Health 2020 Apr 11;17(8):11278 [FREE Full text] [doi: 10.3390/ijerph17082618] [Medline: 32290367]
7. Wang X, Yao S, Wang M, Cao G, Chen Z, Huang Z, et al. Multimorbidity among two million adults in China. Int J Environ Res Public Health 2020 May 13;17(10):1 [FREE Full text] [doi: 10.3390/ijerph17103395] [Medline: 32414117]
8. Kim E, Lee T, Ji Y. Predictors of multimorbidity among Korean older adults: longitudinal secondary data analysis. Innovation in Aging 2021;5(S1):163. [doi: 10.1093/geroni/igab046.626]
9. Cezard G, McHale CT, Sullivan F, Bowles JKF, Keenan K. Studying trajectories of multimorbidity: a systematic scoping review of longitudinal approaches and evidence. BMJ Open 2021 Nov 22;11(11):e048485 [FREE Full text] [doi: 10.1136/bmjopen-2020-048485] [Medline: 34810182]
10. Sum G, Salisbury C, Koh G, Atun R, Oldenburg B, McPake B, et al. Implications of multimorbidity patterns on health care utilisation and quality of life in middle-income countries: cross-sectional analysis. J Glob Health 2019 Dec;9(2):020413 [FREE Full text] [doi: 10.7189/jogh.09.020413] [Medline: 31448114]
11. Zhang R, Lu Y, Shi L, Zhang S, Chang F. Prevalence and patterns of multimorbidity among the elderly in China: a cross-sectional study using national survey data. BMJ Open 2019 Aug 18;9(8):e024268 [FREE Full text] [doi: 10.1136/bmjopen-2018-024268] [Medline: 31427309]
12. Guisado-Clavero M, Roso-Llorach A, López-Jimenez T, Pons-Vigués M, Foguet-Boreu Q, Muñoz MA, et al. Multimorbidity patterns in the elderly: a prospective cohort study with cluster analysis. BMC Geriatr 2018 Jan 16;18(1):16 [FREE Full text] [doi: 10.1186/s12877-018-0705-7] [Medline: 29338690]
13. Marengoni A, Roso-Llorach A, Vetrano D, Fernández-Bertolín S, Guisado-Clavero M, Violán C, et al. Patterns of Multimorbidity in a Population-Based Cohort of Older People: Sociodemographic, Lifestyle, Clinical, and Functional Differences. J Gerontol A Biol Sci Med Sci 2020 Mar 09;75(4):798-805. [doi: 10.1093/gerona/glz137] [Medline: 31125398]
14. Olaya B, Moneta MV, Caballero FF, Tyrovolas S, Bayes I, Ayuso-Mateos JL, et al. Latent class analysis of multimorbidity patterns and associated outcomes in Spanish older adults: a prospective cohort study. BMC Geriatr 2017 Aug 18;17(1):186 [FREE Full text] [doi: 10.1186/s12877-017-0586-1] [Medline: 28821233]
15. Simões D, Araújo FA, Severo M, Monjardino T, Cruz I, Carmona L, et al. Patterns and Consequences of Multimorbidity in the General Population: There is No Chronic Disease Management Without Rheumatic Disease Management. Arthritis Care Res (Hoboken) 2017 Jan 17;69(1):12-20 [FREE Full text] [doi: 10.1002/acr.22996] [Medline: 27482954]
16. Menditto E, Gimeno Miguel A, Moreno Juste A, Poblador Plou B, Aza Pascual-Salcedo M, Orlando V, et al. Patterns of multimorbidity and polypharmacy in young and adult population: Systematic associations among chronic diseases and drugs using factor analysis. PLoS One 2019 Feb 6;14(2):e0210701. [doi: 10.1371/journal.pone.0210701] [Medline: 30726245]
17. Araujo MEA, Silva MT, Galvao TF, Nunes BP, Pereira MG. Prevalence and patterns of multimorbidity in Amazon Region of Brazil and associated determinants: a cross-sectional study. BMJ Open 2018 Nov 03;8(11):e023398 [FREE Full text] [doi: 10.1136/bmjopen-2018-023398] [Medline: 30391918]
18. Kalgotra P, Sharda R, Croff JM. Examining health disparities by gender: A multimorbidity network analysis of electronic medical record. Int J Med Inform 2017 Dec;108:22-28. [doi: 10.1016/j.ijmedinf.2017.09.014] [Medline: 29132627]
19. Shi X, Nikolic G, Van Pottelbergh G, van den Akker M, Vos R, De Moor B. Development of multimorbidity over time: an analysis of Belgium primary care data using Markov chains and weighted association rule mining. J Gerontol A Biol Sci Med Sci 2021 Jun 14;76(7):1234-1241 [FREE Full text] [doi: 10.1093/gerona/glaa278] [Medline: 33159204]
20. Guerrero MD, Vanderloo LM, Rhodes RE, Faulkner G, Moore SA, Tremblay MS. Canadian children's and youth's adherence to the 24-h movement guidelines during the COVID-19 pandemic: A decision tree analysis. J Sport Health Sci 2020 Jul;9(4):313-321 [FREE Full text] [doi: 10.1016/j.jshs.2020.06.005] [Medline: 32525098]
21. Aguiar FS, Almeida LL, Ruffino-Netto A, Kritski AL, Mello FC, Werneck GL. Classification and regression tree (CART) model to predict pulmonary tuberculosis in hospitalized patients. BMC Pulm Med 2012 Aug 07;12(1):40 [FREE Full text] [doi: 10.1186/1471-2466-12-40] [Medline: 22871182]
22. Ganggayah MD, Taib NA, Har YC, Lio P, Dhillon SK. Predicting factors for survival of breast cancer patients using machine learning techniques. BMC Med Inform Decis Mak 2019 Mar 22;19(1):48 [FREE Full text] [doi: 10.1186/s12911-019-0801-4] [Medline: 30902088]
23. Brims FJ, Meniawy TM, Duffus I, de Fonseka D, Segal A, Creaney J, et al. A novel clinical prediction model for prognosis in malignant pleural mesothelioma using decision tree analysis. J Thorac Oncol 2016 Apr;11(4):573-582 [FREE Full text] [doi: 10.1016/j.jtho.2015.12.108] [Medline: 26776867]
24. Shimose S, Kawaguchi T, Iwamoto H, Tanaka M, Miyazaki K, Ono M, et al. Controlling Nutritional Status (CONUT) score is associated with overall survival in patients with unresectable hepatocellular carcinoma treated with lenvatinib: a multicenter cohort study. Nutrients 2020 Apr 13;12(4):1076 [FREE Full text] [doi: 10.3390/nu12041076] [Medline: 32295043]
25. Vallée A, Safar ME, Blacher J. Application of a decision tree to establish factors associated with a nomogram of aortic stiffness. J Clin Hypertens (Greenwich) 2019 Oct 03;21(10):1484-1492 [FREE Full text] [doi: 10.1111/jch.13662] [Medline: 31479194]

26. Gu J, Chao J, Chen W, Xu H, Wu Z, Chen H, et al. Multimorbidity in the community-dwelling elderly in urban China. Arch Gerontol Geriatr 2017 Jan;68:62-67. [doi: 10.1016/j.archger.2016.09.001] [Medline: 27654809]
27. Wang HHX, Wang JJ, Wong SYS, Wong MCS, Li FJ, Wang PX, et al. Epidemiology of multimorbidity in China and implications for the healthcare system: cross-sectional survey among 162,464 community household residents in southern China. BMC Med 2014 Oct 23;12:188 [FREE Full text] [doi: 10.1186/s12916-014-0188-0] [Medline: 25338506]
28. Garin N, Koyanagi A, Chatterji S, Tyrovolas S, Olaya B, Leonardi M, et al. Global multimorbidity patterns: a cross-sectional, population-based, multi-country study. J Gerontol A Biol Sci Med Sci 2016 Feb 29;71(2):205-214 [FREE Full text] [doi: 10.1093/gerona/glv128] [Medline: 26419978]
29. Ma H, Ding J, Liu M, Liu Y. Connections between various disorders: combination pattern mining using apriori algorithm based on diagnosis Information from electronic medical records. Biomed Res Int 2022;2022:2199317 [FREE Full text] [doi: 10.1155/2022/2199317] [Medline: 35601156]
30. Ramezankhani A, Pournik O, Shahrabi J, Azizi F, Hadaegh F. An application of association rule mining to extract risk pattern for type 2 diabetes using tehran lipid and glucose study database. Int J Endocrinol Metab 2015 Apr 30;13(2):e25389 [FREE Full text] [doi: 10.5812/ijem.25389] [Medline: 25926855]
31. Held FP, Blyth F, Gnjidic D, Hirani V, Naganathan V, Waite LM, et al. Association rules analysis of comorbidity and multimorbidity: the Concord Health and Aging in Men Project. J Gerontol A Biol Sci Med Sci 2016 May 27;71(5):625-631. [doi: 10.1093/gerona/glv181] [Medline: 26508296]
32. Feng Y, Wang J, Shao Z, Chen Z, Yao T, Dong S, et al. Predicting related factors of immunological response to hepatitis B vaccine in hemodialysis patients based on integration of decision tree classification and logistic regression. Hum Vaccin Immunother 2021 Sep 02;17(9):3214-3220 [FREE Full text] [doi: 10.1080/21645515.2021.1895603] [Medline: 33989106]
33. Vallée A, Petruescu L, Kretz S, Safar M, Blacher J. Added value of aortic pulse wave velocity index in a predictive diagnosis decision tree of coronary heart disease. Am J Hypertens 2019 Mar 16;32(4):375-383. [doi: 10.1093/ajh/hpz004] [Medline: 30624553]
34. Amini P, Maroufizadeh S, Samani RO, Hamidi O, Sepidarkish M. Prevalence and determinants of preterm birth in Tehran, Iran: a comparison between logistic regression and decision tree methods. Osong Public Health Res Perspect 2017 Jun 30;8(3):195-200 [FREE Full text] [doi: 10.24171/j.phrp.2017.8.3.06] [Medline: 28781942]
35. Ahmed AM, Rizaner A, Ulusoy AH. A novel decision tree classification based on post-pruning with Bayes minimum risk. PLoS One 2018 Apr 4;13(4):e0194168 [FREE Full text] [doi: 10.1371/journal.pone.0194168] [Medline: 29617369]
36. Thottam GE, Krasnokutsky S, Pillinger MH. Gout and metabolic syndrome: a tangled web. Curr Rheumatol Rep 2017 Aug 26;19(10):60. [doi: 10.1007/s11926-017-0688-y] [Medline: 28844079]
37. Diehm C, Allenber J, Pittrow D. Mortality and vascular morbidity in older adults with asymptomatic versus symptomatic peripheral artery disease. Journal of Vascular Surgery 2010 Jul;52(1):253-254. [doi: 10.1016/j.jvs.2010.05.078]
38. Chen X, Zhou D, Shen J, Wu Y, Sun Q, Dong J, et al. Prevalence and causes of visual impairment in adults in Binhu District, Wuxi, China. Med Sci Monit 2018 Jan 16;24:317-323 [FREE Full text] [doi: 10.12659/msm.908218] [Medline: 29335399]
39. Zhao C, Cun Q, Tao Y, Yang W, Zhong H, Li F, et al. Effect of intraocular lens implantation on visual field in glaucoma and comorbid cataracts. Int J Ophthalmol 2020;13(4):580-586. [doi: 10.18240/ijo.2020.04.08] [Medline: 32399408]
40. Muhammad A, Mada SB, Malami I, Forcados GE, Erukainure OL, Sani H, et al. Postmenopausal osteoporosis and breast cancer: The biochemical links and beneficial effects of functional foods. Biomed Pharmacother 2018 Nov;107:571-582. [doi: 10.1016/j.biopha.2018.08.018] [Medline: 30114641]
41. Barzi A, Hershman DL, Till C, Barlow WE, Ramsey S, Lenz H, et al. Osteoporosis in colorectal cancer survivors: analysis of the linkage between SWOG trial enrollees and Medicare claims. Arch Osteoporos 2019 Jul 28;14(1):83 [FREE Full text] [doi: 10.1007/s11657-019-0629-7] [Medline: 31352608]
42. Papaleontiou M, Banerjee M, Reyes-Gastelum D, Hawley S, Haymart M. Risk of osteoporosis and fractures in patients with thyroid cancer: a case-control study in U.S. veterans. Oncologist 2019 Sep;24(9):1166-1173 [FREE Full text] [doi: 10.1634/theoncologist.2019-0234] [Medline: 31164453]
43. Shapiro CL. Osteoporosis: a long-term and late-effect of breast cancer treatments. Cancers (Basel) 2020 Oct 23;12(11):3094 [FREE Full text] [doi: 10.3390/cancers12113094] [Medline: 33114141]
44. Batchelor EK, Kapitsinou P, Pergola PE, Kovesdy CP, Jalal DI. Iron deficiency in chronic kidney disease: updates on pathophysiology, diagnosis, and treatment. JASN 2020 Feb 10;31(3):456-468. [doi: 10.1681/asn.2019020213]
45. Hanna RM, Streja E, Kalantar-Zadeh K. Burden of anemia in chronic kidney disease: beyond erythropoietin. Adv Ther 2021 Jan 29;38(1):52-75 [FREE Full text] [doi: 10.1007/s12325-020-01524-6] [Medline: 33123967]
46. Babitt JL, Lin HY. Mechanisms of anemia in CKD. JASN 2012 Aug 30;23(10):1631-1634. [doi: 10.1681/asn.2011111078]
47. Glassock RJ, Warnock DG, Delanaye P. The global burden of chronic kidney disease: estimates, variability and pitfalls. Nat Rev Nephrol 2017 Feb 12;13(2):104-114. [doi: 10.1038/nrneph.2016.163] [Medline: 27941934]
48. Stauder R, Valent P, Theurl I. Anemia at older age: etiologies, clinical implications, and management. Blood 2018 Feb 01;131(5):505-514 [FREE Full text] [doi: 10.1182/blood-2017-07-746446] [Medline: 29141943]

Abbreviations

ARM: association rule mining
CBD: cerebrovascular disease
CKD: chronic kidney disease
CLD: chronic liver disease
DM: diabetes mellitus
HD: heart disease
HT: hypertension
ICD-10: International Classification of Diseases version 10
LMD: lipoprotein metabolism disorder
MT: malignant tumor
OR: odds ratio
PVD: peripheral vascular disease
SC: senile cataract

Edited by Y Khader; submitted 22.03.22; peer-reviewed by Y He, G Kolostoumpis, Y Chu, J Farzi; comments to author 21.05.22; revised version received 13.07.22; accepted 10.09.22; published 24.11.22.

Please cite as:
Zhang Y, Chen C, Huang L, Liu G, Lian T, Yin M, Zhao Z, Xu J, Chen R, Fu Y, Liang D, Zeng J, Ni J
Associations Among Multimorbid Conditions in Hospitalized Middle-aged and Older Adults in China: Statistical Analysis of Medical Records
JMIR Public Health Surveill 2022;8(11):e38182
URL: https://publichealth.jmir.org/2022/11/e38182
doi:10.2196/38182
PMID:36422885

Original Paper

The Association of Midday Napping With Hypertension Among Chinese Adults Older Than 45 Years: Cross-sectional Study

Dongfeng Tang[1,2], MA; Yiheng Zhou[3], MSc; Chengxu Long[4], PhD; Shangfeng Tang[1,5], PhD

[1]School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Institute for Hospital Management, Tsinghua University, Shenzhen, China

[3]College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow, United Kingdom

[4]Department of Global Health and Social Medicine, King's College London, London, United Kingdom

[5]Research Center for Rural Health Service, Key Research Institute of Humanities and Social Sciences of Hubei Provincial Department of Education, Wuhan, China

Corresponding Author:
Shangfeng Tang, PhD
School of Medicine and Health Management
Tongji Medical College
Huazhong University of Science and Technology
13 Hangkong Road
Wuhan, 430030
China
Phone: 86 133 4989 5639
Email: Sftang2018@hust.edu.cn

Abstract

Background: Hypertension is one of the main public health issues around worldwide, and midday napping is a popular habit. The association between the two remains to be explored.

Objective: The goal of the research was to explore the association of midday napping with hypertension.

Methods: This study separately selected 11,439, 12,689, and 9464 Chinese respondents aged over 45 years from the China Health and Retirement Longitudinal Study 2011, 2015, and 2018 data sets. Binary logistic regression was used to explore the association of midday napping with hypertension, and the 3-step method was used to test the mediation effect of BMI.

Results: Among all respondents, the prevalence rates of hypertension were 24.6% (2818/11439) in 2011, 21.1% (2683/12689) in 2015, and 22.1% (2092/9464) in 2018. Midday napping was positively correlated with hypertension. In 2011 and 2015, napping 60 to 90 minutes had the greatest odds ratios [OR] (OR_{2011} 1.705, OR_{2015} 1.494). In 2018, the biggest OR came from the group napping 30 to 60 minutes (OR 1.223), and ORs of different napping durations decreased from 2011 to 2018. In addition, BMI had a partial mediation effect in 2015 and 2018.

Conclusions: Midday napping is a potential risk factor for hypertension with BMI acting as a mediator. To prevent hypertension, avoiding prolonged duration of midday napping and taking action to maintain a normal BMI level are recommended.

(JMIR Public Health Surveill 2022;8(11):e38782) doi:10.2196/38782

KEYWORDS

hypertension; risk factor; midday napping; BMI; mediation effect

Introduction

Hypertension is one of the main public health issues worldwide, and it has been identified as one of the main risks for stroke, heart failure, and cerebrovascular disease [1-3]. As of 2019, 1.3 billion people, or more than 16% of the world's population, are living with hypertension [4]. It has been estimated to contribute to 50% of coronary heart disease cases and two-thirds of the cerebrovascular disease burden [5]. Successive population surveys conducted in China over the last 30 years have revealed an increasing prevalence of hypertension [6,7]. Now there are 270 million hypertensive patients in China, and it has become the main culprit for disability-adjusted life years, contributing to 24.6% of all-cause mortality [8,9].

Considering the high prevalence and enormous health toll, a series of actions have been taken in China. In 2009, the New

Health Care System Reform was introduced, and hypertension management was made a vital public health service free for all patients [10]. It was stipulated that primary health care facilities must provide residents with free screening, management, and follow-up services [11]. Additionally, the Chinese central government has constructed many national demonstration areas for community-based hypertension management and comprehensive prevention and control of hypertension to improve the lifestyle and health literacy of the population [12]. The turning point came when the Primary Health Care, Medicine, and Health Promotion Law, pioneering legislation for health promotion in China, was implemented in 2020. It established the legality and necessity of a population-wide hypertension prevention and control approach [13]. As a result, the long ignored prevention of hypertension is being addressed, emphasizing the improvement of modifiable risk factors as a public priority.

Previous studies have identified some modifiable risk factors related to hypertension, including excessive drinking, smoking, unhealthy diet, and lack of exercise [14-16]. Some researchers spotted the link between sleep and hypertension and concluded that sleep duration and quality were strongly associated with the risk of hypertension [17-19]. However, the effect of midday napping, another popular sleep activity, has rarely been addressed. Although some studies indicated an independent association between midday napping and the incidence of hypertension [20-22], study results conflicted. Additionally, the association of hypertension with overweight and obesity has been extensively proven, and the prevalence of hypertension among the obese population may range from 60% to 77%, increasing with BMI [23,24]. Prolonged midday napping duration was found to elevate cortisol levels, resulting in abnormal fat distribution [25]. In addition, decreased thermogenesis and energy expenditure and an activated sympathetic nervous system caused by midday napping may also contribute to obesity [26,27]. Therefore, BMI seems an appropriate mediator to explore the association between midday napping and hypertension and help understand the underlying mechanism. Thus, 3 samples (2011, 2015, 2018) from the China Health and Retirement Longitudinal Study were used in this study to examine the relationship between midday napping and hypertension among middle-aged and older Chinese people and test the mediation effect of BMI. By identifying the potential risk modifiable factors, this study aimed to influence individual lifestyles and public policy to control hypertension.

Methods

Sample and Data Collection

The primary data used in this study are from the China Health and Retirement Longitudinal Study, a longitudinal national study conducted in 450 neighborhoods and village committees in 150 counties across 28 provinces. A 4-stage, stratified, cluster probability sampling design was adopted in the baseline survey, and detailed sampling procedures were shown in the study by Wang et al [28]. Data regarding individual demographic and socioeconomic status, health conditions, and related behavior information were collected among residents aged 45 years and older in China. Participants were excluded for the following reasons: aged younger than 45 years, values missing for BMI or height and weight, and information on hypertension missing. The final sample sizes are 11,439 in 2011, 12,689 in 2015, and 9464 in 2018.

Ethics Approval

The study was approved by the institutional review board of Peking University Health Science Center (IRB approval number for the main household survey, including anthropometrics: IRB00001052-11015; IRB approval number for biomarker collection: IRB00001052-11014). All participants provided their written informed consent before completing the interview.

Variables

Primary Dependent Variable

The dependent variable was a binary variable indicating whether a resident suffered from hypertension. Hypertension is defined in accordance with the national guidelines for primary hypertension prevention and management [29,30]: currently taking antihypertensive drugs, previously diagnosed as hypertensive by a clinician, or systolic blood pressure over 140 mm Hg or diastolic blood pressure over 90 mm Hg without antihypertensive drugs.

Primary Independent Variable

Midday napping was set as the independent variable, grouped by napping duration, which was appraised using a self-reported questionnaire [31] that asked, “During the past month, how long did you take a nap after lunch on average?” According to existing literature, categories ranging from no napping to napping longer than 90 minutes were defined (see Multimedia Appendix 1 for data) [32,33].

Control Variables

Sociodemographic characteristics (gender, age, education, residential status, marital status, household annual income per capita), health-related variables (self-reported health status, activities of daily living [ADL], mental health, personal medical histories, BMI, lifestyles (smoking status, drinking status, and night sleep duration) were included in this study (see Multimedia Appendix 1 for data) [34-38]. The information was collected by using a structured questionnaire. Age and household annual income per capita were set as continuous variables, and household annual income per capita was log transformed [39]. Participants were categorized as ADL impaired if they reported difficulty or inability performing any activity item [40]. Mental health was appraised using the 10-item Center for Epidemiological Studies Depression Scale (<10=no depressive symptoms and ≥=depressive symptoms) [41]. Cardiovascular diseases were self-reported as chronic heart problems, stroke, or both [42]. BMI was categorized as underweight (BMI <18.5), normal (18.5≤BMI<25.0), overweight (25.0≤BMI<30.0), and obese (BMI ≥30) [43].

Data Analysis

The disparity in hypertension across different groups was examined by chi-square and independent sample *t* test. After adjusting for control variables, binary logistic regression was

used to explore the relationship between midday napping and hypertension. Variables in the regression model were selected using the Enter method. The association between midday napping and hypertension was quantified using odds ratios (ORs) having 95% confidence intervals, with other variables controlled. To verify whether BMI played a role in the influence of midday napping on hypertension, the 3-step method proposed by Baron and Kenny [44] was used to test the mediating effect of BMI. The judging criteria for whether there was a mediation effect were taken as follows: statistically significant relationship between independent variable (X, coefficient=a) and mediator (M), significant relationship between independent variable (X, coefficient=c) and dependent variable (Y), and coefficient of mediator (coefficient=b) in the regression model that contained independent variable, mediator, and dependent variable is statistically significant [45]. Mediator was defined as complete if the coefficient of X was not significant in the regression model including X, M, and Y and partial if the coefficient of X was still significant in the regression model, indicating other remaining factors in the path from X to Y. The mediation effect value was calculated as a*b, and the ratio of the mediating effect with the total effect was calculated as a*b/c [46]. $P<.05$ (2-tailed) was regarded as statistically significant. The data were described and analyzed using SPSS (version 24.0, IBM Corp).

Results

Sample Characteristics

There was a reasonably steady percentage of participants overall who had hypertension: 24.64% (2818/11,439) in 2011, 21.14% (2683/12,689) in 2015, and 22.10% (2092/9464) in 2018. Participants who regularly took midday naps were 54.19% (6166/11,439) in 2011, 58.39% (7409/12,689) in 2015, and 60.51% (5727/9464) in 2018. Among all midday nappers, those who napped between 60 and 90 minutes were the largest group, with 23.88% (2717/11,439) in 2011, 27.43% (3480/12,689) in 2015, and 23.63% (1717/9464) in 2018. There were slightly more female participants than male (6018 vs 5421 in 2011, 6424 vs 6265 in 2015, and 4971 vs 4493 in 2018), with an average age of 59.5 years in 2011, 61.0 years in 2015, and 60.6 years in 2018. More information can be found in Multimedia Appendix 1.

Association Between Midday Napping and Hypertension

Midday napping was found to be positively correlated with hypertension. In 2011 and 2015, after being adjusted by other control variables, groups of nappers (considering the napping durations 0-30, 30-60, 60-90, and >90 minutes) were all positively correlated with hypertension. Napping 60 to 90 minutes had the greatest ORs (2011: OR 1.705, 95% CI 1.346-2.159; 2015: OR 1.494, 95% CI 1.227-1.818) compared with nonnappers. In 2018, except for the group napping 0 to 30 minutes, participants were positively correlated with hypertension, and the greatest OR came from the group napping 30 to 60 minutes (OR 1.223, 95% CI 1.016-1.473). See Figure 1.

From the longitudinal perspective, the ORs of each group of nappers decreased from 2011 to 2018. ORs of napping 60 to 90 minutes decreased the most, from 1.705 in 2011 to 1.338 in 2015 and 1.163 in 2018. The ORs of napping more than 90 minutes decreased from 1.412 in 2011 to 1.220 in 2018. The ORs of napping 30 to 60 minutes decreased from 1.319 in 2011 to 1223 in 2018. Last, the ORs of napping 0 to 30 minutes decreased from 1.458 in 2011 to 1.331 in 2015. See Figure 2.

Figure 1. Influence of midday napping on hypertension in different years. OR: odds ratio. *P<.05, **P<.01, ***P<.001. The horizontal line at the end of each line represents the 95% confidence interval, the square in the middle line represents the OR value, and the line segment intersects with the middle vertical line (=1), which means that the result is not significant (P>.05). Non-intersection means that the result is significant (P<.05). Unit: minute.

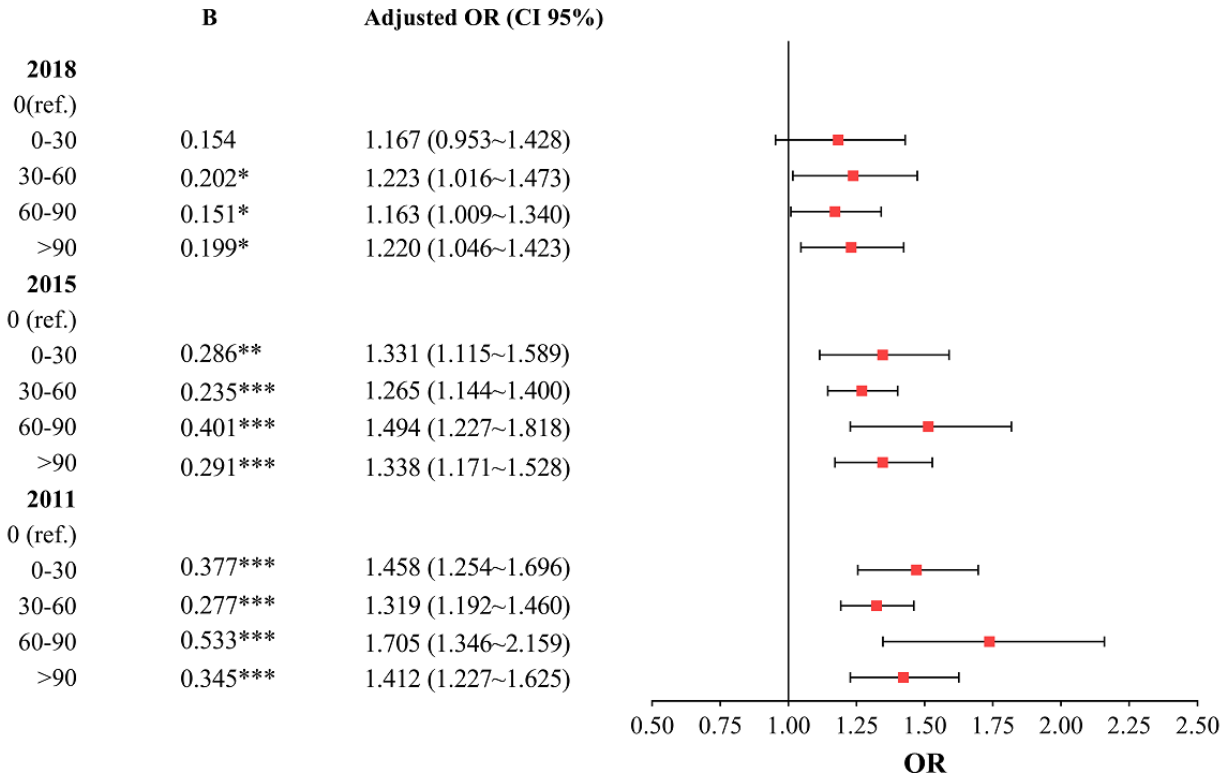
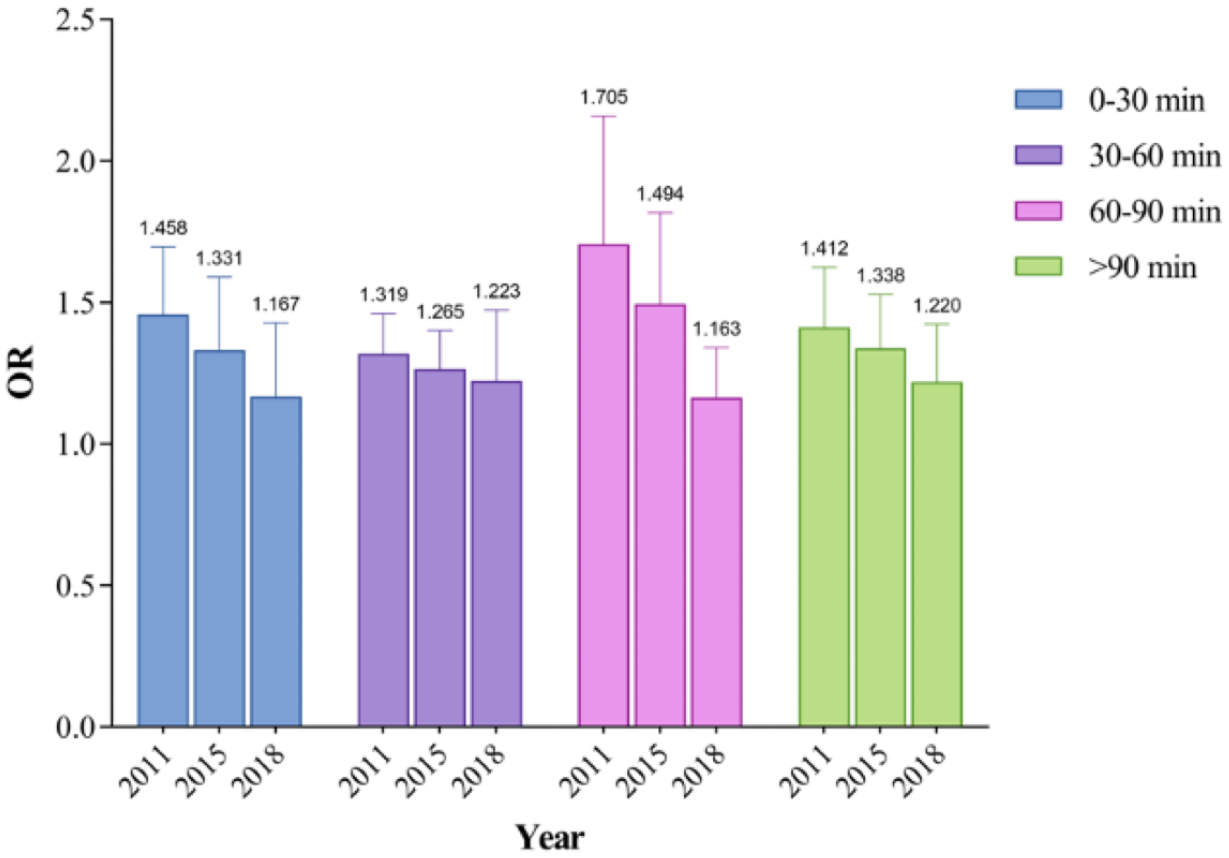


Figure 2. Trajectories of odds ratios from 2011 to 2018. The odds ratio of napping 0 to 30 minutes was not significant in 2018. OR: odds ratio.



Mediation Effect of BMI

The data revealed that the mediation effect of BMI existed in 2015 and 2018 but not in 2011. In 2015, the total effect of midday napping was found to be significant on hypertension (path c: B 0.012, *P*<.001). Midday napping had a positive effect on BMI in path a (B 0.011, *P*<.01), and BMI had a positive effect on hypertension in path b (B 0.022, *P*<.001). In path c', the effect of midday napping was also significant (B 0.012, *P*<.001), so the BMI was identified as a partial mediator. The mediation effect was 0.000242, with a ratio of 2.01% to the total effect. In 2018, the coefficients in path a, path b, and path c were also found to be significant, and the mediation effect of BMI was identified as partial, reaching 0.003058. The ratio of the mediation effect over the total effect increased to 23.52% (see Table 1).

Table 1. Coefficient (B) in testing model of mediation effect.

Year	X[a]–Y[b](path c[c])	X–M[d](path a[e])	(X+M)–Y	
			X(path c'[f])	M(path b[g])
2011	0.018[h]	–0.023[h]	0.017[h]	–0.005
2015	0.012[h]	0.011[i]	0.012[h]	0.022[h]
2018	0.013[h]	0.022[h]	0.010[h]	0.139[h]

[a]X: midday napping time.

[b]Y: hypertension.

[c]path c: regression between X and Y.

[d]M: BMI.

[e]path a: regression between X and M.

[f]path c: regression between X and Y with M controlled.

[g]path b: regression between M and Y with X controlled.

[h]*P*<.001.

[i]*P*<.01.

Discussion

Principal Findings

This cross-sectional study found midday napping positively associated with hypertension among 3 sectional samples in China, indicating that midday napping may represent a potential causal risk factor. Meanwhile, the ORs of various napping duration decreased over time. The BMI was found to be a partial mediator between midday napping and hypertension.

Although napping has long been regarded as a healthy habit, this study suggests that it may be a potential risk factor for hypertension. Evidence from the UK Biobank [47] and cohort studies in China [33] also supported the results of this study. A meta-analysis concluded that the pooled relative risk of hypertension in nappers was 1.13 based on 9 observational studies [48]. However, disparities between this study and existing literature also exist, indicating the need for a cautious interpretation of the results. Some other studies found midday napping to have a protective effect for habitual nappers compared with those who never napped [49] or to decrease the risk of hypertension in specific napping durations [50,51], contradictory to the results in this study. Meanwhile, in this study, different durations of midday napping were all found to be positively associated with hypertension (except napping for 0 to 30 minutes in 2018). However, the associations of midday napping duration with hypertension differed in various studies. For example, the significantly increased odds for hypertension were only found in participants napping over 90 minutes in some studies [32,35]. Another cohort study conducted in China including 13,706 participants found no significant associations of napping for less than 30 minutes with hypertension [36]. The inconsistency might be explained by study designs and samples, different characteristics of participants, disparity in included confounders, and measurements of napping behaviors and other confounders across studies. Therefore, it is important to be cautious about the results, and long-term follow-up and experimental studies are needed to determine the exact impacts of midday napping.

From 2011 to 2018, decreases in ORs were seen in different napping durations (napping 0-30 minutes: 1.458 to 1.331; napping 30-60 minutes: 1.319 to 1.223; napping 60-90 minutes: 1.705 to 1.338; napping over 90 minutes: 1.412 to 1.220). Some speculations were made to understand the results. First, there were only 4 variables significant (including education, marital status, drinking, and napping duration) in the regression model of 2011, but the corresponding number was 8 in 2018 (including gender, age, health status, ADL, diabetes, cardiovascular disease, BMI, and napping duration). The increasing correlation between significant variables might decrease the value of ORs. Second, the great socioeconomic and environmental transformations related to hypertension during 2011-2018, such as dietary patterns, exposure to fine particulate matter ($PM_{2.5}$), built environment factors, and some other confounders, were not controlled in our study [52-54]. Third, the association of midday napping with hypertension might be moderated by other variables such as physical conditions and night sleep duration [49,55]. Therefore, current evidence was not enough to conclude that the impact of midday napping decreased, and this can only be explained after determining the hidden specific mechanism. However, the results deserve our attention because they indicate

the possibility that the potential risk of midday napping might be mitigated or even eliminated if we can control other confounders well.

The mediation effect of BMI was identified in this study. Previous studies found that nappers were more likely to be overweight or obese [34,56,57], so it could be inferred that midday napping contributes to hypertension by elevating the risk of obesity or overweight, which is an acknowledged risk factor for hypertension [58-60]. However, the ratio of mediation effect over the total effect was 2.01% in 2015 and 23.5% in 2018, indicating the existence of other mediators. It was suggested that midday napping could result in sympathetic surge and an increase in nighttime cortisol, elevating blood pressure [25]. Midday napping was also regarded as a symptom of sleep apnea, and it was concluded that the sleep apnea and not the napping itself resulted in cardiovascular diseases [61]. Furthermore, prolonged midday napping may have an impact on the duration and quality of evening sleep [62]. All these factors can indirectly increase the risk of hypertension.

To prevent hypertension, prolonged midday napping should be avoided, and actions related to losing weight such as increased physical activity and a balanced diet are also needed, especially for nappers. Additionally, further research is needed to define the vulnerable population and develop corresponding interventions.

Limitations

In this study, there were some limitations that should be mentioned for cautious interpretation. First, despite the positive correlation observed, the regression model and cross-sectional study design used were not robust enough to conduct the causal inference, which weakened the evidence. Second, the use of self-reported midday napping duration and some other variables might introduce recall bias. Third, although some confounders were adjusted in the model, potential residual covariates might remain due to the absence of information such as genetic factors, family history of hypertension, biomarkers, and environmental factors. In addition, time-dependent covariates were not included in our study, which made comparisons across years difficult. Fourth, all participants were aged 45 years and older, and it remains uncertain whether the conclusion can be applied to other age groups. Additionally, although we added night sleep duration as a control variable, the potential interaction effect of midday napping and night sleep was not analyzed in this study.

Conclusion

In this study, it was found that midday napping was positively associated with hypertension in Chinese people middle-aged and older. Although the causal effects were hard to prove, BMI was found to play the role of mediator. Therefore, avoiding prolonged midday napping and taking action to maintain a normal BMI level are recommended. For future research, the specific mechanism of interaction between midday napping and hypertension deserves more attention as does investigating of other implications of midday napping considering its high prevalence.

Acknowledgments

This work was funded by grant 2022YFE0133000 from the National Key R&D Program of China, grant 72004073 from the National Natural Science Foundation of China, and grant 20YJC630134 from the Chinese Ministry of Education of Humanities and Social Science project.

Authors' Contributions

DT was responsible for the study design, data analysis, interpretation of the data, and writing the manuscript. YZ was responsible for the study design, data analysis, and writing the manuscript. CL and ST were responsible for the study design and writing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant characteristics.
[DOCX File , 31 KB - publichealth_v8i11e38782_app1.docx]

References

1. Cipolla MJ, Liebeskind DS, Chan S. The importance of comorbidities in ischemic stroke: impact of hypertension on the cerebral circulation. J Cereb Blood Flow Metab 2018 Sep 10;38(12):2129-2149. [doi: 10.1177/0271678x18800589]
2. Hamrahian S, Falkner B. Hypertension in chronic kidney disease. In: Islam MS, editor. Hypertension: From Basic Research to Clinical Practice. Advances in Experimental Medicine and Biology. Cham: Springer; 2017:307-325.
3. Kjeldsen SE. Hypertension and cardiovascular risk: General aspects. Pharmacol Res 2018 Mar;129:95-99. [doi: 10.1016/j.phrs.2017.11.003] [Medline: 29127059]

4. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. Lancet 2021 Sep 11;398(10304):957-980 [FREE Full text] [doi: 10.1016/S0140-6736(21)01330-1] [Medline: 34450083]
5. Wang J, Zhang L, Wang F, Liu L, Wang H, China National Survey of Chronic Kidney Disease Working Group. Prevalence, awareness, treatment, and control of hypertension in China: results from a national survey. Am J Hypertens 2014 Nov;27(11):1355-1361 [FREE Full text] [doi: 10.1093/ajh/hpu053] [Medline: 24698853]
6. Lewington S, Lacey B, Clarke R. Uncontrolled Hypertension and Risk of Cardiovascular Mortality in China-Reply. JAMA Intern Med 2016 Aug 01;176(8):1234. [doi: 10.1001/jamainternmed.2016.3825] [Medline: 27479679]
7. Ma L, Chen W, Gao R, Liu L, Zhu M, Wang Y, et al. China cardiovascular diseases report 2018: an updated summary. J Geriatr Cardiol 2020 Jan;17(1):1-8 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2020.01.001] [Medline: 32133031]
8. Zhou M, Wang H, Zeng X, Yin P, Zhu J, Chen W, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2019 Sep;394(10204):1145-1158. [doi: 10.1016/s0140-6736(19)30427-1]
9. Liu M, Li Y, Liu S, Wang W, Zhou M. [Burden on blood-pressure-related diseases among the Chinese population, in 2010]. Zhonghua Liu Xing Bing Xue Za Zhi 2014 Jun;35(6):680-683. [Medline: 25174471]
10. Tang S, Bishwajit G, Ji L, Feng D, Fang H, Fu H, et al. Improving the blood pressure control with the proactive attitude of hypertensive patients seeking follow-up services: evidence from china. Medicine (Baltimore) 2016 Apr;95(14):e3233 [FREE Full text] [doi: 10.1097/MD.0000000000003233] [Medline: 27057859]
11. Wang Y, Hu X, Wang HHX, Duan H, Chen Y, Li Y, et al. Follow-up care delivery in community-based hypertension and type 2 diabetes management: a multi-centre, survey study among rural primary care physicians in China. BMC Fam Pract 2021 Nov 13;22(1):224 [FREE Full text] [doi: 10.1186/s12875-021-01564-z] [Medline: 34774003]
12. Hou L, Chen B, Ji Y, Wang B, Wu J. China CDC in action: hypertension prevention and control. China CDC Wkly 2020 Oct 02;2(40):783-786 [FREE Full text] [doi: 10.46234/ccdcw2020.212] [Medline: 34594767]
13. Hou L. Tamping legal basis of "Health First, Prevention First?": deliberation and advices on the primary health care, medicine and health promotion law. Chin J Soc Med 2020;37(3):238-241.
14. Hall JE, do Carmo JM, da Silva AA, Wang Z, Hall ME. Obesity-induced hypertension interaction of neurohumoral and renal mechanisms. Circ Res 2015 Mar 13;116(6):991-1006. [doi: 10.1161/circresaha.116.305697]
15. Cherfan M, Vallée A, Kab S, Salameh P, Goldberg M, Zins M, et al. Unhealthy behaviors and risk of uncontrolled hypertension among treated individuals-the constances population-based study. Sci Rep 2020 Feb 05;10(1):1925 [FREE Full text] [doi: 10.1038/s41598-020-58685-1] [Medline: 32024888]
16. Gao K, Shi X, Wang W. The life-course impact of smoking on hypertension, myocardial infarction and respiratory diseases. Sci Rep 2017 Jun 28;7(1):4330 [FREE Full text] [doi: 10.1038/s41598-017-04552-5] [Medline: 28659608]
17. Grandner M, Mullington JM, Hashmi SD, Redeker NS, Watson NF, Morgenthaler TI. Sleep duration and hypertension: analysis of > 700,000 adults by age and sex. J Clin Sleep Med 2018 Jun 15;14(6):1031-1039 [FREE Full text] [doi: 10.5664/jcsm.7176] [Medline: 29852916]
18. Lo K, Woo B, Wong M, Tam W. Subjective sleep quality, blood pressure, and hypertension: a meta-analysis. J Clin Hypertens (Greenwich) 2018 Mar 19;20(3):592-605 [FREE Full text] [doi: 10.1111/jch.13220] [Medline: 29457339]
19. Zhang H, Zhao X, Li Y, Mao Z, Huo W, Jiang J, et al. Night sleep duration and sleep initiation time with hypertension in Chinese rural population: the Henan Rural Cohort. Eur J Public Health 2020 Feb 01;30(1):164-170. [doi: 10.1093/eurpub/ckz142] [Medline: 31504445]
20. Liu R, Qian Z, Trevathan E, Chang J, Zelicoff A, Hao Y, et al. Poor sleep quality associated with high risk of hypertension and elevated blood pressure in China: results from a large population-based study. Hypertens Res 2016 Jan 03;39(1):54-59. [doi: 10.1038/hr.2015.98] [Medline: 26333359]
21. Uchmanowicz I, Markiewicz K, Uchmanowicz B, Kołtuniuk A, Rosińczuk J. The relationship between sleep disturbances and quality of life in elderly patients with hypertension. Clin Interv Aging 2019 Jan;Volume 14:155-165. [doi: 10.2147/cia.s188499]
22. Shang X. Meta-analysis of self-reported daytime napping and risk of cardiovascular or all-cause mortality. Med Sci Monit 2015;21:1269-1275. [doi: 10.12659/msm.893186]
23. Hall JE, do Carmo JM, da Silva AA, Wang Z, Hall ME. Obesity-Induced Hypertension. Circ Res 2015 Mar 13;116(6):991-1006. [doi: 10.1161/circresaha.116.305697]
24. Bramlage P, Pittrow D, Wittchen H, Kirch W, Boehler S, Lehnert H, et al. Hypertension in overweight and obese primary care patients is highly prevalent and poorly controlled. Am J Hypertens 2004 Oct;17(10):904-910. [doi: 10.1016/j.amjhyper.2004.05.017] [Medline: 15485752]
25. Woods DL, Kim H, Yefimova M. To nap or not to nap: excessive daytime napping is associated with elevated evening cortisol in nursing home residents with dementia. Biol Res Nurs 2013 Apr 13;15(2):185-190. [doi: 10.1177/1099800411420861] [Medline: 21998447]
26. Sun K, Li F, Qi Y, Lin D, Ren M, Xu M, et al. Sex difference in the association between habitual daytime napping and prevalence of diabetes: a population-based study. Endocrine 2016 May 20;52(2):263-270. [doi: 10.1007/s12020-015-0772-x] [Medline: 26487615]

27. Lin D, Sun K, Li F, Qi Y, Ren M, Huang C, et al. Association between habitual daytime napping and metabolic syndrome: a population-based study. Metabolism 2014 Dec;63(12):1520-1527. [doi: 10.1016/j.metabol.2014.08.005] [Medline: 25249445]
28. Wang R, Chen Z, Zhou Y, Shen L, Zhang Z, Wu X. Melancholy or mahjong? Diversity, frequency, type, and rural-urban divide of social participation and depression in middle- and old-aged Chinese: a fixed-effects analysis. Soc Sci Med 2019 Oct;238:112518. [doi: 10.1016/j.socscimed.2019.112518] [Medline: 31473574]
29. Sugianto RI, Schmidt BMW, Memaran N, Duzova A, Topaloglu R, Seeman T, et al. Sex and age as determinants for high blood pressure in pediatric renal transplant recipients: a longitudinal analysis of the CERTAIN Registry. Pediatr Nephrol 2020 Mar 07;35(3):415-426. [doi: 10.1007/s00467-019-04395-4] [Medline: 31811541]
30. Zou G, Zhang Z, Walley J, Gong W, Yu Y, Hu R, et al. Use of medications and lifestyles of hypertensive patients with high risk of cardiovascular disease in rural China. PLoS One 2015 May 1;10(5):e0124484 [FREE Full text] [doi: 10.1371/journal.pone.0124484] [Medline: 25932640]
31. China Health and Retirement Longitudinal Study. URL: http://charls.pku.edu.cn/en/index.htm [accessed 2022-11-14]
32. Fu J, Zhang X, Moore JB, Wang B, Li R. Midday nap duration and hypertension among middle-aged and older Chinese adults: a nationwide retrospective cohort study. Int J Environ Res Public Health 2021 Apr 01;18(7):3680 [FREE Full text] [doi: 10.3390/ijerph18073680] [Medline: 33916042]
33. Cao Z, Shen L, Wu J, Yang H, Fang W, Chen W, et al. The effects of midday nap duration on the risk of hypertension in a middle-aged and older Chinese population: a preliminary evidence from the Tongji-Dongfeng Cohort Study, China. J Hypertens 2014 Oct;32(10):1993-1998. [doi: 10.1097/HJH.0000000000000291] [Medline: 25023156]
34. Ciren W, Nima Q, Li Y, He R, Suolang D, Ciren Z, et al. Association of daytime napping with chronic diseases among Tibetan people in China: a cross-sectional study. BMC Public Health 2021 Oct 08;21(1):1810 [FREE Full text] [doi: 10.1186/s12889-021-11871-w] [Medline: 34625060]
35. Yang Y, Liu W, Ji X, Ma C, Wang X, Li K, et al. Extended afternoon naps are associated with hypertension in women but not in men. Heart Lung 2020 Jan;49(1):2-9 [FREE Full text] [doi: 10.1016/j.hrtlng.2019.09.002] [Medline: 31521340]
36. Wang L, Wang K, Liu L, Zhang Y, Shu H, Wang K, et al. Associations of daytime napping with incident cardiovascular diseases and hypertension in chinese adults: a nationwide cohort study. Biomed Environ Sci 2022 Jan 20;35(1):22-34 [FREE Full text] [doi: 10.3967/bes2022.004] [Medline: 35078559]
37. Li Z, Fu C, Yang F, Mao Z. Prevalence and risk factors of hypertension for the middle-aged population in China: results from the China Health and Retirement Longitudinal Study (CHARLS). Clin Exp Hypertens 2019 Mar 19;41(1):80-86. [doi: 10.1080/10641963.2018.1445751] [Medline: 29553846]
38. Zhao Y, Mahal A, Tang S, Haregu T, Oldenburg B. Effective coverage for hypertension treatment among middle-aged adults and the older population in China, 2011 to 2013: A nationwide longitudinal study. J Glob Health 2020 Jun;10(1):010805 [FREE Full text] [doi: 10.7189/jogh.10.010805] [Medline: 32257169]
39. Savoldi A, Carrara E, Gladstone B, Azzini A, Göpel S, Tacconelli E. Gross national income and antibiotic resistance in invasive isolates: analysis of the top-ranked antibiotic-resistant bacteria on the 2017 WHO priority list. J Antimicrob Chemother 2019 Dec 01;74(12):3619-3625. [doi: 10.1093/jac/dkz381] [Medline: 31730162]
40. Li X, Jiang Q, Li S, Feldman MW. Female fertility history and mid-late-life health: findings from China. J Women Aging 2018 Feb 02;30(1):62-74 [FREE Full text] [doi: 10.1080/08952841.2016.1259445] [Medline: 28151095]
41. Tu R, Inoue Y, Yazawa A, Hao X, Cai G, Li Y, et al. Social participation and the onset of hypertension among the middle-aged and older population: evidence from the China Health and Retirement Longitudinal Study. Geriatr Gerontol Int 2018 Jul 30;18(7):1093-1099. [doi: 10.1111/ggi.13317] [Medline: 29602268]
42. Sun J, Ma Y, Liu H, Qu Q, Cheng C, Kong X, et al. High waist circumference is a risk factor of new-onset hypertension: evidence from the China Health and Retirement Longitudinal Study. J Clin Hypertens (Greenwich) 2022 Mar 21;24(3):320-328 [FREE Full text] [doi: 10.1111/jch.14446] [Medline: 35188335]
43. https://apps.who.int/iris/bitstream/handle/10665/42330/WHO_TRS_894.pdf?sequence=1&isAllowed=y. Geneva: World Health Organization; 2000. URL: https://apps.who.int/iris/bitstream/handle/10665/42330/WHO_TRS_894.pdf?sequence=1&isAllowed=y [accessed 2022-11-08]
44. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Personality Soc Psychol 1986;51(6):1173-1182. [doi: 10.1037/0022-3514.51.6.1173]
45. Deng Z, Lu Y, Wei KK, Zhang J. Understanding customer satisfaction and loyalty: an empirical study of mobile instant messages in China. Int J Inf Manag 2010 Aug;30(4):289-300. [doi: 10.1016/j.ijinfomgt.2009.10.001]
46. Shen Z, Shi S, Ding S, Zhong Z. Mediating Effect of Self-Efficacy on the Relationship Between Medication Literacy and Medication Adherence Among Patients With Hypertension. Front Pharmacol 2020;11:569092 [FREE Full text] [doi: 10.3389/fphar.2020.569092] [Medline: 33364943]
47. Yang M, Zhang Z, Wang Y, Li J, Guo Q, Chen X, et al. Association of nap frequency with hypertension or ischemic stroke supported by prospective cohort data and mendelian randomization in predominantly middle-aged European subjects. Hypertension 2022 Sep;79(9):1962-1970. [doi: 10.1161/hypertensionaha.122.19120]

48. Cheungpasitporn W, Thongprayoon C, Srivali N, Vijayvargiya P, Andersen CA, Kittanamongkolchai W, et al. The effects of napping on the risk of hypertension: a systematic review and meta-analysis. J Evid Based Med 2016 Nov 27;9(4):205-212. [doi: 10.1111/jebm.12211] [Medline: 27376587]
49. Zhao H, Gui W, Huang H, Liu Y, Ding H, Fan W, et al. Association of long-term sleep habits and hypertension: a cross-sectional study in Chinese adults. J Hum Hypertens 2020 May 20;34(5):378-387. [doi: 10.1038/s41371-019-0225-8] [Medline: 31431681]
50. Huang M, Yang Y, Huang Z, Yuan H, Lu Y. The association of nighttime sleep duration and daytime napping duration with hypertension in Chinese rural areas: a population-based study. J Hum Hypertens 2021 Oct 24;35(10):896-902. [doi: 10.1038/s41371-020-00419-x] [Medline: 32973309]
51. Cao Y, Li D, Li K, Yu H, Xin W, Miao D, et al. [Epidemiological study on the relationship between the siesta and blood pressure]. Zhonghua Yi Xue Za Zhi 2016 Jun 07;96(21):1699-1701. [doi: 10.3760/cma.j.issn.0376-2491.2016.21.018] [Medline: 27290714]
52. Xie H, Wang Q, Zhou X, Yang Y, Mao Y, Zhang X. Built environment factors influencing prevalence of hypertension at community level in China: the case of Wuhan. Sustainability 2021 May 17;13(10):5580. [doi: 10.3390/su13105580]
53. Huang K, Yang X, Liang F, Liu F, Li J, Xiao Q, et al. Long-term exposure to fine particulate matter and hypertension incidence in China. Hypertension 2019 Jun;73(6):1195-1201. [doi: 10.1161/hypertensionaha.119.12666]
54. Shang X, Flehr A, Fang Y, He M. Meal patterns and incident hypertension in community-dwelling middle-aged adults: an 11-year follow-up cohort study. J Hypertens 2021 Jul 01;39(7):1393-1401. [doi: 10.1097/HJH.0000000000002794] [Medline: 33470737]
55. Léger D, Torres MJ, Bayon V, Hercberg S, Galan P, Chennaoui M, et al. The association between physical and mental chronic conditions and napping. Sci Rep 2019 Feb 11;9(1):1795 [FREE Full text] [doi: 10.1038/s41598-018-37355-3] [Medline: 30741949]
56. Papandreou C, Díaz-López A, Babio N, Martínez-González M, Bulló M, Corella D, et al. Long daytime napping is associated with increased adiposity and type 2 diabetes in an elderly population with metabolic syndrome. J Clin Med 2019 Jul 19;8(7):1053 [FREE Full text] [doi: 10.3390/jcm8071053] [Medline: 31330940]
57. Ghazizadeh H, Mobarra N, Esmaily H, Seyedi SMR, Amiri A, Rezaeitalab F, et al. The association between daily naps and metabolic syndrome: evidence from a population-based study in the Middle-East. Sleep Health 2020 Oct;6(5):684-689. [doi: 10.1016/j.sleh.2020.03.007] [Medline: 32482574]
58. Juonala M, Magnussen CG, Berenson GS, Venn A, Burns TL, Sabin MA, et al. Childhood adiposity, adult adiposity, and cardiovascular risk factors. N Engl J Med 2011 Nov 17;365(20):1876-1885. [doi: 10.1056/NEJMoa1010112] [Medline: 22087679]
59. Lloyd-Sherlock P, Beard J, Minicuci N, Ebrahim S, Chatterji S. Hypertension among older adults in low- and middle-income countries: prevalence, awareness and control. Int J Epidemiol 2014 Feb 06;43(1):116-128 [FREE Full text] [doi: 10.1093/ije/dyt215] [Medline: 24505082]
60. Ryu S, Frith E, Pedisic Z, Kang M, Loprinzi PD. Secular trends in the association between obesity and hypertension among adults in the United States, 1999-2014. Eur J Intern Med 2019 Apr;62:37-42. [doi: 10.1016/j.ejim.2019.02.012] [Medline: 30826171]
61. Masa J, Rubio M, Pérez P, Mota M, de Cos J, Montserrat J. Association between habitual naps and sleep apnea. Sleep 2006 Nov;29(11):1463-1468. [doi: 10.1093/sleep/29.11.1463] [Medline: 17162994]
62. Léger D, Torres MJ, Bayon V, Hercberg S, Galan P, Chennaoui M, et al. The association between physical and mental chronic conditions and napping. Sci Rep 2019 Feb 11;9(1):1795 [FREE Full text] [doi: 10.1038/s41598-018-37355-3] [Medline: 30741949]

Abbreviations

ADL: activities of daily living
CESD-10: Center for Epidemiological Studies Depression Scale
OR: odds ratio

Edited by Y Khader; submitted 15.04.22; peer-reviewed by HH Wang, Z Ni; comments to author 26.08.22; revised version received 21.10.22; accepted 22.10.22; published 22.11.22.

Please cite as:
Tang D, Zhou Y, Long C, Tang S
The Association of Midday Napping With Hypertension Among Chinese Adults Older Than 45 Years: Cross-sectional Study
JMIR Public Health Surveill 2022;8(11):e38782
URL: https://publichealth.jmir.org/2022/11/e38782
doi:10.2196/38782
PMID:36279195

Original Paper

Outcomes of COVID-19 Infection in People Previously Vaccinated Against Influenza: Population-Based Cohort Study Using Primary Health Care Electronic Records

Maria Giner-Soriano[1,2*], PharmD, PhD; Vanessa de Dios[3*], MD; Dan Ouchi[1,2*], MSc; Carles Vilaplana-Carnerero[1,2*], PharmD, PhD; Mònica Monteagudo[1,2*], MD, PhD; Rosa Morros[1,2,4,5*], MD, PhD

[1]Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain

[2]Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Spain

[3]Department of Clinical Pharmacology, Medicines Area, Hospital Clinic of Barcelona, Barcelona, Spain

[4]Plataforma Spanish Clinical Research Network, Unidad de Investigación Clínica, Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain

[5]Institut Català de la Salut, Barcelona, Spain

[*]all authors contributed equally

Corresponding Author:
Maria Giner-Soriano, PharmD, PhD
Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)
Gran Via de les Corts Catalanes 587, àtic
Barcelona, 08007
Spain
Phone: 34 934824110
Fax: 34 934824174
Email: mginer@idiapjgol.info

Abstract

Background: A possible link between influenza immunization and susceptibility to the complications of COVID-19 infection has been previously suggested owing to a boost in the immunity against SARS-CoV-2.

Objective: This study aimed to investigate whether individuals with COVID-19 could have benefited from vaccination against influenza. We hypothesized that the immunity resulting from the previous influenza vaccination would boost part of the immunity against SARS-CoV-2.

Methods: We performed a population-based cohort study including all patients with COVID-19 with registered entries in the primary health care (PHC) electronic records during the first wave of the COVID-19 pandemic (March 1 to June 30, 2020) in Catalonia, Spain. We compared individuals who took an influenza vaccine before being infected with COVID-19, with those who had not taken one. Data were obtained from Information System for Research in Primary Care, capturing PHC information of 5.8 million people from Catalonia. The main outcomes assessed during follow-up were a diagnosis of pneumonia, hospital admission, and mortality.

Results: We included 309,039 individuals with COVID-19 and compared them on the basis of their influenza immunization status, with 114,181 (36.9%) having been vaccinated at least once and 194,858 (63.1%) having never been vaccinated. In total, 21,721 (19%) vaccinated individuals and 11,000 (5.7%) unvaccinated individuals had at least one of their outcomes assessed. Those vaccinated against influenza at any time (odds ratio [OR] 1.14, 95% CI 1.10-1.19), recently (OR 1.13, 95% CI 1.10-1.18), or recurrently (OR 1.10, 95% CI 1.05-1.15) before being infected with COVID-19 had a higher risk of presenting at least one of the outcomes than did unvaccinated individuals. When we excluded people living in long-term care facilities, the results were similar.

Conclusions: We could not establish a protective role of the immunity conferred by the influenza vaccine on the outcomes of COVID-19 infection, as the risk of COVID-19 complications was higher in vaccinated than in unvaccinated individuals. Our results correspond to the first wave of the COVID-19 pandemic, where more complications and mortalities due to COVID-19 had occurred. Despite that, our study adds more evidence for the analysis of a possible link between the quality of immunity and COVID-19 outcomes, particularly in the PHC setting.

(JMIR Public Health Surveill 2022;8(11):e36712) doi:10.2196/36712

KEYWORDS

SARS-CoV-2; COVID-19; influenza vaccines; pneumonia; electronic health records; primary health care; vaccination; public health; cohort study; epidemiology; eHeatlh; health outcome; mortality

Introduction

COVID-19 is caused by SARS-CoV-2, a novel coronavirus that emerged in China in 2019, which became the primary agent of a new pandemic that rapidly spread worldwide [1], with an average global infection fatality rate of approximately 0.15%, depending on the data analyzed [2]. SARS-CoV-2 mainly affects the respiratory tract and uses surface proteins in order to infect the host [3].

Although new variants of SARS-CoV-2 have emerged since December 2020, the coronavirus' genome is composed of RNA and depends on the RNA polymerase to generate its proteins, with a mechanism of error correction that results in a lower mutation rate than the influenza virus [4]. This low mutation rate may suggest that the vaccines developed against SARS-CoV-2, as well as the immunity generated in those patients who were infected, could represent a long-lasting immunity [5,6].

COVID-19, similar to influenza A and B, is caused by RNA virus and produce similar symptoms. The influenza virus needs the hemagglutinin and neuraminidase surface proteins to infect cells, whereas SARS-CoV-2 needs the S protein [5]. Previous in vitro and animal studies suggest an induction pathway of indirect etiological immunity between the influenza vaccine and SARS-CoV-2. Animal models suggest that some influenza subtypes might lead to regulation of the angiotensin-converting enzyme-2, with protective properties against SARS [7]. An unspecific effect of infection and vaccination on the immune system and susceptibility to other infections has also been reported, albeit with discordant data [8-10]. Some modeling studies have suggested a possible association between influenza immunization and COVID-19 [11-14].

A study conducted in Australia assessed the cellular and humoral immune responses during and after disease occurrence in a patient with a mild COVID-19 infection. They found that the immune response in different cell types is associated with clinical recovery. These results are coincident with similar findings among patients with influenza reported by the same authors [15,16].

Other studies observed differences in the susceptibility to COVID-19 in children of different ages with a lower infection rate than that in adults and older individuals [17]. Although the mechanism underlying these differences in severity and susceptibility is unclear, a possible explanation might be the difference in the quantity and quality of the immune function determined by the history of infections and the recent vaccines administered [18].

Consequently, a link between the quality of the immunity and recovery from COVID-19 may exist. Thus, we hypothesized that the immunity resulting from the previous influenza vaccination would boost part of the immunity against SARS-CoV-2, and we aimed to investigate whether individuals with COVID-19 could have benefited from vaccination against influenza.

Methods

Study Design

We performed a population-based cohort study including all adults with COVID-19 in Catalonia, Spain, who were registered as confirmed cases (through the polymerase chain reaction [PCR]) or as probable cases (not confirmed through PCR but with International Classification of Diseases (ICD)-10 codes registered that are compatible with COVID-19) in the primary health care (PHC) system. All individuals with COVID-19 were diagnosed from the pandemic's onset (March 2020) to June 30, 2020. Participants were compared on the basis of their influenza vaccination status between those having received the influenza vaccine before having COVID-19 (vaccinated in the previous influenza seasonal campaign in 2019-2020 or before) [19] with those who were not vaccinated.

Data Source

Our data source is the Information System for Research in Primary Care [20], which captures clinical information of approximately 5.8 million people from Catalonia, Spain (approximately 80% of the Catalan population). This information is pseudonymized, having originated from different data sources: (1) electronic health records in PHC system of the Catalan Health Institute, including sociodemographic characteristics, residents in nursing homes or long-term care facilities (LTCFs), comorbidities registered as ICD-10 codes [21], specialist referrals, clinical parameters, toxic habits, sickness leave, date of death, laboratory test data, and drug prescriptions issued in the PHC system, registered in accordance with the anatomical therapeutic chemical classification system [22]; (2) pharmacy invoice data corresponding to the PHC drug prescriptions; (3) database of diagnoses upon hospital discharge [23]; and (4) COVID-19 data from the Catalan Agency of Health Quality and Evaluation (AQuAS) [24].

COVID-19 Classification

Participants were classified in accordance with the following criteria: *confirmed cases* are those with a confirmed COVID-19 diagnosis record, positive PCR outcome, or a positive serology test result. Those with an unconfirmed diagnosis or test (possible or unclear) along with any individual with a record of hospitalization, pneumonia, or death related to COVID-19 were considered *probable cases*. During the first wave of the COVID-19 pandemic in Catalonia, PCR tests were not routinely conducted for all individuals with compatible symptoms owing to the unavailability of laboratory kits to carry out the tests. Thus, we needed to capture those patients with a possible

diagnosis of COVID-19, such as those admitted to hospital with pneumonia or other COVID-19 symptoms, who were not tested. We designed an algorithm to classify patients as "COVID possible" when a test result was unavailable along with registered entries from different databases: PCR tests or serology tests conducted in different settings, discharge diagnoses of pneumonia from Catalan hospitals or from emergency departments, and ICD-10 diagnoses related to COVID-19 coded in PHC. The date of COVID-19 diagnosis was set to be the first of all records used per patient. To guarantee that our algorithm is not far from the Catalan population, the resulting cohort was compared to the official COVID-19 cases reported by the AQuAS during the pandemic [24].

Influenza Immunization

Patients were classified as having taken the influenza vaccine if they had been vaccinated at any time before having COVID-19, and grouped in accordance with the seasonal vaccination campaign: the immediate previous campaign (2019-2020) or other vaccination campaigns (2018-2019 and before) [19,25].

Variables

At baseline, the following variables were captured: sex, age, geographical area, MEDEA (Mortalidad en áreas pequeñas Españolas y Desigualdades Económicas y Ambientales [Mortality in small Spanish areas and economic and environmental inequalities]) socioeconomic index (deprivation index based on 5 indicators of socioeconomic position; it helps analyze health inequity, and higher the MEDEA socioeconomic index, worse the deprivation) [26], BMI, residence in nursing homes, smoking habits, comorbidities, and taking influenza vaccines and pneumococcal and tuberculosis vaccines.

The main outcomes assessed during follow-up (up to June 2020) were at least one of the following variables: diagnosis of pneumonia, hospital admission, and mortality. The risk of these events was analyzed in those people who had been vaccinated against influenza at any time before having COVID-19, in those who were recently vaccinated (campaign of 2019-2020), and in those systematically vaccinated (who had been vaccinated at least during 3 different campaigns). We analyzed the same outcomes excluding those of people living in LTCFs, where vaccination is nearly universal in our country [27].

Statistical Analysis

Quantitative variables were described as mean (SD) values, whereas categorical variables were described as the proportion of vaccinated and unvaccinated individuals. Univariate analyses were based on the Student *t* test or chi-square test depending on the variable.

For each outcome, we fitted a logistic regression model to estimate an odds ratio (OR) comparing the prevalence of each outcome among individuals given the influenza vaccine to that of unvaccinated individuals. The logistic model was fitted along with other covariables such as smoking habits, age, comorbidities (asthma, autoimmune disorders, prior cerebrovascular disease, chronic kidney disease, chronic pulmonary obstructive disease, diabetes, heart failure, hypertension, ischemic heart disease, mental-behavioral disorders, obesity, organ transplant, and other respiratory diseases), concomitant drugs, and previous vaccines (pneumococcal and tuberculosis). As a sensitivity analysis, we conducted the same analysis on a matched population. Individuals vaccinated against influenza and unvaccinated controls were matched 1:2 in accordance with their age and gender at the time of infection or on an index date, and the reported ORs were obtained by fitting a conditional logistic regression model (clogit) accounting for matched pairs and adjusted using the same covariables as in the logistic model. We used the Wald test on the fitted coefficient to determine whether the log-odds were significantly different from 0 at a threshold of .05. All analyses were performed in R (version 4.1.0 or above; The R Foundation).

Ethical Considerations

The study protocol was approved by the Research Ethics Committee of Institut Universitari d'Investigació en Atenció Primària (June 3, 2020). This is a database research study that has been conducted in accordance with the tenets of the Declaration of Helsinki (Fortaleza, Brazil 2013) and does not require consent from the study participants for the purpose of publication. The need for consent was waived by the Research Ethics Committee of Institut Universitari d'Investigació en Atenció Primària as it is deemed unnecessary according to the European legislation (Regulation [EU] 2016/679).

Results

We included 309,039 individuals with COVID-19 during the first wave of the pandemic in accordance with their influenza immunization status (Table 1, Multimedia Appendix 1); 114,181 (36.9%) participants had received the influenza vaccine at least once before having COVID-19 and 194,858 (63.1%) had not been vaccinated, with more women in both groups, especially in the vaccinated cohort (61.0% women vs 39.0% men). The mean age was higher for vaccinated individuals (64.3 years, with 52.3% of them being older than 65 years). Vaccinated individuals had more comorbidities than unvaccinated individuals.

Of those who received the influenza vaccine, 66,611 (58.3%) had been recently vaccinated (2019-2020) and 75,311 (66%) had been systematically vaccinated against influenza at least during 3 different years (Table 2).

Of the participants with COVID-19, 11,000 (5.7%) unvaccinated and 21,721 (19%) vaccinated participants presented at least one of the following events: hospital admission, pneumonia, or death. For those who received the influenza vaccine at any time before having COVID-19, the risks of hospitalization (adjusted OR 1.14, 95% CI 1.10-1.19) and death (OR 1.32, 1.23-1.42) were higher than those among unvaccinated participants. For the recently vaccinated participants, the risk was higher for hospitalization (OR 1.16, 95% CI 1.1-1.23), pneumonia (OR 1.12, 95% CI 1.02-1.23), and death (OR 1.14, 95% CI 1.04-1.24). For people with recurrent vaccination, the risk of the 3 outcomes was also higher that among unvaccinated participants (OR 1.07, 1.16, and 1.24, respectively; Table 3).

We have also analyzed the results in a matched population of vaccinated versus unvaccinated participants, revealing a higher risk of pneumonia and mortality, with an adjusted OR of 1.11 (95% CI 1.01-1.23) and 1.28 (95% CI 1.07-1.53), respectively (Multimedia Appendix 2).

The risks of the outcomes based on influenza vaccination status and excluding those patients living in LTCFs are shown in Figure 1. For non-LTCF residents, the results are similar to those for the whole population, except that there was no significant increase in mortality (OR 0.93, 95% CI 0.85-1.03).

Table 1. Sociodemographic and clinical characteristics of the study population (N=309,039).

Characteristics	Overall	Not vaccinated against influenza (n=194,858)	Vaccinated against influenza at least once before having COVID-19 (n=114,181)	*P* value
COVID-19 status, n (%)				<.001
Confirmed	164,557 (53.2)	105,788 (54.3)	58,769 (51.5)	
Possible	144,482 (46.8)	89,070 (45.7)	55,412 (48.5)	
Gender, n (%)				<.001
Female	173,071 (56.0)	103,413 (53.1)	69,658 (61.0)	
Male	135,968 (44.0)	91,445 (46.9)	44,523 (39.0)	
Age (years), mean (SD)	49.3 (22.3)	40.6 (17.5)	64.3 (21.7)	<.001
Age groups (years), n (%)				<.001
≤40	108,950 (35.3)	90,894 (46.6)	18,056 (15.8)	
41-65	129,576 (41.9)	93,116 (47.8)	36,460 (31.9)	
>65	70,513 (22.8)	10,848 (5.6)	59,665 (52.3)	
Smoker status, n (%)	119,554 (38.7)	72,806 (37.4)	46,748 (40.9)	<.001
Having obesity, n (%)	78,882 (25.5)	36,973 (19.0)	41,909 (36.7)	<.001
Residents of long-term care facilities, n (%)	28,360 (9.2)	3146 (1.6)	25,214 (22.1)	<.001
Geographical information (MEDEA)				<.001
Unknown	278 (0.1)	201 (0.1)	77 (0.1)	
Urban	252,014 (81.5)	159,859 (82.0)	92,155 (80.7)	
Rural	56,747 (18.4)	34,798 (17.9)	21,949 (19.2)	
Comorbidities, n (%)				<.001
Asthma	22,734 (7.4)	9029 (4.6)	13,705 (12.0)	
Autoimmune disorders	30,783 (10.0)	14,005 (7.2)	16,778 (14.7)	
Cancer	23,600 (7.6)	6832 (3.5)	16,768 (14.7)	
Cerebrovascular disease	6937 (2.2)	1053 (0.5)	5884 (5.2)	
Chronic kidney disease	18,450 (6.0)	2088 (1.1)	16,362 (14.3)	
Chronic obstructive pulmonary disease	21,771 (7.0)	6155 (3.2)	15,616 (13.7)	
Diabetes	30,513 (9.9)	5886 (3.0)	24,627 (21.6)	
Heart failure	8307 (2.7)	693 (0.4)	7614 (6.7)	
Hypertension	75,346 (24.4)	21,624 (11.1)	53,722 (47.0)	
Ischemic heart disease	10,049 (3.3)	1837 (0.9)	8212 (7.2)	
Mental-behavioral disorders	9010 (2.9)	685 (0.4)	8325 (7.3)	
Organ transplant	893 (0.3)	213 (0.1)	680 (0.6)	
Other respiratory diseases	16,476 (5.3)	6407 (3.3)	10,069 (8.8)	
Other vaccines, n (%)				<.001
Pneumococcal	78,104 (25.3)	17,617 (9.0)	60,487 (53.0)	
Tuberculosis	2974 (1.0)	2412 (1.2)	562 (0.5)	

Table 2. Taking influenza vaccines before having COVID-19.

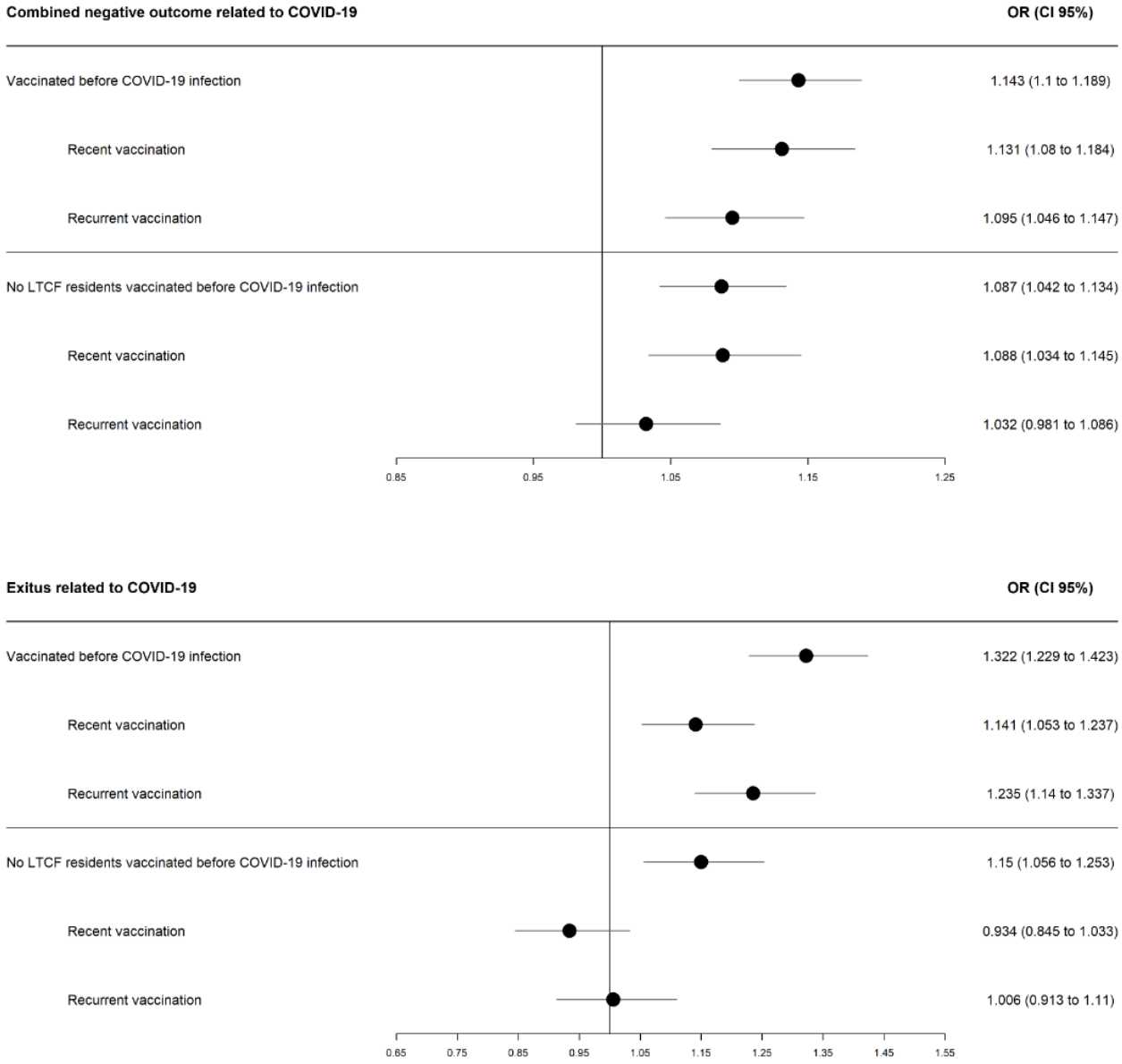
	Vaccinated before having COVID-19 (n=114,181)
Campaign of 2019-2020 (recent immunization), n (%)	66,611 (58.3)
Days from vaccination to infection, median (IQR)	146.0 (127.0-169.0)
Campaign of 2018-2019, n (%)	60,161 (52.7)
Days from vaccination to infection, median (IQR)	515.0 (495.0-539.0)
Campaign of 2017-2018 or before, n (%)	102,235 (89.5)
Days from vaccination to infection, median (IQR)	931.0 (875.0-2018.0)
Campaigns during which participants were vaccinated before having COVID-19, n (%)	
1	26,786 (23.5)
2	12,084 (10.6)
≥3 (recurrent immunization)	75,311 (66.0)
3	7931 (6.9)
4-5	11,146 (9.8)
6-10	18,945 (16.6)
>10	37,289 (32.7)

Table 3. Logistic regression model of COVID-19 outcomes based on influenza immunization status.

Any vaccination	Influenza immunization status prior to having COVID-19, n (%)		Multivariable logistic model[a]	
	Unvaccinated (n=194,858)	Vaccinated (n=114,181)	Adjusted odds ratio (95% CI)	*P* value
≥1 outcome	11,000 (5.7)	21,721 (19.0)	1.14 (1.10-1.19)	<.001
Hospitalization	7848 (4.0)	10,592 (9.3)	1.10 (1.05-1.15)	<.001
Pneumonia	3011 (1.6)	2740 (2.4)	1.08 (1.00-1.16)	.07
Death	1899 (0.97)	11,835 (10.4)	1.32 (1.23-1.42)	<.001
Recent vaccination (with 66,611 vaccinated participants)				
≥1 outcome	11,000 (5.7)	15,129 (22.7)	1.13 (1.10-1.18)	<.001
Hospitalization	7848 (4.0)	7009 (10.5)	1.16 (1.10-1.23)	<.001
Pneumonia	3011 (1.6)	1731 (2.6)	1.12 (1.02-1.23)	.02
Death	1899 (0.97)	8800 (13.2)	1.14 (1.05-1.24)	.001
Recurrent vaccination (with 75,311 vaccinated participants)				
≥1 outcome	11,000 (5.7)	17,798 (23.6)	1.10 (1.05-1.15)	<.001
Hospitalization	7848 (4.0)	8122 (10.8)	1.07 (1.02-1.14)	.01
Pneumonia	3011 (1.6)	1942 (2.6)	1.16 (1.06-1.27)	.002
Death	1899 (0.97)	10,561 (14.0)	1.24 (1.14-1.34)	<.001

[a]A logistic regression model adjusted with the following relevant covariables was fitted: smoking habits, age, comorbidities (asthma, autoimmune disorders, prior cerebrovascular disease, chronic kidney disease, chronic pulmonary obstructive disease, diabetes, heart failure, hypertension, ischemic heart disease, mental-behavioral disorders, obesity, organ transplant, and other respiratory diseases), co-medication, and previous vaccines (pneumococcal and tuberculosis).

Figure 1. Risk of death and of combined COVID-19 complications in all the vaccinated population and excluding people living in long-term care facilities (LTCF).



Discussion

Principal Findings

We analyzed the negative outcomes among people with COVID-19 (N=309,039) and compared those who had received the influenza vaccine with those who were never vaccinated. Those who received the vaccine any time before having COVID-19 were at a higher risk of complications than those who were unvaccinated. We obtained similar results for those who were recently vaccinated (2019-2020 campaign) and for those who were systematically vaccinated (at least 3 years), and the same comparisons were carried out after excluding individuals living in LTCFs. We also obtained similar results on matching vaccinated and unvaccinated individuals. Thus, we did not find a possible link between receiving the influenza vaccine and presenting better clinical outcomes after a COVID-19 infection.

Comparison With Prior Work

Some researchers have studied this possible association. Massoudi and Mohit [28] conducted a study in a hospital in Iran including health care workers, with 80 of them COVID-19 cases confirmed through PCR or on the basis of their symptoms, and 181 of them were controls. They concluded that individuals who were confirmed cases were less likely to have received the 2019 influenza vaccine (OR 0.04, 95% CI 0.01-0.14), suggesting a protective association between the influenza vaccine and COVID-19. Their study had several limitations, such as the lack of availability of COVID-19 test kits or the samples limited to the workers of a single hospital [28].

Candelli et al [29] assessed 602 patients with COVID-19 enrolled at the emergency department in a hospital in Italy, of whom 24.9% had been previously vaccinated against influenza. They found that influenza immunization was independently associated with a lower risk of death at 60 days (OR 0.20, 95%

CI 0.08-0.51), but not with a reduced need of endotracheal intubation (OR 0.73, 95% CI 0.35-1.56) [29].

A study conducted in Brazil [30] included 92,664 confirmed cases of COVID-19, of whom 31.1% had been recently vaccinated against influenza. They found that the vaccinated individuals were at a lower risk of needing intensive care for COVID-19 (OR 0.92, 95% CI 0.86-0.99), a lower risk of needing respiratory support (OR 0.81, 95% CI 0.74-0.88), and lower odds of mortality (OR 0.82, 95% CI 0.75-0.89) [30].

In a systematic review [31] including 12 studies, the authors examined whether influenza vaccination affects the risk of being infected with SARS-CoV-2 and the risk of complicated illness or poor outcomes among patients with COVID-19, all of whom having been confirmed cases through PCR testing. They concluded that influenza vaccination is unlikely to be associated with an increase in the risk of COVID-19 infection or severity and the risk of associated death [31].

There are reports from some countries with high influenza vaccination rates and high incidences of COVID-19 and mortality [32,33]. For instance, Kline et al [33] compared people vaccinated against influenza with unvaccinated individuals admitted to hospital for COVID-19, and they found no differences in the rate of admission to the intensive care unit, intubation, or other complications [33]. Our results follow these same trends in a cohort of the general population attended to in the PHC system and not only hospitalized patients.

Limitations

We need to consider that our results correspond to the first wave of the COVID-19 pandemic, when there were more negative outcomes and mortalities due to COVID-19 than in the subsequent waves in our setting; thus, this higher statistical power allowed us to detect differences. Furthermore, in subsequent waves, more confounders might have been present, such as COVID-19 vaccination or effects of the different SARS-CoV-2 variants, making it more difficult to manage their potential effect in the analysis of the outcomes of the infection.

We also need to bear in mind that the target population for the influenza vaccine in our country are people older than 60 years, individuals with chronic comorbidities or immunodeficiency, and health care workers among others [34], some of them being at a high risk of COVID-19 complications, which is why confounding variables were used to adjust the logistic regression model [35]. Nevertheless, estimates of the effectiveness of the influenza vaccine have been frequently confounded, indicating that a different approach should be used with alternative study designs, different from the typical methods used to study drug exposure [36-38].

Among other limitations of our study is the reliability of the COVID-19 diagnoses; we included individuals without a confirmed result, as during the first wave of the pandemic in our setting, PCR tests were not always performed. This limitation has been described in other studies including those conducted at the beginning of the pandemic when diagnostic tests for COVID-19 were not widely available and clinical algorithms were used to assess COVID-19 diagnoses [39]. We compared our number of COVID-19 cases with the official COVID-19 case numbers provided by the AQuAS during the pandemic [24]. Another limitation is the lack of hospital information: we could not capture ICU admissions, ventilation, or treatments administered upon admission, which clearly have an influence on the prognosis and outcomes of COVID-19. Finally, we have not conducted any subgroup analysis that could have indicated any condition potentially resulting in any benefit or harm from influenza vaccination.

Conclusions

In conclusion, we were not able to establish a protective role of the immunity conferred by the influenza vaccine on the outcomes of COVID-19 infection. Nonetheless, our study adds more evidence to the analysis of the possible link between the quality of the conferred immunity and outcomes of COVID-19 infection, and it has some strengths, such as the large cohort size, its representativeness with respect to the general population, and the completeness of its sociodemographic data. We have already highlighted that our cohort comprises individuals who received care from the PHC system; hence, we have estimated the risk of complications for a different population from the hospitalized individuals who are usually assessed in multiple studies.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline characteristics of the population included by gender.
[DOCX File , 21 KB - publichealth_v8i11e36712_app1.docx]

Multimedia Appendix 2
Conditional logistic regression model for the age and gender matched population. Matching performed in patients ≤ 65 years old to correct for age distribution.
[DOCX File , 14 KB - publichealth_v8i11e36712_app2.docx]

References

1. Coronavirus disease 2019 (COVID-19) : situation report, 51. World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/331475 [accessed 2022-10-31]
2. Ioannidis JPA. Reconciling estimates of global spread and infection fatality rates of COVID-19: an overview of systematic evaluations. Eur J Clin Invest 2021 May;51(5):e13554 [FREE Full text] [doi: 10.1111/eci.13554] [Medline: 33768536]
3. Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. Nat Rev Mol Cell Biol 2022 Jan 05;23(1):3-20 [FREE Full text] [doi: 10.1038/s41580-021-00418-x] [Medline: 34611326]
4. Boehm E, Kronig I, Neher R, Eckerle I, Vetter P, Kaiser L, Geneva Centre for Emerging Viral Diseases. Novel SARS-CoV-2 variants: the pandemics within the pandemic. Clin Microbiol Infect 2021 Aug;27(8):1109-1117 [FREE Full text] [doi: 10.1016/j.cmi.2021.05.022] [Medline: 34015535]
5. Manzanares-Meza LD, Medina-Contreras O. SARS-CoV-2 and influenza: a comparative overview and treatment implications. Bol Med Hosp Infant Mex 2020;77(5):262-273 [FREE Full text] [doi: 10.24875/BMHIM.20000183] [Medline: 33064680]
6. Bar-On YM, Flamholz A, Phillips R, Milo R. SARS-CoV-2 (COVID-19) by the numbers. Elife 2020 Apr 02;9 [FREE Full text] [doi: 10.7554/eLife.57309] [Medline: 32228860]
7. Yang P, Gu H, Zhao Z, Wang W, Cao B, Lai C, et al. Angiotensin-converting enzyme 2 (ACE2) mediates influenza H7N9 virus-induced acute lung injury. Sci Rep 2014 Dec 13;4:7027 [FREE Full text] [doi: 10.1038/srep07027] [Medline: 25391767]
8. Cowling B, Fang V, Nishiura H, Chan K, Ng S, Ip D, et al. Increased risk of noninfluenza respiratory virus infections associated with receipt of inactivated influenza vaccine. Clin Infect Dis 2012 Jul;54(12):1778-1783 [FREE Full text] [doi: 10.1093/cid/cis307] [Medline: 22423139]
9. Feng S, Fowlkes AL, Steffens A, Finelli L, Cowling BJ. Assessment of virus interference in a test-negative study of influenza vaccine effectiveness. Epidemiology 2017 Jul;28(4):514-524 [FREE Full text] [doi: 10.1097/EDE.0000000000000670] [Medline: 28362642]
10. Jehi L, Ji X, Milinovich A, Erzurum S, Rubin BP, Gordon S, et al. Individualizing risk prediction for positive coronavirus disease 2019 testing: results from 11,672 patients. Chest 2020 Oct;158(4):1364-1375 [FREE Full text] [doi: 10.1016/j.chest.2020.05.580] [Medline: 32533957]
11. Gurwitz D. Angiotensin receptor blockers as tentative SARS-CoV-2 therapeutics. Drug Dev Res 2020 Aug 04;81(5):537-540 [FREE Full text] [doi: 10.1002/ddr.21656] [Medline: 32129518]
12. Koo J, Cook A, Park M, Sun Y, Sun H, Lim J, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. Lancet Infect Dis 2020 Jun;20(6):678-688. [doi: 10.1016/s1473-3099(20)30162-6]
13. Li Q, Tang B, Bragazzi NL, Xiao Y, Wu J. Modeling the impact of mass influenza vaccination and public health interventions on COVID-19 epidemics with limited detection capability. Math Biosci 2020 Jul;325:108378 [FREE Full text] [doi: 10.1016/j.mbs.2020.108378] [Medline: 32507746]
14. Salem ML, El-Hennawy D. The possible beneficial adjuvant effect of influenza vaccine to minimize the severity of COVID-19. Med Hypotheses 2020 May 22;140:109752 [FREE Full text] [doi: 10.1016/j.mehy.2020.109752] [Medline: 32361099]
15. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med 2020 Apr;26(4):450-452 [FREE Full text] [doi: 10.1038/s41591-020-0820-9] [Medline: 32284615]
16. Koutsakos M, Wheatley AK, Loh L, Clemens EB, Sant S, Nüssing S, et al. Circulating T cells, serological memory, and tissue compartmentalization shape human influenza-specific B cell immunity. Sci Transl Med 2018 Feb 14;10(428). [doi: 10.1126/scitranslmed.aan8405] [Medline: 29444980]
17. Mehta N, Mytton O, Mullins E, Fowler T, Falconer C, Murphy O, et al. SARS-CoV-2 (COVID-19): what do we know about children? A systematic review. SSRN Preprint posted online April 1, 2020. [doi: 10.2139/ssrn.3558015]
18. Zheng J, Perlman S. Immune responses in influenza A virus and human coronavirus infections: an ongoing battle between the virus and host. Curr Opin Virol 2018 Feb;28:43-52 [FREE Full text] [doi: 10.1016/j.coviro.2017.11.002] [Medline: 29172107]
19. Departament DSGDC. Pla d'informació de les infeccions respiratòries agudes a Catalunya. Balanç temporada gripal18-20. URL: https://canalsalut.gencat.cat/ca/professionals/vigilancia-epidemiologica/pla-dinformacio-de-les-infeccions-respiratories-agudes-a-catalunya-pidirac/ [accessed 2022-11-03]
20. Information system for the development of research in primary care. SIDIAP. 2022. URL: http://www.sidiap.org/index.php/en [accessed 2022-10-31]
21. International Statistical Classification of Diseases and Related Health Problems 10th Revision. ICD-10 Version:2019. URL: https://icd.who.int/browse10/2019/en [accessed 2022-10-31]
22. ATC/DDD Index 2022. WHO Collaborating Centre for Drug Statistics Methodology. URL: https://www.whocc.no/atc_ddd_index/ [accessed 2022-10-31]
23. Conjunt mínim bàsic de dades (CMBD). CatSalut. Servei Català de la Salut. URL: http://catsalut.gencat.cat/ca/proveidors-professionals/registres-catalegs/registres/cmbd/ [accessed 2022-10-31]
24. COVID-19 - Diagnòstics AP. Generalitat de Catalunya. URL: https://aquas.gencat.cat/ca/actualitat/ultimes-dades-coronavirus/index.html#googtrans(ca%7Cen) [accessed 2022-10-31]

25. Pla d'informació de les infeccions respiratòries agudes a Catalunya (PIDIRAC) 2019-2020. Generalitat de Catalunya. Agència de Salut Pública de Catalunya. 2019. URL: https://scientiasalut.gencat.cat/bitstream/handle/11351/4318/pla_informacio_infeccions_respiratories_agudes_catalunya_2019_2020.pdf?sequence=1&isAllowed=y [accessed 2022-10-31]
26. Domínguez-Berjón MF, Borrell C, Cano-Serral G, Esnaola S, Nolasco A, Pasarín MI, et al. [Constructing a deprivation index based on census data in large Spanish cities (the MEDEA project)]. Gac Sanit 2008;22(3):179-187 [FREE Full text] [doi: 10.1157/13123961] [Medline: 18579042]
27. Vacunas y Programa de Vacunación. Ministerio de Sanidad. URL: https://www.mscbs.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/programasDeVacunacion/gripe/home.htm [accessed 2022-10-31]
28. Massoudi N, Mohit B. A case-control study of the 2019 influenza vaccine and incidence of COVID-19 among healthcare workers. J Clin Immunol 2021 Feb 26;41(2):324-334 [FREE Full text] [doi: 10.1007/s10875-020-00925-0] [Medline: 33244671]
29. Candelli M, Pignataro G, Torelli E, Gullì A, Nista EC, Petrucci M, et al. Effect of influenza vaccine on COVID-19 mortality: a retrospective study. Intern Emerg Med 2021 Oct;16(7):1849-1855 [FREE Full text] [doi: 10.1007/s11739-021-02702-2] [Medline: 33743150]
30. Fink G, Orlova-Fink N, Schindler T, Grisi S, Ferrer APS, Daubenberger C, et al. Inactivated trivalent influenza vaccination is associated with lower mortality among patients with COVID-19 in Brazil. BMJ Evid Based Med 2020 Dec 11 [FREE Full text] [doi: 10.1136/bmjebm-2020-111549] [Medline: 33310766]
31. Del Riccio M, Lorini C, Bonaccorsi G, Paget J, Caini S. The association between influenza vaccination and the risk of SARS-CoV-2 infection, severe illness, and death: a systematic review of the literature. Int J Environ Res Public Health 2020 Oct 27;17(21):7870 [FREE Full text] [doi: 10.3390/ijerph17217870] [Medline: 33121070]
32. Influenza vaccination rates. Organisation for Economic Co-operation and Development. URL: https://data.oecd.org/healthcare/influenza-vaccination-rates.htm [accessed 2022-10-31]
33. Kline A, Trinh LN, Hussein MH, Elshazli RM, Toraih EA, Duchesne J, et al. Annual flu shot: does it help patients with COVID-19? Int J Clin Pract 2021 Dec 27;75(12):e14901 [FREE Full text] [doi: 10.1111/ijcp.14901] [Medline: 34547161]
34. Agència DSPDC(. Campanya de vacunació antigripal 2021-2022. Vacunació antigripal. URL: https://canalsalut.gencat.cat/ca/professionals/vigilancia-epidemiologica/pla-dinformacio-de-les-infeccions-respiratories-agudes-a-catalunya-pidirac/ [accessed 2022-11-03]
35. Rothman K, Greenland S, Lash T. Modern Epidemiology (3rd edition). Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
36. Fireman B, Lee J, Lewis N, Bembom O, van der Laan M, Baxter R. Influenza vaccination and mortality: differentiating vaccine effects from bias. Am J Epidemiol 2009 Sep 01;170(5):650-656 [FREE Full text] [doi: 10.1093/aje/kwp173] [Medline: 19625341]
37. McGrath LJ, Kshirsagar AV, Cole SR, Wang L, Weber DJ, Stürmer T, et al. Influenza vaccine effectiveness in patients on hemodialysis: an analysis of a natural experiment. Arch Intern Med 2012 Apr 09;172(7):548-554 [FREE Full text] [doi: 10.1001/archinternmed.2011.2238] [Medline: 22493462]
38. McGrath LJ, Ellis AR, Brookhart MA. Controlling time-dependent confounding by health status and frailty: restriction versus statistical adjustment. Am J Epidemiol 2015 Jul 01;182(1):17-25 [FREE Full text] [doi: 10.1093/aje/kwu485] [Medline: 25868551]
39. Almalki YE, Qayyum A, Irfan M, Haider N, Glowacz A, Alshehri FM, et al. A novel method for COVID-19 diagnosis using artificial intelligence in chest x-ray images. Healthcare (Basel) 2021 Apr 29;9(5) [FREE Full text] [doi: 10.3390/healthcare9050522] [Medline: 33946809]

Abbreviations

AQuAS: Catalan Agency of Health Quality and Evaluation
ICD-10: International Classification of Diseases, version 10
LTCF: long-term care facility
MEDEA: Mortalidad en áreas pequeñas Españolas y Desigualdades Económicas y Ambientales (Mortality in small Spanish areas and economic and environmental inequalities)
OR: odds ratio
PCR: polymerase chain reaction
PHC: primary health care

Edited by A Mavragani, G Eysenbach; submitted 25.01.22; peer-reviewed by N Shcherbakova, A Cook; comments to author 03.05.22; revised version received 11.05.22; accepted 18.10.22; published 11.11.22.

Please cite as:
Giner-Soriano M, de Dios V, Ouchi D, Vilaplana-Carnerero C, Monteagudo M, Morros R
Outcomes of COVID-19 Infection in People Previously Vaccinated Against Influenza: Population-Based Cohort Study Using Primary Health Care Electronic Records
JMIR Public Health Surveill 2022;8(11):e36712
URL: https://publichealth.jmir.org/2022/11/e36712
doi:10.2196/36712
PMID:36265160

Original Paper

Dual Sensory Impairment as a Predictor of Loneliness and Isolation in Older Adults: National Cohort Study

Qiong Wang[1,2*], PhD; Shimin Zhang[1,2*], MSc; Yi Wang[1,2], PhD; Dan Zhao[1,2], PhD; Chengchao Zhou[1,2], PhD

[1]Centre for Health Management and Policy Research, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, China

[2]National Health Commission of China Key Lab of Health Economics and Policy Research, Shandong University, Jinan, China

[*]these authors contributed equally

Corresponding Author:
Chengchao Zhou, PhD
Centre for Health Management and Policy Research
School of Public Health, Cheeloo College of Medicine
Shandong University
Number 44, Wen-hua-xi Road
Jinan, 250012
China
Phone: 86 0531 8838 1567
Email: zhouchengchao@sdu.edu.cn

Abstract

Background: Loneliness and social isolation are global public health challenges. Sensory impairments (SIs) are highly prevalent among older adults but are often ignored as a part of normal aging. Identifying the role of SIs in loneliness and social isolation could provide insight into strategies for improving public health among older adults.

Objective: This study aims to analyze the effects of SIs on loneliness and social isolation among older adults in rural and urban China.

Methods: This cohort study of 3069 older adults (aged 60+) used data from 4 waves (2011, 2013, 2015, and 2018) of the China Health and Retirement Longitudinal Study (CHARLS), a nationally representative survey of adults aged 45 years or older. SIs include hearing impairment (HI), vision impairment (VI), and dual sensory impairment (DSI). DSI is defined as the co-occurrence of VI and HI. Participants with complete data on hearing, vision, social isolation, and loneliness were included in the analysis. Generalized estimating equation models adjusted for covariates were used to examine the relationships of DSI with loneliness and social isolation among older adults.

Results: Older adults in rural areas have higher prevalence of DSI, loneliness, and social isolation than their urban counterparts. In rural areas, participants with VI only (odds ratio [OR] 1.34, 95% CI 1.12-1.62; *P*=.002), HI only (OR 1.32, 95% CI 1.02-1.71; *P*=.03), and DSI (OR 1.84, 95% CI 1.56-2.18; *P*<.001) were more likely to experience loneliness compared with participants without SIs. DSI showed a statistically significant association with loneliness compared with VI only (OR 1.37, 95% CI 1.22-1.54; *P*<.001) and HI only (OR 1.39, 95% CI 1.13-1.72; *P*=.002). In urban areas, participants with VI only (OR 2.44, 95% CI 1.57-3.80; *P*<.001), HI only (OR 2.47, 95% CI 1.41-4.32; *P*=.002), and DSI (OR 1.88, 95% CI 1.24-2.85; *P*=.003) were more likely to experience loneliness compared with participants without SIs. DSI was not associated with the increased likelihood of loneliness compared with HI only or VI only. SIs were not associated with social isolation among older adults in urban and rural areas. Until 2018, 86.97% (2669/3069) reported VI, but only 27.11% (832/3069) and 9.45% (290/3069) were treated with glasses and cataract surgery, respectively; besides, 75 individuals received both glasses and cataract surgery treatment. The prevalence of HI was 74.39% (2283/3069) in 2018, but only 0.72% (22/3069) were treated with a hearing aid.

Conclusions: SIs are associated with an increased risk of loneliness rather than social isolation. A compounded risk of DSI on loneliness exists in rural areas rather than in urban areas. These findings expand our knowledge about the effects of SIs on loneliness and social isolation in non-Western populations. Interventions targeting HI only and DSI might be particularly effective for mitigating loneliness of older adults in urban and rural areas, respectively. Considering the high prevalence and low treatment rate of SIs, measures should be taken to make treatment more accessible.

(JMIR Public Health Surveill 2022;8(11):e39314) doi:10.2196/39314

KEYWORDS

loneliness; social isolation; dual sensory impairment; vision impairment; hearing impairment; mental health

Introduction

Sensory impairments (SIs), comprising hearing impairment (HI), vision impairment (VI), and dual sensory impairment (DSI), are highly prevalent among older adults and increases with age [1]. The World Health Organization (WHO) reported that at least 2.2 billion people have VI or blindness [2]. It is estimated that approximately 1 in 10 people globally will have HI in 2050 [3]. With the increasing number of people with HI or VI, the prevalence of the co-occurrence of VI and HI, termed DSI, is expected to increase rapidly [4]. Previous studies have found that HI and VI have negative effects on mortality [5] and DSI worsens the effect of single SI [6]. Moreover, the negative effects of SIs on health outcomes among older adults are often overshadowed by the negative effects of chronic diseases and functional impairment [7,8]. The effects of SIs for older adults cannot be ignored and merit in-depth exploration.

SIs impose communication difficulties [9,10], difficulties with activities of daily living (ADL) [11], and decreased social participation [4,12], which may lead to impoverished social relationships, such as loneliness and social isolation. Loneliness is a subjective measure of an individual's perceived discrepancy between desired and actual social interactions [13]. By contrast, social isolation refers to the objective state of estrangement, in which social connections are limited or absent [14]. Both loneliness and social isolation have become grand challenges of particular concerns for older adults given their independent association with a wide range of adverse health outcomes such as cognitive decline [15], depression [16,17], and mortality [14]. Reducing loneliness and social isolation in older adults is an important public health goal, which might be achieved by tackling modifiable risk factors or increasing social participation [18]. As a major obstacle of communication but a modifiable factor of aging, SIs merit more attention and addressing them may protect older adults against loneliness and social isolation.

Previous studies have examined the relationship between DSI and loneliness, but the results were inconsistent. For example, studies conducted in Western countries found an association between DSI and loneliness [19,20], whereas a study conducted in Malaysia did not support this association [21]. A paucity of research focused on the relationship between DSI and social isolation. Hajek and König [20] found a cross-sectional association between DSI and social isolation among Germans aged 40 and older. However, there is a complete lack of studies investigating the longitudinal relationship between DSI and social isolation. Given the geographic, racial, and cultural differences between Western and Asian populations, the relationships of DSI with loneliness and social isolation merit further research based on local conditions. China is changing rapidly in population aging and internal migration [22]. Awareness of DSI and its impact on loneliness and social isolation may be important to help Chinese older adults maintain a good quality of life and promote healthy aging. Moreover, due to great disparity existing in socioeconomic status and health care resources between rural and urban areas [23], older adults in rural areas might be at a higher risk of loneliness, social isolation, and DSI than those in urban areas. Thus, it is necessary to stratify the analyses by region of residence in this study.

To our knowledge, the longitudinal relationships of DSI with loneliness and social isolation in Chinese older adults have not been studied. Moreover, the relative relationship between older adults with DSI and those with single SI was less clear. In addition, whether the effect of DSI on loneliness and social isolation among older adults is similar between rural and urban areas remains unclear. Therefore, this study aims to assess the longitudinal relationships of DSI with loneliness and social isolation and examine whether these associations differ in rural and urban China.

Methods

Participants

The data used in this study were from the 2011, 2013, 2015, and 2018 waves of the China Health and Retirement Longitudinal Study (CHARLS). It is a nationally representative longitudinal study that surveys Chinese residents aged 45 years or older since 2011 (wave 1). It covers not only personal information and environmental information, but also factual information and attitude information, such as sociodemographic characteristics, socioeconomic status, health status, and psychological conditions. In the sampling method, a stratified (by per capita GDP of urban districts and rural counties) multistage (county/district-village/community household) random sampling strategy was adopted, and finally a total of 150 counties in 28 provinces of China were sampled [24]. The CHARLS baseline survey in 2011 included 17,708 respondents aged 45 years or older. Up to wave 4, 5587 respondents were lost to follow-up. For this study, we excluded older adults aged less than 60 years (n=7217) at baseline. Then, older adults with missing data (n=987) on main variables (SIs, loneliness, and social isolation) and who moved between urban and rural areas during follow-up (n=848) were excluded. Finally, a total of 3069 older adults who participated in all follow-up waves were included in our study, with 530 urban residents and 2539 rural residents (Multimedia Appendix 1).

Measures

Loneliness

Loneliness was measured by a single item of the Center for Epidemiological Studies Depression Scale (CESD), "In the last week, how often did you feel lonely?" [25]. This single measurement is highly correlated with multi-item loneliness scales, such as the University of California Los Angeles Loneliness Scale, and has been used in many previous studies [26,27]. Based on previous research experience [25,26], respondents were considered lonely if they feel lonely on some days (1-2 days), occasionally (3-4 days), or most of the time (5-7 days). Respondents were considered not lonely if they feel lonely rarely or none of the time (<1 day).

Social Isolation

We created an index of social isolation by giving 1 point for each of being unmarried, living alone, having less than weekly contact (by phone, in person, or by email) with children, and not participating in any social activities over the last month (eg, interacted with friends; played chess or cards; went to sport, social, or other clubs). Scores ranged from 0 to 4, with higher scores indicating a higher level of social isolation. Because of the positively skewed distribution of social isolation scores, we categorized participants according to the top quintile (>1 for social isolation) [28].

Sensory Impairments

The self-reported data on VI were composed of 2 questions: (1) "How good is your vision for seeing things at a distance (with glasses or corrective lenses), like recognizing a friend from across the street?" and (2) "How good is your vision for seeing things up close (with glasses or corrective lenses), like reading ordinary newspaper print?" For each question, the responses included "excellent," "very good," "good," "fair," or "poor." We identified respondents as having VI if they reported fair or poor vision (for either long distance or near vision). One question was used to assess HI, "Is your hearing excellent, very good, good, fair, or poor (with a hearing aid if you normally use it and without if you normally don't)." Participants were identified as having HI if they reported fair or poor hearing. When HI and VI were both present, participants were regarded as having DSI. SIs assessment and categorization was in accordance with previous studies [29,30].

Covariates

Sociodemographic characteristics, lifestyle factors, and health-related variables were considered as potential confounding variables. Demographic characteristics included gender (male/female), age (mean and SD), educational level (lower than primary school/primary school/middle school, or above), and household income. The household total annual income is the sum of all income at the household level including that from earning income, capital income, pension income, income from government transfers, other income, and the total income from other household members (eg, from parents, children, relatives). Lifestyle factors included smoking status (yes/no) and alcohol drinking (yes/no). Health-related variables were collected by asking their chronic health conditions and functional impairment. We categorized the self-reported chronic diseases into 3 groups (no chronic disease/1 chronic disease/multimorbidity). Functional impairment was assessed using the ADL scale which consists of 6 items including dressing, bathing or showering, eating, getting in or out of bed, toileting, and controlling urination and defecation. The ADL score ranges from 6 to 24, with a higher score indicating the worse ability of daily living activities [31].

Statistical Analysis

Descriptive statistics were used to describe the baseline characteristics of the sample. Continuous variables were summarized using means and SDs. Categorical variables were reported using numbers and percentages. The generalized estimating equation model assuming an independent working correlation structure was used to examine the associations of SIs with loneliness and social isolation among older adults in rural and urban areas during the follow-up period. We calculated the estimate odds ratios (ORs) and 95% CIs while adjusting for all identified confounders. Furthermore, in sensitivity analysis, we treated social isolation as a continuous variable to assess the robustness of the relationship between SIs and social isolation. A 2-sided $P<.05$ was considered statistically significant. Stata version 14.2 (StataCorp) was used for the data analyses.

Ethical Considerations

This study protocol was approved and organized by Peking University's Institutional Review Board (IRB00001052-11015). All procedures were in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. This survey was anonymous, and the answers were protected by privacy law. Written informed consents clarifying the study purposes were obtained from each participant before the survey.

Results

Tables 1 and 2 show the baseline characteristics of the urban and rural respondents, respectively. Mean (SD) age of the 3069 respondents was 66.02 (5.15) years, and 51.28% (1574/3069) were female. Of all participants, being more socially isolated or lonelier was associated with being older, being female, less educated, lower income, being a smoker, a higher level of functional impairment, more kinds of chronic diseases, and having 1 of the SIs. Until 2018, 83.8% (444/530) reported VI in urban areas, 45.5% (241/530) were treated with glasses, and 10.6% (56/530) were treated with cataract surgery. The prevalence of VI is 87.63% (2225/2539) in rural areas, but only 23.28% (591/2539) and 9.22% (234/2539) were treated with glasses and cataract surgery, respectively. Until 2018, 70% (371/530) and 75.31% (1912/2539) reported HI in urban and rural areas, but only 1.13% (6/530) and 0.63% (16/2539) were treated with a hearing aid, respectively (Multimedia Appendix 2).

Table 3 displays the prevalence of loneliness, social isolation, and SIs over time in urban and rural areas, respectively. The prevalence rates of loneliness, social isolation, and DSI are on the rise overall. At baseline (2011), 5.60% (172/3069) reported HI only, 25.48% (782/3069) reported VI only, and 54.90% (1685/3069) reported DSI. Participants in rural areas were more likely to report DSI (1442/2539, 56.79% vs 243/530, 45.85%; $P<.001$) and loneliness (847/2539, 33.36% vs 116/530, 21.89%; $P<.001$) than those in urban areas. There was no statistical difference in the prevalence of social isolation among urban and rural older adults at baseline. In 2018, the prevalence of social isolation in rural areas was higher than that in urban areas (739/2539, 29.11% vs 125/530, 23.58%; $P=.01$).

Table 1. Baseline characteristics of older adults in urban China (n=530).

	No loneliness	Loneliness	*P* value[a]	Low social isolation	High social isolation	*P* value[a]
Gender, n (%[b])			*.03*			*<.001*
Male	215 (82.06)	47 (17.94)		237 (90.46)	25 (9.54)	
Female	199 (74.25)	69 (25.75)		203 (75.75)	65 (24.25)	
Age, mean (SD)	66.37 (5.54)	66.67 (5.37)	.92	66.09 (5.11)	69.80 (6.25)	*<.001*
Education, n (%)			*.005*			*<.001*
Lower than primary school	40 (68.97)	18 (31.03)		38 (65.52)	20 (34.48)	
Primary school	141 (73.06)	52 (26.94)		154 (79.79)	39 (20.21)	
Middle school or above	233 (83.51)	46 (16.49)		248 (88.89)	31 (11.11)	
Marital status, n (%)			*<.001*			__[c]
Couple	357 (82.07)	78 (17.93)		413 (94.94)	22 (5.06)	
Single	57 (60.00)	38 (40.00)		27 (28.42)	68 (71.58)	
Household annual income, n (%)[d]			*<.001*			*<.001*
Q1	85 (63.91)	48 (36.09)		80 (60.15)	53 (39.85)	
Q2	176 (78.92)	47 (21.08)		193 (86.55)	30 (13.45)	
Q3	39 (92.86)	3 (7.14)		40 (95.24)	2 (4.76)	
Q4	114 (86.36)	18 (13.64)		127 (96.21)	5 (3.79)	
Smoking status, n (%)			.52			*.047*
No	329 (78.71)	89 (21.29)		340 (81.34)	78 (18.66)	
Yes	85 (75.89)	27 (24.11)		100 (89.29)	12 (10.71)	
Alcohol consumption, n (%)			.17			*.02*
No	287 (76.53)	88 (23.47)		302 (80.53)	73 (19.47)	
Yes	127 (81.94)	28 (18.06)		138 (89.03)	17 (10.97)	
Activities of daily living, mean (SD)	6.24 (1.16)	6.33 (0.84)	.45	6.23 (0.95)	6.39 (1.65)	.22
Chronic disease, n (%)			*.045*			.78
None	103 (83.06)	21 (16.94)		101 (81.45)	23 (18.55)	
One	108 (82.44)	23 (17.56)		111 (84.73)	20 (15.27)	
≥2	203 (73.82)	72 (26.18)		228 (82.91)	47 (17.09)	
Sensory impairments, n (%)			*.01*			.77
No sensory impairments	80 (88.89)	10 (11.11)		73 (81.11)	17 (18.89)	
Hearing impairment only	31 (83.78)	6 (16.22)		31 (83.78)	6 (16.22)	
Vision impairment only	114 (71.25)	46 (28.75)		130 (81.25)	30 (18.75)	
Dual sensory impairment	189 (77.78)	54 (22.22)		206 (84.77)	37 (15.23)	

[a]Italicized values denote statistical significance (*P*<.05) between the groups.

[b]Percentages were estimated over cases with valid data in every group.

[c]Chi-square test was not performed.

[d]Q1 was the poorest and Q4 was the richest. Q1: ≦US $415; Q2: US $416-1981; Q3: US $1982-3391; Q4: >US $3391.

Table 2. Baseline characteristics of older adults in rural China (n=2539).

	No loneliness	Loneliness	*P* value[a]	Low social isolation	High social isolation	*P* value[a]
Gender, n (%[b])			*<.001*			*.003*
Male	879 (71.29)	354 (28.71)		1038 (84.18)	195 (15.82)	
Female	813 (62.25)	493 (37.75)		1041 (79.71)	265 (20.29)	
Age, mean (SD)	65.69 (5.00)	66.26 (5.18)	*.008*	65.50 (4.83)	67.60 (5.72)	*<.001*
Education, n (%)			*<.001*			*<.001*
Lower than primary school	695 (62.39)	419 (37.61)		873 (78.37)	241 (21.63)	
Primary school	809 (69.15)	361 (30.85)		990 (84.62)	180 (15.38)	
Middle school or above	188 (73.73)	67 (26.27)		216 (84.71)	39 (15.29)	
Marital status, n (%)			*<.001*			__[c]
Couple	1504 (70.98)	615 (29.02)		1968 (92.87)	151 (7.13)	
Single	188 (44.76)	232 (55.24)		111 (26.43)	309 (73.57)	
Household annual income, n (%)[d]			*.001*			*<.001*
Q1	412 (64.88)	223 (35.12)		446 (70.24)	189 (29.76)	
Q2	390 (61.42)	245 (38.58)		504 (79.37)	131 (20.63)	
Q3	508 (69.59)	222 (30.41)		635 (86.99)	95 (13.01)	
Q4	382 (70.87)	157 (29.13)		494 (91.65)	45 (8.35)	
Smoking status, n (%)			.32			.34
No	1119 (65.98)	577 (34.02)		1380 (81.37)	316 (18.63)	
Yes	573 (67.97)	270 (32.03)		699 (82.92)	144 (17.08)	
Alcohol consumption, n (%)			.87			.52
No	1139 (66.53)	573 (33.47)		1396 (81.54)	316 (18.46)	
Yes	553 (66.87)	274 (33.13)		683 (82.59)	144 (17.41)	
Activities of daily living, mean (SD)	6.41 (1.16)	6.81 (1.68)	*<.001*	6.53 (1.37)	6.59 (1.35)	.40
Chronic disease, n (%)			*<.001*			.40
None	494 (72.43)	188 (27.57)		563 (82.55)	119 (17.45)	
One	547 (70.13)	233 (29.87)		647 (82.95)	133 (17.05)	
≥2	651 (60.45)	426 (39.55)		869 (80.69)	208 (19.31)	
Sensory impairments, n (%)			*<.001*			.09
No sensory impairments	264 (77.65)	76 (22.35)		277 (81.47)	63 (18.53)	
Hearing impairment only	100 (74.07)	35 (25.93)		100 (74.07)	35 (25.93)	
Vision impairment only	441 (70.90)	181 (29.10)		507 (81.51)	115 (18.49)	
Dual sensory impairment	887 (61.51)	555 (38.49)		1195 (82.87)	247 (17.13)	

[a]Italicized values indicate statistical significance (*P*<.05) between the groups.

[b]Percentages were estimated over cases with valid data in every group.

[c]Chi-square test was not performed.

[d]Q1 was the poorest and Q4 was the richest. Q1: ≦US $415; Q2: US $416-1981; Q3: US $1982-3391; Q4: >US $3391.

Table 3. The prevalence of loneliness, social isolation, and SIs[a] in urban and rural China (N=3069).

Setting	2011, n (%)	2013, n (%)	2015, n (%)	2018, n (%)
Urban areas (n=530)				
Loneliness				
No	414 (78.11)	441 (83.21)	429 (80.94)	402 (75.85)
Yes	116 (21.89)	89 (16.79)	101 (19.06)	128 (24.15)
Social isolation				
Low	440 (83.02)	455 (85.85)	440 (83.02)	405 (76.42)
High	90 (16.98)	75 (14.15)	90 (16.98)	125 (23.58)
SIs				
No SIs	90 (16.98)	74 (13.96)	65 (12.26)	56 (10.57)
HI[b] only	37 (6.98)	55 (10.38)	32 (6.04)	30 (5.66)
VI[c] only	160 (30.19)	131 (24.72)	121 (22.83)	103 (19.43)
DSI[d]	243 (45.85)	270 (50.94)	312 (58.87)	341 (64.34)
Rural areas (n=2539)				
Loneliness				
No	1692 (66.64)	1876 (73.89)	1732 (68.22)	1584 (62.39)
Yes	847 (33.36)[e,f]	663 (26.11)[e,f]	807 (31.78)[e,f]	955 (37.61)[e,f]
Social isolation				
Low	2079 (81.88)	2071 (81.57)	2023 (79.68)	1800 (70.89)
High	460 (18.12)[e]	468 (18.43)[e,g]	516 (20.32)[e]	739 (29.11)[e,g]
SIs				
No SIs	340 (13.39)	284 (11.19)	207 (8.15)	209 (8.23)
HI only	135 (5.32)	163 (6.42)	123 (4.84)	105 (4.14)
VI only	622 (24.50)	568 (22.37)	460 (18.12)	418 (16.46)
DSI	1442 (56.79)[e,f]	1524 (60.02)[e,f]	1749 (68.89)[e,f]	1807 (71.17)[e,g]

[a]SI: sensory impairments.

[b]HI: hearing impairment.

[c]VI: vision impairment.

[d]DSI: dual sensory impairment.

[e]Compared with urban areas.

[f]*P*<.001.

[g]*P*<.05.

As shown in Figure 1, among older adults in urban areas, participants with VI only (OR 2.44, 95% CI 1.57-3.80; *P*<.001), HI only (OR 2.47, 95% CI 1.41-4.32; *P*=.002), and DSI (OR 1.88, 95% CI 1.24-2.85; *P*=.003) were more likely to experience loneliness compared with participants without SIs. Compared with HI only (OR 0.76, 95% CI 0.49-1.17; *P*=.22) or VI only (OR 0.77, 95% CI 0.59-1.01; *P*=.06), DSI showed an insignificant association with loneliness. Among older adults in rural areas, participants with DSI were more likely to experience loneliness compared with those without SIs (OR 1.84, 95% CI 1.56-2.18; *P*<.001), those with VI only (OR 1.37, 95% CI 1.22-1.54; *P*<.001), and HI only (OR 1.39, 95% CI 1.13-1.72; *P*=.002). Both VI only (OR 1.34, 95% CI 1.12-1.62; *P*=.002) and HI only (OR 1.32, 95% CI 1.02-1.71; *P*=.03) were associated with increased feelings of loneliness. Regardless of the rural or urban location, the effects of SIs on social isolation were not statistically significant (urban areas: DSI vs no SIs: *P*=.50; VI only vs no SIs: *P*=.92; HI only vs no SIs: *P*=.87; rural areas: DSI vs no SIs: *P*=.10; VI only vs no SIs: *P*=.80; HI only vs no SIs: *P*=.25; Figure 2). The results of the sensitivity analysis were consistent with the main analysis (Multimedia Appendix 3).

Figure 1. Association between SIs and loneliness among older adults in urban and rural China. DSI: dual sensory impairment; HI: hearing impairment; SIs: sensory impairments; VI: vision impairment.

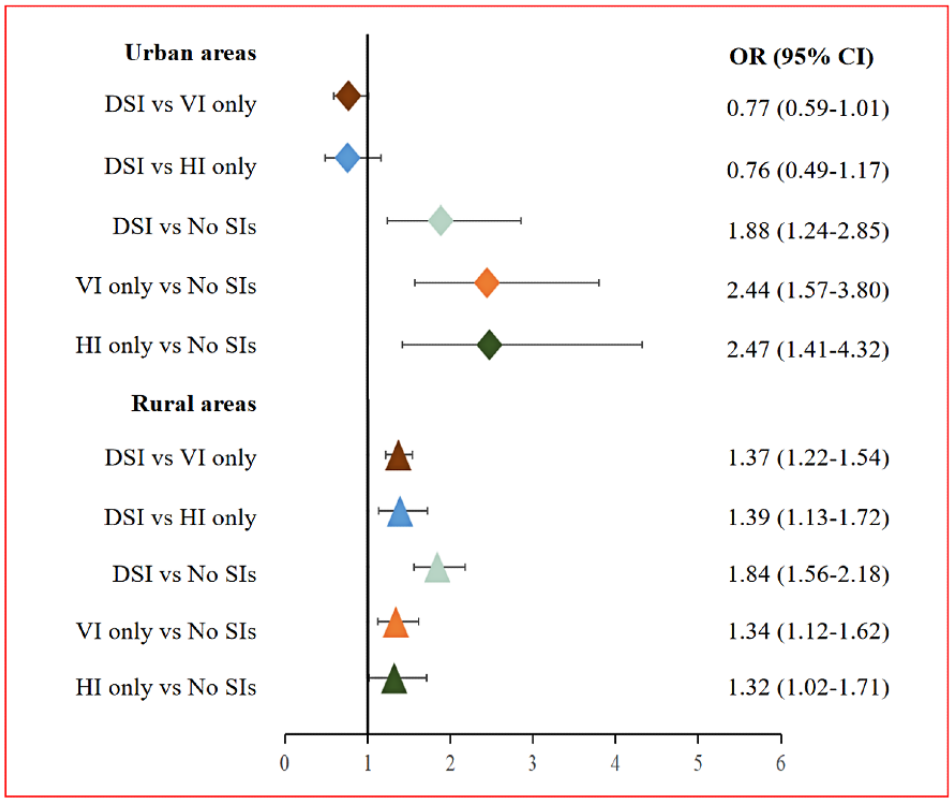
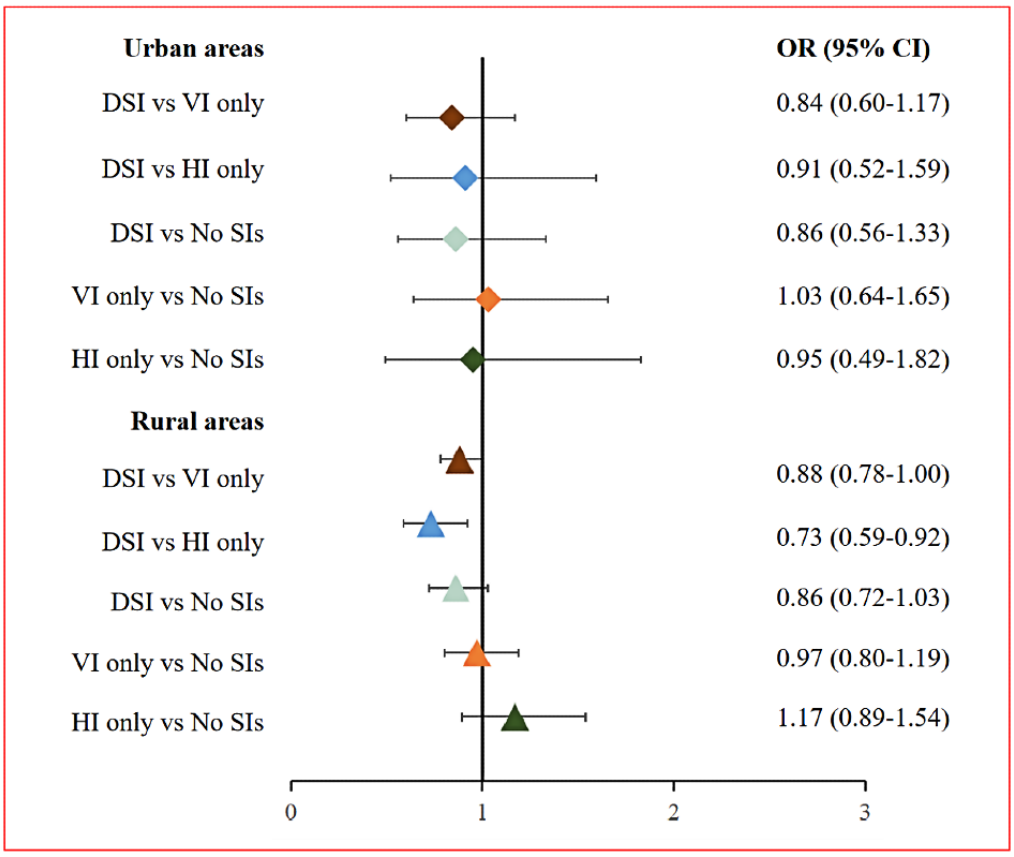


Figure 2. Association between SIs and social isolation among older adults in urban and rural China. DSI: dual sensory impairment; HI: hearing impairment; SIs: sensory impairments; VI: vision impairment.



Discussion

Principal Findings

Using longitudinal data from a nationally representative sample, this study provides an insight into the effect of DSI on loneliness and social isolation among Chinese older adults living in urban and rural areas. Results show that HI only, VI only, and DSI are each independently associated with a higher risk of loneliness in both urban and rural areas. HI and VI have a synergistic relationship with loneliness in rural areas, but we did not observe such an effect in urban areas. Our findings supported the view that SIs did not affect social isolation.

There are multiple reasons why older adults in rural areas have higher prevalence of DSI, loneliness, and social isolation than their urban counterparts. Older adults in rural areas may rarely receive appropriate treatment measures when HI or VI appears, as they have fewer health care resources, lower access to several health information sources, and poorer socioeconomic status compared with their urban counterparts [32-34]. With rapid industrialization and urbanization, there is massive migration of younger and middle-aged people from rural to urban areas, leaving older adults to either live alone or with their spouse in rural areas. Older adults with SIs who lived in rural areas had poorer perceived tangible support than their urban counterparts [35]. Furthermore, public services and voluntary organizations in rural areas are less developed than those in urban areas; consequently, older adults in rural areas have less choices for participating in social activities [36]. Therefore, older adults in rural areas might be more likely to experience loneliness and social isolation.

Our study found that individuals with SIs experience more loneliness compared with those without SIs, which was consistent with other studies [19,37]. Older adults with SIs are often lonely but they are not necessarily socially isolated. One possible reason is that SIs influence loneliness by affecting communication with their closest relatives or friends, as loneliness is not caused by being alone but rather by the unmet affective gain of their closest relationship [38]. Previous studies have indicated that older adults with DSI were more prone to experiencing a breakdown in communication compared with those without SIs [39,40]. Older adults with HI may feel frustrated or embarrassed over their difficulty communicating, resulting in loneliness [41]. Another reason may be neural changes associated with SIs. HI may contribute to changes in the frontal lobe, which alter the regulation of emotion [42], that may contribute to the likelihood of loneliness.

Surprisingly, older adults with DSI are at a compounded risk of loneliness compared with those with HI only or VI only in rural areas, but not in urban areas. Previous studies reported similar results, but there was no difference in mental health between older adults with DSI and single SI [43-45]; however, these studies did not consider the regional difference. A typical cultural and personal value among Chinese older adults was that they do not tend to bother others or even their adult children as they perceived that people are very busy with their own lives; therefore, they scarcely bother to seek help even though they may be sick [46]. In other words, older adults almost do not depend on their children until they are unable to take care of themselves, although children are the main caregivers for their older parents in China. Older adults with HI only or VI only were able to compensate for their impairment in one sense with the other so that they could take care of themselves to some extent [37]. As a particularly vulnerable group with challenges, older adults with DSI probably have to count on their children [47]. In urban areas, older adults with DSI could get more attention and care from their children in a timely manner, which might help them relieve loneliness. Likewise, older adults with DSI also deserve more attention and care in rural areas. However, most of the rural adults out-migrate to the cities for work and cannot take care of their parents by their side. Phone calls appear to be the main way of communication between older adults and their children, but this may be limited by HI unfortunately. Moreover, older parents with DSI have to reduce social activities due to the lack of their caregivers' help. Consequently, older parents with DSI who lived in rural areas may be at a higher risk to be lonely.

Previous cross-sectional studies have yielded conflicting results on the relationship between single SI and social isolation. A systematic review indicated that most studies found a relationship between HI and social isolation [41]. Kotwal et al [48] also found that HI rather than VI was associated with social isolation among older adults. However, significant correlations were found between VI and social isolation in another study [20]. Interestingly, in this longitudinal study, social isolation of older adults with SIs were not fundamentally different from those without SIs. According to the Socioemotional Selectivity Theory, older adults may selectively narrow their range of social partners and focus more on their closest relationships [38]. Reduction in social networks is likely normal in older adults, regardless of their health status including SIs and ADL limitation. Notably, undesired social isolation represents a low quality of social relationships and is very closely related to loneliness [49,50]. This led us to speculate that SIs may affect undesired social isolation. Future research is needed to shed more light on this issue. In addition, given the high prevalence of social isolation found in this study, observation should continuously be made for any signs of negative health outcomes in older adults with social isolation. Besides, active steps should be taken to prevent older adults from being socially isolated. Maintaining ties with family members and friends is important for preventing social isolation among American older adults [50]. In the context of Chinese culture, social ties to children may be an important priority in older adults.

Implications and Contribution

Our findings provide some new inspiration for older adults to relieve loneliness. Many prior interventions were conducted to mitigate loneliness by stimulating socialization [51], but they rarely considered that SIs may prevent older adults from enrolling or adhering to an intervention. Furthermore, an individual may be lonely without being socially isolated [49]. Future interventions considering SIs might be more accessible and effective for reducing loneliness among older adults. Recently, efforts to improve hearing have already shown beneficial effects on loneliness [52,53]. However, 74.39% (2283/3069) of older adults reported HI in 2018, but only 0.72%

(22/3069) were treated with a hearing aid in this study. This study emphasized that preventing HI among older adults is of high priority to reduce their risk of loneliness. In addition, we propose that parents with single SI in urban areas and those with DSI in rural areas should be get more attention from their children.

Strengths and Limitations

A particular strength of this study is the longitudinal examination of the effects of DSI on loneliness and social isolation by region of residence in Chinese older adults. However, several limitations of this study should be noted. First, SIs were self-reported, which might result in some bias. Participants may overestimate or underestimate their abilities to see and hear. Further research should use clinical diagnostic measurements to verify the data. Second, due to data limitation, the severity of SIs was not measured. Although the prevalence of SIs in urban areas was lower than that in rural areas, whether the severity of SIs in urban areas is higher than that in rural areas remains unknown, as urban people are more likely to experience occupational noise exposure. Future studies are therefore needed to consider these issues as the more serious SIs older adults have, the more likely they are to have physical and mental health problems [37,54]. Finally, loneliness was assessed with only 1 question. Although this measure was widely used in the literature [25,26], it might be less reliable than a composite measure.

Conclusion

Overall, this study found that SIs were significantly associated with loneliness rather than social isolation among older adults living in both urban and rural China. A synergistic effect of HI and VI on loneliness was observed in rural areas, but such an effect was not found in urban areas. A better understanding of the longitudinal effect of SIs on loneliness by region of residence could help policy makers to allocate health resources and conduct targeted interventions accordingly.

Acknowledgments

We thank the China Health and Retirement Longitudinal Study (CHARLS) team for providing data and all respondents for their contribution. This work was supported by the National Science Foundation of China (grant numbers 72274109, 71774104 and 71974117), the China Medical Board (grant number 16–257), Cheeloo Youth Scholar Grant, and Shandong University (grant numbers IFYT1810 and 2012DX006). The sponsor played no role in the design, methods, subject recruitment, data collection, analysis, or preparation of article.

Data Availability

The original data sets are publicly available at [55]. The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

CZ contributed to conceptualization and critical revision. QW and SZ performed formal analysis, drafting of the manuscript, and critical revision. YW and DZ performed formal analysis. All authors approved the final approval of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The flow chart of study respondent.
[DOCX File , 69 KB - publichealth_v8i11e39314_app1.docx]

Multimedia Appendix 2
The prevalence of loneliness and social isolation in treated and untreated participants in 2018.
[DOCX File , 17 KB - publichealth_v8i11e39314_app2.docx]

Multimedia Appendix 3
Association between SIs and social isolation among older adults in urban and rural China.
[DOCX File , 15 KB - publichealth_v8i11e39314_app3.docx]

References

1. Pinto J, Kern D, Wroblewski K, Chen R, Schumm L, McClintock M. Sensory function: insights from Wave 2 of the National Social Life, Health, and Aging Project. J Gerontol B Psychol Sci Soc Sci 2014 Nov;69 Suppl 2(Suppl 2):S144-S153 [FREE Full text] [doi: 10.1093/geronb/gbu102] [Medline: 25360015]

2. World Health Organization (WHO). Blindness and vision impairment. World Health Organization (WHO). Geneva, Switzerland: World Health Organization (WHO); 2021. URL: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment [accessed 2021-12-25]
3. World Health Organization (WHO). Deafness and hearing loss. World Health Organization (WHO). Geneva, Switzerland: World Health Organization (WHO); 2020. URL: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss [accessed 2021-12-25]
4. Heine C, Browning CJ, Gong CH. Sensory Loss in China: Prevalence, Use of Aids, and Impacts on Social Participation. Front Public Health 2019 Jan 24;7:5 [FREE Full text] [doi: 10.3389/fpubh.2019.00005] [Medline: 30733938]
5. Zhang Y, Ge M, Zhao W, Liu Y, Xia X, Hou L, et al. Sensory Impairment and All-Cause Mortality Among the Oldest-Old: Findings from the Chinese Longitudinal Healthy Longevity Survey (CLHLS). J Nutr Health Aging 2020 Jan 09;24(2):132-137. [doi: 10.1007/s12603-020-1319-2] [Medline: 32003401]
6. Gopinath B, Schneider J, McMahon CM, Burlutsky G, Leeder SR, Mitchell P. Dual sensory impairment in older adults increases the risk of mortality: a population-based study. PLoS One 2013 Mar 4;8(3):e55054 [FREE Full text] [doi: 10.1371/journal.pone.0055054] [Medline: 23469161]
7. Tay T, Wang JJ, Lindley R, Chia E, Landau P, Ingham N, et al. Sensory impairment, use of community support services, and quality of life in aged care clients. J Aging Health 2007 Apr 30;19(2):229-241. [doi: 10.1177/0898264307299243] [Medline: 17413133]
8. Wang Q, Gao T, Zhang S, Jing Z, Wang Y, Zhao D, et al. Association of Sensory Impairment and Health Care Utilization Among Chinese Older Adults With and Without Functional Impairment. J Am Med Dir Assoc 2021 Nov;22(11):2397-2398. [doi: 10.1016/j.jamda.2021.06.024] [Medline: 34270951]
9. Crowe K, Hovaldt HB, Dammeyer J. Communication participation in older adults with dual sensory loss. Speech Lang Hear 2019 May 30;23(4):232-242. [doi: 10.1080/2050571x.2019.1623457]
10. Guthrie DM, Davidson JGS, Williams N, Campos J, Hunter K, Mick P, et al. Combined impairments in vision, hearing and cognition are associated with greater levels of functional and communication difficulties than cognitive impairment alone: Analysis of interRAI data for home care and long-term care recipients in Ontario. PLoS One 2018 Feb 15;13(2):e0192971 [FREE Full text] [doi: 10.1371/journal.pone.0192971] [Medline: 29447253]
11. Haymes SA, Johnston AW, Heyes AD. Relationship between vision impairment and ability to perform activities of daily living. Ophthalmic Physiol Opt 2002 Mar;22(2):79-91. [doi: 10.1046/j.1475-1313.2002.00016.x] [Medline: 12014491]
12. Shukla A, Cudjoe T, Lin F, Reed N. Functional Hearing Loss and Social Engagement Among Medicare Beneficiaries. J Gerontol B Psychol Sci Soc Sci 2021 Jan 01;76(1):195-200 [FREE Full text] [doi: 10.1093/geronb/gbz094] [Medline: 31359056]
13. Hawkley LC, Cacioppo JT. Loneliness and pathways to disease. Brain Behav Immun 2003 Feb;17(1):98-105. [doi: 10.1016/s0889-1591(02)00073-9]
14. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
15. Lara E, Caballero FF, Rico-Uribe LA, Olaya B, Haro JM, Ayuso-Mateos JL, et al. Are loneliness and social isolation associated with cognitive decline? Int J Geriatr Psychiatry 2019 Nov 25;34(11):1613-1622. [doi: 10.1002/gps.5174] [Medline: 31304639]
16. Domènech-Abella J, Mundó J, Haro JM, Rubio-Valera M. Anxiety, depression, loneliness and social network in the elderly: Longitudinal associations from The Irish Longitudinal Study on Ageing (TILDA). J Affect Disord 2019 Mar 01;246:82-88. [doi: 10.1016/j.jad.2018.12.043] [Medline: 30578950]
17. Luo F, Guo L, Thapa A, Yu B. Social isolation and depression onset among middle-aged and older adults in China: Moderating effects of education and gender differences. J Affect Disord 2021 Mar 15;283:71-76. [doi: 10.1016/j.jad.2021.01.022] [Medline: 33524661]
18. CATTAN M, WHITE M, BOND J, LEARMOUTH A. Preventing social isolation and loneliness among older people: a systematic review of health promotion interventions. Aging Soc 2005 Jan 10;25(01):41-67. [doi: 10.1017/S0144686X04002594]
19. Mick P, Parfyonov M, Wittich W, Phillips N, Guthrie D, Kathleen Pichora-Fuller M. Associations between sensory loss and social networks, participation, support, and loneliness: Analysis of the Canadian Longitudinal Study on Aging. Can Fam Physician 2018 Jan;64(1):e33-e41 [FREE Full text] [Medline: 29358266]
20. Hajek A, König HH. Dual sensory impairment and psychosocial factors. Findings based on a nationally representative sample. Arch Gerontol Geriatr 2020 Aug 18;91:104234. [doi: 10.1016/j.archger.2020.104234] [Medline: 32835870]
21. Harithasan D, Mukari SZS, Ishak WS, Shahar S, Yeong WL. The impact of sensory impairment on cognitive performance, quality of life, depression, and loneliness in older adults. Int J Geriatr Psychiatry 2020 Apr 05;35(4):358-364. [doi: 10.1002/gps.5237] [Medline: 31736109]
22. Yang F, Gu D. Predictors of loneliness incidence in Chinese older adults from a life course perspective: a national longitudinal study. Aging Ment Health 2020 Jun 09;24(6):879-888. [doi: 10.1080/13607863.2018.1558174] [Medline: 30621448]

23. Zhang X, Dupre ME, Qiu L, Zhou W, Zhao Y, Gu D. Urban-rural differences in the association between access to healthcare and health outcomes among older adults in China. BMC Geriatr 2017 Jul 19;17(1):151 [FREE Full text] [doi: 10.1186/s12877-017-0538-9] [Medline: 28724355]
24. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol 2014 Feb 12;43(1):61-68 [FREE Full text] [doi: 10.1093/ije/dys203] [Medline: 23243115]
25. Yu B, Steptoe A, Niu K, Jia X. Social Isolation and Loneliness as Risk Factors for Grip Strength Decline Among Older Women and Men in China. J Am Med Dir Assoc 2020 Dec;21(12):1926-1930. [doi: 10.1016/j.jamda.2020.06.029] [Medline: 32723532]
26. Zhang J, Xu L, Li J, Sun L, Ding G, Qin W, et al. Loneliness and Health Service Utilization among the Rural Elderly in Shandong, China: A Cross-Sectional Study. Int J Environ Res Public Health 2018 Jul 11;15(7):1468 [FREE Full text] [doi: 10.3390/ijerph15071468] [Medline: 29997379]
27. Luo Y, Waite LJ. Loneliness and mortality among older adults in China. J Gerontol B Psychol Sci Soc Sci 2014 Jul 18;69(4):633-645 [FREE Full text] [doi: 10.1093/geronb/gbu007] [Medline: 24550354]
28. Steptoe A, Shankar A, Demakakos P, Wardle J. Social isolation, loneliness, and all-cause mortality in older men and women. Proc Natl Acad Sci U S A 2013 Apr 09;110(15):5797-5801 [FREE Full text] [doi: 10.1073/pnas.1219686110] [Medline: 23530191]
29. Xie T, Liu D, Guo J, Zhang B. The longitudinal effect of sensory loss on depression among Chinese older adults. J Affect Disord 2021 Mar 15;283:216-222. [doi: 10.1016/j.jad.2021.01.081] [Medline: 33561802]
30. Wang Q, Zhang S, Wang Y, Zhao D, Chen X, Zhou C. The Effect of Dual Sensory Impairment and Multimorbidity Patterns on Functional Impairment: A Longitudinal Cohort of Middle-Aged and Older Adults in China. Front Aging Neurosci 2022 Apr 8;14:807383 [FREE Full text] [doi: 10.3389/fnagi.2022.807383] [Medline: 35462686]
31. Zhao D, Li J, Hao W, Yuan Y, Yu C, Jing Z, et al. The relationship between activities of daily living and suicidal ideation among Chinese rural older adults: a multiple mediation model through sleep quality and psychological distress. Aging (Albany NY) 2020 Nov 17;12(22):22614-22625 [FREE Full text] [doi: 10.18632/aging.103857] [Medline: 33202378]
32. 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]
33. Chen X, Orom H, Hay JL, Waters EA, Schofield E, Li Y, et al. Differences in Rural and Urban Health Information Access and Use. J Rural Health 2019 Jun;35(3):405-417 [FREE Full text] [doi: 10.1111/jrh.12335] [Medline: 30444935]
34. Zhao Y, Xu X, Dupre ME, Xie Q, Qiu L, Gu D. Individual-level factors attributable to urban-rural disparity in mortality among older adults in China. BMC Public Health 2020 Sep 29;20(1):1472 [FREE Full text] [doi: 10.1186/s12889-020-09574-9] [Medline: 32993592]
35. Hay-McCutcheon MJ, Hyams A, Yang X, Parton J. Hearing loss and social support in urban and rural communities. Int J Audiol 2018 Aug 19;57(8):610-617. [doi: 10.1080/14992027.2018.1461262] [Medline: 29671659]
36. Sun J, Lyu S. Social participation and urban-rural disparity in mental health among older adults in China. J Affect Disord 2020 Sep 01;274:399-404. [doi: 10.1016/j.jad.2020.05.091] [Medline: 32663969]
37. Davidson JGS, Guthrie DM. Older Adults With a Combination of Vision and Hearing Impairment Experience Higher Rates of Cognitive Impairment, Functional Dependence, and Worse Outcomes Across a Set of Quality Indicators. J Aging Health 2019 Jan 13;31(1):85-108. [doi: 10.1177/0898264317723407] [Medline: 28805100]
38. Carstensen LL. Social and emotional patterns in adulthood: Support for socioemotional selectivity theory. Psychol Aging 1992;7(3):331-338. [doi: 10.1037/0882-7974.7.3.331]
39. Vreeken HL, van Nispen RMA, Kramer SE, van Rens GHMB. 'Dual Sensory Loss Protocol' for Communication and Wellbeing of Older Adults With Vision and Hearing Impairment - A Randomized Controlled Trial. Front Psychol 2020 Nov 26;11:570339 [FREE Full text] [doi: 10.3389/fpsyg.2020.570339] [Medline: 33324283]
40. Crews JE, Campbell VA. Vision impairment and hearing loss among community-dwelling older Americans: implications for health and functioning. Am J Public Health 2004 May;94(5):823-829. [doi: 10.2105/ajph.94.5.823] [Medline: 15117707]
41. Shukla A, Harper M, Pedersen E, Goman A, Suen JJ, Price C, et al. Hearing Loss, Loneliness, and Social Isolation: A Systematic Review. Otolaryngol Head Neck Surg 2020 May 10;162(5):622-633 [FREE Full text] [doi: 10.1177/0194599820910377] [Medline: 32151193]
42. Rutherford BR, Brewster K, Golub JS, Kim AH, Roose SP. Sensation and Psychiatry: Linking Age-Related Hearing Loss to Late-Life Depression and Cognitive Decline. Am J Psychiatry 2018 Mar 01;175(3):215-224 [FREE Full text] [doi: 10.1176/appi.ajp.2017.17040423] [Medline: 29202654]
43. Lehane CM, Hofsöe SM, Wittich W, Dammeyer J. Mental Health and Spouse Support Among Older Couples Living With Sensory Loss. J Aging Health 2018 Sep 14;30(8):1205-1223. [doi: 10.1177/0898264317713135] [Medline: 28613091]
44. Capella-McDonnall ME. The effects of single and dual sensory loss on symptoms of depression in the elderly. Int J Geriatr Psychiatry 2005 Sep;20(9):855-861. [doi: 10.1002/gps.1368] [Medline: 16116571]
45. Chou K, Chi I. Combined effect of vision and hearing impairment on depression in elderly Chinese. Int J Geriatr Psychiatry 2004 Sep 26;19(9):825-832. [doi: 10.1002/gps.1174] [Medline: 15352139]

46. Leung AY, Molassiotis A, Carino DA. A Challenge to Healthy Aging: Limited Social Participation in Old Age. Aging Dis 2021;12(7):1536. [doi: 10.14336/ad.2021.02018]
47. Kuriakose RK, Khan Z, Almeida DRP, Braich PS. Depression and burden among the caregivers of visually impaired patients: a systematic review. Int Ophthalmol 2017 Jun 29;37(3):767-777. [doi: 10.1007/s10792-016-0296-2] [Medline: 27473225]
48. Kotwal AA, Cenzer IS, Waite LJ, Covinsky KE, Perissinotto CM, Boscardin WJ, et al. The epidemiology of social isolation and loneliness among older adults during the last years of life. J Am Geriatr Soc 2021 Nov 11;69(11):3081-3091 [FREE Full text] [doi: 10.1111/jgs.17366] [Medline: 34247388]
49. Wright-St Clair VA, Neville S, Forsyth V, White L, Napier S. Integrative review of older adult loneliness and social isolation in Aotearoa/New Zealand. Australas J Ageing 2017 Jun 04;36(2):114-123 [FREE Full text] [doi: 10.1111/ajag.12379] [Medline: 28258607]
50. Taylor HO, Taylor RJ, Nguyen AW, Chatters L. Social Isolation, Depression, and Psychological Distress Among Older Adults. J Aging Health 2018 Feb 17;30(2):229-246 [FREE Full text] [doi: 10.1177/0898264316673511] [Medline: 28553785]
51. Huang AR, Deal JA, Rebok GW, Pinto JM, Waite L, Lin FR. Hearing Impairment and Loneliness in Older Adults in the United States. J Appl Gerontol 2021 Oct 04;40(10):1366-1371. [doi: 10.1177/0733464820944082] [Medline: 32749194]
52. Contrera KJ, Sung YK, Betz J, Li L, Lin FR. Change in loneliness after intervention with cochlear implants or hearing aids. Laryngoscope 2017 Aug 06;127(8):1885-1889 [FREE Full text] [doi: 10.1002/lary.26424] [Medline: 28059448]
53. Applebaum J, Hoyer M, Betz J, Lin FR, Goman AM. Long-term subjective loneliness in adults after hearing loss treatment. Int J Audiol 2019 Aug 31;58(8):464-467. [doi: 10.1080/14992027.2019.1593523] [Medline: 30929531]
54. Ye X, Zhu D, Chen S, He P. The association of hearing impairment and its severity with physical and mental health among Chinese middle-aged and older adults. Health Qual Life Outcomes 2020 May 26;18(1):155 [FREE Full text] [doi: 10.1186/s12955-020-01417-w] [Medline: 32456646]
55. CHARLS - China Health and Retirement Longitudinal Study. URL: http://charls.pku.edu.cn/index/en.html [accessed 2022-11-09]

Abbreviations

ADL: activities of daily living
CESD: Center for Epidemiological Studies Depression Scale
CHARLS: China Health and Retirement Longitudinal Study
DSI: dual sensory impairment
HI: hearing impairment
SI: sensory impairment
VI: vision impairment
OR: odds ratio
WHO: World Health Organization

Edited by A Mavragani, G Eysenbach; submitted 06.05.22; peer-reviewed by J Ye, P Sooful; comments to author 09.09.22; revised version received 28.09.22; accepted 18.10.22; published 14.11.22.

Please cite as:
Wang Q, Zhang S, Wang Y, Zhao D, Zhou C
Dual Sensory Impairment as a Predictor of Loneliness and Isolation in Older Adults: National Cohort Study
JMIR Public Health Surveill 2022;8(11):e39314
URL: https://publichealth.jmir.org/2022/11/e39314
doi:10.2196/39314
PMID:36374533

Original Paper

Stage-Specific Survival in Breast Cancer in Chinese and White Women: Comparative Data Analysis

Jun Wang[1,2*], MD; Juan Zhou[3*], MD; Lei Liu[1,2], MD; San-Gang Wu[4,5], MD

[1]Department of Head and Neck Oncology, Cancer Center, West China Hospital, Sichuan University, Chengdu, China

[2]Department of Radiation Oncology, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, China

[3]Department of Obstetrics and Gynecology, the First Affiliated Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, China

[4]Department of Radiation Oncology, the First Affiliated Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, China

[5]Xiamen Key Laboratory of Radiation Oncology, Xiamen Cancer Center, Xiamen, China

[*]these authors contributed equally

Corresponding Author:
San-Gang Wu, MD
Department of Radiation Oncology
the First Affiliated Hospital of Xiamen University
School of Medicine, Xiamen University
55 Zhenhai Road
Xiamen, 361003
China
Phone: 86 5922139531
Email: wusg@xmu.edu.cn

Abstract

Background: Stage-specific survival, according to the eighth edition of the American Joint Committee on Cancer (AJCC) pathological prognostic staging (PPS) on breast cancer (BC), between Chinese and White American women remains unclear.

Objective: This study aimed to assess stage-specific survival in BC between Chinese and White American women according to the eighth AJCC PPS.

Methods: We included Chinese and White American women with BC diagnosed between 2010 and 2018 from the Surveillance, Epidemiology, and End Results database. A chi-square test, the Kaplan–Meier method, a receiver operating characteristic (ROC) curve, and multivariate Cox proportional hazards models were used for data analysis.

Results: We included 376,818 individuals in this study: 369,522 White American and 7296 Chinese. Of them, 149,452 (39.7%) migrated from the seventh AJCC anatomic staging (AS) to the eighth AJCC PPS, 22,516 (6.0%) were upstaged, and 126,936 (33.7%) were downstaged. With a median follow-up duration of 44 months, the 5-year overall survival and cancer-specific survival (CSS) for the entire group were 87.4% and 95.9%, respectively. The seventh AJCC AS (P<.001) and the eighth AJCC PPS (P<.001) could significantly predict the survival outcomes of BC, and multivariate analysis revealed that both staging systems were significant prognostic indicators of CSS. The ROC curve revealed that the PPS had a better discriminating ability than the AS (area under the curve [AUC] 0.769 vs 0.753, P<.001). Similar trends were observed after stratification by the 2 ethnic groups. The eighth AJCC PPS had better discriminating ability than the seventh AJCC AS among both White American (AUC 0.769 vs 0.753, P<.001) and Chinese patients (AUC 0.790 vs 0.776, P<.001). In the seventh AJCC AS, Chinese women had better CSS in stage IA (P=.02), stage IIA (P=.005), and stage IIIB (P=.04) disease than White American women, but no significant CSS was observed in stage IB, IIB, IIIA, and IIIC disease between the 2 ethnic groups. Regarding the eighth AJCC PPS, Chinese women had better CSS in stage IA (P=.002) and IIIA (P=.046) disease than White American women, and CSS was similar in Chinese and White American women in other substages.

Conclusions: The eighth AJCC PPS has a similar discriminative ability between White American and Chinese individuals with BC compared with the seventh AJCC AS. Therefore, the eighth AJCC PPS is also applicable to Chinese individuals with BC.

(JMIR Public Health Surveill 2022;8(11):e40386) doi:10.2196/40386

KEYWORDS

breast cancer; AJCC; American Joint Committee on Cancer; Chinese; White American; survival; surveillance; epidemiology; staging; pathological prognostic staging; AJCC stage; overall survival; cancer-specific survival

Introduction

Breast cancer (BC) is the most commonly diagnosed cancer among women, and 2,261,419 new cases were estimated to occur worldwide in 2020 [1]. The incidence of BC varies among different countries and racial and ethnic groups with age-adjusted rates of 129 per 100,000 population for White American and 75.1 per 100,000 population for Chinese women in the United States, while the rate was 30.69 per 100,000 population for Chinese women based on the Chinese population [2,3]. The survival outcomes are also discrepant in different races: White American individuals with BC have better cancer-specific survival (CSS) than African American individuals but poorer survival than Chinese individuals [4,5].

Traditional American Joint Committee on Cancer (AJCC) anatomic staging (AS) of BC includes tumor size (T), lymph nodes (N), and distant metastasis (M), which has been extensively used for predicting prognosis and guiding treatment in BC [6,7]. The eighth edition of the AJCC pathological prognostic staging (PPS) integrates the estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor-2 (HER2), and tumor grade into AS, which facilitates better precise prognostic stratification than AS [8-10]. However, the vast majority of the included patients for establishing the initial model of the eighth AJCC PPS were White American and Hispanic American [11], and whether the new PPS is also applicable to the Chinese individuals with BC remains unclear. As the country with the largest population worldwide, China has approximately 400,000 new BC cases annually, and previous studies have shown that Chinese patients have different morbidity and survival outcomes than White American patients [3,4,12]. Therefore, more ethnic-based studies are needed to explore the value of the PPS in Chinese individuals with BC to make this new staging more widely available.

Several studies have attempted to validate the value of the new staging in the Chinese population [13-15]. However, these studies included specific subgroups, such as T1-2N1 and triple-negative breast cancer (TNBC), which could not represent patients with BC in general [13-15]. In addition, the small sample size and the heterogeneity of treatment also made it difficult to accurately assess the value of the new staging in Chinese patients. In this study, we used a population-based cohort to compare stage-specific survival in BC between White American and Chinese women in accordance with the eighth AJCC PPS, and to expand the applicability of the new staging system in more races.

Methods

Data Source and Patient Selection

The patient data in this study were extracted from the Surveillance, Epidemiology, and End Results (SEER) database between 2010 and 2018. The SEER database collects information on cancer statistics, treatment, and survival, covering approximately 48% of the population in the United States [16]. Patients with the following criteria were included: (1) pathologically diagnosed with invasive BC; (2) White American or Chinese individuals with BC; (3) having detailed information on age, TNM stage, ER status, PR status, HER2 status, histology subtype, tumor grade, surgery, radiotherapy, and chemotherapy administration. Male patients, patients with contralateral BC, and those diagnosed with a distant metastatic stage were excluded from this study.

Variables and Endpoints

We selected the variables including age (<50, 50-70, and >70 years), race and ethnicity (White American and Chinese), grade (well-, moderately-, poorly differentiated, and undifferentiated), histological subtype (infiltrating duct carcinoma, lobular carcinoma, mixed, and other carcinomas), molecular subtype (luminal A, luminal B, HER2-enriched, and triple-negative BC), T stage (T0-T4), N stage (N0-N3), and the seventh and eighth AJCC staging (IA-IIIC). The races and ethnicities of White American (code 01) and Chinese (code 04) were chosen for analysis using “Race/ethnicity” codes in the SEER database. The end point of this study was CSS, which was calculated as the time from BC diagnosis to the occurrence of BC-related death.

Ethical Considerations

This study was approved by the ethics committee of the First Affiliated Hospital of Xiamen University (Xiamen) and West China Hospital, Sichuan University (Chengdu; 2021GGB027). Informed consent is not required because the data were extracted from the SEER database after obtaining permission from the administrator. In addition, the privacy of the participants was well protected through anonymization and deidentification of their information.

Statistical Analysis

The Pearson chi-square test was used to compare the differences in baseline characteristics and stage migration changes between groups of Chinese and White American individuals with BC. The receiver operating characteristic (ROC) curve was used to identify the discriminating ability of the AS and PPS. The Kaplan–Meier method was used to plot the survival curves, and the log-rank test was used to compare the differences. Multivariate Cox proportional hazards models were used to calculate the independent risk predictors of CSS. Sensitivity analyses were used to investigate the effect of race on CSS after stratification by different AJCC substages. SPSS (version 22.0; IBM Corp) was used for analyzing all the data. A *P* value less than .05 was defined as the threshold for statistical significance.

Results

Cohort Characteristics

In total, 376,818 individuals with BC were identified in this study. Of them, 369,522 (98.1%) were White American and 7296 (1.9%) were Chinese. The proportions of patients aged <50, 50-70, and ≥70 years were 18.3% (n=68,833), 50.9% (n=191,831), and 30.8% (n=116,154), respectively. The majority of the patients had infiltrating ductal carcinoma (n=294,359, 78.1%), a low-grade tumor (n=279,429, 74.2%), luminal A subtype BC (n=288,545, 76.6%), T1-2 stage BC (n=347,931, 92.3%), and N0-1 stage BC (n=351,666, 93.3%).

Regarding the distribution of baseline characteristics between White American and Chinese individuals, the latter were more likely to be younger (P<.001) and have a lower-grade tumor (P<.001), infiltrating duct carcinoma (P<.001), luminal B or HER2-enriched subtype (P<.001), and a lower-stage tumor (P<.001) than White American participants. Among all participants, stage IA, IB, IIA, IIB, IIIA, IIIB, and IIIC disease accounted for 53.3% (n=200,926), 2.3% (n=8,569), 23.0% (n=86,651), 11.1% (n=41,837), 5.9% (n=22,237), 2.0% (n=7649), and 2.4% (n=8949) of cases in the seventh AJCC AS and for 66.1% (n=249,173), 15.5% (n=58,405), 7.8% (n=29,379), 3.2% (n=12,116), 3.9% (n=14,800), 1.9% (n=7175), and 1.5% (n=5770) of cases in the eighth AJCC PPS, respectively.

With regard to the treatments, 96.1% (n=362,189) of the participants received surgical intervention, 52.7% (n=198,399) of them received radiotherapy, and 36.7% (n=139,160) of them received chemotherapy. White American participants were more likely to undergo breast-conserving surgery, while more Chinese participants were treated with mastectomy (P<.001). In addition, Chinese participants were more prone to receiving chemotherapy (P<.001), while White American patients were more likely to be treated with radiotherapy (P<.001). Detailed information on the study population is presented in Table 1.

Table 1. Participants' characteristics (N=376,818).

Variables	Total, n (%)	Race and ethnicity, n (%)		*P* value
		White American (n=369,522)	Chinese (n=7296)	
Age groups (years)				<.001
<50	68,833 (18.3)	66,803 (18.1)	2030 (27.8)	
50-69	191,831 (50.9)	188,051 (50.9)	3780 (51.8)	
≥70	116,154 (30.8)	114,668 (31.0)	1486 (20.4)	
Grade				<.001
Well differentiated	97,389 (25.8)	95,820 (25.9)	1569 (21.5)	
Moderately differentiated	173,649 (46.1)	170,182 (46.1)	3467 (47.5)	
Poorly differentiated or undifferentiated	105,780 (28.1)	103,520 (28.0)	2260 (31.0)	
Histology				<.001
Infiltrating duct carcinoma	294,359 (78.1)	288,243 (78.0)	6116 (83.8)	
Lobular carcinoma	39,292 (10.4)	38,880 (10.5)	412 (5.6)	
Mixed	22,473 (6.0)	22,186 (6.0)	287 (3.9)	
Other	20,694 (5.5)	20,213 (5.5)	481 (6.6)	
Molecular subtype				<.001
Luminal A	288,545 (76.6)	283,251 (76.7)	5294 (72.6)	
Luminal B	37,325 (9.9)	36,449 (9.9)	876 (12.0)	
HER2-enriched	14,065 (3.7)	13,611 (3.7)	454 (6.2)	
Triple-negative	36,757 (9.8)	36,086 (9.8)	671 (9.2)	
T (tumor size) stage				<.001
T0	164 (0.0)	160 (0.0)	4 (0.0)	
T1	237,926 (63.1)	233,569 (63.2)	4357 (59.7)	
T2	110,005 (29.2)	107,560 (29.1)	2445 (33.5)	
T3	19,784 (5.3)	19,443 (5.3)	341 (4.7)	
T4	8939 (2.4)	8790 (2.4)	149 (2.0)	
N (lymph nodes) stage				.63
N0	272,990 (72.4)	267,686 (72.4)	5304 (72.7)	
N1	78,676 (20.9)	77,149 (20.9)	1527 (20.9)	
N2	16,203 (4.3)	15,895 (4.3)	308 (4.2)	
N3	8949 (2.4)	8792 (2.4)	157 (2.2)	
Seventh version of the American Joint Committee on Cancer (AJCC) staging				<.001
IA	200,926 (53.3)	197,231 (53.4)	3695 (50.6)	
IB	8569 (2.3)	8422 (2.3)	147 (2.0)	
IIA	86,651 (23.0)	84,765 (22.9)	1886 (25.8)	
IIB	41,837 (11.1)	40,962 (11.1)	875 (12.0)	
IIIA	22,237 (5.9)	21,827 (5.9)	410 (5.6)	
IIIB	7649 (2.0)	7523 (2.0)	126 (1.7)	
IIIC	8949 (2.4)	8792 (2.4)	157 (2.2)	
Eighth version of the AJCC staging				<.001
IA	249,173 (66.1)	244,477 (66.2)	4696 (64.4)	
IB	58,405 (15.5)	57,236 (15.5)	1169 (16.0)	

Variables	Total, n (%)	Race and ethnicity, n (%)		*P* value
		White American (n=369,522)	Chinese (n=7296)	
IIA	29,379 (7.8)	28,722 (7.8)	657 (9.0)	
IIB	12,116 (3.2)	11,843 (3.2)	273 (3.7)	
IIIA	14,800 (3.9)	14,538 (3.9)	262 (3.6)	
IIIB	7175 (1.9)	7023 (1.9)	152 (2.1)	
IIIC	5770 (1.5)	5683 (1.5)	87 (1.2)	
Surgery				<.001
No surgery	14,145 (3.8)	13,867 (3.8)	278 (3.8)	
Breast-conserving surgery	213,329 (56.6)	209,671 (56.7)	3658 (50.1)	
Mastectomy	148,860 (39.5)	145,509 (39.4)	3351 (45.9)	
Unknown	484 (0.1)	475 (0.1)	9 (0.1)	
Radiotherapy				<.001
No	167,183 (44.4)	163,677 (44.3)	3506 (48.1)	
Beam radiation	187,640 (49.8)	184,138 (49.8)	3502 (48.0)	
Radioactive implants	10,759 (2.9)	10,631 (2.9)	128 (1.80)	
Unknown	11,236 (3.0)	11,076 (3.0)	160 (2.2)	
Chemotherapy				<.001
No	237,658 (63.1)	233,283 (63.1)	4375 (60.0)	
Yes	139,160 (36.9)	136,239 (36.9)	2921 (40.0)	

Stage Migration

According to the seventh AJCC AS, 197,231 (53.4%), 8422 (2.3%), 84,765 (22.9%), 40,962 (11.1%), 21,827 (5.9%), 7523 (2.0%), and 8792 (2.4%) White American participants versus 3695 (50.6%), 147 (2.0%), 1886 (25.8%), 875 (12.0%), 410 (5.6%), 126 (1.7%), and 157 (2.2%) Chinese participants had stage IA, IB, IIA, IIB, IIIA, IIIB, and IIIC disease, respectively (*P*<.001). According to the eighth AJCC PPS, 244,477 (66.2%), 57,236 (15.5%), 28,722 (7.8%), 11,843 (3.2%), 14,538 (3.9%), 7023 (1.9%), and 5683 (1.5%) White American participants versus 4696 (64.4%), 1169 (16.0%), 657 (9.0%), 273 (3.7%), 262 (3.6%), 152 (2.1%), and 87 (1.2%) Chinese participants had stage IA, IB, IIA, IIB, IIIA, IIIB, and IIIC disease, respectively (*P*<.001; Table 1). A total of 149,452 (39.7%) participants migrated from the seventh AJCC AS to the eighth AJCC PPS (n=22,516, 6.0% upstaged and n=126,936, 33.7% downstaged). Among the upstaged participants, 22,127 (6.0%) were White American and 389 (6.3%) were Chinese, while among the downstaged participants, 124,368 (33.6%) were White American and 2586 (35.2%) were Chinese (*P*=.004). Furthermore, the disease stages of 223,027 (60.4%) White American and 4339 (59.5%) Chinese participants remained unchanged. There was a significant difference in stage migration (upstaging, downstaging, and unchanging stage) between White American and Chinese participants (*P*=.004). The frequencies of stage discrepancies between White American and Chinese participants are shown in Table 2.

Table 2. The frequencies of stage discrepancies between White American and Chinese participants.

Seventh AJCC[a] AS[b]	Eighth AJCC PPS[c], n (%)							Total, n (%)
	IA	IB	IIA	IIB	IIIA	IIIB	IIIC	
White American								
IA	183,587 (49.7)	13,644 (3.7)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	197,231 (53.4)
IB	7984 (2.2)	438 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	8422 (2.3)
IIA	48,507 (13.1)	14,764 (4.0)	21,494 (5.8)	0 (0)	0 (0)	0 (0)	0 (0)	84,765 (22.9)
IIB	4399 (1.2)	17,924 (4.9)	6058 (1.6)	8379 (2.3)	4202 (1.1)	0 (0)	0 (0)	40,962 (11.1)
IIIA	0 (0)	10,466 (2.8)	1170 (0.3)	3464 (0.9)	4245 (1.1)	399 (0.1)	2083 (0.6)	21,827 (5.9)
IIIB	0 (0)	0 (0)	0 (0)	0 (0)	2641 (0.7)	3083 (0.8)	1799 (0.5)	7523 (2.0)
IIIC	0 (0)	0 (0)	0 (0)	0 (0)	3450 (0.9)	3541 (1.0)	1801 (0.5)	8792 (2.4)
Total	244,477 (66.2)	57,236 (15.5)	28,722 (7.8)	11,843 (3.2)	14,538 (3.9)	7023 (1.9)	5683 (1.5)	369,522 (100)
Chinese								
IA	3447 (47.2)	248 (3.4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3695 (50.6)
IB	143 (2.0)	4 (0.1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	147 (2.0)
IIA	1040 (14.3)	350 (4.8)	496 (6.8)	0 (0)	0 (0)	0 (0)	0 (0)	1886 (25.8)
IIB	66 (0.1)	404 (5.5)	133 (1.8)	194 (2.7)	78 (1.1)	0 (0)	0 (0)	875 (12.0)
IIIA	0 (0)	163 (2.2)	28 (0.4)	79 (1.1)	99 (1.4)	9 (0.1)	32 (0.4)	410 (5.6)
IIIB	0 (0)	0 (0)	0 (0)	0 (0)	38 (0.5)	66 (0.9)	22 (0.3)	126 (1.7)
IIIC	0 (0)	0 (0)	0 (0)	0 (0)	47 (0.6)	77 (1.1)	33 (0.5)	157 (2.2)
Total	4696 (64.4)	1169 (16.0)	657 (9.0)	273 (3.7)	262 (3.6)	152 (2.1)	87 (1.2)	7296 (100)

[a]AJCC: American Joint Committee on Cancer.

[b]AS: anatomic staging.

[c]PPS: pathological prognostic staging.

Survival and Prognostic Analyses by Race and Ethnicity

With a median follow-up duration of 44 months (range 0-107 months), 42,522 deaths and 17,807 breast cancer–related deaths occurred. The Kaplan–Meier curves showed that 5-year overall survival and CSS for the entire group were 87.4% and 95.9%, respectively. The seventh AJCC AS (P<.001; Figure 1A) and the eighth AJCC PPS (P<.001; Figure 1B) could significantly predict the survival outcome of BC. Multivariate Cox proportional hazards analysis revealed that the AS and PPS both had significant prognostic predicting value in the study population (Table 3). Kaplan–Meier curves (Figure 1) and ROC curve (area under the curve [AUC] 0.769 vs 0.753, P<.001; Figure 2A) indicated that the eighth AJCC PPS had better discriminating ability than the seventh AJCC AS.

In White American participants, the AJCC AS (P<.001; Figure 3A) and the AJCC PPS (P<.001; Figure 3B) could also significantly predict survival and prognosis consistent with stages, and multivariate analysis showed that the 2 staging systems were significant prognostic predictors of CSS (Table 4). In addition, the eighth AJCC PPS had better discriminating ability than the seventh AJCC AS in White American participants (AUC 0.769 vs 0.753, P<.001; Figure 2B). Similar results were obtained for Chinese participants, in that the seventh AJCC AS (P<.001; Figure 4A) and the eighth AJCC PPS (P<.001; Figure 4B) both had significant prognostic values (Table 4), and the eighth AJCC PPS still showed better discriminating ability (AUC 0.790 vs 0.776, P<.001; Figure 2C) in this population.

Figure 1. Cancer-specific survival according to the anatomic staging (A) and pathological prognostic staging (B) systems for the entire cohort.

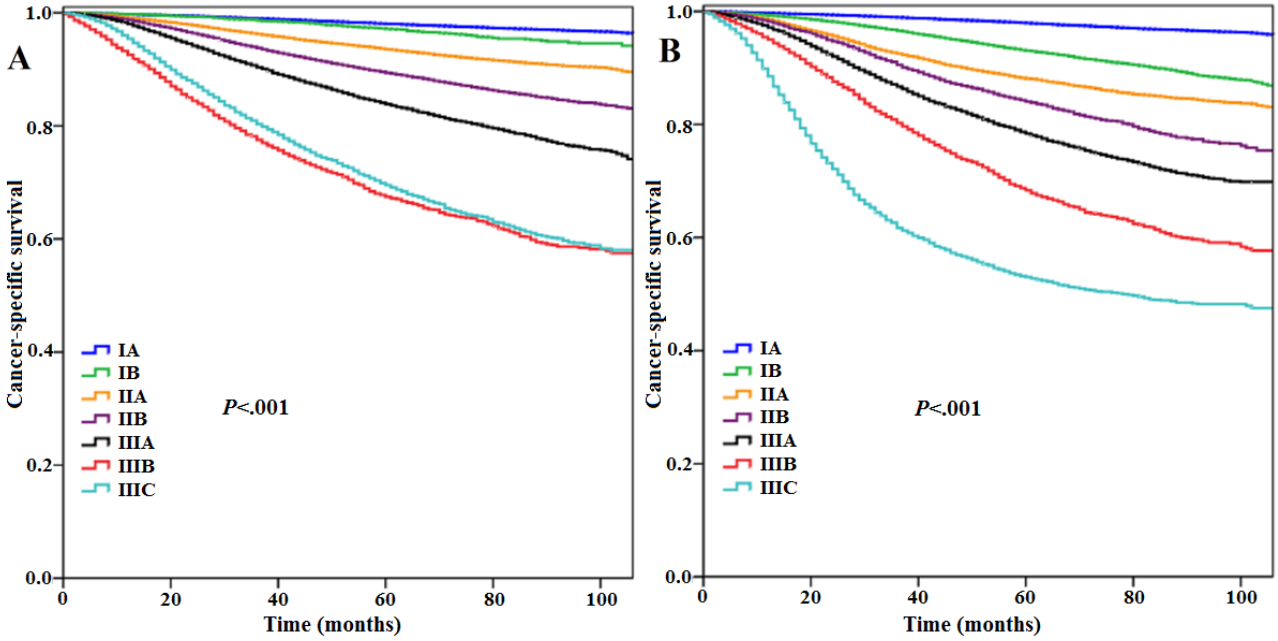


Table 3. Cox multivariate analysis for cancer-specific survival according to different staging systems.

Variables	Cancer-specific survival			
	Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
Age (year)				
<50	1 (reference)	N/A[a]	1 (reference)	N/A
50-69	1.115 (1.070-1.162)	<.001	1.100 (1.055-1.146)	<.001
≥70	2.458 (2.357-2.564)	<.001	2.413 (2.314-2.517)	<.001
Race and ethnicity				
White	1 (reference)	N/A	1 (reference)	N/A
Chinese	0.827 (0.732-0.934)	.002	0.827 (0.732-0.934)	.002
Histology				
Infiltrating duct carcinoma	1 (reference)	N/A	1 (reference)	N/A
Lobular carcinoma	1.037 (0.984-1.093)	.18	1.064 (1.009-1.121)	.02
Mixed	0.950 (0.888-1.016)	.13	0.978 (0.914-1.045)	.51
Other	0.919 (0.862-0.979)	.009	0.924 (0.867-0.984)	.01
Grade				
Well differentiated	1 (reference)	N/A	1 (reference)	N/A
Moderately differentiated	1.594 (1.503-1.690)	<.001	1.544 (1.454-1.640)	<.001
Poorly differentiated or undifferentiated	2.862 (2.693-3.041)	<.001	2.272 (2.126-2.427)	<.001
Molecular subtype				
Luminal A	1 (reference)	N/A	1 (reference)	N/A
Luminal B	0.894 (0.849-0.941)	<.001	0.946 (0.898-0.996)	.03
HER2-enriched	1.164 (1.092-1.240)	<.001	1.040 (0.975-1.110)	.23
Triple negative	2.224 (2.139-2.313)	<.001	1.560 (1.482-1.641)	<.001
T (tumor size) stage				
T0	1 (reference)	N/A	1 (reference)	N/A
T1	0.735 (0.477-1.133)	.16	0.623 (0.405-0.958)	.03
T2	1.007 (0.655-1.548)	.98	1.136 (0.739-1.745)	.56
T3	1.428 (0.927-2.201)	.11	1.555 (1.011-2.393)	.045
T4	2.072 (1.335-3.217)	.001	2.160 (1.402-3.329)	<.001
N (lymph nodes) stage				
N0	1 (reference)	N/A	1 (reference)	N/A
N1	1.202 (1.132-1.276)	<.001	1.369 (1.311-1.429)	<.001
N2	1.618 (1.477-1.774)	<.001	1.821 (1.714-1.935)	<.001
N3	8.945 (8.212-9.745)	<.001	2.261 (2.093-2.442)	<.001
Seventh edition of the American Joint Committee on Cancer (AJCC) staging				
IA	1 (reference)	N/A	—[b]	N/A
IB	1.389 (1.196-1.612)	<.001	—	—
IIA	2.029 (1.882-2.188)	<.001	—	—
IIB	2.791 (2.493-3.124)	<.001	—	—
IIIA	3.399 (2.973-3.888)	<.001	—	—
IIIB	4.162 (3.575-4.843)	<.001	—	—
IIIC	8.945 (8.212-9.745)	<.001	—	—

Variables	Cancer-specific survival			
	Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
Eighth edition of the AJCC pathologic prognostic staging				
IA	—	N/A	1 (reference)	N/A
IB	—	—	1.541 (1.449-1.639)	<.001
IIA	—	—	2.043 (1.894-2.202)	<.001
IIB	—	—	2.534 (2.313-2.754)	<.001
IIIA	—	—	2.773 (2.533-3.036)	<.001
IIIB	—	—	3.060 (2).735-3.425	<.001
IIIC	—	—	4.380 (3.872-4.955)	<.001

[a]N/A: not applicable.

[b]Not available.

Figure 2. Receiver operating characteristic curve for predicting the discriminating value of the seventh and the eighth editions of the American Joint Committee on Cancer (AJCC) staging system in the entire cohort (A), White American women (B), and Chinese women (C).

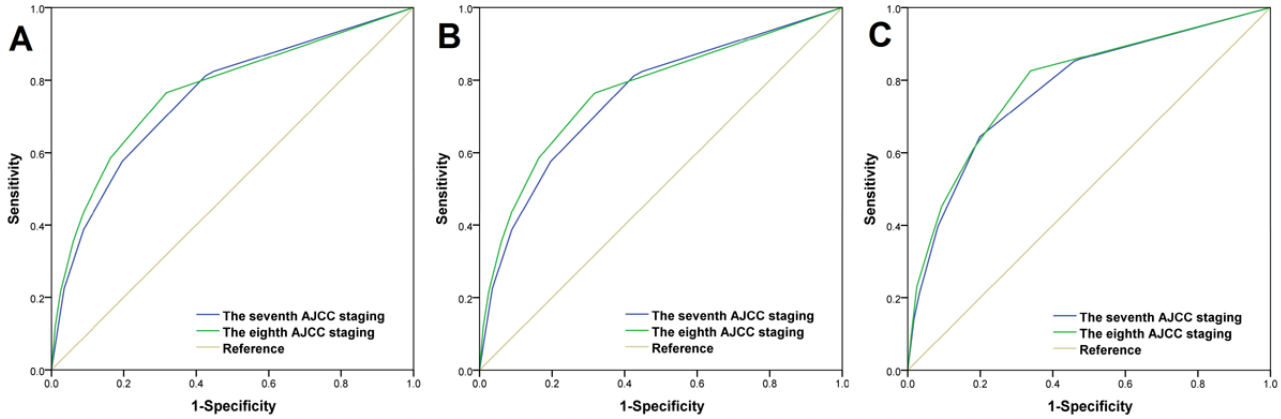


Figure 3. Kaplan–Meier curve for cancer-specific survival according to the seventh (A) and eighth editions of the American Joint Committee on Cancer (B) staging system in White American women.

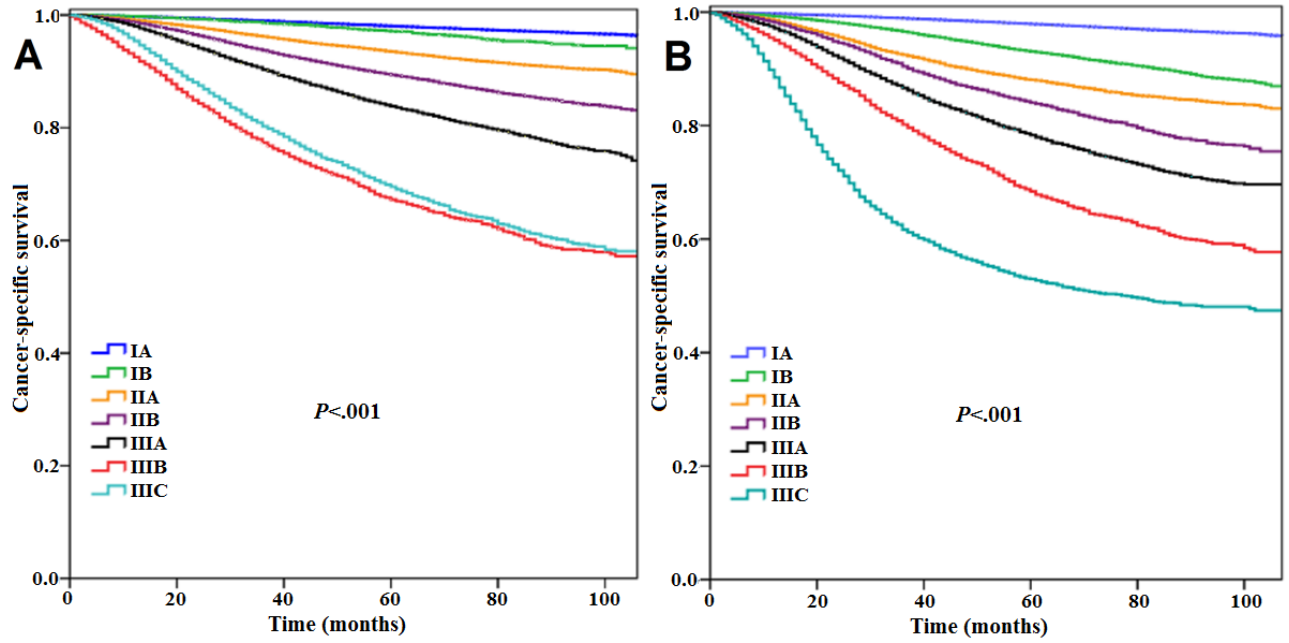


Figure 4. Kaplan–Meier curve for cancer-specific survival according to the seventh (A) and eighth editions of the American Joint Committee on Cancer (B) staging system in Chinese women.

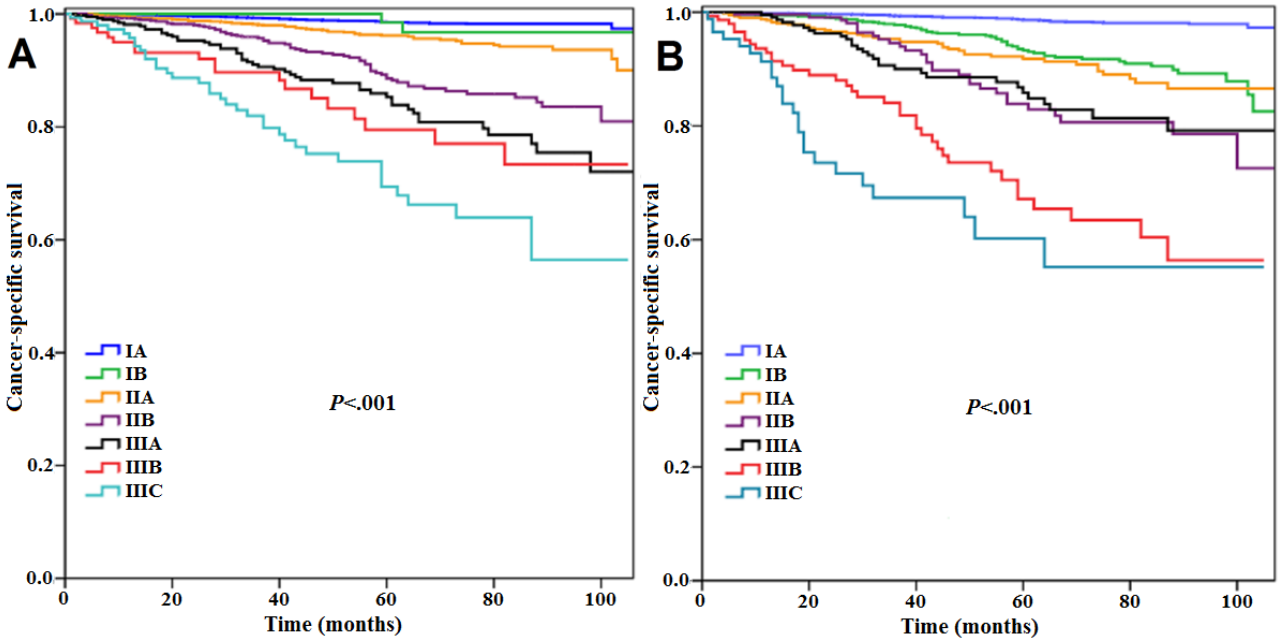


Table 4. Cox multivariate analysis for cancer-specific survival according to race.

Variables	Cancer-specific survival	
	Hazard ratio (95% CI)	*P* value
White American		
Seventh edition of the American Joint Committee on Cancer (AJCC) anatomic staging (AS)	1.358 (1.325-1.391)	<.001
Eighth edition of the AJCC pathological prognostic staging (PPS)	1.244 (1.223-1.265)	<.001
Chinese		
Seventh edition of the AJCC AS	1.493 (1.225-1.818)	<.001
Eighth edition of the AJCC PPS	1.242 (1.082-1.427)	.002

Survival and Prognostic Analyses According to All Substages

According to the seventh AJCC AS, Chinese women had a better 5-year CSS in stage IA (98.6% vs 99.1%, *P*=.01), stage IIA (95.8% vs 97.7%, *P*<.001), and stage IIIB (78.0% vs 86.5%, *P*=.01) disease than White American women, while no significant 5-year CSS was observed in those with stage IB (98.1% vs 99.3%, *P*=.37), IIB (93.1% vs 93.8%, *P*=.69), IIIA (88.9% vs 90.7%, *P*=.92), and IIIC (78.4% vs 79.6%, *P*=.93) disease (Figure 5). Multivariate Cox proportional hazards analysis revealed that race was an independent predictor in stage IA (hazard ratio [HR] 0.679, 95% CI 0.491-0.939, *P*=.02), IIA (HR 0.683, 95% CI 0.524-0.892, *P*=.005), and IIIB (HR 0.616, 95% CI 0.392-0.969, *P*=.04) disease but not in stage IB (*P*=.56), IIB (*P*=.94), IIIA (*P*=.82), and IIIC (*P*=.70) disease (Table 5).

When further stratified by the eighth AJCC PPS, Chinese women had a better 5-year CSS in stage IA (98.7% vs 99.2%, *P*<.001), stage IIA (92.0% vs 94.8%, *P*=.049), and IIIA (85.5% vs 90.8%, *P*=.02) disease than White American women, and the 5-year CSS was not significant in those with stage IB (95.6% vs 96.3%, *P*=.44), IIB (89.4% vs 91.2%, *P*=.42), IIIB (78.2% vs 78.2%, *P*=.89), and IIIC (64.5% vs 73.5%, *P*=.42) disease between White American and Chinese women (Figure 6). Cox multivariate analysis revealed that Chinese women had a better CSS in stage IA (HR 0.673, 95% CI 0.476-0.853, *P*=.002) and IIIA (HR 0.689, 95% CI 0.478-0.994, *P*=.046) disease than White American women, but race was not a prognostic factor in those with stage IB (*P*=.85), IIA (*P*=.27), IIB (*P*=.90), IIIB (*P*=.63), and IIIC (*P*=.25) disease (Table 5).

Figure 5. Survival curves for cancer-specific survival between White American and Chinese women with breast cancer according to the seventh edition of the American Joint Committee on Cancer substages. (A) Stage IA, (B) stage IB, (C) stage IIA, (D) stage IIB, (E) stage IIIA, (F) stage IIIB, and (G) stage IIIC.

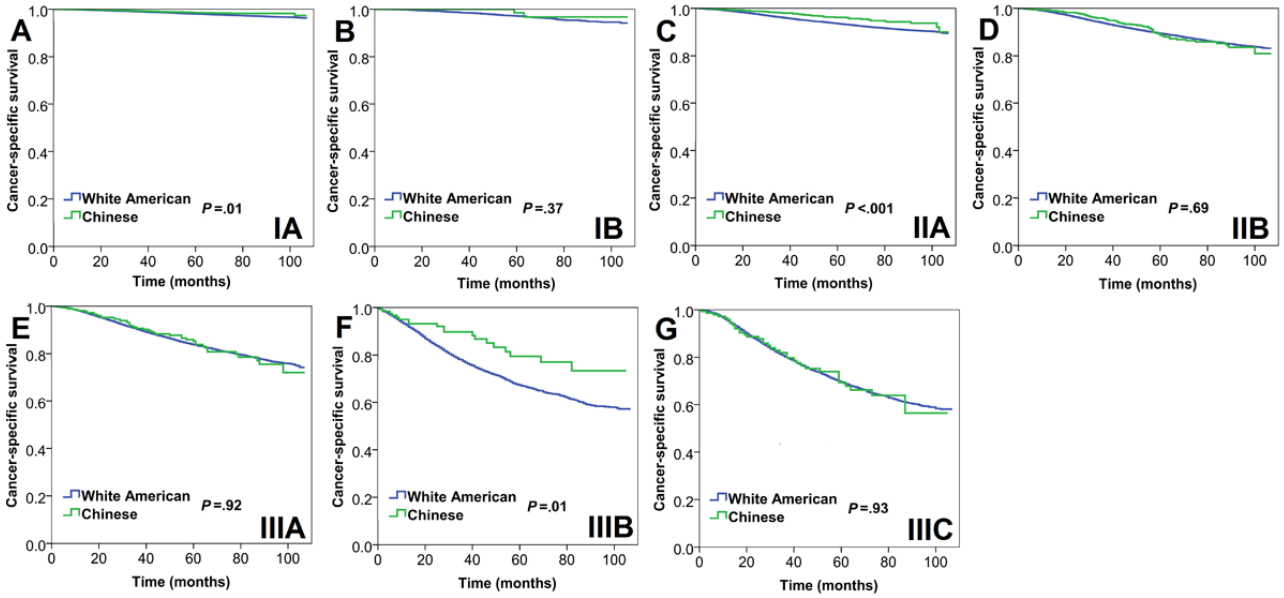
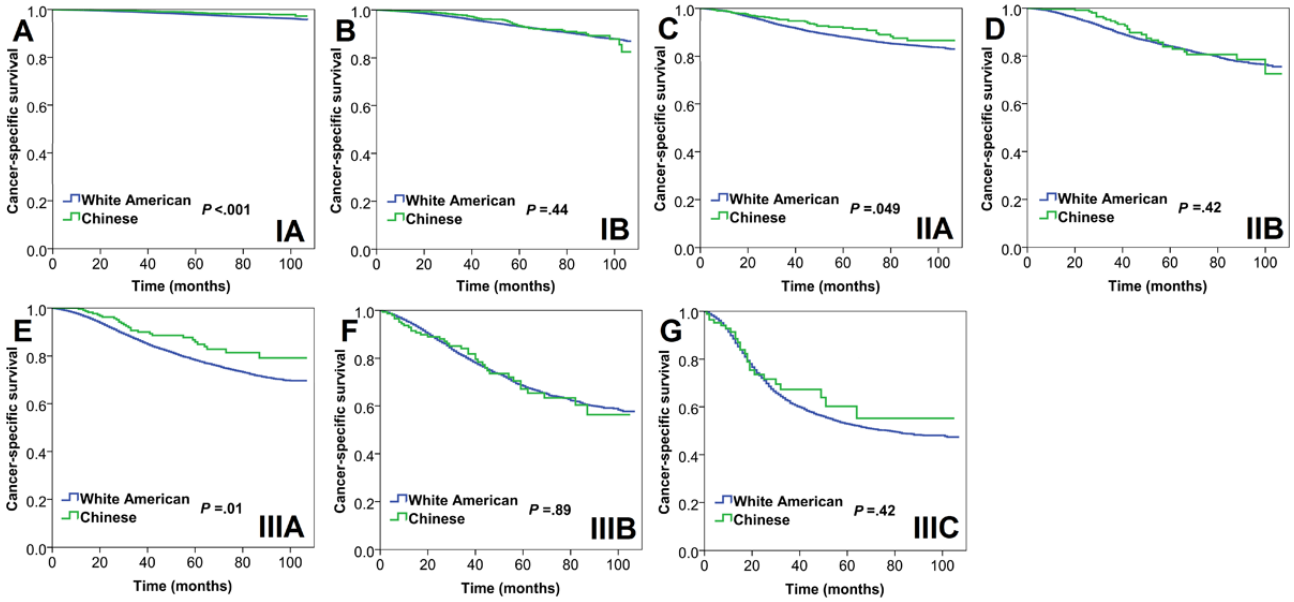


Table 5. Cox multivariate analysis for cancer-specific survival between White American and Chinese women according to the seventh and eighth editions of the American Joint Committee on Cancer (AJCC) substages.

Variables	Seventh edition of the AJCC staging		Eighth edition of the AJCC staging	
	Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
Stage IA				
White American	1 (reference)	N/A[a]	1 (reference)	N/A
Chinese	0.679 (0.491-0.939)	.02	0.637 (0.476-0.853)	.002
Stage IB				
White American	1 (reference)	N/A	1 (reference)	N/A
Chinese	0.659 (0.163-2.663)	.56	0.975 (0.750-1.268)	.85
Stage IIA				
White American	1 (reference)	N/A	1 (reference)	N/A
Chinese	0.683 (0.524-0.892)	.005	0.872 (0.620-1.142)	.27
Stage IIB				
White American	1 (reference)	N/A	1 (reference)	N/A
Chinese	1.010 (0.790-1.292)	.94	0.976 (0.675-1.410)	.90
Stage IIIA				
White American	1 (reference)	N/A	1 (reference)	N/A
Chinese	1.034 (0.780-1.372)	.82	0.689 (0.478-0.994)	.046
Stage IIIB				
White American	1 (reference)	N/A	1 (reference)	N/A
Chinese	0.616 (0.392-0.969)	.04	1.083 (0.781-1.501)	.63
Stage IIIC				
White American	1 (reference)	N/A	1 (reference)	N/A
Chinese	1.066 (0.769-1.477)	.70	0.790 (0.528-1.183)	.25

[a]N/A: not applicable.

Figure 6. Survival curves for cancer-specific survival between White American and Chinese women with breast cancer according to the eighth edition of the American Joint Committee on Cancer substages. (A) Stage IA, (B) stage IB, (C) stage IIA, (D) stage IIB, (E) stage IIIA, (F) stage IIIB, and (G) stage IIIC.



Discussion

Principal Findings

This study aimed to evaluate stage-specific survival in BC between Chinese and White American women in accordance with the eighth AJCC PPS. Our results show that the eighth AJCC PPS had a similar discriminating ability between White American and Chinese participants with BC compared with the seventh AJCC AS. Our study provides additional data on the use of new PPS in different races based on current real-world practices.

Advances in molecular biomarkers (ER, PR, and HER2) and their close relationship with treatment responses and prognosis rendered the traditional AS unable to meet the trend of individualized treatment [6,7,17]. The eighth AJCC PPS, which integrates the aforementioned biomarkers and grade, facilitates more precise prognosis prediction than the seventh AJCC AS [8-10]. However, the small sample size and treatment heterogeneity in their studies limited the application of the eighth AJCC PPS in BC. In our real-world study with a large sample size (n=376,818), the eighth AJCC PPS revealed better prognostic accuracy than the seventh AJCC AS (P<.001) and performed well with discriminating ability consistent with disease stages. Therefore, the new AJCC staging system could better predict prognosis and guide the treatment of BC.

Our study shows that 149,452 (39.7%) individuals with BC migrated from the AJCC AS to AJCC PPS, which was similar to the rates observed in previous studies (20.7%-52.8%) [10,18-22]. The downstaging rate was significantly higher than upstaging rate (33.7% vs 6.0%) in this study, and the results are consistent with those of previous studies (downstaging: 15.2%-42.1%; upstaging: 5.5%-41.0%) [10,18-22]. Change in stage leads to diverse therapeutic decisions. The new AJCC staging enabled 126,936 (33.7%) participants to be downstaged, and these patients might be exempt from the therapies, such as chemotherapy and radiotherapy, which could ensure efficacy and reduce the treatment burden of patients [23]. In our previous studies, we found that the new AJCC staging can accurately guide individualized treatment of patients with BC in clinical decision-making. Patients who were downstaged from the eighth AJCC PPS can safely avoid adjuvant chemotherapy or radiotherapy [24-26]. Therefore, the 8th AJCC staging better reflects the trend of personalized treatment. In addition, among the patients with stage changes in this study, White American participants had a higher upstaging rate than Chinese participants (6.0% vs 5.3%), while the latter had a higher downstaging rate than the former (35.2% vs 33.6%). The Will Rogers phenomenon might explain the differences in “stage migration” in individuals with cancer who are of different races. Differences in culture, education, and diet lead to differences in migration rates between different ethnic groups [27-29]. In addition, socioeconomic status might be another critical factor affecting stage distribution and survival outcomes in different races. A study by Kantor et al [30] included 259,852 individuals with BC who are of different races and reported that non-Hispanic Black individuals and those of lower socioeconomic status had a lower disease-specific survival, even in all substages of the PPS [30].

The initial model for establishing the new AJCC PPS in BC was based on 305,519 patients from National Cancer Database between 2010 and 2012 [11]. However, the majority of the participants were White American and Hispanic American [11]. Therefore, the applicability of the eighth edition of the AJCC staging in Asian individuals, especially in Chinese individuals, remains unclear. Several retrospective studies explored the value of new AJCC staging in Chinese individuals with BC. However, only partial subgroups of BC, such as T1-2N1 and TNBC were included [13-15,31]. The cohort study conducted by He et al [15] recruited patients with TNBC from Sun Yat-sen University Cancer Center (n=611) and the SEER database (n=31,941) to examine the prognostic value of the eighth AJCC PPS in

comparison with the seventh AJCC AS. However, no significant discriminatory ability was observed between the 2 staging systems in Chinese individuals with BC in this study and patients from the SEER database [15]. The opposite result was obtained in another study conducted by Yang et al [31], which included 1556 Chinese individuals with BC and compared the prognostic value of the 2 staging systems. They found that the new AJCC PPS had better accuracy of prognosis prediction than AS in Chinese individuals with BC [28]. However, their sample size was relatively small, especially in the eighth AJCC PPS of stage IIB (n=83) and IIIC (n=22). Therefore, their result might not accurately reflect the value of new staging [31]. In addition, most of the studies assessing the effect of the eighth AJCC PPS in the Chinese population lacked a comparison with the standard population, and their applicability may not be adequate [13-15,31]. In our study, we used a much larger sample to evaluate the new AJCC staging, and we observed a better discriminating value than that of the AJCC AS regardless of race. Therefore, our study better verified the applicability of the new staging in Chinese individuals with BC.

In a previous SEER study, Lim et al [4] reported that Chinese women with BC in the United States have better CSS than White American women, and the largest survival differences between Chinese and White American women were observed for stage I and node-negative cancers [4]. In this study, using the new AJCC staging, we found that Chinese women had superior CSS among those with stage IA and stage IIIA disease compared to White American women. The main reasons for this difference are not clear. The differences in treatment compliance and inherent genetic predisposition may lead to differences in survival between the 2 ethnic groups [32-35].

Limitations

There are several limitations to be acknowledged in this study. First, we extracted the patient data from the SEER database, and selection biases inherently existing in retrospective studies should not be disregarded. Second, although the sample size of the group of Chinese individuals was much larger than that in previous studies, the number of individuals in some substages, such as stage IIIC (n=87), was still small. Therefore, the value of new AJCC staging in Chinese individuals with BC should be further explored. Third, details of treatment were not collected in the SEER database, including radiotherapy (technique, target volume, and radiation dose), chemotherapy regimens, endocrine therapy (regimen and duration), and targeted therapy, which may potentially affect the final analysis. Even with these limitations, our study reflects real-world practices and extends the applicability of the new staging.

Conclusions

In conclusion, our study suggests that the eighth AJCC PPS has a similar discriminating ability in White American and Chinese individuals with BC than the AJCC AS. Therefore, the new staging is also applicable to Chinese individuals with BC. Further studies are needed to explore the value of the PPS in Chinese individuals with BC.

Acknowledgments

We acknowledge the efforts of the SEER Program's tumor registries in the creation of the SEER database. This work was partly supported by the Commission Young and Middle-aged Talents Training Project of Fujian Health Commission (2019-ZQNB-25 and 2021GGB027) and the Natural Science Foundation of Fujian Province (2020J011240).

Authors' Contributions

SGW and LL conceived the study design. SGW participated in performed data cleaning, and statistical analysis. JW and JZ wrote the original draft. SGW and LL were responsible for resources and supervision. SGW and LL edited the manuscript. Authors SGW and LL are cocorresponding authors for this paper.

Conflicts of Interest

None declared.

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. DeSantis C, Ma J, Bryan L, Jemal A. Breast cancer statistics, 2013. CA Cancer J Clin 2014 Oct 01;64(1):52-62 [FREE Full text] [doi: 10.3322/caac.21203] [Medline: 24114568]
3. Li H, Zheng RS, Zhang SW, Zeng HM, Sun KX, Xia CF, et al. [Incidence and mortality of female breast cancer in China, 2014]. Zhonghua Zhong Liu Za Zhi 2018 Mar 23;40(3):166-171. [doi: 10.3760/cma.j.issn.0253-3766.2018.03.002] [Medline: 29575833]
4. Lim DW, Giannakeas V, Narod SA. Survival differences in Chinese versus White women with breast cancer in the United States: a SEER-based analysis. JCO Global Oncology 2020 Nov(6):1582-1592. [doi: 10.1200/go.20.00316]
5. Ben Ramadan AA, Jackson-Thompson J, Schmaltz CL. Improving visualization of female breast cancer survival estimates: analysis using interactive mapping reports. JMIR Public Health Surveill 2018 May 03;4(2):e42 [FREE Full text] [doi: 10.2196/publichealth.8163] [Medline: 29724710]

6. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol 2010 Jun 24;17(6):1471-1474. [doi: 10.1245/s10434-010-0985-4] [Medline: 20180029]
7. Cserni G, Chmielik E, Cserni B, Tot T. The new TNM-based staging of breast cancer. Virchows Arch 2018 May;472(5):697-703. [doi: 10.1007/s00428-018-2301-9] [Medline: 29380126]
8. Bagaria SP, Ray PS, Sim M, Ye X, Shamonki JM, Cui X, et al. Personalizing breast cancer staging by the inclusion of ER, PR, and HER2. JAMA Surg 2014 Feb 01;149(2):125-129. [doi: 10.1001/jamasurg.2013.3181] [Medline: 24306257]
9. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin 2017 Mar 17;67(2):93-99 [FREE Full text] [doi: 10.3322/caac.21388] [Medline: 28094848]
10. Weiss A, Chavez-MacGregor M, Lichtensztajn DY, Yi M, Tadros A, Hortobagyi GN, et al. Validation study of the American Joint Committee on Cancer eighth edition prognostic stage compared with the anatomic stage in breast cancer. JAMA Oncol 2018 Feb 01;4(2):203-209 [FREE Full text] [doi: 10.1001/jamaoncol.2017.4298] [Medline: 29222540]
11. Amin MB, Edge SB, Greene FL, Byrd DR, Brookland RK, Washington MK, et al. AJCC Cancer Staging Manual. New York, NY: Springer International Publishing; 2018.
12. Lei S, Zheng R, Zhang S, Wang S, Chen R, Sun K, et al. Global patterns of breast cancer incidence and mortality: a population-based cancer registry data analysis from 2000 to 2020. Cancer Commun (Lond) 2021 Nov 16;41(11):1183-1194 [FREE Full text] [doi: 10.1002/cac2.12207] [Medline: 34399040]
13. Sun G, Wang S, Tang Y, Yang Y, Fang H, Wang J, et al. [The 8th edition of the American Joint Committee on Cancer staging system provide improved prognostic accuracy in T1-2N1M0 postmastectomy breast cancer patients]. Zhonghua Zhong Liu Za Zhi 2019 Aug 23;41(8):615-623. [doi: 10.3760/cma.j.issn.0253-3766.2019.08.011] [Medline: 31434454]
14. Li JP, Zhang XM, Zhang YS, Zheng LH, Liu YJ. The prognostic value of the 8th edition of the American Joint Committee on Cancer (AJCC) staging system in triple-negative breast cancer. Neoplasma 2019 Sep;66(5):810-817. [doi: 10.4149/neo_2019_190107N26] [Medline: 31129969]
15. He J, Tsang JY, Xu X, Li J, Li M, Chao X, et al. AJCC 8th edition prognostic staging provides no better discriminatory ability in prognosis than anatomical staging in triple negative breast cancer. BMC Cancer 2020 Jan 06;20(1):18 [FREE Full text] [doi: 10.1186/s12885-019-6494-3] [Medline: 31906874]
16. SEER is an authoritative source for cancer statistics in the United States. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. URL: http://www.seer.cancer.gov/ [accessed 2021-12-01]
17. Edge SB, Hortobagyi GN, Giuliano AE. New and important changes in breast cancer TNM: incorporation of biologic factors into staging. Expert Rev Anticancer Ther 2019 Apr 26;19(4):309-318. [doi: 10.1080/14737140.2019.1582335] [Medline: 30759347]
18. Luo S, Wu Q, Chen H, Wang X, Chen Q, Zhang J, et al. Validation of the prognostic significance of the prognostic stage group according to the eighth edition of American Cancer Joint Committee on Cancer Staging System in Triple-Negative Breast Cancer: an analysis from surveillance, epidemiology, and end results 18 database. J Surg Res 2020 Mar;247:211-219 [FREE Full text] [doi: 10.1016/j.jss.2019.09.072] [Medline: 31706539]
19. Jang N, Choi JE, Kang SH, Bae YK. Validation of the pathological prognostic staging system proposed in the revised eighth edition of the AJCC staging manual in different molecular subtypes of breast cancer. Virchows Arch 2019 Feb 24;474(2):193-200. [doi: 10.1007/s00428-018-2495-x] [Medline: 30474738]
20. Zhou J, Lei J, Wang J, Lian C, Hua L, Yang L, et al. Validation of the 8 edition of the American Joint Committee on Cancer Pathological Prognostic Staging for young breast cancer patients. Aging (Albany NY) 2020 Apr 22;12(8):7549-7560 [FREE Full text] [doi: 10.18632/aging.103111] [Medline: 32320950]
21. Wu S, Shi J, Zhang W, Wang J, Lian C, Lei J, et al. Prognostic validation and treatment decision making of the 8 edition of the American Joint Committee on Cancer pathological staging system for elderly women with early-stage breast cancer. Aging (Albany NY) 2020 Jul 25;12(14):15077-15090 [FREE Full text] [doi: 10.18632/aging.103574] [Medline: 32710731]
22. Kantor O, Bao J, Jaskowiak N, Yao K, Tseng J. The prognostic value of the AJCC 8th edition staging system for patients undergoing neoadjuvant chemotherapy for breast cancer. Ann Surg Oncol 2020 Feb 2;27(2):352-358. [doi: 10.1245/s10434-019-07636-w] [Medline: 31376037]
23. Plichta J, Ren Y, Thomas S, Greenup R, Fayanju O, Rosenberger L, et al. Implications for breast cancer restaging based on the 8th edition AJCC staging manual. Ann Surg 2020 Jan;271(1):169-176 [FREE Full text] [doi: 10.1097/SLA.0000000000003071] [Medline: 30312199]
24. Nittala MR, Mundra EK, Packianathan S, Mehta D, Smith ML, Woods WC, et al. The Will Rogers phenomenon, breast cancer and race. BMC Cancer 2021 May 17;21(1):554 [FREE Full text] [doi: 10.1186/s12885-021-08125-8] [Medline: 34001038]
25. Wu S, Wang J, Lei J, Lian C, Hua L, Zhou J, et al. Prognostic validation and therapeutic decision-making of the AJCC eighth pathological prognostic staging for T3N0 breast cancer after mastectomy. Clin Transl Med 2020 Jan 23;10(1):125-136 [FREE Full text] [doi: 10.1002/ctm2.3] [Medline: 32508053]
26. Lian C, Li G, Zhou P, Wang J, He Z, Wu S. Triple-negative breast cancer outcomes: Does AJCC 8th staging improve chemotherapy decision-making. Breast 2021 Oct;59:117-123 [FREE Full text] [doi: 10.1016/j.breast.2021.06.009] [Medline: 34229126]

27. Yang S, Zhou P, Lian C, He Z, Wu S. The predictive effect of the 8th AJCC pathological prognostic staging on the benefit of postmastectomy radiotherapy in N2/N3 breast cancer. BCTT 2022 May;Volume 14:133-144. [doi: 10.2147/bctt.s362355]
28. Song C, Hu Z, Wu J, Luo J, Shen Z, Huang W, et al. The prevalence of BRCA1 and BRCA2 mutations in eastern Chinese women with breast cancer. J Cancer Res Clin Oncol 2006 Oct 12;132(10):617-626. [doi: 10.1007/s00432-006-0105-9] [Medline: 16835750]
29. Warner ET, Tamimi RM, Hughes ME, Ottesen RA, Wong Y, Edge SB, et al. Racial and ethnic differences in breast cancer survival: mediating effect of tumor characteristics and sociodemographic and treatment factors. JCO 2015 Jul 10;33(20):2254-2261. [doi: 10.1200/jco.2014.57.1349]
30. Kantor O, Wang ML, Bertrand K, Pierce L, Freedman RA, Chavez-MacGregor M, et al. Racial and socioeconomic disparities in breast cancer outcomes within the AJCC pathologic prognostic staging system. Ann Surg Oncol 2022 Jan 30;29(1):686-696. [doi: 10.1245/s10434-021-10527-8] [Medline: 34331158]
31. Yang P, Peng Q, Lian W, Fu F, Wang C, Chen D. Validation of the eighth American Joint Committee on Cancer Anatomic and Prognostic Staging System for Breast Cancer. J Surg Res 2022 Feb;270:539-546 [FREE Full text] [doi: 10.1016/j.jss.2021.09.025] [Medline: 34808473]
32. Zhang G, Wang Y, Chen B, Guo L, Cao L, Ren C, et al. Characterization of frequently mutated cancer genes in Chinese breast tumors: a comparison of Chinese and TCGA cohorts. Ann Transl Med 2019 Apr;7(8):179-179 [FREE Full text] [doi: 10.21037/atm.2019.04.23] [Medline: 31168460]
33. Felder TM, Heiney SP, Hebert JR, Friedman DB, Elk R, Franco R, et al. Improving adherence to adjuvant hormonal therapy among disadvantaged women diagnosed with breast cancer in South Carolina: proposal for a multimethod study. JMIR Res Protoc 2020 Sep 03;9(9):e17742 [FREE Full text] [doi: 10.2196/17742] [Medline: 32880374]
34. Donevant S, Heiney SP, Wineglass C, Schooley B, Singh A, Sheng J. Perceptions of endocrine therapy in African-American breast cancer survivors: mixed methods study. JMIR Form Res 2021 Jun 11;5(6):e23884 [FREE Full text] [doi: 10.2196/23884] [Medline: 34114955]
35. Yu J, Wu J, Huang O, Chen X, Shen K. A smartphone-based app to improve adjuvant treatment adherence to multidisciplinary decisions in patients with early-stage breast cancer: observational study. J Med Internet Res 2021 Sep 16;23(9):e27576 [FREE Full text] [doi: 10.2196/27576] [Medline: 34528890]

Abbreviations

AJCC: American Joint Committee on Cancer
AS: anatomic staging
AUC: area under the curve
BC: breast cancer
CSS: cancer-specific survival
ER: estrogen receptor
HR: hazard ratio
M: distant metastasis
N: lymph nodes
PPS: pathological prognostic staging
PR: progesterone receptor
SEER: Surveillance, Epidemiology, and End Results
T: tumor size
TNBC: triple-negative breast cancer

Edited by A Mavragani, G Eysenbach; submitted 18.06.22; peer-reviewed by A Qing, CF Cai; comments to author 09.09.22; revised version received 29.09.22; accepted 18.10.22; published 15.11.22.

Please cite as:
Wang J, Zhou J, Liu L, Wu SG
Stage-Specific Survival in Breast Cancer in Chinese and White Women: Comparative Data Analysis
JMIR Public Health Surveill 2022;8(11):e40386
URL: https://publichealth.jmir.org/2022/11/e40386
doi:10.2196/40386
PMID:36378507

Original Paper

Modeling the Potential Impact of Missing Race and Ethnicity Data in Infectious Disease Surveillance Systems on Disparity Measures: Scenario Analysis of Different Imputation Strategies

Bahareh Ansari[1,2], MA, PhD; Rachel Hart-Malloy[2,3,4], MPH, PhD; Eli S Rosenberg[2,3,4], PhD; Monica Trigg[5], MPH; Erika G Martin[2,6], MPH, PhD

[1]Center for Policy Research, Rockefeller College of Public Affairs and Policy, University at Albany, Albany, NY, United States

[2]Center for Collaborative HIV Research in Practice and Policy, School of Public Health, University at Albany, Albany, NY, United States

[3]New York State Department of Health, Albany, NY, United States

[4]Department of Epidemiology and Biostatistics, School of Public Health, University at Albany, Albany, NY, United States

[5]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[6]Department of Public Administration and Policy, Rockefeller College of Public Affairs and Policy, University at Albany, Albany, NY, United States

Corresponding Author:
Erika G Martin, MPH, PhD
Department of Public Administration and Policy
Rockefeller College of Public Affairs and Policy
University at Albany
300 Milne Hall
135 Western Ave
Albany, NY, 12203
United States
Phone: 1 518 442 5243
Email: emartin@albany.edu

Abstract

Background: Monitoring progress toward population health equity goals requires developing robust disparity indicators. However, surveillance data gaps that result in undercounting racial and ethnic minority groups might influence the observed disparity measures.

Objective: This study aimed to assess the impact of missing race and ethnicity data in surveillance systems on disparity measures.

Methods: We explored variations in missing race and ethnicity information in reported annual chlamydia and gonorrhea diagnoses in the United States from 2007 to 2018 by state, year, reported sex, and infection. For diagnoses with incomplete demographic information in 2018, we estimated disparity measures (relative rate ratio and rate difference) with 5 imputation scenarios compared with the base case (no adjustments). The 5 scenarios used the racial and ethnic distribution of chlamydia or gonorrhea diagnoses in the same state, chlamydia or gonorrhea diagnoses in neighboring states, chlamydia or gonorrhea diagnoses within the geographic region, HIV diagnoses, and syphilis diagnoses.

Results: In 2018, a total of 31.93% (560,551/1,755,510) of chlamydia and 22.11% (128,790/582,475) of gonorrhea diagnoses had missing race and ethnicity information. Missingness differed by infection type but not by reported sex. Missing race and ethnicity information varied widely across states and times (range across state-years: from 0.0% to 96.2%). The rate ratio remained similar in the imputation scenarios, although the rate difference differed nationally and in some states.

Conclusions: We found that missing race and ethnicity information affects measured disparities, which is important to consider when interpreting disparity metrics. Addressing missing information in surveillance systems requires system-level solutions, such as collecting more complete laboratory data, improving the linkage of data systems, and designing more efficient data collection procedures. As a short-term solution, local public health agencies can adapt these imputation scenarios to their aggregate data to adjust surveillance data for use in population indicators of health equity.

(JMIR Public Health Surveill 2022;8(11):e38037) doi:10.2196/38037

KEYWORDS

missing data; sexually transmitted diseases; imputation; surveillance; health equity

Introduction

Background

Infectious disease surveillance systems are important information technologies used to identify outbreaks of infectious diseases, describe the current burden of the diseases, and monitor trends and disparities among populations [1]. However, many surveillance systems have data quality issues [2-4] that must be understood for the correct interpretation of data. Although informatics solutions exist for dealing with data quality issues in surveillance systems [3,5], the optimal solution for a specific surveillance system requires a deeper understanding of the contributing factors and the consequences of data quality issues in interpreting surveillance data. In this study, we focused on missing race and ethnicity information in surveillance systems and explored the effect of missingness on the calculated disparity measures to guide future informatics solutions.

We focused on health equity because racial and ethnic minority populations in the United States continue to experience a disproportionately high burden of poor health outcomes. These disparities can be attributed to persistent systemic racism against African American people in health care settings and medical research throughout the US history [6] and a range of social and structural factors such as residential segregation, lower opportunities for education, unemployment, and lower income [7]. Robust measures of population health using high-quality data are needed for a complete understanding of disparities in health outcomes [8]. Moreover, the data should be representative of the population without coverage bias. A systematic undercounting of communities of color in surveillance data [9], one type of coverage bias, is an example of systematic racism built into government databases, which may skew public health decision-making.

Public health surveillance systems are critical sources of information for measuring and monitoring disparities and evaluating public health initiatives to improve equity [10]. However, incomplete information on race and ethnicity may affect disparity measures. Missing race and ethnicity information has been a major limitation in different health care databases, such as birth certificate records in a large US health care system [11], Veterans' health administration records [12], reported COVID-19 cases, and persons who received COVID-19 vaccinations in the United States [13]. A previous study found that incomplete race and ethnicity information in COVID-19 data resulted in an underrepresentation of disparities among racial and ethnic population groups [9]. The use of biased disparity measures in policy and funding decisions can perpetuate the legacy of systemic racism.

Objectives

We examined missing race and ethnicity information in chlamydia and gonorrhea surveillance data from 2007 to 2018 and used 5 imputation strategies to explore how missing demographic information could have impacted our measurement of racial and ethnic disparities. We chose chlamydia and gonorrhea for our exploration because they are among the most common notifiable conditions in the United States [14] and had an estimated total lifetime cost of US $1.0 billion in 2018 [15], and it is well established that non-Hispanic Black and Hispanic populations have persistently higher rates of diagnosed sexually transmitted infections (STIs) than White populations [16]. Our findings highlight the importance of understanding and addressing missing demographic data in surveillance systems to reduce systematic biases in the measures of racial and ethnic disparities.

Methods

Study Population and Data Sources

We conducted 2 sets of analyses: (1) a descriptive trend analysis to investigate the extent of missing race and ethnicity information across the 2 infections by reported sex (hereafter, sex) and year and (2) a scenario analysis to assess how the rate ratios (RRs) and rate differences (RDs) changed under different methods to redistribute diagnoses with incomplete demographic data to specific racial and ethnic groups. The study population differed in the descriptive trend analysis and the scenario analysis. For the descriptive trend analysis, we used aggregated state-level counts of all reported chlamydia and gonorrhea cases among male and female patients aged ≥15 years for 50 states and the District of Columbia from 2007 to 2018 (n=612 state-year observations for each infection in male or female patients). For the scenario analysis, we restricted the analysis to 2018 (n=51 state-level observations).

The counts of chlamydia and gonorrhea diagnoses were obtained from the Centers for Disease Control and Prevention's (CDC) National Center for HIV Viral Hepatitis, STD, and TB Prevention AtlasPlus [17]. The underlying data are from the National Notifiable Disease Surveillance System, a complex surveillance system that is a collaboration among numerous local, state, and federal partners. Gonorrhea and chlamydia are reportable and nationally notifiable conditions. As such, states and territories have set requirements for laboratories and medical providers to report case information to public health departments. In turn, states voluntarily transmit case report data to the CDC, which secures and processes deidentified data that are then provided to disease-specific programs across the CDC [18,19]. This process is complex for several reasons. First, jurisdictions use various surveillance information systems [20]. Adding to the complexity of data collection is that not all newly identified cases are contacted by disease intervention specialists; jurisdictions follow state and federal guidelines regarding which STIs to prioritize for partner services. Chlamydia and gonorrhea cases generally receive a lower priority for follow-up than HIV and syphilis cases [21], which may lead to missing demographic and other information if the surveillance record is based exclusively on laboratory data that are automatically sent to the public health authority without an accompanying case report from the provider.

To establish rates and disparity measures, we used the 5-year American Community Survey 2018 [22] to determine the population in the United States by state, sex, and race and ethnicity. We limited our analysis to non-Hispanic Black, Hispanic, and non-Hispanic White persons because other racial and ethnic groups, including persons with multiple races, had small numbers of reported cases (59,687/1,755,510, 3.4% and 22,134/582,475, 3.8% of the total reported cases for chlamydia and gonorrhea during 2018, respectively). Although these other racial and ethnic groups are important, their small counts impeded our ability to produce stable rates and disparity measures. Male versus female sex was defined as a binary variable, which might represent sex at birth or current identity, as current gender identity is not systematically recorded in the surveillance data.

Ethical Considerations

The data used in this study were publicly available for direct download from the CDC in an aggregate and anonymized format without use restrictions (ie, number of cases per state by stratum). The granularity of the strata renders it impossible to reidentify the respondents. We did not need to seek a review from our Human Subjects Committee because the nature of the data and the research question were not considered human subjects research by University at Albany policy guidance.

Statistical Methods

The statistical methods had 4 parts. First, we conducted a descriptive trend analysis of the percentage of diagnoses with unknown race and ethnicity information for chlamydia and gonorrhea in male and female patients in each state from 2007 to 2018. This analysis produced descriptive statistics to explore variations by state, year, sex, and infection, and the Cochran-Armitage test [23] was used to explore the trends of chlamydia and gonorrhea among male and female patients. Second, we calculated the rates and 2 disparity measures based on the available demographic information. To measure racial and ethnic disparities, we chose both RR and RD, following best practices for reporting disparities using multiple measures [21]. Third, we redistributed the diagnoses with unknown race and ethnicity information in 2018 using 5 imputation scenarios. Fourth, we compared the disparity measures under different scenarios with the base case (disparity measures calculated using only available data and no adjustment for missing data) to evaluate the potential impact of missing data. Weights were not applied in the analyses because AtlasPlus provides the total number of known reported cases (ie, the full population) rather than a sample of cases.

Table 1 summarizes the 5 scenarios. The first scenario (scenario 1) was redistributed according to the reported chlamydia and gonorrhea diagnoses with known race and ethnicity information in the same state. We used 2 other methodologies that used available demographic data for chlamydia and gonorrhea diagnoses and redistributed diagnoses with unknown race and ethnicity to population groups based on known diagnoses in neighboring states (scenario 2) or the same region (scenario 3). Neighboring and regional data have been used in previous studies to impute aggregate-level spatial data [24]. Our fourth and fifth scenarios were based on available demographic information from HIV and syphilis in the same state in the same year (2018). These are 2 other common STIs with more complete racial and ethnic information because people with newly reported diagnoses of HIV and syphilis are prioritized for follow-up by disease intervention specialists as part of partner services programs for HIV and STI [25].

For the fourth and fifth scenarios, the racial and ethnic distributions of all 4 infections were not identical. For example, the number of chlamydia and gonorrhea diagnoses is larger among female patients than male patients, whereas the number of HIV and syphilis cases is larger among male than among female patients. However, HIV and syphilis surveillance data are commonly under the purview of surveillance staff and are likely to be accessible to data analysts who calculate disparity measures. Therefore, we added these scenarios as alternative methods for considering the impact of missing race and ethnicity information.

To measure disparities, we used both an absolute measure (RD) and relative measure (RR). Finally, we created visualizations to compare disparity measures produced in each scenario to the base case in which diagnoses with missing race and ethnicity information were excluded from the calculations.

Table 1. Summary of imputation scenarios to assign race and ethnicity to reported diagnoses with incomplete demographic information.

Scenario[a]	Description
Base case	No adjustments. This scenario includes reported diagnoses with available race and ethnicity information in National Center for HIV, Viral Hepatitis, STD, and TB Prevention (NCHHSTP) AtlasPlus. Counts of diagnoses with missing race and ethnicity information are omitted from analysis.
Scenario 1	Reallocation based on reported chlamydia and gonorrhea diagnoses with known race and ethnicity from the same state. Within a state, the diagnoses with missing race and ethnicity information are reapportioned to a racial and ethnic group based on their distribution among known diagnoses. For example, if 50% of diagnoses have missing race and ethnicity information and among the remaining diagnoses, 40%, 20%, and 40% are recorded as Black, Hispanic, or White race and ethnicity, then the unknown diagnoses will be reassigned following the 40%-20%-40% distribution. This will not change the *distribution* of cases in terms of the percentage in each racial and ethnic group, but it does increase the *number* of diagnoses within each group.
Scenario 2	Reallocation based on reported chlamydia and gonorrhea diagnoses with known race and ethnicity from neighboring states. Within a state, the diagnoses with missing race and ethnicity information are reapportioned to a racial and ethnic group based on the distribution of known diagnoses in the states that share a contiguous border. In the case of Alaska and Hawaii, which do not have any neighboring states, this scenario does not adjust the rate.
Scenario 3	Reallocation based on information from states in the geographic region. Within a state, the diagnoses with missing race and ethnicity information are reapportioned to a racial and ethnic group based on their distribution in all states within the 4-level US Census region (Northeast, Midwest, South, and West).
Scenario 4	Reallocation based on information from HIV diagnoses within a state. Within a state, diagnoses with missing race and ethnicity information are reapportioned to a racial and ethnic group based on the distribution of HIV diagnoses, which do not have missing race and ethnicity data in NCHHSTP AtlasPlus.
Scenario 5	Reallocation based on information from syphilis diagnoses within a state. Within a state, diagnoses with missing race and ethnicity information are reapportioned to a racial and ethnic group based on the distribution of syphilis diagnoses, which do not have missing race and ethnicity data in NCHHSTP AtlasPlus.

[a]In 2018, the number of chlamydia and gonorrhea diagnoses stratified by race and ethnicity was not available for Connecticut, and this state was excluded from all scenarios. In addition, the number of HIV diagnoses by race and ethnicity is suppressed for New Hampshire in 2018, and the rates for New Hampshire were not adjusted under scenario 4.

Results

Descriptive Trends

Figure 1 shows the annual trend of the percentage of missing race and ethnicity information among reported chlamydia and gonorrhea diagnoses by sex in 50 states and the District of Columbia from 2007 to 2018. The 2 solid lines represent the trends among male and female patients for reported chlamydia diagnoses, and the 2 dashed lines represent the trends among male and female patients for gonorrhea diagnoses. The percentage of missing race and ethnicity information was higher for chlamydia compared with gonorrhea. For each infection, female participants had a higher percentage of missing race and ethnicity data in 2007. The percentage of missing race and ethnicity information in reported gonorrhea diagnoses among female patients decreased over time (Cochran-Armitage 2-sided test for trend: Z=28.71; P<.001), but the corresponding indicator increased for reported gonorrhea diagnoses among male patients (Z=−29.21; P<.001). This resulted in a higher percentage of missing race and ethnicity information in reported gonorrhea diagnoses among male patients than among female patients in 2018. The percentage of missing race and ethnicity information in reported chlamydia diagnoses increased among both male patients (Z=−127.97; P<.001) and female patients (Z=−74.08; P<.001). However, the increasing trend was sharper for male patients, which resulted in closing the gap between the percentage of missing race and ethnicity information in reported chlamydia diagnoses among male and female patients in 2018.

Table 2 presents summary statistics of the percentage of missing race and ethnicity information among reported chlamydia and gonorrhea diagnoses for male and female patients in 50 states and the District of Columbia from 2007 to 2018. The results are stratified by female and male patients. For each year, the percentages reflect summary statistics of missingness across the 50 states and the District of Columbia. Overall, the reported chlamydia diagnoses had a higher frequency of missing race and ethnicity information than gonorrhea diagnoses, but differences in missingness between male and female patients were not remarkable. There was no clear trend when examining the median values of the percentage of missing racial and ethnic information across states. The range of missing data changed across states, with the minimum values remaining near 0% in all years for both infections but the maximum values increasing over time.

Figure 1. Percentage of reported chlamydia and gonorrhea diagnoses with missing race and ethnicity information in the United States (2007-2018). All 50 states and the District of Columbia are included. The national counts were developed by summing all counts from the states.

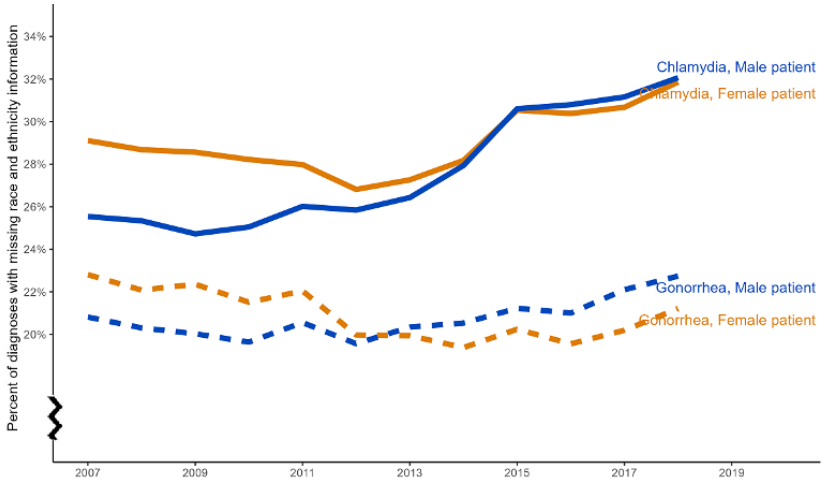


Table 2. Percentage of reported chlamydia and gonorrhea diagnoses with missing race and ethnicity information in 50 states and District of Columbia (2007-2018).

Sex and year	Chlamydia[a]		Gonorrhea[a]	
	Value, median (range; %)	IQR (Q1[b]-Q3[b]; %)	Value, median (range; %)	IQR (Q1[b]-Q3[b]; %)
Female patients				
2007	23.6 (1.2-73.6)	(13.4-36.1)	18.5 (0-59.0)	(9.5-29.6)
2008	25.5 (0-64.1)	(13.7-36.9)	22.6 (0-44.0)	(8.2-30.2)
2009	24.3 (3.1-64.3)	(15.0-35.2)	19.6 (0-45.5)	(11.1-30.4)
2010	29.2 (3.5-60.7)	(17.3-34.7)	21.0 (0-45.0)	(11.7-26.5)
2011	27.5 (0.9-57.0)	(15.7-34.5)	19.7 (0-45.7)	(9.4-27.9)
2012	23.5 (0-59.6)	(14.1-34.2)	15.9 (0-44.1)	(8.3-29.1)
2013	22.8 (0.4-62.6)	(13.8-36.7)	16.5 (0-45.8)	(8.6-27.5)
2014	27.4 (1.9-64.6)	(15.8-37.3)	18.2 (0-61.2)	(10.4-26.3)
2015	29.2 (1.4-88.8)	(14.2-40.6)	17 (0-92.9)	(9.1-26.4)
2016	27.2 (0.1-76.7)	(14.5-37.6)	17.8 (0-70.6)	(8.4-25.6)
2017	26.5 (0.2-92.5)	(14.0-39.3)	17.4 (0.1-91.8)	(7.9-26.1)
2018	26.9 (0.1-96.2)	(15.3-38.1)	16.9 (0.1-94.1)	(9.7-25.8)
Male patients				
2007	19.9 (0.8-65.3)	(12.3-32.8)	17.4 (1.3-62.1)	(8.8-27.5)
2008	20.7 (0-51.3)	(12.4-33.2)	22.4 (0-41.4)	(9.3-27.9)
2009	21.4 (3.9-53.9)	(14.9-32.6)	18.2 (0-43.1)	(11.7-27.0)
2010	24.3 (3.5-50.2)	(17.7-32.4)	21.2 (0-48.7)	(11.9-26.2)
2011	23.9 (0.9-51.0)	(15.2-33.1)	18.2 (0-39.8)	(9.9-25.5)
2012	23.5 (0-55.8)	(13.4-31.0)	16.1 (0-43.4)	(9.5-25.2)
2013	24.6 (0.4-65.7)	(13.4-32.9)	18.6 (0-48.0)	(8.9-25.5)
2014	25.2 (1.7-70.2)	(15.4-33.3)	20.1 (0-67.7)	(9.9-25.4)
2015	26.8 (1.4-88.5)	(14.7-37.5)	17.6 (0.2-92.3)	(9.9-25.7)
2016	27.3 (0.2-77.0)	(14.6-36.0)	18.2 (0.1-70.1)	(9.9-24.9)
2017	25.9 (0.2-89.2)	(15.1-36.8)	17.4 (0.1-86.7)	(9.6-26.8)
2018	29.5 (0.1-92.2)	(16.3-37.0)	17 (0-91.3)	(10.8-25.7)

[a]The observations are the percentage of diagnoses with missing racial and ethnic information in the 50 states and the District of Columbia.

[b]Q1 and Q3 are the first and third quartiles (the 25th and 75th percentiles, respectively).

Scenario Analysis

Figure 2 shows how the absolute Black-White and Hispanic-White disparity measures changed under each imputation scenario for the 2018 data, with the calculated disparity measures presented in Table 3. The numerators comprise all reported diagnoses regardless of the mode of transmission. National counts were developed by summing all counts from the states, except Connecticut, for which the number of chlamydia and gonorrhea diagnoses were not available by race and ethnicity in 2018. The denominator is the population of the United States aged ≥15 years in the 50 states and the District of Columbia, except Connecticut. The dashed line represents the value for the base case. The 4 charts display the RDs for chlamydia (left), gonorrhea (right), female patients (top), and male patients (bottom). The orange bars represent Black-White RDs, and the blue bars represent Hispanic-White RDs. There are 6 bars for each RD to represent the base case and the 5 imputation scenarios. In the base case, the Hispanic-White RDs for both chlamydia and gonorrhea are smaller than the Black-White RDs (chlamydia, RD: 284.1 per 100,000 for female patients and 119.4 per 100,000 for male patients; gonorrhea, RD: 27.5 per 100,000 for female patients and 71.8 per 100,000 for male patients). Under each imputation scenario, the RD disparity measure was higher compared with the base case. For chlamydia, the Black-White RD increased by up to 789.1 per 100,000 among female patients and up to 394.3 per 100,000 among male patients. The Hispanic-White RD increased by up to 210.1 per 100,000 among female patients and up to 168.2 per 100,000 among male patients. For gonorrhea, the Black-White RD increased by up to 114.2 per 100,000 among female patients and up to 182.2 per 100,000

among male patients. The Hispanic-White RD increased by up to 25.9 per 100,000 among female vs and up to 60.5 per 100,000 among male patients.

Figure 2. Estimated Black-White and Hispanic-White rate differences (RDs) for chlamydia and gonorrhea under 5 scenarios to impute race and ethnicity for reported diagnoses with incomplete demographic data (2018).

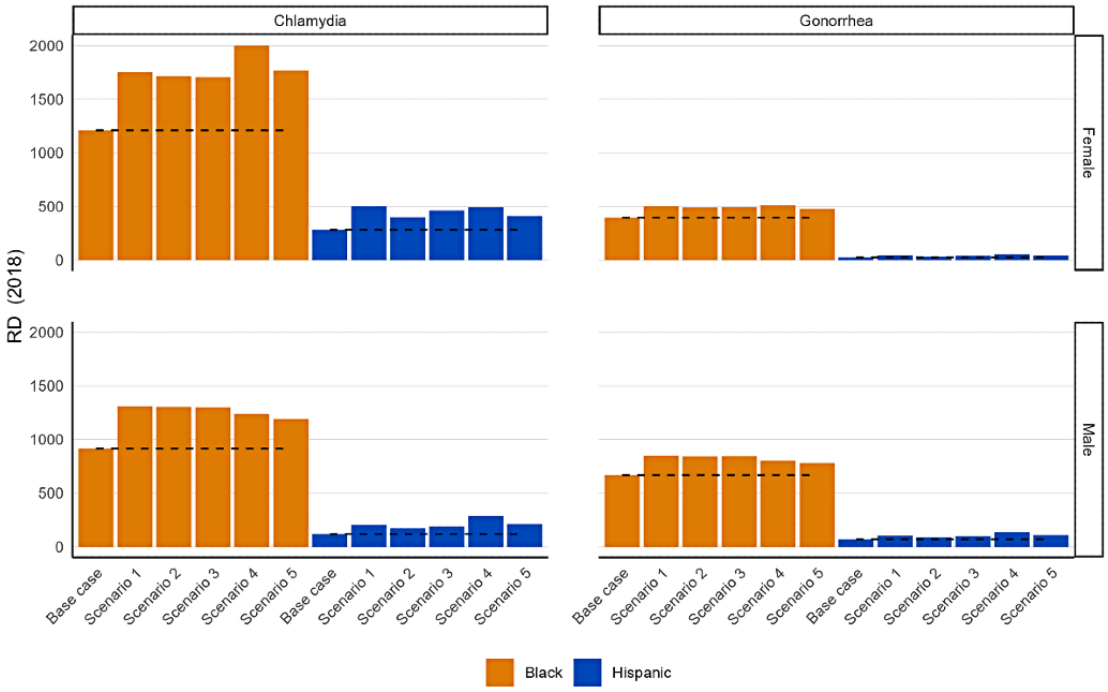


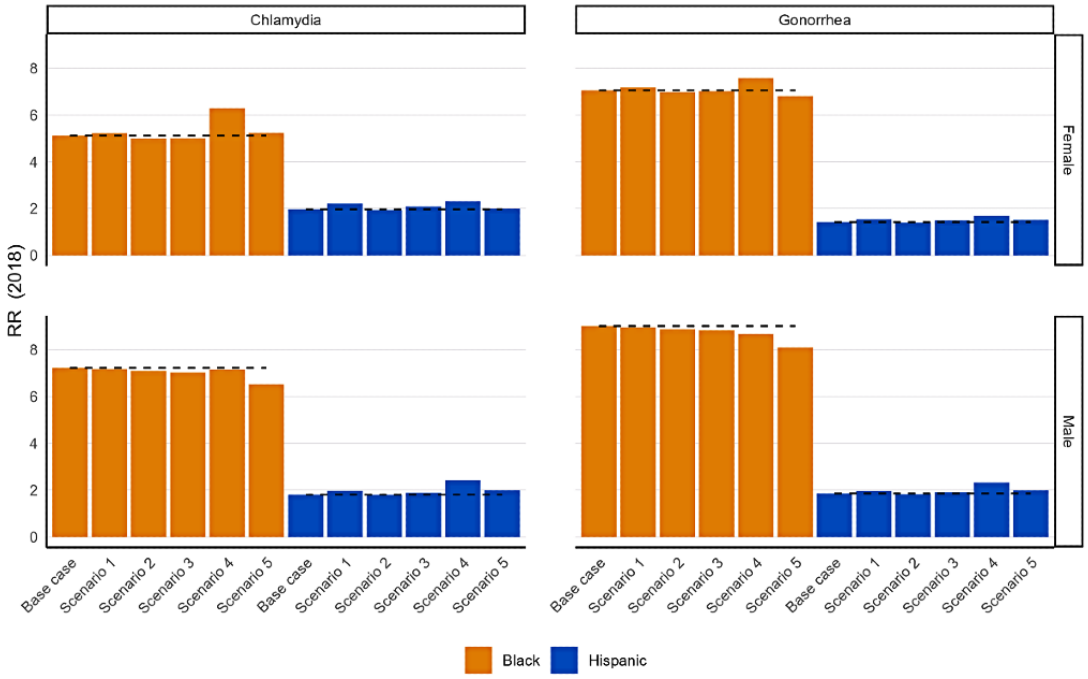
Table 3. Estimated Black-White and Hispanic-White rate differences (RDs) and rate ratios (RRs) for chlamydia and gonorrhea under 5 scenarios to impute race and ethnicity for reported diagnoses with incomplete demographic data (2018).

	Chlamydia				Gonorrhea			
	Female		Male		Female		Male	
	RD	RR	RD	RR	RD	RR	RD	RR
Black-White								
Base case	1210.8	5.1	917.3	7.2	396.2	7.1	669.4	9.0
Scenario 1	1750.9	5.2	1311.6	7.2	502.0	7.2	851.6	9.0
Scenario 2	1715.6	5.0	1305.5	7.1	492.6	7.0	844.5	8.9
Scenario 3	1705.1	5.0	1300.4	7.0	494.8	7.0	846.3	8.8
Scenario 4	1999.9	6.3	1241.1	7.2	510.4	7.6	803.3	8.7
Scenario 5	1766.5	5.2	1192.6	6.5	479.9	6.8	783.2	8.1
Hispanic-White								
Base case	284.1	2.0	119.4	1.8	27.5	1.4	71.8	1.9
Scenario 1	501.8	2.2	205.2	2.0	44.0	1.5	102.8	2.0
Scenario 2	400.3	1.9	171.9	1.8	32.7	1.4	88.2	1.8
Scenario 3	462.9	2.1	192.0	1.9	40.1	1.5	98.0	1.9
Scenario 4	494.2	2.3	287.6	2.4	53.4	1.7	138.3	2.3
Scenario 5	412.4	2.0	212.5	2.0	42.3	1.5	108.4	2.0

Figure 3 displays the changes in the Black-White and Hispanic-White RRs under each scenario as a relative disparity measure, with the calculated disparity measures shown in Table 3. Its layout is similar to Figure 2, except that Figure 3 shows the RR outcome and a value of 1.0 would indicate there is no observed disparity. Without any adjustment for missing race and ethnicity information (base-case scenario), the Black-White RR for chlamydia in 2018 was 5.1 for female and 7.2 for male patients. The Black-White RR for gonorrhea was 7.1 for female and 9.0 for male patients. In the base case, the Hispanic-White RRs for both chlamydia and gonorrhea were smaller than the Black-White RRs (chlamydia, RR: 2.0 for female and 1.8 for

male patients; gonorrhea: RR: 1.4 for female and 1.9 per for male patients). Under each imputation scenario, the RR remained stable compared with the base case. For chlamydia, the Black-White RR did not change remarkably in any scenario among the female or male patients. The Hispanic-White RR did not change remarkably in any scenario among female or male patients. Similarly, for gonorrhea, the Black-White RR did not change remarkably in any scenario among female or male patients. Moreover, the Hispanic-White RR showed no remarkable changes among female or male patients.

Figure 3. Estimated Black-White and Hispanic-White relative rate ratios (RRs) for chlamydia and gonorrhea under 5 scenarios to impute race and ethnicity for reported diagnoses with incomplete demographic data (2018).



At the state level, there was variation in how disparity measures changed under each scenario compared with the base case, with no adjustments for missing race and ethnicity information. Figure 4 presents dumbbell charts to illustrate how RDs for Black-White and Hispanic-White disparities among reported chlamydia diagnoses differ for each state under scenario 3 compared with the base case. This scenario and infection are presented for illustration, and all figures corresponding to other scenarios for each infection are available in the Multimedia Appendix 1. There is a dumbbell per state, excluding Connecticut and the District of Columbia. States were grouped into 3 categories based on their percentage of missing race and ethnicity information (0%-14% missing, 15%-29% missing, and ≥30% missing). The rate difference (x-axis) refers to the difference between the 2 diagnosis rates and is measured as diagnoses per 100,000 individuals. The gray dot is the base case, and the colored dot (orange or blue) is the scenario 3 value. The top and bottom panels display the RDs for the Black-White and Hispanic-White disparities, respectively. These patients were stratified according to sex. Each dumbbell represents the difference between the observed RD in the base-case scenario (gray dot) and the estimated RDs in scenario 3 (orange or blue dots). States that had larger discrepancies in their RDs after scenario 3 missing data adjustment had longer dumbbells. Under scenario 3, larger changes occurred in states with ≥30% of diagnoses with missing race and ethnicity information. The differences in RDs in scenario 3 versus the base case were more pronounced among female diagnoses and Black-White disparities. These qualitative conclusions were consistent when considering the other scenarios and gonorrhea (Multimedia Appendix 1).

Figure 5 illustrates the changes in state-level RRs for chlamydia diagnoses, comparing scenario 3 with the base case. This is the same interpretation as that shown in Figure 4. Similar to the findings from Figure 4 (RDs), there were larger differences in RRs among states with a higher percentage of missing race and ethnicity information, and RR differences were more pronounced for Black-White disparities. However, there was no clear pattern based on sex. These qualitative conclusions were consistent when considering the other scenarios and gonorrhea (Multimedia Appendix 1).

Figure 4. Illustration of changes in rate differences (RDs) as an absolute disparity measure for chlamydia across states with varying levels of missing race and ethnicity information using scenario 3 (reallocation based on information from states in the geographic region, for 2018).

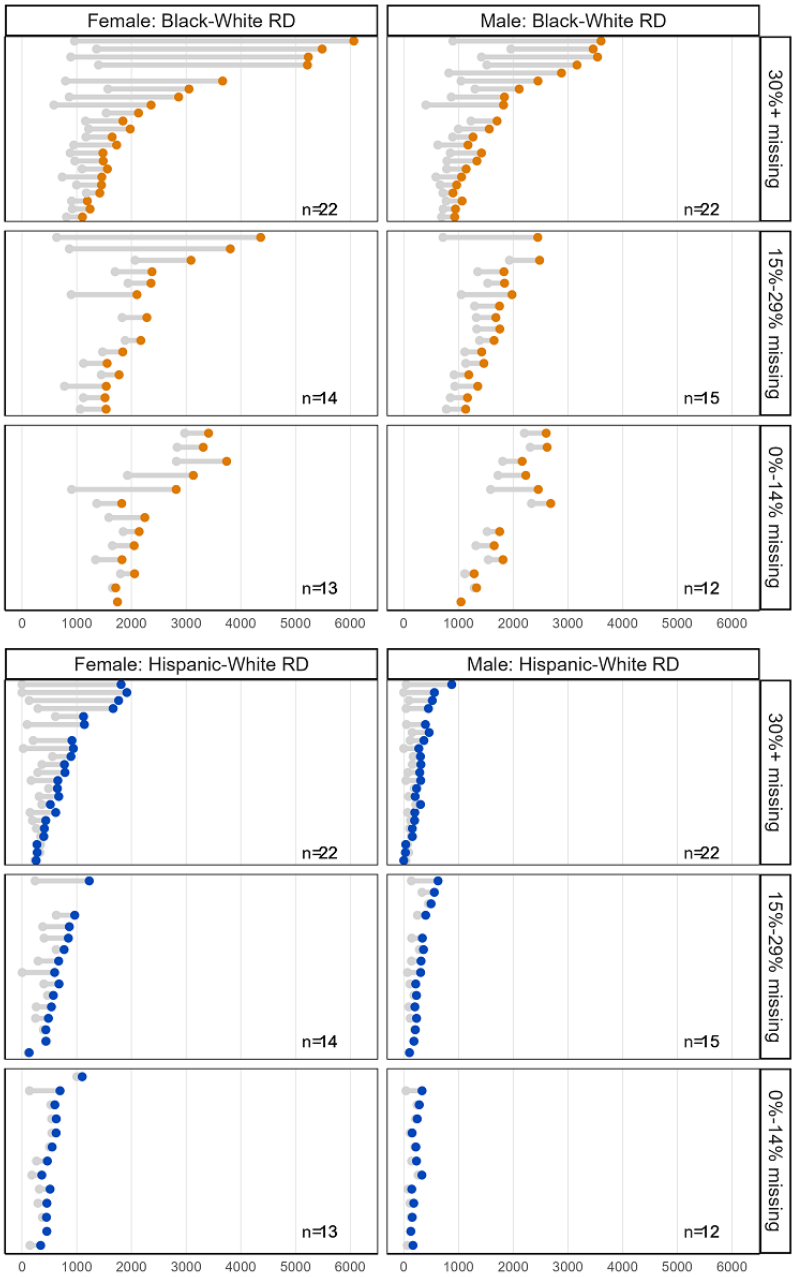
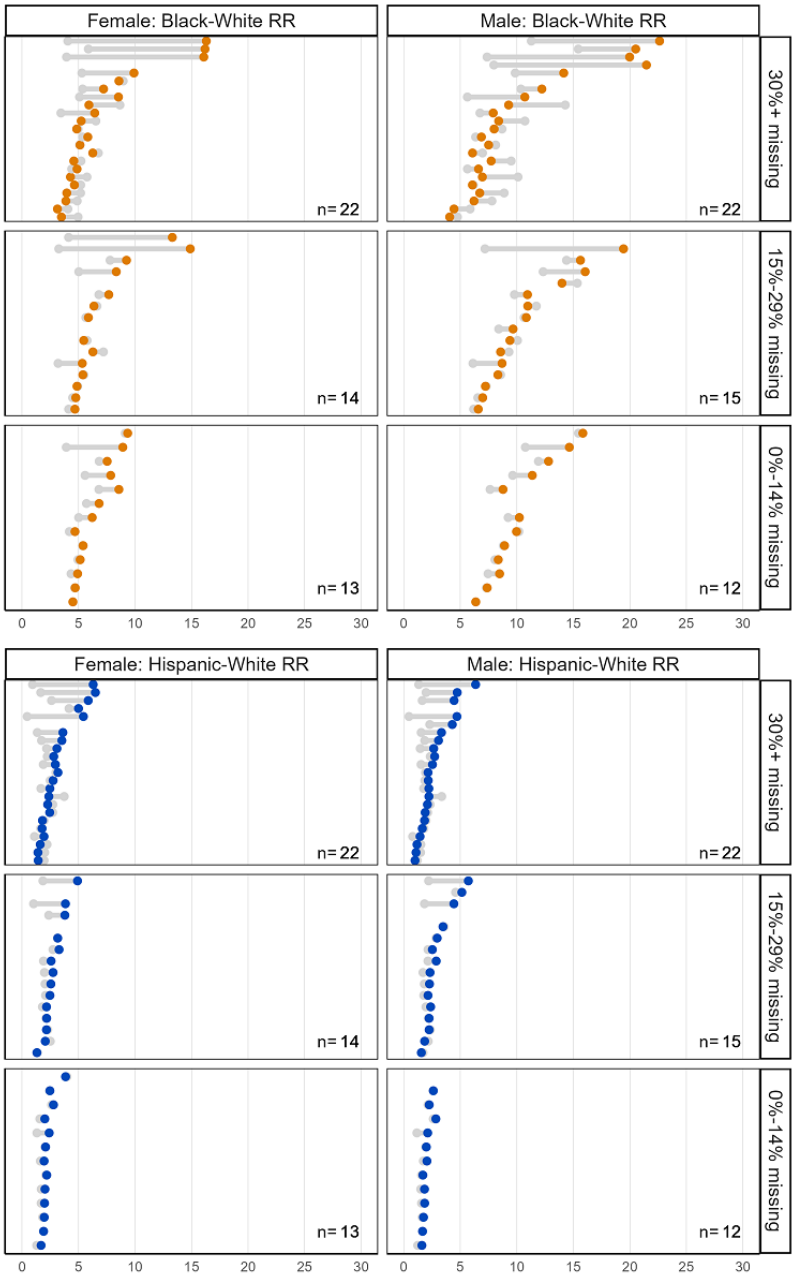


Figure 5. Illustration of changes in relative rate ratios (RRs) as relative disparity measures for chlamydia across states with varying levels of missing race and ethnicity information using scenario 3 (reallocation based on information from states in the geographic region, for 2018).



Discussion

Principal Findings

To explore the impact of missing race and ethnicity information on disparity measures, we used 5 scenarios to redistribute diagnoses with missing race and ethnicity information and compared our estimated disparity measures to the base-case scenario that excluded diagnoses with missing demographic data. Nationally, the absolute disparity measures notably increased in the 5 imputation scenarios for both the infections and sexes. By contrast, at the national level, the relative disparity measures did not change notably under the 5 scenarios. States with higher percentages of missing race and ethnicity information experienced larger changes in their disparity measures when the information was imputed [26]. Our analysis provides several solutions to assess potential bias from missing demographic information. Choosing the best approach depends on the contextual factors of the affected population. For example, scenarios 4 and 5 may not be the best solutions for chlamydia and gonorrhea because of the differences between the race and ethnicity distributions of the chosen infections. However, these scenarios might be appropriate for other diseases that have similar race and ethnicity distributions. Similarly, scenarios 2 and 3 may not be appropriate for geographic regions that have a very different distribution of race and ethnicity than the population in their neighboring or regional states.

Prior research on cancer has shown how absolute and relative disparity measures can yield different conclusions about trends in population disparities and that the lack of a framework for measuring disparities can yield inconsistent communications about cancer-related health disparities and measuring progress toward national goals [15]. Absolute and relative measures take

different perspectives on which aspects of population health to assess, and selecting a measure requires careful thinking about methodological issues as well as ethical and conceptual matters [15]. For example, a population health perspective prioritizes an absolute measure as a method to reflect the number of cases that would be averted from an intervention [15]. Our finding that the absolute measure was more sensitive to missing racial and ethnic information than the relative measure confirms that careful consideration is needed to select an appropriate disparity measure and interpret situations in which absolute and relative disparity measures diverge.

There are several reasons why the demographic data for reported chlamydia and gonorrhea diagnoses may be incomplete. Although standardized recommendations exist for collecting race and ethnicity information [27], demographic data collection is incomplete and inconsistent across jurisdictions and health care systems [28,29]. Incomplete collection of race and ethnicity information might result from individuals not disclosing information about their race and ethnicity because of mistrust or if they are provided with limited response options that do not match their self-identity [30]. Local health agencies' efforts to follow up on reported diagnoses to collect additional demographic information can be costly and inefficient [31].

Implications for Practice

In our experience and based on conversations with practitioners in the field, there are 3 primary sources of race and ethnicity information for newly diagnosed chlamydia and gonorrhea infections. First, diagnostic data may be obtained from laboratory reports that are automatically submitted to health departments, which frequently omit race and ethnicity information. Second, providers may submit case reports of notifiable conditions. Although these case reports should have race and ethnicity information, they may be incomplete, and passive surveillance systems based on laboratory data and case reports may have missing demographic information unless states can do active surveillance to obtain case reports on laboratory reports for which there is no matched case report. Third, race and ethnicity information may have been collected by disease intervention specialists through partner service interviews. However, interviews are less frequently conducted for gonorrhea and chlamydia following the CDC guidelines to prioritize HIV and syphilis for outreach [21]; furthermore, the high number of gonorrhea and chlamydia cases makes it infeasible to interview all individuals. Promising strategies for improving data quality include strengthening relationships with providers to improve the completeness of reporting, focusing on large-volume providers, and updating surveillance systems to use standardized electronic reporting.

Upstream and system-level solutions, such as enhanced electronic reporting, are needed to improve the availability of race and ethnicity information in public health surveillance systems, particularly when it is infeasible for public health workers to interview all cases. A past assessment of race and ethnicity information across different disease registries found that inconsistencies occurred more frequently among Hispanic populations and populations categorized as being in an "other" racial and ethnic group, suggesting that a more granular coding system for collecting demographic information might improve data completeness [32]. Furthermore, requiring race and ethnicity information in the initial data collection and simultaneously working with communities to improve surveillance instruments has been previously recommended to reduce the incompleteness of race and ethnicity information [30].

Informatics specialists can play important roles in designing cost-effective and interoperable solutions by defining standardized data elements, designing validation procedures, and automatically populating registries to enhance electronic reporting systems [5]. A recent case study showed that the automatic transfer of clinical data from an electronic health record system to public health surveillance improved the timeliness and quality of data with minimal manual intervention [26]. Moreover, collaboration with informatics specialists can improve the design and efficiency of data-entry systems for collecting more complete data. For example, systems can prevent progression until all required elements are filled out, and some aspects of the data entry can be automatically filled to avoid frustrating users with too many questions [3]. These types of informatics solutions could help enhance the electronic reporting of information required by public health agencies. Ultimately, obtaining more complete and accurate information on the front end is more efficient in terms of time and cost than assigning health department staff to locate persons with incomplete information for follow-up, particularly for high-morbidity diseases.

Our analysis highlights the importance of addressing missing data when calculating population rates and disparity measures. Although we focused on reported chlamydia and gonorrhea diagnoses among Black, Hispanic, and White populations, our findings likely apply to other outcomes or other population group comparisons. Missing data may lead to biased conclusions, especially if data are not missing at random across subpopulations [33]. When individual-level data are available, maximum likelihood and Bayesian multiple imputation methods are recommended to handle missing data [34]. For aggregate data, if spatial-level data are available, simple approaches, such as our 5 scenarios, can be used to impute missing data.

Limitations

Our study has several limitations. First, there may be other approaches to impute the missing race and ethnicity information. Second, with our aggregate data, our analysis was not designed to assess the best imputation scenario but to illustrate the potential impacts of missing race and ethnicity information on health disparity measures. Finding the best imputation scenario is an important area for future research using individual-level data from medical records, claims data, or other sources matched with surveillance data. Third, we examined a limited number of racial and ethnic disparities because the number of reported chlamydia and gonorrhea diagnoses was small in the population groups other than those recorded as Black, Hispanic, or White, making it difficult to calculate stable estimates.

Conclusions

Our analysis showed that the observed disparities are likely underestimated because of missing race and ethnicity information, particularly when using an absolute disparity measure. More complete race and ethnicity information is important to better understand the contributing causes of inequities and to monitor progress toward policy initiatives to reduce disparities. Addressing the missing demographic information in surveillance systems requires system-level solutions. However, as a short-term solution, local public health agencies can adapt imputation scenarios to adjust surveillance data for use in population indicators of health equity. Imputation scenarios can be integrated with the existing public health informatics infrastructure. Using these scenarios requires data analytics staff with knowledge of statistical analysis software, and there may be a limited ability to prioritize human resources for scenario analysis, particularly in local health departments or during public health emergencies such as COVID-19 or monkeypox. However, they do not require additional data or changes to the system design, making them useful short-term solutions for situations in which human resources are available.

Acknowledgments

The authors would like to thank Thomas Bertrand, Gregory Felzien, Nanette Benbow, and Mary Ann Chiasson from our research advisory group and Travis O'Donnell (New York State Department of Health) for helpful feedback on an earlier draft. John Angles (University of Albany) conducted a quality assurance review of the code. This work was funded by the US Centers for Disease Control and Prevention, National Center for HIV, Viral Hepatitis, STD, and TB Prevention Epidemiologic and Economic Modeling Agreement (NEEMA #NU38PS004650). The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the funding agency or the New York State Department of Health.

Authors' Contributions

BA, RH-M, and EGM conceived the study. BA directed the research, conducted the statistical analyses, and wrote the initial draft. All authors contributed to the interpretation of data and revised the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Technical appendix with complete results for all scenarios and infections.
[DOCX File , 4894 KB - publichealth_v8i11e38037_app1.docx]

References

1. Murray J, Cohen A. Infectious disease surveillance. In: International Encyclopedia of Public Health (Second Edition). Amsterdam, Netherlands: Elsevier Science; 2017.
2. Basit MA, Lehmann CU, Medford RJ. Managing pandemics with health informatics: successes and challenges. Yearb Med Inform 2021 Aug 21;30(1):17-25 [FREE Full text] [doi: 10.1055/s-0041-1726478] [Medline: 33882594]
3. Costa-Santos C, Neves AL, Correia R, Santos P, Monteiro-Soares M, Freitas A, et al. COVID-19 surveillance data quality issues: a national consecutive case series. BMJ Open 2021 Dec 06;11(12):e047623 [FREE Full text] [doi: 10.1136/bmjopen-2020-047623] [Medline: 34872992]
4. Dixon BE, Wen C, French T, Williams JL, Duke JD, Grannis SJ. Extending an open-source tool to measure data quality: case report on Observational Health Data Science and Informatics (OHDSI). BMJ Health Care Inform 2020 Mar 29;27(1):e100054 [FREE Full text] [doi: 10.1136/bmjhci-2019-100054] [Medline: 32229499]
5. Rastegar-Mojarad M, Sohn S, Wang L, Shen F, Bleeker TC, Cliby WA, et al. Need of informatics in designing interoperable clinical registries. Int J Med Inform 2017 Dec;108:78-84 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.004] [Medline: 29132635]
6. Prather C, Fuller TR, Marshall KJ, Jeffries WL. The impact of racism on the sexual and reproductive health of African American women. J Womens Health (Larchmt) 2016 Jul;25(7):664-671 [FREE Full text] [doi: 10.1089/jwh.2015.5637] [Medline: 27227533]
7. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep 2014;129 Suppl 2:19-31 [FREE Full text] [doi: 10.1177/00333549141291S206] [Medline: 24385661]
8. Penman-Aguilar A, Talih M, Huang D, Moonesinghe R, Bouye K, Beckles G. Measurement of health disparities, health inequities, and social determinants of health to support the advancement of health equity. J Public Health Manag Pract 2016;22 Suppl 1:S33-S42 [FREE Full text] [doi: 10.1097/PHH.0000000000000373] [Medline: 26599027]
9. Labgold K, Hamid S, Shah S, Gandhi NR, Chamberlain A, Khan F, et al. Estimating the unknown: greater racial and ethnic disparities in COVID-19 burden after accounting for missing race and ethnicity data. Epidemiology 2021 Mar 01;32(2):157-161 [FREE Full text] [doi: 10.1097/EDE.0000000000001314] [Medline: 33323745]

10. Dover DC, Belon AP. The health equity measurement framework: a comprehensive model to measure social inequities in health. Int J Equity Health 2019 Feb 19;18(1):36 [FREE Full text] [doi: 10.1186/s12939-019-0935-0] [Medline: 30782161]
11. Smith N, Iyer RL, Langer-Gould A, Getahun DT, Strickland D, Jacobsen SJ, et al. Health plan administrative records versus birth certificate records: quality of race and ethnicity information in children. BMC Health Serv Res 2010 Nov 23;10:316 [FREE Full text] [doi: 10.1186/1472-6963-10-316] [Medline: 21092309]
12. Long JA, Bamba MI, Ling B, Shea JA. Missing race/ethnicity data in Veterans Health Administration based disparities research: a systematic review. J Health Care Poor Underserved 2006 Feb;17(1):128-140. [doi: 10.1353/hpu.2006.0029] [Medline: 16520522]
13. Painter EM, Ussery EN, Patel A, Hughes MM, Zell ER, Moulia DL, et al. Demographic characteristics of persons vaccinated during the first month of the COVID-19 vaccination program - United States, December 14, 2020-January 14, 2021. MMWR Morb Mortal Wkly Rep 2021 Feb 05;70(5):174-177 [FREE Full text] [doi: 10.15585/mmwr.mm7005e1] [Medline: 33539333]
14. Kreisel KM, Spicknall IH, Gargano JW, Lewis FM, Lewis RM, Markowitz LE, et al. Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2018. Sexual Trans Dis 2021 Jan 26;48(4):208-214. [doi: 10.1097/olq.0000000000001355]
15. Chesson HW, Spicknall IH, Bingham A, Brisson M, Eppink ST, Farnham PG, et al. The estimated direct lifetime medical costs of sexually transmitted infections acquired in the United States in 2018. Sexual Trans Dis 2021 Jan 26;48(4):215-221. [doi: 10.1097/olq.0000000000001380]
16. Sexually transmitted disease surveillance 2019. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/std/statistics/2019/default.htm [accessed 2021-10-12]
17. NCHHSTP AtlasPlus. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchhstp/atlas/ [accessed 2020-07-30]
18. How NNDSS conducts case surveillance. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nndss/about/conduct.html [accessed 2022-07-31]
19. What is case surveillance? Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nndss/about/index.html [accessed 2022-08-01]
20. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/std/program/data-mgmt.htm [accessed 2022-07-31]
21. Harper S, Lynch J, Meersman SC, Breen N, Davis WW, Reichman ME. An overview of methods for monitoring social disparities in cancer with an example using trends in lung cancer incidence by area-socioeconomic position and race-ethnicity, 1992-2004. Am J Epidemiol 2008 Apr 15;167(8):889-899 [FREE Full text] [doi: 10.1093/aje/kwn016] [Medline: 18344513]
22. Ruggles S, Flood S, Goeken R, Grover J, Meyer E, Pacas J, et al. IPUMS USA: Version 10. IPUMS. URL: https://doi.org/10.18128/D010.V10.0 [accessed 2020-05-09]
23. Agresti A. Categorical Data Analysis, 2nd Edition. Hoboken, New Jersey, United States: Wiley; 2002.
24. Amitha P, Binu V, Seena B. Estimation of missing values in aggregate level spatial data. Clin Epidemiol Global Health 2021 Jan;9:304-309. [doi: 10.1016/j.cegh.2020.10.003]
25. Centers for Disease ControlPrevention (CDC). Recommendations for partner services programs for HIV infection, syphilis, gonorrhea, and chlamydial infection. MMWR Recomm Rep 2008 Nov 07;57(RR-9):1-83; quiz CE1 [FREE Full text] [Medline: 18987617]
26. Whipple A, Jackson J, Ridderhoff J, Nakashima AK. Piloting electronic case reporting for improved surveillance of sexually transmitted diseases in Utah. Online J Public Health Inform 2019;11(2):e7 [FREE Full text] [doi: 10.5210/ojphi.v11i2.9733] [Medline: 31632601]
27. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. Federal Register. 1997. URL: https://www.govinfo.gov/content/pkg/FR-1997-10-30/pdf/97-28653.pdf [accessed 2022-10-26]
28. Pittman M, Pierce D, Hasnain-Wynia R. Who, when, and how: the current state of race, ethnicity, and primary language data collection in hospitals. The Commonwealth Fund. 2004 May 1. URL: https://www.commonwealthfund.org/publications/fund-reports/2004/may/who-when-and-how-current-state-race-ethnicity-and-primary [accessed 2022-07-31]
29. Laws M, Heckscher RA. Racial and ethnic identification practices in public health data systems in New England. Public Health Reports 2002 Jan;117(1):50-61. [doi: 10.1016/s0033-3549(04)50108-5]
30. Kader F, Smith CL. Participatory approaches to addressing missing COVID-19 race and ethnicity data. Int J Environ Res Public Health 2021 Jun 18;18(12):6559 [FREE Full text] [doi: 10.3390/ijerph18126559] [Medline: 34207130]
31. Chen J, Etkind P, Coman G, Tang Y, Whelan M. Eliminating missing race/ethnicity data from a sexually transmitted disease case registry. J Community Health 2003 Aug;28(4):257-265. [doi: 10.1023/a:1023986024918] [Medline: 12856795]
32. Lee SJ, Grobe JE, Tiro JA. Assessing race and ethnicity data quality across cancer registries and EMRs in two hospitals. J Am Med Inform Assoc 2016 May;23(3):627-634 [FREE Full text] [doi: 10.1093/jamia/ocv156] [Medline: 26661718]
33. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 2006 Oct;59(10):1087-1091. [doi: 10.1016/j.jclinepi.2006.01.014] [Medline: 16980149]
34. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002 Jun;7(2):147-177. [Medline: 12090408]

Abbreviations

CDC: Centers for Disease Control and Prevention
RD: rate difference
RR: rate ratio
STI: sexually transmitted infection

Edited by A Mavragani; submitted 16.03.22; peer-reviewed by C Khosropour, S Nagavally, R Zhang; comments to author 25.07.22; revised version received 03.08.22; accepted 29.09.22; published 09.11.22.

Please cite as:
Ansari B, Hart-Malloy R, Rosenberg ES, Trigg M, Martin EG
Modeling the Potential Impact of Missing Race and Ethnicity Data in Infectious Disease Surveillance Systems on Disparity Measures: Scenario Analysis of Different Imputation Strategies
JMIR Public Health Surveill 2022;8(11):e38037
URL: https://publichealth.jmir.org/2022/11/e38037
doi:10.2196/38037
PMID:36350701

Original Paper

Mass Screening of SARS-CoV-2 With Rapid Antigen Tests in a Receding Omicron Wave: Population-Based Survey for Epidemiologic Evaluation

Tsz Ho Kwan[1]; Ngai Sze Wong[1,2]; Chin Pok Chan[2]; Eng Kiong Yeoh[2]; Samuel Yeung-shan Wong[2]; Shui Shan Lee[1]

[1]Stanley Ho Centre for Emerging Infectious Diseases, The Chinese University of Hong Kong, Shatin, Hong Kong

[2]Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Shatin, Hong Kong

Corresponding Author:
Shui Shan Lee
Stanley Ho Centre for Emerging Infectious Diseases
The Chinese University of Hong Kong
Room 207, Postgraduate Education Centre
Prince of Wales Hospital
Shatin
Hong Kong
Phone: 852 22528862
Email: sslee@cuhk.edu.hk

Abstract

Background: The COVID-19 Omicron BA.2 epidemic wave in Hong Kong peaked in the first quarter of 2022. Following the implementation of stringent public health measures, the daily number of reported cases fell from over 50,000 to below 2000. Although outbreaks steadily receded, the government rolled out a 3-day "voluntary universal rapid testing" campaign to invite all citizens to self-perform a rapid antigen test (RAT) daily to identify undetected prevalent infections.

Objective: This study aimed to evaluate the uptake and results of RAT mass screening to estimate the population's residual epidemic burden and assess the risk of further transmission.

Methods: A cross-sectional study comprising an open web-based population-based survey was conducted a week after the RAT campaign. Participants were asked to report their COVID-19 vaccination and infection history and the RAT performance and test result during the period. They were also invited to report their coliving individuals' test performance and results. Reasons for nonuptake were enquired. Testing and positive rates were age-adjusted. Determinants of undergoing RAT were identified using univariable and multivariable logistic regression models.

Results: In total, particulars from 21,769 individuals were reported by 8338 participants. The overall age-adjusted testing rate was 74.94% (95% CI 73.71%-76.18%), with over 80% of participants in the age groups between 45-84 years having self-performed RAT during the campaign period. After age-adjustment, 1.03% (95% CI 0.86%-1.21%) of participants tested positive. The positive rates in the age groups between 20-29 years and >84 years exceeded 2%. Taking into account the positive rate and 5819 reported cases during the period, the cases identified in the campaign might account for 7.65% (95% CI 6.47%-9.14%) of all infections. Testers were more likely to be female, older, not previously diagnosed with COVID-19, and have received COVID-19 vaccination. Adjusting for the number of household members, those living with a child aged <12 years and whose household members were also tested were more likely to have self-performed an RAT. Main reasons for not performing an RAT included the absence of symptoms (598/1108, 53.97%), disbelief of the appropriateness of the campaign as an antiepidemic measure (355/1108, 32.04%), and a recent COVID-19 diagnosis (332/1108, 29.96%).

Conclusions: The residual population burden remained substantial in spite of the clear evidence of a receding epidemic wave. Despite caution in generalization to the Hong Kong population, the high participation rate in mass screening indicated that the voluntary RAT was well accepted, making it a feasible option for implementation as a complementary means of public health surveillance.

(JMIR Public Health Surveill 2022;8(11):e40175) doi:10.2196/40175

KEYWORDS

COVID-19; SARS-CoV-2 antigen testing; COVID-19 vaccine; mass screening; antigen test; epidemiology; Omicron; Hong Kong; public health; outbreak; epidemic; screening; transmission; online; vaccination; vaccines; surveillance

Introduction

Worldwide, SARS-CoV-2 transmissions are characterized by repeated outbreak waves of different intensities and amplitudes. In 2020 and 2021, three waves of SARS-CoV-2 transmission in Hong Kong, a densely populated Asia-Pacific city, were brought under control with stringent public health and social measures, comprising case detection, contract tracing, isolation of infected persons, quarantine of close contacts, and widely accessible polymerase chain reaction (PCR) tests in health care facilities and community centers [1]. Social and mobility restrictions were imposed once community transmission had been detected, thereby shifting the epidemic burden to other less restricted exposure settings [2]. By late 2021, no local transmissions were detected for almost 3 months. This enviable record was broken when the first cases of Omicron BA.2 infections became detected in the community, causing a superspreading event [3]. Despite reimposing restrictions to social activities and mobility in anticipation of increased social mix in the Lunar New Year holiday period that followed, Hong Kong was hard hit by the Omicron BA.2 epidemic in February and March 2022, with over 50,000 cases reported daily at its peak [4]. Although the epidemic was receding, the government rolled out a 3-day "voluntary universal rapid testing" campaign between April 8-10, 2022, during which citizens were invited to self-perform a rapid antigen test (RAT) daily in the absence of any lockdown policies while other social distancing measures remained in place [5]. Antiepidemic service bags containing, inter alia, 20 RAT kits were distributed to all households across the city a week in advance. A web-based declaration system was in place to facilitate the statutory reporting of positive cases within 24 hours for issuing isolation and quarantine orders. In the week prior to screening, the daily number of COVID-19 cases reported had decreased to below 5000, and a downward trend was observed [6].

Mass screening is an uncommon control strategy for COVID-19, and only limited studies on its application have been published [7,8], although it has been suggested for developing an exit strategy [9]. Guangzhou's mass-screening exercise in 2021, along with the isolation and city border control policies, had contributed to the suppression of the epidemic in 6 months [8]. A modeling study in Slovakia showed that after the mass-testing campaign, the prevalence could be reduced by 70% [10]. Another modeling study in France demonstrated that, on average, the RAT-based mass-testing campaign could reduce daily incidence by up to 30% [11]. However, these campaigns do not necessarily contribute to the reduction of mortality [12]. A web-based survey conducted in United States showed that, for voluntary testing without a stipulated mass-testing period, the reasons for self-testing included potential exposure and the presence of symptoms [13]. The uptake rate also varied across geographic regions and age groups. The role of RAT is not limited to case detection but also surveillance, particularly in places adopting the "living with the virus" policy, thereby informing public health policies [14]. Against these backgrounds, we conducted a population-based survey on the uptake and results of RAT mass screening to estimate the population's residual SARS-CoV-2 burden and assess the risk of further transmission in the territory.

Methods

Study Design

This was a cross-sectional study. Eligible participants were Hong Kong residents aged ≥18 years whose households had received an antiepidemic service bag distributed by the government. A bilingual (Chinese and English versions), open, self-administered, web-based, and population-based survey was designed, covering demographics (age, sex, and residing district); COVID-19 vaccination history (type, date, and dose of the last vaccination received); COVID-19 infection history; signs and symptoms; RAT performed during the 3 campaign days with result; post–positive result actions (reporting to the government's web-based declaration system, seeking treatment, and isolation); and the number of coliving individuals. Participants could opt to report their coliving individuals' age, sex, RAT performed and result during the campaign period, and their relationships, up to 5 persons. Due to the simplistic nature of this study, the items were not randomized. Adaptive questioning was used on the same page to display questions relating to COVID-19 vaccination history, details about RAT history during the campaign period, and particulars about the coliving individuals. For those who did not undergo an RAT, they were asked to select at least one of the following reasons for not doing so: recent diagnosis, recently tested, regular testing as part of work requirement, no appropriate time and environment, avoiding isolation if tested positive, avoiding compulsory declaration if tested positive, avoiding sampling discomfort, no confidence to self-test, no symptoms, not worried about getting infected, not believing the campaign was an appropriate antiepidemic measure, and others. There were at most 16 questions for each participant, and a maximum of 5 questions for each coliving individual. No personal identifiers were collected. The survey was tested and refined before fielding. After completing the survey, participants were invited to share their location using the HTML5 Geolocation Application Programming Interface if they were at home or in the workplace. Coordinates outside the territory of Hong Kong were removed.

Subjects and Recruitment

Web and newspapers advertisements were placed to recruit Hong Kong residents to join as participants. All responses were collected through the bespoke website designed for this web-based survey. The completion of the survey by participants was voluntary. Completeness checking was done using JavaScript before submission. Incomplete responses were not collected. No data were excluded due to atypical time stamps because of the simplistic nature of the survey. An anonymous

session identifier was set in the cookies, and the IP addresses of participants were collected. Duplicate entries of the same session identifier were removed.

Ethics Approval

The study data collected were anonymous. Web-based informed consents were obtained before participants filled out the questionnaire. No incentives were offered upon the completion of the study. This study was approved by the Survey and Behavioural Research Ethics Committee of The Chinese University of Hong Kong (SBRE-21-0685). The conduct of the study was in compliance with the Declaration of Helsinki.

Statistical Analysis

Determinants of undergoing an RAT during the campaign period were identified using univariable and multivariable logistic regression models. Testing rate and positive rate were determined by aggregating both study participants and their coliving participants. The Wilson score method was used to calculate the 95% CI of age-specific testing rate [15]. The testing rate and positive rate were age-adjusted by groups defined by 5-year windows, except those aged ≥85 years were grouped together, using the provisional figures published by the Census and Statistics Department from the 2021 Population Census [16]. The 95% CI of the directly standardized testing rate was computed using the Byar method [17]. The number of prevalent infections in the territory during the campaign period was estimated by multiplying the population size by the age-adjusted positive rate with the 95% CI. Maps were drawn with the QGIS platform (QGIS Development Team) using 2019 District Council Constituency Areas as the spatial unit. There was a total of 452 District Council Constituency Areas, each of which normally containing about 16,599 residents [18]. Secondary outcomes, including determinants of prior COVID-19 diagnosis and reasons for not getting tested, were evaluated using chi-square test and Mann-Whitney *U* test for categorical and continuous predictors, respectively. Multivariable logistic regression analysis was performed by including variables with a *P* value of <.05 in the univariable analyses. All statistical analyses were conducted in R statistical software (R Foundation for Statistical Computing). Reporting in this manuscript follows the Checklist for Reporting Results of Internet E-Surveys [19].

Results

In total, 8759 responses were collected between April 17-25, 2022, of which 8338 were analyzed after removing duplicate entries. Of the 8338 participants, the median age was 61 (IQR 53-67) years, and 38.89% (n=3243) were male (Table 1). In all, 16.89% (n=1408) reported at least one episode of previous COVID-19 diagnosis. Almost all (8086/8314, 97.26%) participants have received at least one dose of COVID-19 vaccine, with 81.48% (6774/8314) having 3 doses or more. The distribution of the types of vaccine received for the last dose was similar (BNT162b2 by Pfizer-BioNTech: 4566/8105, 56.34%; CoronaVac by Sinovac: 3522/8015, 43.45%; and others: 17/8105, 0.21%). The median number of coliving individuals was 2 (IQR 1-3), totaling 15,243 persons, of whom the particulars of 13,431 (88.11%) household members were complete for analysis. Combining both index respondents and coliving household members (N=21,769), the overall median age was 56 (IQR 38-65) years, and the overall crude RAT self-screening rate was 78.53% (n=17,096), with age-specific rates of over 80% in the age groups between 45-84 years. The overall age-adjusted testing rate was 74.94% (95% CI 73.71%-76.18%; Figure 1). Although geographical variation of the proportion of households who performed the RAT was observed, there were no significant differences among spatial units (n=6949; *P*>.99 by chi-square test; Multimedia Appendix 1).

Among index participants, having performed an RAT was associated with one's sex (reference: female; adjusted odds ratio [aOR] 0.76, 95% CI 0.67-0.87; *P*=.001), older age in years (aOR 1.03, 95% CI 1.03-1.04; *P*<.001), a previous COVID-19 diagnosis (aOR 0.42, 95% CI 0.37-0.49; *P*<.001), and vaccination history (aOR 2.03, 95% CI 1.46-2.78; *P*<.001; Table 2). Among those vaccinated, the number of doses (aOR 1.80, 95% CI 1.58-2.06; *P*<.001) and type of the last vaccine dose received (Sinovac compared to BioNTech: aOR 2.28, 95% CI 1.96-2.66; *P*<.001) were associated with testing during the campaign period. Taking factors regarding coliving individuals into account, after adjusting for the number of household members, household members having been tested (*P*<.001) and living with a child or children aged <12 years (*P*=.002) were additionally associated with RAT performance during the campaign period.

Among the reasons (n=1108) for not getting tested, the 3 most common ones were not having symptoms (n=598, 53.97%), not believing the campaign was an appropriate antiepidemic measure (n=355, 32.04%), and a recent diagnosis (n=332, 29.96%; Table 3). Factors associated with prior diagnosis included not living alone (odds ratio [OR] 1.49, 95% CI 1.23-1.79; *P*<.001), especially with those aged <12 years (OR 1.32, 95% CI 1.08-1.62; *P*=.007); age (*P*<.001, by Mann-Whitney *U* test); and not having been vaccinated with at least 2 doses (OR 2.11, 95% CI 1.68-2.66; *P*<.001). The multivariable logistic regression showed that age in years (aOR 0.98, 95% CI 0.97-0.98; *P*<.001), not living alone (aOR 1.42, 95% CI 1.18-1.72; *P*<.001), and receiving less than 2 vaccination doses (aOR 2.05, 95% CI 1.60-2.60; *P*<.001) were significantly associated with prior diagnosis. The untested respondents who did not believe the campaign was an appropriate antiepidemic measure were more likely to be of younger age in years (aOR 0.99, 95% CI 0.98-0.99; *P*=.001), male (aOR 1.70, 95% CI 1.31-2.20; *P*<.001), and unvaccinated (aOR 1.95, 95% CI 1.12-3.39; *P*=.02). Notably, only a small proportion of participants reported being not confident to perform a self-test (n=32, 2.89%) and not having time and an appropriate environment to test (n=8, 0.72%). Separately, the crude positive rate among participants and including their coliving individuals was 0.62% (45/7226) and 1.19% (117/9870), respectively. Age-specific positive rates exceeded 2% in the population groups between 20-29 years and those aged >85 years (Figure 2). After adjusting for the age structure of the population, 1.03% (95% CI 0.86%-1.21%) of the population could have tested positive during the campaign period. It can be inferred that 76,039 (95% CI 63,663-89,947) persons in the 7.4 million population could have tested positive during the 3 days. In term of compliance, of the 45 participants

reporting a positive RAT result, 62% (n=28) had declared to the government and 73% (n=33) self-isolated. Some 64% (29/45) sought treatments, including self-medications reported in a majority (22/29, 76%) of participants.

Table 1. Demographics and COVID-19–related histories of participants and their household members.

Characteristic	Participants	Participants' household members
Sex, male (participants: n=8338; participants' household members: n=13,431), n (%)	3243 (38.89)	5593 (41.64)
Age (years; participants: n=8338; participants' household members: n=13,431), median (IQR)	61 (53-67)	50 (30-63)
Performed an RAT[a] during the campaign period (participants: n=8338; participants' household members: n=13,431), n (%)	7226 (86.66)	9870 (73.49)
Performed an RAT more than once during the campaign period (n=7226), n (%)	6258 (86.6)	N/A[b]
Tested positive during the campaign period (participants: n=7226; participants' household members: n=9870), n (%)	45 (0.62)	117 (1.19)
Previous COVID-19 diagnosis (n=8338), n (%)	1408 (16.89)	N/A
Number of COVID-19 vaccines received (n=8314), n (%)		
0	228 (2.74)	N/A
1	122 (1.47)	N/A
2	1190 (14.31)	N/A
3	6425 (77.28)	N/A
4	349 (4.2)	N/A
Type of the last COVID-19 vaccine received (n=8105), n (%)		
BNT162b2 by BioNTech	4566 (56.34)	N/A
CoronaVac by Sinovac	3522 (43.45)	N/A
Others	17 (0.21)	N/A

[a]RAT: rapid antigen test.

[b]N/A: not applicable.

Figure 1. Age-specific and age-adjusted rapid antigen testing rates (dots) and 95% CIs (error bars).

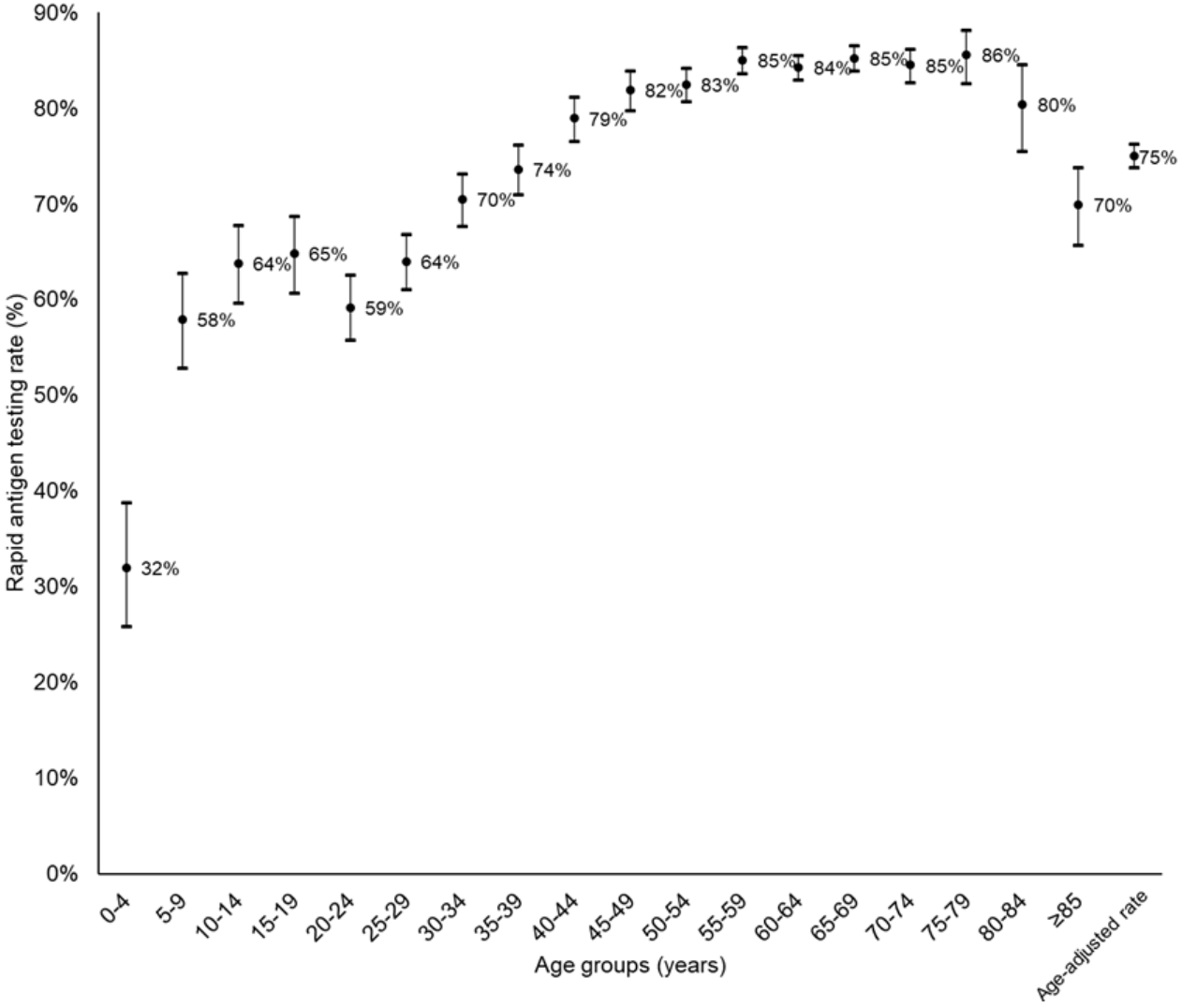


Table 2. Factors associated with performing a rapid antigen test during the campaign period.

Factor	Model 1	*P* value	Model 2	*P* value	Model 3	*P* value
Sex, male (reference: female), aOR[a] (95% CI)	0.76 (0.67-0.87)	<.001	0.75 (0.66-0.86)	<.001	0.75 (0.62-0.89)	.002
Age (years), aOR (95% CI)	1.03 (1.03-1.04)	<.001	1.02 (1.02-1.03)	<.001	1.02 (1.01-1.03)	<.001
Previous diagnosis of COVID-19, aOR (95% CI)	0.42 (0.37-0.49)	<.001	0.47 (0.41-0.56)	<.001	0.45 (0.37-0.56)	<.001
Vaccinated for at least one dose against COVID-19, aOR (95% CI)	2.03 (1.46-2.78)	<.001	N/A[b]	N/A	N/A	N/A
Received Sinovac COVID-19 vaccine (reference: BioNTech vaccine), aOR (95% CI)	N/A	N/A	2.28 (1.96-2.66)	<.001	1.97 (1.62-2.39)	<.001
Number of doses of COVID-19 vaccination, aOR (95% CI)	N/A	N/A	1.80 (1.58-2.06)	<.001	2.02 (1.62-2.39)	<.001
Number of household members, aOR (95% CI)	N/A	N/A	N/A	N/A	0.53 (0.48-0.58)	<.001
Coliving with a person aged <12 years, aOR (95% CI)	N/A	N/A	N/A	N/A	1.81 (1.26-2.61)	.002
Any of the coliving individuals having been tested during the campaign, aOR (95% CI)	N/A	N/A	N/A	N/A	7.28 (6.36-8.36)	<.001
AIC[c]	6208		5739		3336	

[a]aOR: adjusted odds ratio.

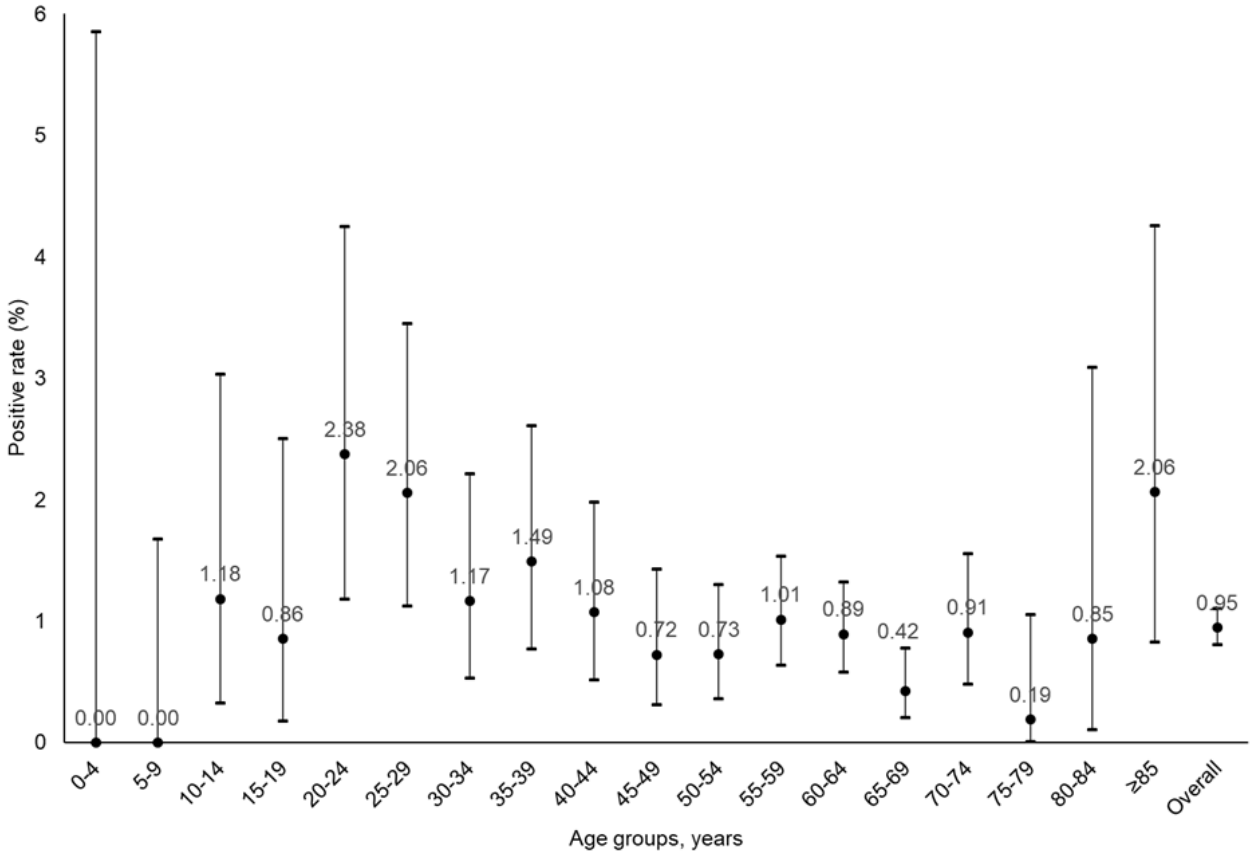
[b]N/A: not applicable.

[c]AIC: Akaike information criterion.

Table 3. Reasons for not having performed rapid antigen testing during the campaign period (n=1108).

Reason	Participant, n (%)
Not necessary because I have been diagnosed recently	332 (29.96)
Unwilling to repeat because I have tested recently	181 (16.34)
Doing testing regularly as part of work requirement, so would not want to do additional tests	95 (8.57)
Did not have the time and the appropriate environment to do the test	8 (0.72)
To avoid isolation due to a positive result	97 (8.75)
To avoid declaration of a positive result to the government	58 (5.23)
To avoid discomfort caused by swab collection	60 (5.52)
Not confident to perform a self-test	32 (2.89)
Not having symptoms	598 (53.97)
Not worried about getting infected	204 (18.41)
Not believing that “voluntary universal rapid testing” campaign is an appropriate antiepidemic measure	355 (32.04)
Other reasons	95 (8.57)

Figure 2. Age-specific and age-adjusted positive rates (dots) and 95% CIs (error bars).



Discussion

Principal Findings

Participation is key to a voluntary mass-screening campaign for SARS-CoV-2 infection, the coverage of which would not be known in the absence of an accompanying surveillance mechanism. An evaluation of the Hong Kong campaign was made possible through a separately conducted ensuing population-based web-based survey. The geographical diversity of the participants in this study supported that a diverse sample population had been recruited. The proportion of participants reporting previous COVID-19 diagnosis corresponded well with that recorded (about 16%) in the published government statistics [6]. Although participation in mass screening was voluntary, three-quarters of respondents in our population-based survey had undergone the testing at least once during the 3-day period, confirming the feasibility of implementing self-RAT screening as a complementary means of public health surveillance.

Our survey results showed that those participating in the RAT campaign were more likely to be older, female, and vaccinated against COVID-19. These characteristics were similar to participants in a previous PCR-based voluntary screening campaign in 2020, which showed that participation was associated with perceived efficacy of the campaign in controlling the epidemic, perceived susceptibility to COVID-19, and their trust of the government [20]. About one-third of nontesters did not believe the RAT campaign could control the epidemic. Such a low perceived efficacy might have prevented some citizens from participating in the campaign. The mandatory reporting of positive results had lowered the willingness of a certain proportion of people to participate. The higher odds of being vaccinated against COVID-19 and receiving more doses among testers could be a result of one's perceived susceptibility. Trust in the government that the policies are efficacious in controlling the epidemic could contribute to their engagement in vaccination and screening [20]. On the contrary, distrust in the government could also contribute to self-regulation to prevent infection and protect one's own interests, leading to passive compliance with antipandemic measures [21]. Separately, in the recent epidemic waves in Hong Kong, children were more likely to be asymptomatic and be infected through household transmissions rather than exogenous acquisition from schools [22], which might have prompted household members living with children to be tested to prevent transmission to the younger members if they tested positive.

Differentiating the epidemic situation between the time of the 2020 PCR-based screening campaign and the 2022 RAT-based one, a greater proportion of population had already been infected prior to the latter campaign, which may have affected the participation rate as some residents may not consider it necessary to undergo testing. The presence of symptoms was one of the indicators of SARS-CoV-2 infection, which prompts one to have testing performed [23]. This testing process was educated publicly to encourage people to get tested when they are presented with symptoms; on the other hand, people without symptoms may not be interested in taking the RAT. Although the figures in the previous study cannot be directly compared with findings from this study due to methodological differences,

it is worth noting that the participation rate of the previous PCR-based screening campaign was only 47%, and one-quarter of nonparticipants noted they did not have time for screening [20]. With three-quarters of respondents having participated in the self-RAT screening and just 1% concerned about spending time on it, the convenience and acceptability of an RAT-based voluntary screening campaign over a PCR-based one was highlighted.

As the official number of locally reported cases was just 5819 during the campaign period [6], the reported cases might have accounted for about 7.65% (95% CI 6.47%-9.14%) of all infections given the 1% positive rate; the rest being not reported despite statutory requirement or not detected because of either the low sensitivity in picking up early infections or that no screening had been performed. As only 62% participants who tested positive declared their results to the government, the number of reported cases could be underestimated. Since a higher proportion self-isolated after receiving a positive result, they were willing to take precaution to prevent onward transmissions in the community, although they did not declare the results to the authority. The high mobility of the younger working population and older adults were not reduced much during the epidemic, which predisposed them to the risk of infection [24]. Evidently, the Omicron wave has rapidly receded after over a million people have reportedly been diagnosed in the preceding 2 months. The size of the residual burden has, however, remained high and could easily be underestimated if statutory reporting statistics alone is used for epidemiologic surveillance. The high vaccination uptake rate and its protective effect might have played a role in minimizing the population risk. RAT mass screening has contributed to the assessment of the epidemiologic situation in a receding Omicron wave in Hong Kong.

Our population-based survey carried some limitations, notably self-selection bias with older and health-conscious adults and those testing positive being attracted to join the survey. The uptake rate may, therefore, be overestimated. In the analysis, we have performed age-adjustment to better reflect the situation in the population. The generalization of the results to the entire Hong Kong population should be cautioned due to the use of nonprobability sampling. Similar to other population-level surveys, recall and social desirability biases were inevitable. The survey was rolled out a week after the campaign to minimize recall bias. We assured participants of the anonymous nature of the survey to ensure the accuracy of the test result reported and compliance. By including proxy participants in the household, duplicate entries from the same household may have happened. We removed entries with the same session identifier to minimize duplicate records. As multiple brands of RAT were distributed and used with different sensitivity and specificity levels, their performance was unlikely to be perfect, so even if all participants were sampled and interpreted and reported the results correctly, the true infection status of all individuals might not have been determined definitively. It should also be noted the positive predictive value could be low in places where the prevalence is low [25]. Previous studies have, however, showed that RAT had a low false-positive rate [26] but an adequate sensitivity to identify asymptomatic and high–viral load cases [27]. The 1.03% positive rate found in this study was similar to the estimated daily point-prevalence on the last day of the campaign (0.76%, 95% CI 0.32-1.56%) [28], demonstrating the reliability of the results from this population-based survey. In conjunction with its low cost, RAT is well positioned to be used should mass screening be adopted as a cost-effective intervention in the public health control of COVID-19 [29]. As a perfect reporting rate of positive results is unlikely to be achieved, an accompanying survey would be needed and could be a feasible and appropriate means to estimate the actual prevalence in the community.

Conclusions

In a receding Omicron wave in 2022, a large proportion of residents in Hong Kong self-performed an RAT during the “voluntary universal rapid testing” campaign promoted by the government. RAT could be a useful adjunct not just for clinical diagnosis but also as a tool for public health surveillance and self-detection of infection. Accompanied with an information system, isolation facilities, and supporting services, voluntary mass RAT screening could support the estimation of the residual population burden and for supplementing risk assessment.

Acknowledgments

We thank Ms Nelly Cheung and Ms Sharon Chung for technical support in coordinating publicity and the recruitment of subjects. This study was funded by the Health and Medical Research Fund Commissioned Research on the Novel Coronavirus Disease (COVID1903008-Project A). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Geographical distribution of participants who self-performed a rapid antigen test (RAT) at least once during the campaign period.
[PNG File , 1397 KB - publichealth_v8i11e40175_app1.png]

References

1. Kwan TH, Wong NS, Yeoh EK, Lee SS. Mining relationships between transmission clusters from contact tracing data: an application for investigating COVID-19 outbreak. J Am Med Inform Assoc 2021 Oct 12;28(11):2385-2392 [FREE Full text] [doi: 10.1093/jamia/ocab175] [Medline: 34498059]
2. Kwan TH, Wong NS, Yeoh EK, Lee SS. Shifts of SARS-CoV-2 exposure settings in the transmission clusters of 2 epidemic waves in Hong Kong. Int J Environ Health Res 2022 Apr 18:1-13. [doi: 10.1080/09603123.2022.2064438] [Medline: 35437073]
3. Guo Z, Zhao S, Lee SS, Mok CKP, Wong NS, Wang J, et al. Superspreading potential of COVID-19 outbreak seeded by Omicron variants of SARS-CoV-2 in Hong Kong. J Travel Med 2022 Sep 17;29(6):049 [FREE Full text] [doi: 10.1093/jtm/taac049] [Medline: 35435992]
4. Smith DJ, Hakim AJ, Leung GM, Xu W, Schluter WW, Novak RT, et al. COVID-19 mortality and vaccine coverage - Hong Kong Special Administrative Region, China, January 6, 2022-March 21, 2022. MMWR Morb Mortal Wkly Rep 2022 Apr 15;71(15):545-548 [FREE Full text] [doi: 10.15585/mmwr.mm7115e1] [Medline: 35421076]
5. CE explains 3-day self-test exercise. The Government of the Hong Kong Special Administrative Region. 2022 Apr 02. URL: https://www.news.gov.hk/eng/2022/04/20220402/20220402_120347_856.html [accessed 2022-04-27]
6. Latest local situation of COVID-19. The Government of the Hong Kong Special Administrative Region. 2022 Apr 27. URL: https://www.coronavirus.gov.hk/eng/ [accessed 2022-04-27]
7. Boďová K, Kollár R. Spatial scales, patterns, and positivity trends of SARS-CoV-2 pandemics in mass rapid antigen testing in Slovakia. PLoS One 2021 Aug 25;16(8):e0256669 [FREE Full text] [doi: 10.1371/journal.pone.0256669] [Medline: 34432845]
8. Zhang C, Liang B, Xiong Z, Liang Z, Gong S, Zhou Z. Mass screening strategy for the SARS-CoV-2 Delta variant outbreak in Guangzhou, May 2021. Clin Microbiol Infect 2022 Jul;28(7):1040-1041. [doi: 10.1016/j.cmi.2022.03.020] [Medline: 35337978]
9. Peto J, Alwan NA, Godfrey KM, Burgess RA, Hunter DJ, Riboli E, 27 signatories. Universal weekly testing as the UK COVID-19 lockdown exit strategy. Lancet 2020 May 02;395(10234):1420-1421 [FREE Full text] [doi: 10.1016/S0140-6736(20)30936-3] [Medline: 32325027]
10. Pavelka M, Van-Zandvoort K, Abbott S, Sherratt K, Majdan M, CMMID COVID-19 working group, Inštitút Zdravotných Analýz, et al. The impact of population-wide rapid antigen testing on SARS-CoV-2 prevalence in Slovakia. Science 2021 May 07;372(6542):635-641 [FREE Full text] [doi: 10.1126/science.abf9648] [Medline: 33758017]
11. Bosetti P, Kiem CT, Yazdanpanah Y, Fontanet A, Lina B, Colizza V, et al. Impact of mass testing during an epidemic rebound of SARS-CoV-2: a modelling study using the example of France. Euro Surveill 2021 Jan;26(1):2001978 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.26.1.2001978] [Medline: 33413741]
12. Johanna N, Citrawijaya H, Wangge G. Mass screening vs lockdown vs combination of both to control COVID-19: a systematic review. J Public Health Res 2020 Oct 14;9(4):2011 [FREE Full text] [doi: 10.4081/jphr.2020.2011] [Medline: 33409247]
13. 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]
14. Peeling RW, Heymann DL, Teo Y, Garcia PJ. Diagnostics for COVID-19: moving from pandemic response to control. Lancet 2022 Feb 19;399(10326):757-768 [FREE Full text] [doi: 10.1016/S0140-6736(21)02346-1] [Medline: 34942102]
15. Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 1927 Jun;22(158):209-212. [doi: 10.1080/01621459.1927.10502953]
16. Census and Statistics Department. Table 1A : population by sex and age group. The Government of the Hong Kong Special Administrative Region. 2022. URL: https://www.censtatd.gov.hk/en/web_table.html?id=1A [accessed 2022-04-27]
17. Breslow NE, Day NE. Statistical methods in cancer research. volume II--the design and analysis of cohort studies. IARC Sci Publ 1987(82):1-406. [Medline: 3329634]
18. Delineation of district council constituency boundaries for the 2019 district council ordinary election. Hong Kong Electoral Affairs Commission. 2018. URL: https://www.eac.hk/pdf/distco/2019dc/en/statutory_criteria(Eng).pdf [accessed 2022-04-27]
19. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
20. Xin M, Lau JT, Lau MMC. Multi-dimensional factors related to participation in a population-wide mass COVID-19 testing program among Hong Kong adults: a population-based randomized survey. Soc Sci Med 2022 Feb;294:114692 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114692] [Medline: 35030396]
21. Ho LKK, Fong CS, Wan TTS. High level of (passive) compliance in a low-trust society: Hong Kong citizens' response towards the COVID-19 lockdown. Policing 2021;15(2):1046-1061. [doi: 10.1093/police/paaa090]
22. Chua GT, Wong JSC, Lam I, Ho PPK, Chan WH, Yau FYS, et al. Clinical characteristics and transmission of COVID-19 in children and youths during 3 waves of outbreaks in Hong Kong. JAMA Netw Open 2021 May 03;4(5):e218824 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.8824] [Medline: 33938934]
23. Tostmann A, Bradley J, Bousema T, Yiek W, Holwerda M, Bleeker-Rovers C, et al. Strong associations and moderate predictive value of early symptoms for SARS-CoV-2 test positivity among healthcare workers, the Netherlands, March

2020. Euro Surveill 2020 Apr;25(16):2000508 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.16.2000508] [Medline: 32347200]
24. Zhang N, Jia W, Wang P, Dung C, Zhao P, Leung K, et al. Changes in local travel behaviour before and during the COVID-19 pandemic in Hong Kong. Cities 2021 May;112:103139. [doi: 10.1016/j.cities.2021.103139] [Medline: 33589850]
25. Wu AHB. Screening the general population for SARS-CoV-2 virus and COVID-19 antibodies: a counterargument. J Appl Lab Med 2020 Sep 01;5(5):1107-1110 [FREE Full text] [doi: 10.1093/jalm/jfaa104] [Medline: 32609341]
26. Yang YP, Huang LL, Pan SJ, Xu D, Jiesisibieke ZL, Tung TH. False-positivity results in rapid antigen tests for SARS-CoV-2: an umbrella review of meta-analyses and systematic reviews. Expert Rev Anti Infect Ther 2022 Jul 29;20(7):1005-1013. [doi: 10.1080/14787210.2022.2070152] [Medline: 35452591]
27. Schwartz KL, McGeer AJ, Bogoch II. Rapid antigen screening of asymptomatic people as a public health tool to combat COVID-19. CMAJ 2021 Mar 29;193(13):E449-E452 [FREE Full text] [doi: 10.1503/cmaj.210100] [Medline: 33658247]
28. COVID-19 daily point-prevalence. School of Public Health, The University of Hong Kong. 2022 Oct 05. URL: https://covid19.sph.hku.hk [accessed 2022-10-05]
29. Du Z, Pandey A, Bai Y, Fitzpatrick MC, Chinazzi M, Pastore Y Piontti A, et al. Comparative cost-effectiveness of SARS-CoV-2 testing strategies in the USA: a modelling study. Lancet Public Health 2021 Mar;6(3):e184-e191 [FREE Full text] [doi: 10.1016/S2468-2667(21)00002-5] [Medline: 33549196]

Abbreviations

aOR: adjusted odds ratio
OR: odds ratio
PCR: polymerase chain reaction
RAT: rapid antigen test

Edited by A Mavragani; submitted 09.06.22; peer-reviewed by C Strong, Q Deng, S Galmiche; comments to author 12.08.22; revised version received 17.08.22; accepted 13.10.22; published 09.11.22.

Please cite as:
Kwan TH, Wong NS, Chan CP, Yeoh EK, Wong SYS, Lee SS
Mass Screening of SARS-CoV-2 With Rapid Antigen Tests in a Receding Omicron Wave: Population-Based Survey for Epidemiologic Evaluation
JMIR Public Health Surveill 2022;8(11):e40175
URL: https://publichealth.jmir.org/2022/11/e40175
doi:10.2196/40175
PMID:36240027

Original Paper

Effect of Comorbidities on the Infection Rate and Severity of COVID-19: Nationwide Cohort Study With Propensity Score Matching

Jiyong Kim[1], MD, PhD; Seong Hun Park[2], MS; Jong Moon Kim[3,4], MD

[1]Department of Rehabilitation, Inje University Ilsanpaik Hospital, Goyang, Republic of Korea

[2]Statistical analysis company, HYMS, Gwangju, Republic of Korea

[3]Department of Rehabilitation Medicine, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea

[4]Department of Medical Informatics Big Data Center, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea

Corresponding Author:
Jong Moon Kim, MD
Department of Rehabilitation Medicine
CHA Bundang Medical Center
CHA University
59 Yatap-ro
Bundang-gu
Seongnam, 13496
Republic of Korea
Phone: 82 31 780 5456
Fax: 82 31 780 3449
Email: jmkim1013@gmail.com

Abstract

Background: A vaccine against COVID-19 has been developed; however, COVID-19 transmission continues. Although there have been many studies of comorbidities that have important roles in COVID-19, some studies have reported contradictory results.

Objective: This study was conducted using real-world data from COVID-19 patients in South Korea and aimed to investigate the impact of patient demographics and comorbidities on the infection rate and severity of COVID-19.

Methods: Data were derived from a nationwide South Korean COVID-19 cohort study with propensity score (PS) matching. We included infected individuals who were COVID-19–positive between January 1, 2020, and May 30, 2020, and PS-matched uninfected controls. PS matching was performed to balance the baseline characteristics of each comorbidity and to adjust for potential confounders, such as age, sex, Charlson Comorbidity Index, medication, and other comorbidities, that were matched with binary variables. The outcomes were the confirmed comorbidities affecting the infection rate and severity of COVID-19. The endpoints were COVID-19 positivity and severe clinical outcomes of COVID-19 (such as tracheostomy, continuous renal replacement therapy, intensive care unit admission, ventilator use, cardiopulmonary resuscitation, and death).

Results: The COVID-19 cohort with PS matching included 8070 individuals with positive COVID-19 test results and 8070 matched controls. The proportions of patients in the severe group were higher for individuals 60 years or older (severe clinical outcomes for those 60 years or older, 16.52%; severe clinical outcomes for those of other ages, 2.12%), those insured with Medicaid (Medicaid, 10.81%; other insurance, 5.61%), and those with disabilities (with disabilities, 18.26%; without disabilities, 5.07%). The COVID-19 infection rate was high for patients with pulmonary disease (odds ratio [OR] 1.88; 95% CI 1.70-2.03), dementia (OR 1.75; 95% CI 1.40-2.20), gastrointestinal disease (OR 1.74; 95% CI 1.62-1.88), stroke (OR 1.67; 95% CI 1.23-2.27), hepatobiliary disease (OR 1.31; 95% CI 1.19-1.44), diabetes mellitus (OR 1.28; 95% CI 1.16-1.43), and cardiovascular disease (OR 1.20; 95% CI 1.07-1.35). In contrast, it was lower for individuals with hyperlipidemia (OR 0.73; 95% CI 0.67-0.80), autoimmune disease (OR 0.73; 95% CI 0.60-0.89), and cancer (OR 0.73; 95% CI 0.62-0.86). The severity of COVID-19 was high for individuals with kidney disease (OR 5.59; 95% CI 2.48-12.63), hypertension (OR 2.92; 95% CI 1.91-4.47), dementia (OR 2.92; 95% CI 1.91-4.47), cancer (OR 1.84; 95% CI 1.15-2.94), pulmonary disease (OR 1.72; 95% CI 1.35-2.19), cardiovascular disease (OR 1.54; 95% CI 1.17-2.04), diabetes mellitus (OR 1.43; 95% CI 1.09-1.87), and psychotic disorders (OR 1.29; 95% CI 1.01-6.52). However, it was low for those with hyperlipidemia (OR 0.78; 95% CI 0.60-1.00).

Conclusions: Upon PS matching considering the use of statins, it was concluded that people with hyperlipidemia could have lower infection rates and disease severity of COVID-19.

(JMIR Public Health Surveill 2022;8(11):e35025) doi:10.2196/35025

KEYWORDS

COVID-19; comorbidity; infection rate; severity of illness index; hyperlipidemia

Introduction

The World Health Organization declared that COVID-19 was a pandemic in March 2020. By August 2022, approximately 600 million individuals had been infected, and more than 6 million had died. Since then, vaccines and therapeutic agents for COVID-19 have been developed. However, the current number of individuals with COVID-19 is still the same as that 1 year ago because it has not yet been eradicated [1]. COVID-19 can result in an asymptomatic presentation or flu-like symptoms. Some patients are admitted to the hospital for conservative treatment, and some require intensive care unit admission. Moreover, some patients may die as a result of COVID-19 [2,3]. As the number of individuals with COVID-19 increases, it is important to identify those who are vulnerable to severe COVID-19 to effectively manage health care resources accordingly and to improve the prognosis [4,5].

Since the COVID-19 outbreak, many studies of the demographic factors that predispose individuals to infection and of the identification of comorbidities of infected individuals have been performed. Most studies have reported similar overall results; however, some results of these studies are contradictory [6-9]. These differences in results may be attributable to the diversity of patients and medical systems in various countries worldwide. Most previous studies on comorbidities analyzed the baseline characteristics of people infected with COVID-19 without considering the bias caused by various factors that influence COVID-19. For instance, to determine whether hyperlipidemia affects the severity of COVID-19, it is necessary to control for statins, which are often used by individuals with hyperlipidemia. Although some studies suggested that statins might have a role in reducing the severity of COVID-19 [10,11], most studies did not confirm the use of statins; they only reported the effect of hyperlipidemia [12-16]. Hence, it is difficult to accurately determine the effect of hyperlipidemia on the severity of COVID-19. We investigated the effects of patient comorbidities on the infection rate and severity of COVID-19. Bias was reduced by propensity score (PS) matching for various variables that may affect COVID-19. We also analyzed the demographic characteristics of patients with COVID-19.

Methods

Study Design and Participants

We conducted a large-scale cohort study using a South Korean National Health Insurance claims database [17]. In South Korea, all citizens are registered in the Korean National Health Insurance Service (KNHIS) database. The KNHIS uses a nationwide, large-scale database system including information regarding the diagnostic codes from the International Classification of Diseases (ICD)-10, the names of the procedures performed, prescription drugs, hospital information, direct medical costs of inpatient and outpatient treatments, and medical insurance premiums. Because all Koreans are given unique identification numbers at birth that are used in the KNHIS, the health records of patients are not duplicated nor omitted [18,19]. For COVID-19 studies, KNHIS provides a COVID-19 cohort that includes people infected with COVID-19 and a control group that had never been infected. From January 1, 2020, to May 31, 2020, disease codes B342, B972, U071, U072, MT043, and 3/02 were used to identify patients with confirmed COVID-19. Data from the control group of individuals who were not previously diagnosed with COVID-19 were adjusted for sex, age, and region of residence. Moreover, the number of participants in the control group was 15 times the number of confirmed COVID-19 cases.

Ethical Considerations

This study was approved by the relevant institutional review board and research ethics committee (ISPAIK 2020-06-048-001). The need for written consent was formally waived by the ethics committee. This study used the NHIS-2020-1-328 database provided by the KNHIS in 2020.

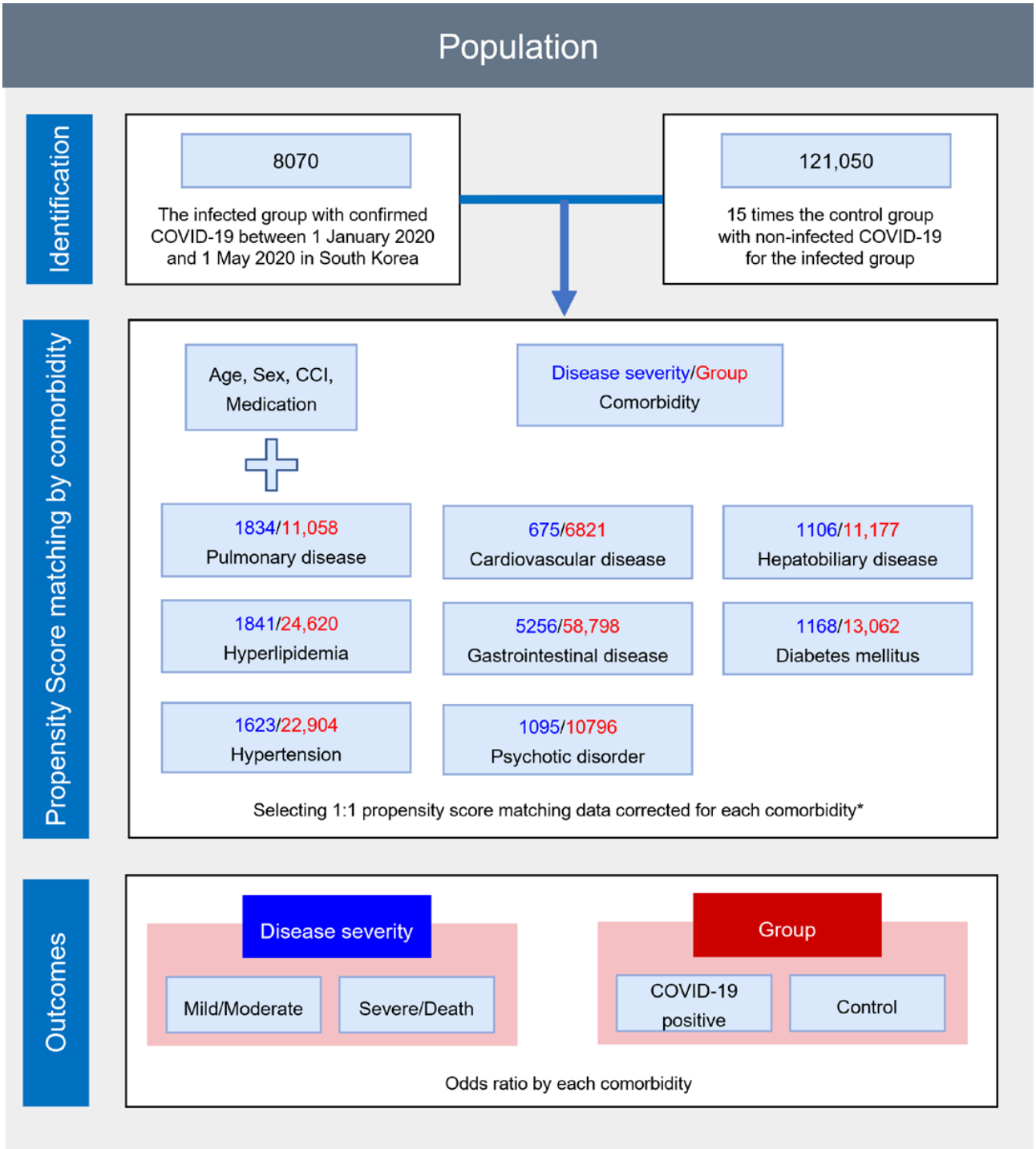
Study Population

In accordance with the World Health Organization guidelines, laboratory confirmation of COVID-19 was defined as a positive result of a real-time reverse-transcription polymerase chain reaction assay using a sample obtained with nasal and pharyngeal swabs [20]. We combined the claims-based data from the KNHIS between January 1, 2015, and May 31, 2020, and extracted information regarding age, sex, and region of residence from the insurance eligibility data (Figure 1). The Charlson Comorbidity Index (CCI) score was calculated using the ICD-10 codes and previously reported methods [21]. Certain underlying medications and diseases with a high risk of serious illness attributable to SARS-CoV-2, which causes COVID-19, were studied and reported by the Centers for Disease Control and Prevention (CDC) and previous meta-analysis studies [6-9,22]. In these studies, we selected factors to use for PS matching in the analysis (Tables S1 and S2 in Multimedia Appendix 1). Only those (pulmonary disease, cardiovascular disease, hepatobiliary disease, hyperlipidemia, gastrointestinal disease, diabetes mellitus, hypertension, and psychotic disorder) with more than 500 people with COVID-19 were selected because a small number of people with corresponding comorbidities might cause statistical bias (Multimedia Appendix 2). A history of underlying diseases (pulmonary disease, cardiovascular disease, kidney disease, hepatobiliary disease, hyperlipidemia, gastrointestinal disease, diabetes mellitus, hypertension, psychotic disorder, dementia, stroke, neurologic

disorder, autoimmune disease, and cancer) was confirmed by the assignment of at least two claims within 1 year using the appropriate ICD-10 code.

We used various PS matching methods for factors affecting COVID-19: (1) matching for age, sex, and CCI; (2) additional matching for comorbidities; and (3) additional matching for medications. Finally, the results from (3) were used (Tables S3 and S4 in Multimedia Appendix 1). The financial revenue of the National Health Insurance of Korea consists of contributions from the insured and government subsidies, which can be used to analyze socioeconomic status. The contributions to the National Health Insurance differ according to the family income level. The higher the income, the greater the contribution to the National Health Insurance. Income was divided into 5 categories for the purpose of statistical analyses. The first category is Medicaid, and the successive categories include progressively higher (by 25%) income groups. Disability grades were categorized as mild or severe based on the KNHIS database information for people registered with the Korean government.

Figure 1. Disposition of patients in the KNHIS-COVID cohort (South Korea; January 1 to May 31, 2020). CCI: Charlson Comorbidity Index; KNHIS: Korean National Health Insurance Service.



Outcomes

To determine the severity of disease according to the demographic factors of COVID-19–infected patients, the severity scale was divided into the following 4 grades: mild, moderate, severe, and death. In South Korea, patients with asymptomatic or mild symptoms are discharged when a negative COVID-19 test result is confirmed 2 weeks after hospitalization. This time period also corresponds to the period of self-isolation. When we checked the hospitalization period of COVID-19–infected patients, the hospitalization period peaked on day 16 and decreased thereafter. Based on this result, a hospitalization period of ≤16 days was defined as the mild grade corresponding to asymptomatic or mild symptoms. The severe grade was defined as the need for tracheostomy, continuous renal replacement therapy, intensive care unit admission, ventilator use, and cardiopulmonary resuscitation. The moderate grade was defined as a hospitalization period >16 days but not requiring treatment corresponding to the severe grade.

The primary aim of this study was to compare the severity grades of the COVID-19–infected and control groups based on demographic factors, comorbidities, and complications. The secondary aim was to perform PS matching for comparisons. We identified the infection rate and severity (severe and death or mild and moderate) of COVID-19 according to the comorbid conditions.

Statistical Analysis

We performed PS matching to balance the baseline characteristics of each comorbidity (existence or nonexistence) and to adjust for potential confounders. Because we focused on each comorbidity, PS matching was performed for each comorbidity. The PS was estimated using a logistic regression model and calculating the predicted probability of covariates. Age and CCI (0, 1, or ≥2) were matched with continuous variables. Sex, medication, and other comorbidities were matched with binary variables. We assessed the PS matching of the comorbidity existence using a 1:1 ratio, the greedy nearest neighbor algorithm, and a scale with a caliper of 0.25 (Multimedia Appendix 2). Data obtained after PS matching were analyzed by calculating the odds ratios (ORs) with 95% CIs for the infection rate and severity (severe and death or mild and moderate) of COVID-19. Statistical analyses were performed using SAS version 9.4 (SAS Institute Inc, Cary, NC).

Patient and Public Involvement

No patient was directly involved in designing the study question or in conducting the study. No patients were asked for advice regarding the interpretation or writing of the results. There are no plans to involve patients or the relevant patient community in the dissemination of study findings at this time.

Results

Clinical Characteristics of the Study Population

A total of 8070 individuals had positive COVID-19 results according to the reverse-transcription polymerase chain reaction assay. We identified 121,050 uninfected individuals as control participants (Multimedia Appendix 2). The demographic characteristics of the entire cohort are displayed in Table 1. The COVID-19 severity grade was mild for 2419 (2419/8070, 29.98%) individuals, moderate for 5160 (5160/8070, 63.94%) individuals, severe for 254 (254/8070, 3.15%) individuals, and death for 237 (237/8070, 2.94%) individuals. Among the total sample of infected individuals, 3236 (3236/8070, 40.10%) were male. Most patients were in their fifth (1567/8070, 19.42%) or sixth (1199/8070, 14.86%) decade of life. In terms of the medical insurance grade, which indicates socioeconomic status, those receiving Medicaid had high rates of severe grade and death. However, there were no obvious trends for the other grades. Individuals with disabilities had more severe infections and a much higher case fatality rate (Table 1). Those with COVID-19 had a medical history of gastrointestinal disease (n=5256), pulmonary disease (n=2539), hyperlipidemia (n=1841), and hypertension (n=1623). The case fatality rate was high for individuals with dementia (74/235, 31.5%), kidney disease (25/86, 29%), and cardiovascular disease (110/675, 16.3%; Table 2). After COVID-19 was confirmed, gastrointestinal disease (n=2912), pulmonary disease (n=2398), and hepatobiliary disease (n=1248) were the most common complications (Table 3).

Table 1. Baseline characteristics of the study population, including those infected (n=8070) and uninfected (n=121,050; controls) with COVID-19 in the Korean National Health Insurance Service (KNHIS)-COVID cohort (South Korea; January 1, 2020, to May 31, 2020).

Variables	Severity of COVID-19, n (%)				Total COVID-19 cases, n (%)	Controls[a,b], n (%)
	Mild	Moderate	Severe	Death		
Sex						
Male	894 (27.63)	2073 (64.06)	135 (4.17)	134 (4.14)	3236 (40.10)	48,540 (40.10)
Female	1525 (31.55)	3087 (63.86)	119 (2.46)	103 (2.13)	4834 (59.90)	72,510 (59.90)
Age (years)						
0-9	32 (39.51)	45 (55.56)	4 (4.94)	0 (0.00)	81 (1.00)	1215 (1.00)
10-19	77 (27.90)	195 (70.65)	4 (1.45)	0 (0.00)	276 (3.42)	4140 (3.42)
20-29	697 (33.88)	1342 (65.24)	18 (0.88)	0 (0.00)	2057 (25.49)	30,855 (25.49)
30-39	273 (32.81)	541 (65.02)	17 (2.04)	1 (0.12)	832 (10.31)	12,480 (10.31)
40-49	358 (34.56)	655 (63.22)	20 (1.93)	3 (0.29)	1036 (12.84)	15,540 (12.84)
50-59	504 (32.16)	1006 (64.20)	43 (2.74)	14 (0.89)	1567 (19.42)	23,505 (19.42)
60-69	322 (26.86)	776 (64.72)	66 (5.50)	35 (2.92)	1199 (14.86)	17,985 (14.86)
70-79	122 (19.77)	389 (63.05)	40 (6.48)	66 (10.70)	617 (7.65)	9255 (7.65)
≥80	34 (8.40)	211 (52.10)	42 (10.37)	118 (29.14)	405 (5.02)	6075 (5.02)
Medical insurance[a]						
Medicaid	186 (27.56)	416 (61.63)	31 (4.59)	42 (6.22)	675 (8.36)	4424 (3.65)
Grade 1	604 (32.95)	1146 (62.52)	44 (2.40)	39 (2.13)	1833 (22.71)	26,258 (21.69)
Grade 2	462 (30.84)	971 (64.82)	36 (2.40)	29 (1.94)	1498 (18.56)	24,270 (20.05)
Grade 3	484 (29.02)	1081 (64.81)	58 (3.48)	45 (2.70)	1668 (20.67)	27,521 (22.74)
Grade 4	637 (28.07)	1475 (65.01)	79 (3.48)	78 (3.44)	2269 (28.12)	37,241 (30.76)
Disability grade						
Mild	65 (20.44)	192 (60.38)	28 (8.81)	33 (10.38)	318 (3.94)	4367 (3.61)
Severe	83 (27.57)	166 (55.15)	18 (5.98)	34 (11.30)	301 (3.73)	2275 (1.88)
Total	2419 (29.98)	5160 (63.94)	254 (3.15)	237 (2.94)	8070 (100)	121,050 (100)

[a]Participants from some specific groups, such as soldiers, were not included.

[b]The uninfected controls were adjusted for sex, age, and region, resulting in a figure equivalent to 15 times the number of confirmed COVID-19 cases in the KNHIS-COVID cohort.

Table 2. Baseline characteristics of comorbidities of the study population, including those infected with (n=8070) and not infected with (n=121,050; controls) COVID-19 in the Korean National Health Insurance Service (KNHIS)-COVID cohort (South Korea; January 1, 2020, to May 31, 2020).

Comorbidities	Severity of COVID-19, n (%)				Total COVID-19 cases, n (%)	Controls, n (%)
	Mild	Moderate	Severe	Death		
Pulmonary disease	501 (27.32)	1129 (61.56)	80 (4.36)	124 (6.76)	1834 (22.73)	11,058 (9.14)
Cardiovascular disease	137 (20.44)	387 (57.33)	40 (5.93)	110 (16.30)	675 (8.36)	6821 (5.63)
Kidney disease	15 (17.44)	37 (43.02)	9 (10.47)	25 (29.07)	86 (1.07)	1018 (0.84)
Hepatobiliary disease	281 (25.41)	690 (62.39)	51 (4.62)	84 (7.59)	1106 (13.71)	10,177 (8.41)
Hyperlipidemia	466 (25.31)	1158 (62.90)	97 (5.27)	120 (6.52)	1841 (22.81)	24,620 (20.34)
Gastrointestinal disease	1559 (29.66)	3334 (63.43)	181 (3.44)	182 (3.46)	5256 (65.13)	58,798 (48.57)
Diabetes mellitus	246 (12.50)	720 (61.64)	69 (5.91)	133 (11.39)	1168 (14.47)	13,062 (10.79)
Hypertension	348 (21.44)	1010 (62.23)	95 (5.85)	170 (10.47)	1623 (20.11)	22,904 (18.92)
Psychotic disorder	256 (23.38)	672 (61.37)	67 (6.12)	100 (9.13)	1095 (13.57)	10,796 (8.92)
Dementia	19 (8.09)	116 (49.36)	26 (11.06)	74 (31.49)	235 (2.91)	1429 (1.18)
Stroke	18 (15.13)	73 (61.34)	10 (8.40)	18 (15.13)	119 (1.47)	809 (0.67)
Neurogenic disorder	25 (21.01)	75 (63.03)	4 (3.36)	15 (12.61)	119 (1.47)	1037 (0.86)
Autoimmune disease	51 (28.33)	115 (63.89)	5 (2.78)	9 (5.00)	180 (2.23)	2423 (2.00)
Cancer	55 (20.45)	158 (58.74)	19 (7.06)	37 (13.75)	269 (3.33)	3275 (2.71)

Table 3. Baseline characteristics of complications of the study population, including those infected with (n=8070) and not infected with (n=121,050; controls) COVID-19 in the Korean National Health Insurance Service (KNHIS)-COVID cohort (South Korea; January 1, 2020, to May 31, 2020).

Complications	Severity of COVID-19, n (%)				Total COVID-19 cases, n (%)	Controls, n (%)
	Mild	Moderate	Severe	Death		
Pulmonary disease	542 (22.60)	1580 (65.89)	180 (7.51)	96 (4.00)	2398 (29.71)	2027 (1.67)
Cardiovascular disease	161 (20.77)	475 (61.29)	85 (10.97)	54 (6.97)	775 (9.60)	1223 (1.01)
Kidney disease	12 (11.01)	50 (45.87)	22 (20.18)	25 (22.94)	109 (1.35)	226 (0.19)
Hepatobiliary disease	321 (25.72)	789 (63.22)	107 (8.57)	31 (2.48)	1248 (15.46)	4359 (3.60)
Gastrointestinal disease	824 (28.30)	1906 (65.45)	129 (4.43)	53 (1.82)	2912 (36.08)	20,477 (16.92)
Stroke	6 (13.33)	27 (60.00)	10 (22.22)	2 (4.44)	45 (0.56)	298 (0.25)
Neurogenic disorder	21 (28.77)	40 (54.79)	12 (16.44)	3 (4.11)	73 (0.90)	169 (0.14)
Sepsis	16 (10.46)	71 (46.41)	38 (24.84)	28 (18.30)	153 (1.90)	22 (0.02)

Risks of COVID-19 Positivity and Disease Severity According to Comorbidities

To identify differences according to comorbidity, predispositions were matched between the COVID-19–infected group and uninfected control group. No significant imbalances in the demographics and clinical characteristics were observed when they were assessed using the standardized mean difference within groups of PS-matched cohorts, which included the standardized mean difference of binary type variables <0.1. PS-matched ORs were checked for age, sex, CCI, medication, and comorbidities. When the control group and COVID-19–infected group were compared, COVID-19 was likely to occur in individuals with a history of the diseases and medical conditions but not for those with a history of hyperlipidemia (OR 0.73; 95% CI 0.67-0.80), autoimmune disease (OR 0.73; 95% CI 0.60-0.89), or cancer (OR 0.73; 95% CI 0.62-0.86; Table 4). The severity grade was high for COVID-19–infected individuals with pulmonary disease (OR 1.72; 95% CI 1.35-2.19), cardiovascular disease (OR 1.54; 95% CI 1.17-2.04), kidney disease (OR 5.59; 95% CI 2.48-12.63), diabetes mellitus (OR 1.43; 95% CI 1.09-1.87), hypertension (OR 1.63; 95% CI 1.23-2.15), psychotic disorder (OR 1.29; 95% CI 1.01-6.52), dementia (OR 2.92; 95% CI 1.91-4.47), or cancer (OR 1.84; 95% CI 1.15-2.94). However, the severity grade was low for COVID-19–infected individuals with hyperlipidemia (OR 0.70; 95% CI 0.55-0.90; Table 5 and Table S5 in Multimedia Appendix 1).

Table 4. Propensity score–matched (age, sex, Charlson Comorbidity Index, medications, and comorbidities) baseline characteristics and COVID-19 infection positivity rates according to comorbidity in the Korean National Health Insurance Service (KNHIS)-COVID cohort (South Korea; January 1, 2020, to May 31, 2020).

Comorbidities	Group[a], n		Odds ratio (95% CI)
	COVID-19	Control	
Pulmonary disease[b]	2880	22,900	1.88 (1.70-2.03)
Cardiovascular disease[b]	1245	13,733	1.20 (1.07-1.35)
Kidney disease	171	2037	1.01 (0.74-1.39)
Hepatobiliary disease[b]	1959	20,505	1.31 (1.19-1.44)
Hyperlipidemia[b]	2375	25,663	0.73 (0.67-0.80)
Gastrointestinal disease[b]	3164	43,370	1.74 (1.62-1.88)
Diabetes mellitus[b]	1544	16,468	1.28 (1.16-1.43)
Hypertension	1483	16,473	1.04 (0.93-1.15)
Psychotic disorder	2122	21,458	1.06 (0.97-1.16)
Dementia[b]	365	2753	1.75 (1.40-2.20)
Stroke[b]	194	1662	1.67 (1.23-2.27)
Neurologic disease	223	2089	1.16 (0.88-1.53)
Autoimmune disease[c]	421	4785	0.73 (0.60-0.89)
Cancer[c]	629	6459	0.73 (0.62-0.86)

[a]We assessed each propensity score–matched comorbidity using a 1:1 ratio for those in the COVID-19 and control groups.

[b]Comorbidity with more susceptibility to COVID-19.

[c]Comorbidity with less susceptibility to COVID-19.

Table 5. Propensity score–matched (age, sex, Charlson Comorbidity Index, medications, and comorbidities) baseline characteristics and clinical outcomes of COVID-19 among patients in the mild or moderate group and those in the severe or death group according to the comorbidity of patients with laboratory-confirmed COVID-19 infection in the Korean National Health Insurance Service (KNHIS)-COVID cohort (South Korea; January 1, 2020, to May 31, 2020).

Comorbidities	Severity[a], n		Odds ratio (95% CI)
	Mild + moderate	Severe + death	
Pulmonary disease[b]	3031	307	1.72 (1.35-2.19)
Cardiovascular disease[b]	1078	252	1.54 (1.17-2.04)
Kidney disease[b]	129	43	5.59 (2.48-12.63)
Hepatobiliary disease	1899	253	1.01 (0.78-1.31)
Hyperlipidemia[c]	2204	270	0.78 (0.60-1.00)
Gastrointestinal disease	3060	206	1.00 (0.75-1.33)
Diabetes mellitus[b]	1465	259	1.43 (1.09-1.87)
Hypertension[b]	1262	248	2.92 (1.91-4.47)
Psychotic disorder[b]	1846	288	1.29 (1.01-6.52)
Dementia[b]	305	137	2.92 (1.91-4.47)
Stroke	188	50	1.36 (0.72-2.54)
Neurologic disease	198	40	0.88 (0.45-1.75)
Autoimmune disease	338	20	2.25 (0.92-6.52)
Cancer[b]	448	88	1.84 (1.15-2.94)

[a]We assessed each propensity score–matched comorbidity using a 1:1 ratio for those in the mild and moderate group and those in the severe and death group.

[b]Comorbidity with increasing COVID-19 severity.

[c]Comorbidity with decreasing COVID-19 severity.

Discussion

Principal Findings

This study was a retrospective cohort study conducted in South Korea from January 2020 to May 2020. It involved confirmed COVID-19 patients with medical insurance. Previous studies of the demographic factors of individuals with COVID-19 showed that male sex, old age, and low income were factors likely associated with COVID-19 with a high severity grade [23,24]. In this study, more women had COVID-19, but the severity grade of COVID-19 was higher for men; this was directly proportional to age, especially for men older than 70 years. All medical expenses for COVID-19 are paid for by the South Korea government; therefore, all patients, including those receiving Medicaid, received the same level of care for COVID-19. Although there was no difference in medical care, those with Medicaid had the lowest income level and a higher severity grade; however, there were no differences between the groups with grades 1 to 4 medical insurance. For individuals with disabilities, the incidence was slightly higher than that of the control group. However, the severity grade was much higher than that of other individuals infected with COVID-19.

Other studies of COVID-19 reported that SARS-CoV-2 binds to the angiotensin-converting enzyme 2 (ACE2) receptor through the viral structural spike protein at the onset of infection [25]. ACE2 is expressed to varying degrees in almost all human organs. ACE2 is highly expressed in cardiomyocytes, proximal tubule cells of the kidney, and bladder urinary tract cells. Additionally, it is abundantly expressed in intestinal cells of the small intestine, especially in the ileum [25-28]. Therefore, most critically ill patients with COVID-19 experience multiple organ injuries, including acute lung injuries, acute kidney injuries, cardiac injuries, hepatobiliary disease, and pneumothorax [29]. Therefore, to analyze the effect of each comorbidity on the COVID-19 infection severity grade, it is necessary to consider other comorbidities.

Each demographic factor, comorbidity, and medication may influence each other, resulting in different outcomes in terms of the infection rate and severity of COVID-19. When analyzing comorbidities with hypertension, the effect of hypertension on the infection rate and severity of COVID-19 experienced by an 80-year-old woman with asthma and that of a 30-year-old man without an underlying medical condition may be different. Accurate results can be obtained for sufficiently studied diseases by controlling for only important factors. However, in the case of understudied diseases, such as COVID-19, various factors should be considered. In this study, PS matching was performed for various factors that could affect COVID-19, to minimize bias. When selecting a factor for PS matching, in order to select objective data, data provided by the CDC and meta-analysis studies were used. However, there was a limit, as data may

change as research on COVID-19 progresses. Most of the results obtained were similar to those of previously published studies; however, some results were conflicting. For people with cancer and autoimmune disease, infection rates were even lower; these results were possibly affected by reducing social contact because of the risk of COVID-19 infection. Exposure to COVID-19 is an important factor that can affect the infection rate of COVID-19. Individuals with hyperlipidemia had a low COVID-19 infection rate and low severity grade. Previous studies reported that hyperlipidemia should be managed to prevent COVID-19 because high cholesterol levels induce inflammation and increase ACE2 availability [30-32]. Moreover, the use of statins for patients with COVID-19 reduced mortality by interfering with the mevalonate pathway and because of their antiviral effects [10,11,33]. However, some studies have shown that people with low lipid levels are more susceptible to and have more severe COVID-19 infection [34-41]. A meta-analysis published in 2022 indicated that patients with severe COVID-19 had lower total cholesterol levels (pooled mean difference –10.4; 95% CI –18.7 to –2.2), low-density lipoprotein cholesterol levels (pooled mean difference –4.4; 95% CI –8.4 to –0.42), and high-density lipoprotein cholesterol levels (pooled mean difference –4.4; 95% CI –6.9 to –1.8) on admission compared with patients with non-severe disease [42]. This may be similar to the "obesity paradox," which states that mild obesity is advantageous to improvements after stroke [43,44]. Mild obesity can withstand the systemic catabolic imbalance with impaired metabolic efficiency and body tissue degradation that occur after stroke. Hyperlipidemia may also have a role in minimizing the severity of COVID-19.

Limitations

Our study has several limitations. As a limitation of most medical data studies, there is bias caused by confounding factors that may affect our results. When selecting factors for PS matching, information from the CDC and meta-analysis studies were used to select objective data, but these data may change as research on COVID-19 progresses. Further, we defined diseases based on the ICD codes provided in the insurance claims data. There may have been additional unmeasured confounders influencing our results, including genetic polymorphisms, smoking, body mass index, and exposure to the virus. In this study, the infection rate of COVID-19 may have been influenced by the degree of exposure to COVID-19, which may be an important factor in addition to the comorbidity factors. However, the influence of COVID-19 itself could be confirmed because the bias was less than that of previous studies. One race in South Korea comprises more than 95% of the population; hence, there was minimal racial bias compared with previous studies. Because the government funds the treatment for COVID-19 in South Korea and because the medical facilities for COVID-19 treatment are ubiquitous, there was minimal economic bias. The PS matching was performed for sex, age, CCI, comorbidity, and medication, including statins (standardized mean difference <0.1). Hence, selection bias was minimized. Therefore, more accurate information regarding the incidence of COVID-19 and its severity according to comorbidities was provided.

This study was based on data from patients who experienced COVID-19 during the early outbreak period; therefore, that strain may differ from the current strain of COVID-19. However, an accurate analysis of recent COVID-19 strains, including Omicron, is difficult because the effects of acquired or natural immunity and vaccination are mixed. Data at the time of its early onset can provide fundamental information, including regarding mutations that may occur in the future.

Conclusions

Although the severity of COVID-19 has decreased, its hospitalization rate has not decreased significantly, and its burden on medical facilities continues; therefore, an analysis of comorbidities is still important. Therefore, many studies of comorbidities that affect COVID-19 have been published; however, some have reported conflicting results. This may be because various factors such as medication and comorbidities, in addition to demographic factors such as age and sex, affect the infection rate and severity of COVID-19. It is necessary to analyze as many factors as possible to obtain more accurate data regarding COVID-19. Based on the results of previous studies, this study tried to derive objective results by considering various factors affecting COVID-19. In conclusion, certain comorbidities known as risk factors in previous studies increase the infection rate and severity of COVID-19. However, hyperlipidemia decreases the infection rate and severity. These results can be utilized to effectively manage COVID-19.

Acknowledgments

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT; number 2016M3A9E8941108).

The NRF had no role in the design and conduct of the study. The authors are responsible for the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication.

Authors' Contributions

JMK and JK had full access to all of the data in the study, take responsibility for the integrity of the data and the accuracy of the data analysis, conceived and designed the study, and drafted the manuscript. JMK obtained the funding. SHP and JMK acquired, analyzed, and interpreted the data. SHP and JK performed the statistical analyses.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 60 KB - publichealth_v8i11e35025_app1.docx]

Multimedia Appendix 2
Statistical analysis method for the baseline characteristics of comorbidity and propensity score–matched infection rate and severity of COVID-19 in the Korean National Health Insurance Service (KNHIS)-COIVD cohort (South Korea; January 1, 2020, to May 31, 2020; pulmonary disease used as the example).
[PNG File , 177 KB - publichealth_v8i11e35025_app2.png]

References

1. Coronavirus disease (COVID-19). World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2022-10-29]
2. Argenziano MG, Bruce SL, Slater CL, Tiao JR, Baldwin MR, Barr RG, et al. Characterization and clinical course of 1000 patients with coronavirus disease 2019 in New York: retrospective case series. BMJ 2020 May 29;369:m1996 [FREE Full text] [doi: 10.1136/bmj.m1996] [Medline: 32471884]
3. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, the Northwell COVID-19 Research Consortium. 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]
4. Moghadas SM, Shoukat A, Fitzpatrick MC, Wells CR, Sah P, Pandey A, et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. Proc Natl Acad Sci U S A 2020 Apr 21;117(16):9122-9126 [FREE Full text] [doi: 10.1073/pnas.2004064117] [Medline: 32245814]
5. Murray CJL. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. medRxiv Preprint posted online March 30, 2020. [FREE Full text] [doi: 10.1101/2020.03.27.20043752]
6. Fang X, Li S, Yu H, Wang P, Zhang Y, Chen Z, et al. Epidemiological, comorbidity factors with severity and prognosis of COVID-19: a systematic review and meta-analysis. Aging (Albany NY) 2020 Jul 13;12(13):12493-12503 [FREE Full text] [doi: 10.18632/aging.103579] [Medline: 32658868]
7. Kim DW, Byeon KH, Kim J, Cho KD, Lee N. The correlation of comorbidities on the mortality in patients with COVID-19: an observational study based on the Korean National Health Insurance big data. J Korean Med Sci 2020 Jul 06;35(26):e243 [FREE Full text] [doi: 10.3346/jkms.2020.35.e243] [Medline: 32627443]
8. Wang B, Li R, Lu Z, Huang Y. Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. Aging (Albany NY) 2020 Apr 08;12(7):6049-6057 [FREE Full text] [doi: 10.18632/aging.103000] [Medline: 32267833]
9. Gold MS, Sehayek D, Gabrielli S, Zhang X, McCusker C, Ben-Shoshan M. COVID-19 and comorbidities: a systematic review and meta-analysis. Postgrad Med 2020 Nov 14;132(8):749-755. [doi: 10.1080/00325481.2020.1786964] [Medline: 32573311]
10. Gupta A, Madhavan MV, Poterucha TJ, DeFilippis EM, Hennessey JA, Redfors B, et al. Association between antecedent statin use and decreased mortality in hospitalized patients with COVID-19. Nat Commun 2021 Feb 26;12(1):1325 [FREE Full text] [doi: 10.1038/s41467-021-21553-1] [Medline: 33637713]
11. Lee SW, Kim SY, Moon SY, Yoo IK, Yoo E, Eom GH, et al. Statin use and COVID-19 infectivity and severity in South Korea: two population-based nationwide cohort studies. JMIR Public Health Surveill 2021 Oct 08;7(10):e29379 [FREE Full text] [doi: 10.2196/29379] [Medline: 34623311]
12. Gayam V, Chobufo MD, Merghani MA, Lamichhane S, Garlapati PR, Adler MK. Clinical characteristics and predictors of mortality in African-Americans with COVID-19 from an inner-city community teaching hospital in New York. J Med Virol 2021 Feb 05;93(2):812-819 [FREE Full text] [doi: 10.1002/jmv.26306] [Medline: 32672844]
13. Goicoechea M, Sánchez Cámara LA, Macías N, Muñoz de Morales A, Rojas, Bascuñana A, Verdalles, et al. COVID-19: clinical course and outcomes of 36 hemodialysis patients in Spain. Kidney Int 2020 Jul;98(1):27-34 [FREE Full text] [doi: 10.1016/j.kint.2020.04.031] [Medline: 32437770]
14. Khalil K, Agbontaen K, McNally D, Love A, Mandalia S, Banya W, et al. Clinical characteristics and 28-day mortality of medical patients admitted with COVID-19 to a central London teaching hospital. J Infect 2020 Sep;81(3):e85-e89 [FREE Full text] [doi: 10.1016/j.jinf.2020.06.027] [Medline: 32562795]
15. Santos CS, Morales CM, Álvarez ED, Castro, Robles AL, Sandoval TP. Determinants of COVID-19 disease severity in patients with underlying rheumatic disease. Clin Rheumatol 2020 Sep 27;39(9):2789-2796 [FREE Full text] [doi: 10.1007/s10067-020-05301-2] [Medline: 32720259]

16. Simonnet A, Chetboun M, Poissy J, Raverdy V, Noulette J, Duhamel A, LICORN and the Lille COVID-19 and Obesity study group. High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. Obesity (Silver Spring) 2020 Jul;28(7):1195-1199 [FREE Full text] [doi: 10.1002/oby.22831] [Medline: 32271993]
17. Social security system. National Health Insurance Service. URL: https://www.nhis.or.kr/english/wbheaa02200m01.do [accessed 2022-10-29]
18. An C, Lim H, Kim D, Chang JH, Choi YJ, Kim SW. Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study. Sci Rep 2020 Oct 30;10(1):18716 [FREE Full text] [doi: 10.1038/s41598-020-75767-2] [Medline: 33127965]
19. Kim JM, Park JH, Kim HS, Lee JW, Lim HS, Choi WA, et al. Epidemiology and diagnostic process of amyotrophic lateral sclerosis as distinct from myelopathy: 5-year cohort study of whole-population in South Korea. Amyotroph Lateral Scler Frontotemporal Degener 2018 Nov;19(7-8):547-554. [doi: 10.1080/21678421.2018.1491600] [Medline: 30421999]
20. Reynolds HR, Adhikari S, Pulgarin C, Troxel AB, Iturrate E, Johnson SB, et al. Renin–angiotensin–aldosterone system inhibitors and risk of Covid-19. N Engl J Med 2020 Jun 18;382(25):2441-2448. [doi: 10.1056/nejmoa2008975]
21. Lee SW, Ha EK, Yeniova A, Moon SY, Kim SY, Koh HY, et al. Severe clinical outcomes of COVID-19 associated with proton pump inhibitors: a nationwide cohort study with propensity score matching. Gut 2021 Jan 30;70(1):76-84. [doi: 10.1136/gutjnl-2020-322248] [Medline: 32732368]
22. People with Certain Medical Conditions. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html [accessed 2022-10-29]
23. Drefahl S, Wallace M, Mussino E, Aradhya S, Kolk M, Brandén M, et al. A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. Nat Commun 2020 Oct 09;11(1):5097 [FREE Full text] [doi: 10.1038/s41467-020-18926-3] [Medline: 33037218]
24. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. JAMA 2020 Aug 25;324(8):782-793. [doi: 10.1001/jama.2020.12839] [Medline: 32648899]
25. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 2020 Apr 16;181(2):271-280.e8 [FREE Full text] [doi: 10.1016/j.cell.2020.02.052] [Medline: 32142651]
26. Ni W, Yang X, Yang D, Bao J, Li R, Xiao Y, et al. Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. Crit Care 2020 Jul 13;24(1):422 [FREE Full text] [doi: 10.1186/s13054-020-03120-0] [Medline: 32660650]
27. Zhang H, Li H, Lyu J, Lei X, Li W, Wu G, et al. Specific ACE2 expression in small intestinal enterocytes may cause gastrointestinal symptoms and injury after 2019-nCoV infection. Int J Infect Dis 2020 Jul;96:19-24 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.027] [Medline: 32311451]
28. Zou X, Chen K, Zou J, Han P, Hao J, Han Z. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. Front Med 2020 Apr 12;14(2):185-192 [FREE Full text] [doi: 10.1007/s11684-020-0754-0] [Medline: 32170560]
29. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. The Lancet Respiratory Medicine 2020 May;8(5):475-481. [doi: 10.1016/s2213-2600(20)30079-5]
30. Radenkovic D, Chawla S, Pirro M, Sahebkar A, Banach M. Cholesterol in relation to COVID-19: should we care about it? J Clin Med 2020 Jun 18;9(6):1909 [FREE Full text] [doi: 10.3390/jcm9061909] [Medline: 32570882]
31. Wang H, Yuan Z, Pavel MA, Jablonski SM, Jablonski J, Hobson R, et al. The role of high cholesterol in age-related COVID19 lethality. bioRxiv 2021 Jun 28:1 [FREE Full text] [doi: 10.1101/2020.05.09.086249] [Medline: 32511366]
32. Iqbal Z, Ho JH, Adam S, France M, Syed A, Neely D, Heart UK's Medical Scientific and Research Committee. Managing hyperlipidaemia in patients with COVID-19 and during its pandemic: An expert panel position statement from HEART UK. Atherosclerosis 2020 Nov;313:126-136 [FREE Full text] [doi: 10.1016/j.atherosclerosis.2020.09.008] [Medline: 33045618]
33. Proto MC, Fiore D, Piscopo C, Pagano C, Galgani M, Bruzzaniti S, et al. Lipid homeostasis and mevalonate pathway in COVID-19: Basic concepts and potential therapeutic targets. Prog Lipid Res 2021 Apr;82:101099 [FREE Full text] [doi: 10.1016/j.plipres.2021.101099] [Medline: 33915202]
34. Ahn C, Hwang Y, Park SK. Predictors of all-cause mortality among 514,866 participants from the Korean National Health Screening Cohort. PLoS One 2017 Sep 28;12(9):e0185458 [FREE Full text] [doi: 10.1371/journal.pone.0185458] [Medline: 28957371]
35. Park CH, Kang EW, Park JT, Han SH, Yoo T, Kang S, et al. Association of serum lipid levels over time with survival in incident peritoneal dialysis patients. J Clin Lipidol 2017 Jul;11(4):945-954.e3. [doi: 10.1016/j.jacl.2017.06.004] [Medline: 28669685]
36. Zuliani G, Volpato S, Dugo M, Vigna GB, Morieri ML, Maggio M, et al. Combining LDL-C and HDL-C to predict survival in late life: The InChianti study. PLoS One 2017 Sep 28;12(9):e0185307 [FREE Full text] [doi: 10.1371/journal.pone.0185307] [Medline: 28957382]

37. Wang M, Hu H, Lin I, Chuang J. Plasma lipid concentrations and survival in geriatric population: A retrospective cohort study. Medicine (Baltimore) 2019 Dec;98(49):e18154 [FREE Full text] [doi: 10.1097/MD.0000000000018154] [Medline: 31804326]
38. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
39. Wei X, Zeng W, Su J, Wan H, Yu X, Cao X, et al. Hypolipidemia is associated with the severity of COVID-19. J Clin Lipidol 2020 May;14(3):297-304 [FREE Full text] [doi: 10.1016/j.jacl.2020.04.008] [Medline: 32430154]
40. Hu X, Chen D, Wu L, He G, Ye W. Low serum cholesterol level among patients with COVID-19 infection in Wenzhou, China. SSRN Journal 2020:1. [doi: 10.2139/ssrn.3544826]
41. Wang G, Zhang Q, Zhao X, Dong H, Wu C, Wu F, et al. Low high-density lipoprotein level is correlated with the severity of COVID-19 patients: an observational study. Lipids Health Dis 2020 Sep 07;19(1):204 [FREE Full text] [doi: 10.1186/s12944-020-01382-9] [Medline: 32892746]
42. Chidambaram V, Shanmugavel Geetha H, Kumar A, Majella MG, Sivakumar RK, Voruganti D, et al. Association of Lipid Levels With COVID-19 Infection, Disease Severity and Mortality: A Systematic Review and Meta-Analysis. Front Cardiovasc Med 2022 Mar 24;9:862999 [FREE Full text] [doi: 10.3389/fcvm.2022.862999] [Medline: 35402531]
43. Scherbakov N, Dirnagl U, Doehner W. Body weight after stroke. Stroke 2011 Dec;42(12):3646-3650. [doi: 10.1161/strokeaha.111.619163]
44. Andersen KK, Olsen TS. The obesity paradox in stroke: lower mortality and lower risk of readmission for recurrent stroke in obese stroke patients. Int J Stroke 2015 Jan 12;10(1):99-104. [doi: 10.1111/ijs.12016] [Medline: 25635277]

Abbreviations

ACE2: angiotensin-converting enzyme 2
CCI: Charlson Comorbidity Index
CDC: Centers for Disease Control and Prevention
ICD: International Classification of Diseases
KNHIS: Korean National Health Insurance Service
NRF: National Research Foundation
OR: odds ratio
PS: propensity score

Edited by T Sanchez, A Mavragani; submitted 17.11.21; peer-reviewed by D Roger, H Jeon, S Jung; comments to author 04.07.22; revised version received 31.08.22; accepted 13.10.22; published 18.11.22.

Please cite as:
Kim J, Park SH, Kim JM
Effect of Comorbidities on the Infection Rate and Severity of COVID-19: Nationwide Cohort Study With Propensity Score Matching
JMIR Public Health Surveill 2022;8(11):e35025
URL: https://publichealth.jmir.org/2022/11/e35025
doi:10.2196/35025
PMID:36265125

Original Paper

The Association Between Clinical Severity and Incubation Period of SARS-CoV-2 Delta Variants: Retrospective Observational Study

Kai Wang[1*], PhD; Zemin Luan[1*], BMed; Zihao Guo[2*], MSc; Jinjun Ran[3], PhD; Maozai Tian[1], PhD; Shi Zhao[2], PhD

[1]Department of Medical Engineering and Technology, Xinjiang Medical University, Urumqi, China

[2]JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[3]School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Shi Zhao, PhD
JC School of Public Health and Primary Care
Chinese University of Hong Kong
Rm 417, Public Health Building
Prince of Wales Hospital, 30-32 Ngan Shing St
Hong Kong
China (Hong Kong)
Phone: 852 22528865
Fax: 852 26467297
Email: zhaoshi.cmsa@gmail.com

Abstract

Background: As of August 25, 2021, Jiangsu province experienced the largest COVID-19 outbreak in eastern China that was seeded by SARS-CoV-2 Delta variants. As one of the key epidemiological parameters characterizing the transmission dynamics of COVID-19, the incubation period plays an essential role in informing public health measures for epidemic control. The incubation period of COVID-19 could vary by different age, sex, disease severity, and study settings. However, the impacts of these factors on the incubation period of Delta variants remains uninvestigated.

Objective: The objective of this study is to characterize the incubation period of the Delta variant using detailed contact tracing data. The effects of age, sex, and disease severity on the incubation period were investigated by multivariate regression analysis and subgroup analysis.

Methods: We extracted contact tracing data of 353 laboratory-confirmed cases of SARS-CoV-2 Delta variants' infection in Jiangsu province, China, from July to August 2021. The distribution of incubation period of Delta variants was estimated by using likelihood-based approach with adjustment for interval-censored observations. The effects of age, sex, and disease severity on the incubation period were expiated by using multivariate logistic regression model with interval censoring.

Results: The mean incubation period of the Delta variant was estimated at 6.64 days (95% credible interval: 6.27-7.00). We found that female cases and cases with severe symptoms had relatively longer mean incubation periods than male cases and those with nonsevere symptoms, respectively. One-day increase in the incubation period of Delta variants was associated with a weak decrease in the probability of having severe illness with an adjusted odds ratio of 0.88 (95% credible interval: 0.71-1.07).

Conclusions: In this study, the incubation period was found to vary across different levels of sex, age, and disease severity of COVID-19. These findings provide additional information on the incubation period of Delta variants and highlight the importance of continuing surveillance and monitoring of the epidemiological characteristics of emerging SARS-CoV-2 variants as they evolve.

(JMIR Public Health Surveill 2022;8(11):e40751) doi:10.2196/40751

KEYWORDS

COVID-19; Delta variant; incubation period; clinical severity; China

Introduction

The ongoing COVID-19 pandemic caused by SARS-CoV-2 has been continuously spreading worldwide, posing significant threat and burden to public health systems. The emergence of SARS-CoV-2 variants has accelerated the global spread of COVID-19 [1]. In February 2021, the SARS-CoV-2 Delta variant (Phylogenetic Assignment of Named Global Outbreak lineage: B.1.617.2) was first detected in India [2]. Subsequently, major outbreaks seeded by the Delta variants have been reported in various regions [3,4]. A comprehensive understanding of the epidemiological characteristics of the Delta variant would help inform targeted interventions for containing the spread of COVID-19 [5].

The continuous evolution of new variants of SARS-CoV-2 since the outbreak has been a great challenge, especially for those in health care and research and development in the areas of diagnosis, prevention and treatment development, as well as policy makers and administrators [6], resulting in rapid changes in the epidemiological information used to plan and evaluate strategies to prevent the spread of COVID-19 [7].

The incubation period, defined as the time delay between the onset of infection and symptoms of a case, is an imperative epidemiological parameter of an infectious disease. From the perspective of epidemic control, estimating the incubation period could help determine the quarantine time, develop control measures, and predict the transmission dynamics [8]. Apart from that, the incubation period also plays an important role in determining the proportion of presymptomatic transmission, which has posed significant challenges in the containment of epidemics [9]. Thus, it is of crucial importance to clarify the distribution of the incubation period especially for the SARS-CoV-2 variant, which could cause large outbreaks.

The current understanding on the incubation period for the SARS-CoV-2 Delta variant is limited. Although estimates of the incubation period of various historical SARS-CoV-2 strains can be found in the literature [10-12], knowledge of the incubation period of Delta variants has been largely scarce. However, recent studies conducted in Guangdong province, China, have shown that the Delta variant has a shorter incubation period than non-Delta variants [13,14]. The incubation period of COVID-19 could vary by age, sex, disease severity, and study settings [15]. The impact of these factors on the incubation period for the circulating Delta variant remains uninvestigated.

From July to August 2021, outbreaks seeded by the Delta variant were reported in Nanjing and Yangzhou, Jiangsu province, China, with a larger scale compared to the Delta outbreak that had occurred in Guangdong province from May to June 2021. The aim of this study was to characterize the incubation period of Delta variants using detailed epidemiological contact tracing data collected during the Delta outbreak in Jiangsu. Subgroup analysis was also conducted to examine the effect of age, sex, and disease severity on the incubation period. Furthermore, by applying a multivariate logistic regression model, we investigated the association between disease severity and incubation period of the Delta cases.

Methods

Data

Epidemiological contact tracing data of the cases infected with the Delta variant were collected from Nanjing Health Committee of Jiangsu Province [16] and Yangzhou Health Committee [17], from July to August 2021. We extracted the demographic and clinical information for each case, including age, sex, home address, exposure and contact history, date of COVID-19 diagnosis, and clinical severity categorized according to the criteria proposed by the National Health Commission of the People's Republic of China (ie, asymptomatic, mild, moderate, severe, and critical). Asymptomatic cases and cases that did not have any information on the exposure were excluded when estimating the incubation period.

On July 27, 2021, according to the Nanjing Centers for Disease Control and Prevention, the outbreaks were seeded by the SARS-CoV-2 Delta variants according to the whole genome sequencing results [18]. All cases included in this study were laboratory confirmed through real-time reverse transcription polymerase chain reaction or antigen test on a nasopharyngeal swab. The incubation period is the time delay between the date of infection and the date of onset of symptoms. A transmission pair was identified if 2 confirmed cases had a clear epidemiological link (clearly identified who is infected by whom through the contact history in the dataset which was confirmed by the official published epidemic reports). The date of infection is identified based on the contact history between each infected-infector transmission pair in the officially reported epidemiological survey reports. The time of symptom onset date is identified based on the time of symptom onset for each infected person in the officially reported epidemiological investigation reports. For cases without information on the exact date of infection, exposure windows (with lower and upper bound for the exact exposure date) were determined according to the trajectory and duration of contact.

Incubation Period

We assumed the incubation period T of the Delta cases was a random variable following a gamma distribution. For case i with known date of infection E and symptom onset S, the likelihood function was given by the following:

Here, f (.) is the probability density function of gamma distribution with parameters denoted by θ. For cases identified with an exposure window (E_{Li}, E_{Ri}), the incubation period was therefore interval-censored and bounded by (T_{Li}, T_{Ri})=($S_i - E_{Ri}$, $S_i - E_{Li}$). The total likelihood function was thus formulated as follows:

Here, F (.) represents the cumulative distribution function, and ω_i represents indicator variable. We have $\omega_i = 1$ if the incubation period was interval-censored and $\omega_i = 0$ if the exact incubation period was observed. The parameters were estimated by Markov

chain Monte Carlo (MCMC) method with uniform prior distribution *U(0,100)*. Marginal posterior distributions were obtained from 10,000 iterations, among which the first 5000 iterations were discarded as burn-in period. The 95% credible interval (CrI) was obtained from marginal posterior distributions. We estimated the incubation period distribution for overall cases and for different stratification of cases including age groups (ie, 0-18 years, 19-39 years, 40-59 years, 60-79 years, and over 80 years), sex, clinical severity, and geographical regions (ie, Nanjing and Yangzhou).

Logistic Regression

Multivariate logistic regression model was applied to examine the associations between the incubation period and disease severity  of the cases infected with the Delta variant. The independent variables including age (*A*), sex (*S*), and incubation period (*T*) were included in the model. For case *i* with known date of infection, the probability *P* that the case's symptom is severe and critical (Y_i = 1) is:

For cases that had a window of exposure, the probability *P* is given by:

Moreover, we define  . Therefore, the likelihood function was constructed as:

We estimated the coefficients' vector  by MCMC with normal prior distribution. The marginal posterior distributions of parameters were obtained from 100,000 iterations, among which the first 50,000 iterations were discarded for the burn-in period. The 95% CrI was obtained from the marginal posterior distributions of unknown parameters.

Ethical Considerations

The collection of specimens as well as epidemiological and clinical data for SARS-CoV-2–infected individuals and their close contacts were a part of a continuing public health investigation of COVID-19 outbreaks, ruled in the Protocol on the Prevention and Control of COVID-19 by the National Health Commission of the People's Republic of China, which was exempt from ethical approval (ie, institutional review board assessment). All data used in this study were collected via public domains without personal identity; thus, institutional ethics review was waived.

Results

A total of 763 COVID-19 cases infected by the Delta variant were reported in Nanjing and Yangzhou from July to August 2021. Of the 763 cases, 410 (53.7%) were excluded due to a lack of exposure history, and the remaining 353 (46.3%) were included in the analysis. Of the 353 included cases, 161 (45.6%) were from Nanjing and 192 (54.4%) were from Yangzhou. In this study, the included cases were divided into two subgroups according to the severity of the disease: age group and sex. The age groups were divided into 5 groups (0-18 years, 19-39 years, 40-59 years, 60-79 years, and over 80 years).

A total of 132 (37.4%) cases aged 40-59 years accounted for a higher proportion than other age groups, with a smaller proportion of children (n=47, 13.3%) and people older than 80 years (n=7, 2%). The proportion of female (n=220, 62.3%) cases was higher than that of male cases (n=133, 37.7%) (Table 1).

Figure 1 shows the exposure to the symptom onset timeline for the included cases. The estimated mean incubation period for the Delta variant was 6.64 days (95% CrI 6.27-7.00) (Figures 2 and 3). There was a trend toward longer incubation period in male cases (7.10 days, 95% CrI 6.52-7.71) compared with female cases (6.36 days, 95% CrI 5.89-6.83; Table 2).

In the age group, the mean incubation period was 6.45 days (95% CrI 5.40-7.56) for cases aged 0-18 years, 6.20 days (95% CrI 5.59-6.89) for cases aged 19-39 years, and 6.85 days (95% CrI 6.17-7.55) for those aged 40-59 years; cases aged 60-79 years had a mean incubation period of 7.02 days (95% CrI 6.34-7.76), and the shortest mean incubation period was 6.45 days (95% CrI 5.40-7.56) for those older than 80 years. The mean incubation period estimates also differed among age groups, with a shorter mean incubation period for cases aged 0-39 years and ≥80 years compared with those aged 40-79 years (Table 2).

The estimated mean incubation was shorter for critical cases (5.73 days, 95% CrI 3.83-8.11), compared with mild cases (6.41 days, 95% CrI 5.67-7.19), moderate cases (6.78 days, 95% CrI 6.34-7.25), and severe cases (6.63 days, 95% CrI 5.10-8.47). There was a trend toward longer mean incubation period for cases in Yangzhou (6.72 days, 95% CrI 6.23-7.23), compared with cases in Nanjing (6.51 days, 95% CrI 5.99-7.07; Table 2).

The duration of incubation period had a weak and negative association with the clinical severity of COVID-19 cases infected Delta variants, with an adjusted odds ratio (OR) of 0.88 (95% CrI 0.71-1.07). After adjusting for age and sex, which implies that a 1-day increase in incubation period was associated with a 12% decrease in the probability of severe illness (Figure 4). Furthermore, age was found to be positively associated with the incubation period, with an adjusted OR of 1.07 (95% CrI 1.05-1.10).

Table 1. Basic characteristics of the confirmed SARS-CoV-2 Delta cases.

Characteristics	All cases (n=353)	Mild cases (n=84)	Moderate cases (n=238)	Severe cases (n=21)	Critical cases (n=10)
Age group					
0-18 years	47 (13.3)	30 (35.7)	17 (7.1)	0 (0)	0 (0)
19-39 years	84 (23.8)	24 (28.6)	59 (24.8)	1 (4.8)	0 (0)
40-59 years	132 (37.4)	26 (31%)	94 (39.5)	10 (47.6)	2 (20)
60-79 years	83 (23.5)	4 (4.8)	64 (26.9)	10 (47.6)	5 (50)
≥80 years	7 (2)	0 (0)	4 (1.7)	0 (0)	3 (30)
Sex					
Male	133 (37.7)	28 (33.3)	92 (38.7)	7 (33.3)	6 (60)
Female	220 (62.3)	56 (66.7)	146 (61.3)	14 (66.7)	4 (40)

Figure 1. Timeline of the course of infection for each case infected by SARS-CoV-2 Delta variants (n=353) from July to August 2021, in Jiangsu province, China.

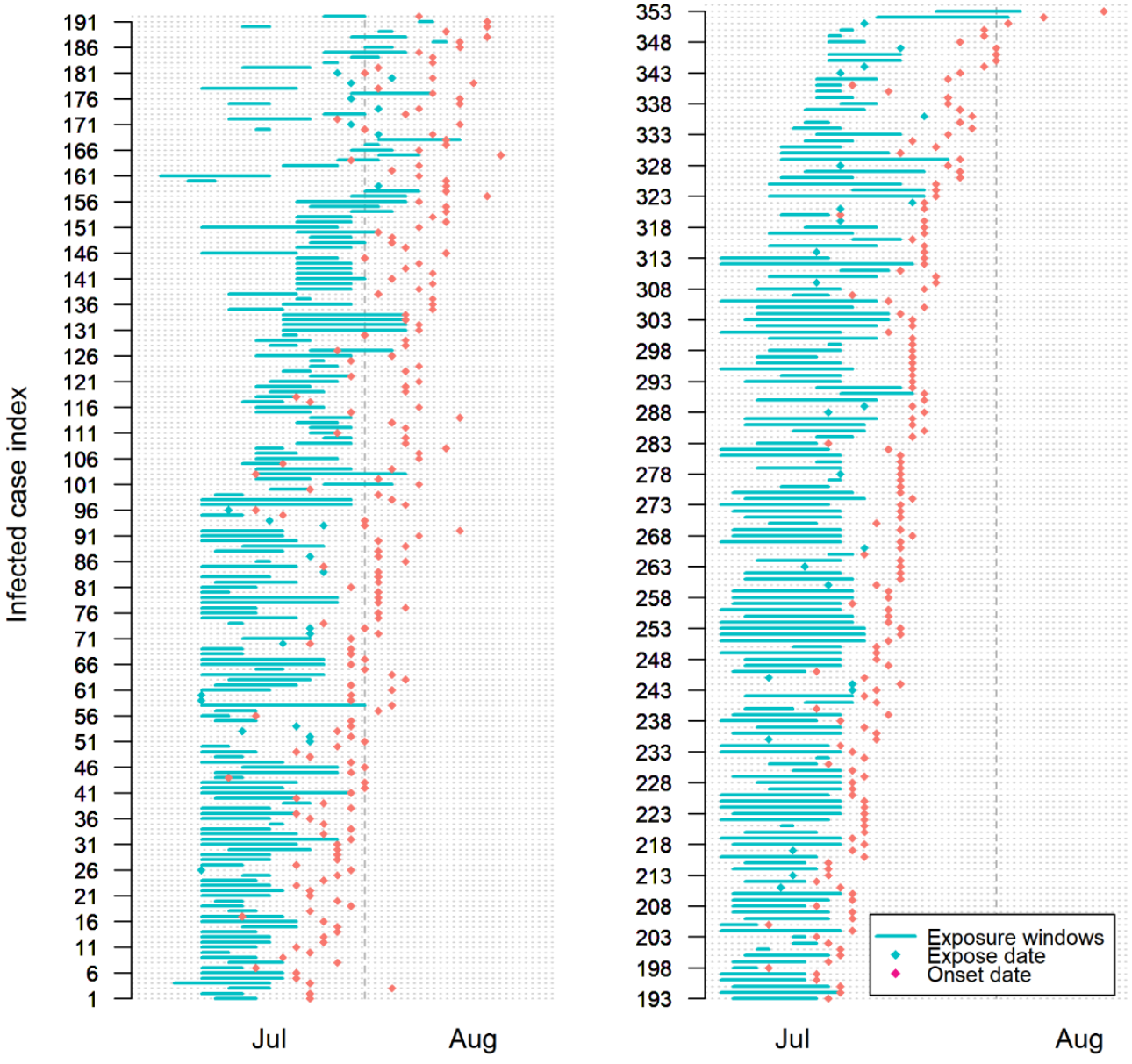


Figure 2. Cumulative distribution of the estimated gamma incubation period for the confirmed SARS- CoV-2 Delta cases (n=353). MCMC: Markov chain Monte Carlo.

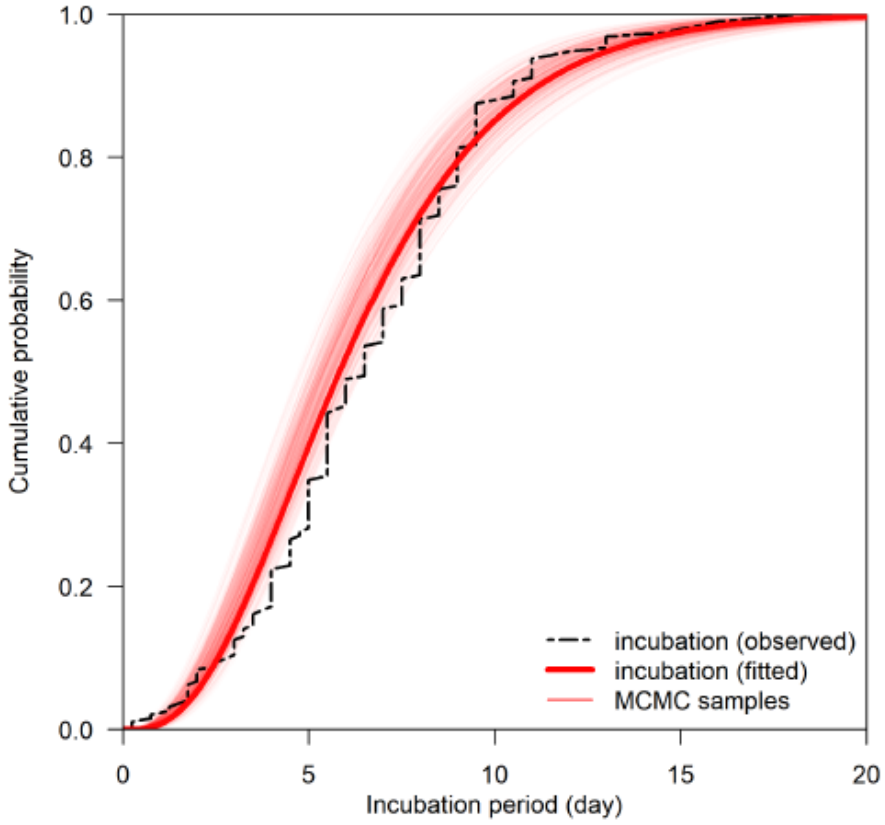


Figure 3. Incubation period distribution for Delta cases with severe diseases (n=31) and Delta cases with nonsevere diseases (n=322).

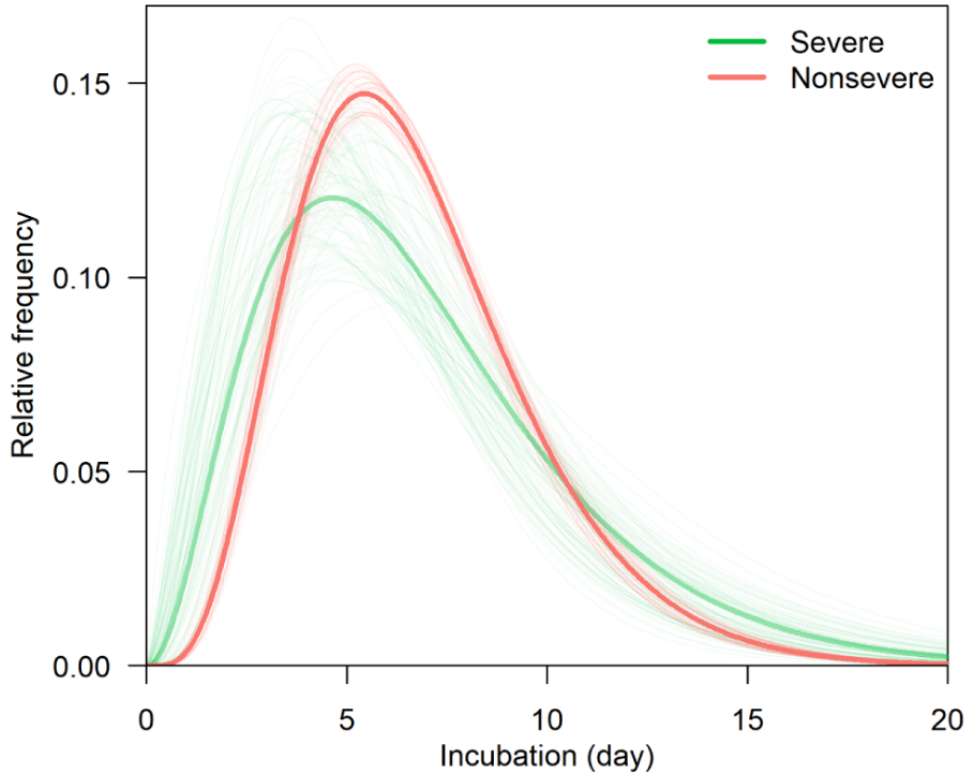
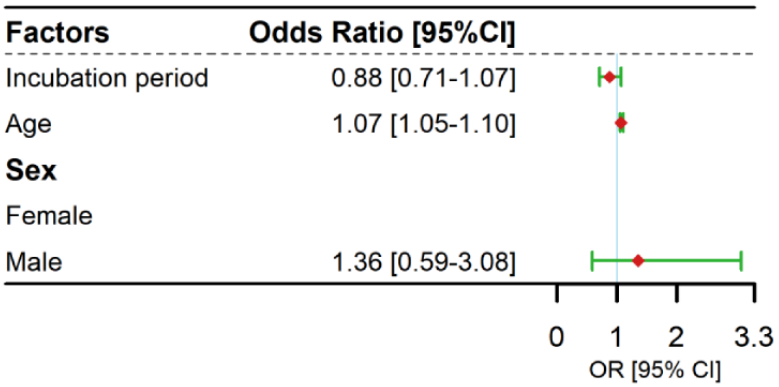


Table 2. Estimated incubation period by sex, age groups, clinical severity, and 2 geographical regions of the Delta cases.

Characteristics	Mean (days)	Median (days)	The lower bound of 95% CrI	The upper bound of 95% CrI
Overall (n=353)	6.64	6.63	6.27	7.00
Sex				
Male (n=133)	7.10	7.09	6.52	7.71
Female (n=220)	6.36	6.36	5.89	6.83
Age group				
0-18 years (n=47)	6.45	6.42	5.40	7.65
19-39 years (n=84)	6.20	6.19	5.59	6.89
40-59 years (n=132)	6.85	6.85	6.17	7.55
60-79 years (n=83)	7.02	7.02	6.34	7.76
≥80 years (n=7)	6.05	5.96	4.50	8.07
Clinical severity				
Mild cases (n=84)	6.41	6.41	5.67	7.19
Moderate cases (n=238)	6.78	6.78	6.34	7.25
Severe cases (n=21)	6.63	6.58	5.10	8.47
Critical cases (n=10)	5.73	5.66	3.83	8.11
Geographical region				
Yangzhou (n=192)	6.72	6.72	6.23	7.23
Nanjing (n=161)	6.51	6.51	5.99	7.07

Figure 4. Risk factors associated with the disease severity of the Delta cases. OR: odds ratio.



Discussion

Principal Findings

In this study, the mean incubation period of the Delta variant was estimated to be 6.64 days (95% CrI 6.27-7.00) by using the MCMC method for the interval-censored data based on uniform prior distribution *U(0,100)*. We found that a 1-day increase in the incubation period for the Delta variant was associated with a 12% decrease in the probability of severe disease after adjusting for age and sex (OR=0.88, 95% CI 0.71-1.07).

Characterizing the epidemiological features of the SARS-CoV-2 variants could provide insights into the transmission potential of COVID-19. Based on detailed contact tracing data, we estimated the incubation period of the SARS-CoV-2 Delta variant and examined the association between the incubation period and disease severity. Subgroup analysis was also conducted to investigate the difference in the incubation period distribution between age groups, sex, disease severity, and 2 geographical regions.

Our mean (6.64 days) and median (6.63 days) incubation period estimates for overall Delta cases are slightly longer than the pooled point estimates (mean: 6.3 days; median: 5.4 days) from a previous meta-analysis on the incubation period of the historical wild-type COVID-19 strains [19]. The mean incubation estimates were relatively longer than those of Grant et al [20] (6.64 days vs 4.3 days). The mean incubation estimates are also larger than previous findings from Guangdong Province, China, with a mean estimate range of 3.9 to 5.8 days [21,22]. Moreover, the mean incubation estimates were relatively longer than those by Ogata et al [23] (6.64 days vs 3.7 days). These discrepancies may be attributed to not only the biological

difference between the SARS-CoV-2 strains but also the definition of the date of infection and symptom onset, as well as the estimation methodology.

The incubation period is considered as a function of the initial infectious dose, the rate of pathogen replication in the host, and intrahost defense mechanisms [24,25]. The result of mean incubation period was estimated to have a tendency for severe Delta cases to be shorter than nonsevere Delta cases (5.73 days vs 6.78 days) and for female cases to be longer than male cases (7.10 days vs 6.36 days), consistent with an earlier study using a larger sample size [15]. The multivariate logistic regression model also suggested a negative association between disease severity and incubation period even after control for age and sex. Early studies suggested that a shorter incubation period is related to a higher viral load of the initial infection, which may give rise to a more rapid pathogen replication rate that outpaces the adaptive immune system, thereby resulting in a more severe disease [26,27]. Although the biological pathway behind the incubation period and the clinical severity of COVID-19 has not been well established, a shorter incubation period may serve as an indicator of a more severe outcome for patients. Apart from that, the differences in incubation period estimates between sexes could be due to female cases exhibiting stronger innate and adaptive immune responses than male cases, which may result in a faster clearance of the in-host pathogens [28].

During the ongoing COVID-19 pandemic, estimating the distribution of the incubation period under the context of the local epidemics is essential to inform the local public health interventions such as the quarantine and isolation period [29]. With more recent data, the 95th and 97.5th percentile of the estimated distribution of the incubation period could give policy makers hints on how to adjust and improve the current control measures to effectively use the limited public health resources and at the same time to minimize the risk of permitting infectious persons into the community. Therefore, to mitigate current epidemics and prevent future outbreaks, it is crucial to obtain the incubation period estimates based on more updated epidemiological data of novel SARS-CoV-2 variants [30]. The estimates of the incubation period, across demographic and clinical features of cases, for the Delta variants added additional information to the existent evidence, which could potentially improve the policy making process.

This study has some limitations. First, the epidemiological contact tracing data were subjected to recall bias. When the confirmed cases recall their exposure window, some activities may be omitted due to unclear memory, which may lead to extra uncertainty or bias in the incubation period estimates. Second, it is possible that incubation distribution varies by vaccination status (ie, whether on is vaccinated or not), which may act as a potential confounding factor in the logistic regression model. However, because complete information on vaccination was not available, it was not included in the model. Future studies with more data could further investigate the effect of the vaccine on the incubation period of an emerging variant and explore additional factors affecting the latency of SARS-CoV-2 with more subgroup analyses. Finally, there were significantly fewer severe cases than nonsevere cases. Such imbalance may deviate the coefficient estimates.

In the future, it is necessary to further study the transmission dynamics and viral shedding, particularly for vaccinated cases with the Delta infection. Given that the current pandemic is dominated by the SARS-CoV-2 Omicron variants, it is also necessary to continue the surveillance of epidemiological characteristics of the Omicron variants.

For the novelties of this study, we adopted a state-of-the-art statistical approach to examine the association between incubation period of Delta variants and potential factors, including sex, age, and clinical severity of COVID-19 illness. The samples of incubation period observations were collected from the largest COVID-19 epidemics in eastern China seeded by Delta variants with well-traced and individual-level information of each laboratory-confirmed case. Interval censoring of incubation period observations was adjusted in the likelihood-based statistical inference framework to approach the intrinsic characteristics of incubation period. As such, the estimated associations between case characteristics and incubation period reflected evidence of the intrinsic feature of COVID-19, rather than being unauthentic due to observational or sampling bias.

Conclusions

In conclusion, this study estimated the incubation period distribution of Delta variants according to detailed contact tracing data of COVID-19 cases in eastern China. The incubation period was found varied across sex, age, and disease severity of cases. A mild negative association between incubation period of Delta variants and clinical severity of COVID-19 was reported. These findings provided additional information on the incubation period of Delta variants and highlighted the importance of continuing surveillance and monitoring of the epidemiological characteristics of emerging SARS-CoV-2 variants as they evolve.

This study uncovered differences in incubation period between age, sex, and severe disease for patients with the Delta variant, and it will help researchers uncover key areas of the combination of incubation period with the disease severity for SARS-CoV-2 Delta variants, which many researchers have not been able to explore. Thus, a new theory on the prevention of transmission of different variants of SARS-CoV-2 may be arrived at.

Acknowledgments

The authors would like to thank the editor and the reviewers for their helpful comments. This work was supported in part by the program for Tianshan Innovative Research Team of Xinjiang Uygur Autonomous Region, China (2020D14020), and the Natural Science Foundation of China (11961071).

Data Availability

The processed data sets used for the analyses in this study may be available on request to the corresponding author.

Authors' Contributions

All authors contributed equally to this work.

Conflicts of Interest

None declared.

References

1. Hamamoto Y. The COVID-19 world - Are we there yet? J Diabetes Investig 2021 Jul;12(7):1125-1127 [FREE Full text] [doi: 10.1111/jdi.13605] [Medline: 34056843]
2. Ranjan R, Sharma A, Verma M. Characterization of the second wave of COVID-19 in India. medRxiv 2021:5665. [doi: 10.1101/2021.04.17.21255665]
3. Shitrit P, Zuckerman NS, Mor O. Nosocomial outbreak caused by the SARS-CoV-2 Delta variant in a highly vaccinated population, Israel, July. Eurosurveillance 2021:1-4. [doi: 10.2807/1560-7917.es.2021.26.39.2100822]
4. Alizon S, Haim-Boukobza S, Foulongne V, Verdurme L, Trombert-Paolantoni S, Lecorche E, et al. Rapid spread of the SARS-CoV-2 Delta variant in some French regions, June 2021. J. Eurosurveillance 2021;26(28):1-5. [doi: 10.2807/1560-7917.es.2021.26.28.2100573]
5. Li L, Han Z, Qin P, Liu W, Yang Z, Chen Z, et al. Transmission and containment of the SARS-CoV-2 Delta variant of concern in Guangzhou, China: A population-based study. PLoS Negl Trop Dis 2022 Jan 5;16(1):e0010048 [FREE Full text] [doi: 10.1371/journal.pntd.0010048] [Medline: 34986169]
6. Math RK, Mudennavar N, Javaregowda PK, Savanur A. In Silico Comparative Analysis of the Functional, Structural, and Evolutionary Properties of SARS-CoV-2 Variant Spike Proteins. JMIR Bioinform Biotech 2022 May 30;3(1):e37391 [FREE Full text] [doi: 10.2196/37391] [Medline: 35669291]
7. Hamadeh A, Feng Z, Niergarth J, Wong WW. Estimation of COVID-19 Period Prevalence and the Undiagnosed Population in Canadian Provinces: Model-Based Analysis. JMIR Public Health Surveill 2021 Sep 09;7(9):e26409 [FREE Full text] [doi: 10.2196/26409] [Medline: 34228626]
8. Wang Z, Fu Y, Guo Z, Li J, Li J, Cheng H, et al. Transmission and prevention of SARS-CoV-2. Biochem Soc Trans 2020 Oct 30;48(5):2307-2316. [doi: 10.1042/BST20200693] [Medline: 33084885]
9. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 2020 May 15;26(5):672-675. [doi: 10.1038/s41591-020-0869-5] [Medline: 32296168]
10. McAloon C, Collins, Hunt K, Barber A, Byrne AW, Butler F, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. BMJ Open 2020 Aug 16;10(8):e039652 [FREE Full text] [doi: 10.1136/bmjopen-2020-039652] [Medline: 32801208]
11. Xin H, Li Y, Wu P, Li Z, Lau EHY, Qin Y, et al. Estimating the latent period of Coronavirus Disease 2019 (COVID-19). Clin Infect Dis 2022 May 03;74(9):1678-1681. [doi: 10.1093/cid/ciab746] [Medline: 34453527]
12. Liu Y, Zhao S, Ryu S, Ran J, Fan J, He D. Estimating the incubation period of SARS-CoV-2 Omicron BA.1 variant in comparison with that during the Delta variant dominance in South Korea. One Health 2022 Dec;15:100425 [FREE Full text] [doi: 10.1016/j.onehlt.2022.100425] [Medline: 35942477]
13. Wang Y, Chen R, Hu F, Lan Y, Yang Z, Zhan C, et al. Transmission, viral kinetics and clinical characteristics of the emergent SARS-CoV-2 Delta VOC in Guangzhou, China. EClinicalMedicine 2021 Oct;40:101129 [FREE Full text] [doi: 10.1016/j.eclinm.2021.101129] [Medline: 34541481]
14. Zhang M, Xiao J, Deng A, Zhang Y, Zhuang Y, Hu T, et al. Transmission Dynamics of an Outbreak of the COVID-19 Delta Variant B.1.617.2 - Guangdong Province, China, May-June 2021. China CDC Wkly 2021 Jul 02;3(27):584-586 [FREE Full text] [doi: 10.46234/ccdcw2021.148] [Medline: 34594941]
15. Huang S, Li J, Dai C, Tie Z, Xu J, Xiong X, et al. Incubation period of coronavirus disease 2019: new implications for intervention and control. Int J Environ Health Res 2022 Aug 04;32(8):1707-1715. [doi: 10.1080/09603123.2021.1905781] [Medline: 33818217]
16. Public Announcement, Public Government Information. Nanjing Health Committee of Jiangsu Province. 2021. URL: http://wjw.nanjing.gov.cn/njswshjhsywyh/?id=xxgk_228 [accessed 2022-11-11]
17. Highlights, News. Yangzhou Health Committee of Jiangsu Province. 2021. URL: http://wjw.yangzhou.gov.cn/yzwshjh/ywkd/wjw_list_37.shtml [accessed 2022-11-11]
18. Interface News. Baidu. URL: https://baijiahao.baidu.com/s?id=1706426781822522062&wfr=spider&for=pc [accessed 2022-11-11]
19. Xin H, Wong JY, Murphy C, Yeung A, Taslim Ali S, Wu P, et al. The Incubation Period Distribution of Coronavirus Disease 2019: A Systematic Review and Meta-analysis. Clin Infect Dis 2021 Dec 16;73(12):2344-2352. [doi: 10.1093/cid/ciab501] [Medline: 34117868]

20. Grant R, Charmet T, Schaeffer L, Galmiche S, Madec Y, Von Platen C, et al. Impact of SARS-CoV-2 Delta variant on incubation, transmission settings and vaccine effectiveness: Results from a nationwide case-control study in France. Lancet Reg Health Eur 2022 Feb;13:100278 [FREE Full text] [doi: 10.1016/j.lanepe.2021.100278] [Medline: 34849500]
21. Li B, Deng A, Li K, Hu Y, Li Z, Shi Y, et al. Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. Nat Commun 2022 Jan 24;13(1):460 [FREE Full text] [doi: 10.1038/s41467-022-28089-y] [Medline: 35075154]
22. Kang M, Xin H, Yuan J. Transmission dynamics and epidemiological characteristics of SARS-CoV-2 Delta variant infections in Guangdong, China, May to June 2021. Euro Surveill 2022;27(10):1-10. [doi: 10.2807/1560-7917.es.2022.27.10.2100815]
23. Ogata T, Tanaka H, Irie F, Hirayama A, Takahashi Y. Shorter Incubation Period among Unvaccinated Delta Variant Coronavirus Disease 2019 Patients in Japan. Int J Environ Res Public Health 2022 Jan 20;19(3):1127 [FREE Full text] [doi: 10.3390/ijerph19031127] [Medline: 35162151]
24. Fine PEM. The interval between successive cases of an infectious disease. Am J Epidemiol 2003 Dec 01;158(11):1039-1047. [doi: 10.1093/aje/kwg251] [Medline: 14630599]
25. Nishiura H. Early efforts in modeling the incubation period of infectious diseases with an acute course of illness. Emerg Themes Epidemiol 2007 May 11;4(1):2 [FREE Full text] [doi: 10.1186/1742-7622-4-2] [Medline: 17466070]
26. Ho M, Chen W, Chen H, Lin S, Wang M, Di J, et al. Neutralizing antibody response and SARS severity. Emerg Infect Dis 2005 Nov;11(11):1730-1737 [FREE Full text] [doi: 10.3201/eid1111.040659] [Medline: 16318725]
27. Gandhi M, Beyrer C, Goosby E. Masks Do More Than Protect Others During COVID-19: Reducing the Inoculum of SARS-CoV-2 to Protect the Wearer. J Gen Intern Med 2020 Oct 31;35(10):3063-3066 [FREE Full text] [doi: 10.1007/s11606-020-06067-8] [Medline: 32737790]
28. Klein SL, Flanagan KL. Sex differences in immune responses. Nat Rev Immunol 2016 Oct;16(10):626-638. [doi: 10.1038/nri.2016.90] [Medline: 27546235]
29. Nishiura H. Determination of the appropriate quarantine period following smallpox exposure: an objective approach using the incubation period distribution. Int J Hyg Environ Health 2009 Jan;212(1):97-104. [doi: 10.1016/j.ijheh.2007.10.003] [Medline: 18178524]
30. Gussow AB, Auslander N, Wolf YI, Koonin EV. Prediction of the incubation period for COVID-19 and future virus disease outbreaks. BMC Biol 2020 Nov 30;18(1):186. [doi: 10.1186/s12915-020-00919-9] [Medline: 33256718]

Abbreviations

CrI: credible interval
MCMC: Markov chain Monte Carlo
OR: odds ratio

Edited by A Mavragani, T Sanchez; submitted 04.07.22; peer-reviewed by M Mozaffari, C Herbert, S Pesälä, M Raimi, SR Sebastian; comments to author 05.10.22; revised version received 22.10.22; accepted 27.10.22; published 18.11.22.

Please cite as:
Wang K, Luan Z, Guo Z, Ran J, Tian M, Zhao S
The Association Between Clinical Severity and Incubation Period of SARS-CoV-2 Delta Variants: Retrospective Observational Study
JMIR Public Health Surveill 2022;8(11):e40751
URL: https://publichealth.jmir.org/2022/11/e40751
doi:10.2196/40751
PMID:36346940

Original Paper

The Risk of Hospitalization and Mortality After Breakthrough SARS-CoV-2 Infection by Vaccine Type: Observational Study of Medical Claims Data

Meghana Kshirsagar[1], BS, MS, PhD; Md Nasir[1], BS, PhD; Sumit Mukherjee[2], BS, PhD; Nicholas Becker[1,3], BS, MS; Rahul Dodhia[1], BS, PhD; William B Weeks[1], MD, PhD; Juan Lavista Ferres[1], BS, MS; Barbra Richardson[4], BS, MS, PhD

[1]Microsoft, Redmond, WA, United States

[2]Insitro Inc, South San Francisco, CA, United States

[3]Paul G Allen School of Computer Science & Engineering, University of Washington, Seattle, WA, United States

[4]Department of Biostatistics and Global Health, University of Washington, Seattle, WA, United States

Corresponding Author:
Meghana Kshirsagar, BS, MS, PhD
Microsoft
1 Microsoft Way
Redmond, WA, 98052
United States
Phone: 1 425 638 7777
Email: meghana.ksagar@gmail.com

Abstract

Background: Several risk factors have been identified for severe COVID-19 disease by the scientific community. In this paper, we focus on understanding the risks for severe COVID-19 infections after vaccination (ie, in breakthrough SARS-CoV-2 infections). Studying these risks by vaccine type, age, sex, comorbidities, and any prior SARS-CoV-2 infection is important to policy makers planning further vaccination efforts.

Objective: We performed a comparative study of the risks of hospitalization (n=1140) and mortality (n=159) in a SARS-CoV-2 positive cohort of 19,815 patients who were all fully vaccinated with the Pfizer, Moderna, or Janssen vaccines.

Methods: We performed Cox regression analysis to calculate the risk factors for developing a severe breakthrough SARS-CoV-2 infection in the study cohort by controlling for vaccine type, age, sex, comorbidities, and a prior SARS-CoV-2 infection.

Results: We found lower hazard ratios for those receiving the Moderna vaccine (P<.001) and Pfizer vaccine (P<.001), with the lowest hazard rates being for Moderna, as compared to those who received the Janssen vaccine, independent of age, sex, comorbidities, vaccine type, and prior SARS-CoV-2 infection. Further, individuals who had a SARS-CoV-2 infection prior to vaccination had some increased protection over and above the protection already provided by the vaccines, from hospitalization (P=.001) and death (P=.04), independent of age, sex, comorbidities, and vaccine type. We found that the top statistically significant risk factors for severe breakthrough SARS-CoV-2 infections were age of >50, male gender, moderate and severe renal failure, severe liver disease, leukemia, chronic lung disease, coagulopathy, and alcohol abuse.

Conclusions: Among individuals who were fully vaccinated, the risk of severe breakthrough SARS-CoV-2 infection was lower for recipients of the Moderna or Pfizer vaccines and higher for recipients of the Janssen vaccine. These results from our analysis at a population level will be helpful to public health policy makers. Our result on the influence of a previous SARS-CoV-2 infection necessitates further research into the impact of multiple exposures on the risk of developing severe COVID-19.

(JMIR Public Health Surveill 2022;8(11):e38898) doi:10.2196/38898

KEYWORDS

breakthroughs; vaccines; Pfizer; Moderna; Janssen; SARS-CoV-2; COVID-19; coronavirus; infectious disease; viral infection; vaccination; breakthrough infection; public health; health policy; decision making; booster vaccine; mortality; hospitalization; healthcare system

Introduction

Despite widespread COVID-19 vaccination, high community levels of SARS-CoV-2 circulating throughout the United States have led to many breakthrough SARS-CoV-2 infections [1-3]. Breakthrough infections, where fully vaccinated individuals who are exposed to SARS-CoV-2 get infected, are generally uncommon (0.02% of fully vaccinated individuals reported developing breakthrough infections in a Washington state cohort [4]) and are generally less severe than infections in unvaccinated individuals [5,6]. There now exists a large body of literature studying the risk factors for severe COVID-19 disease, much of which has involved studies in unvaccinated populations [7-10], prior to the large-scale availability of vaccines. Studies on how these risks vary after vaccination are fewer in comparison, mostly focused on vaccine effectiveness in preventing SARS-CoV-2 infections or the influence of specific variants on vaccine effectiveness [11].

The impact of underlying factors on breakthrough infections are quite challenging to understand outside of randomized, placebo-controlled, double-blind field trials due to variation in their severity, distribution in the population, and contribution to transmission [11,12]. Early studies have found that a third dose of vaccine reduces the viral load in breakthrough infections, even for newer variants such as delta and omicron [13,14]. To understand the comparative advantages of the various vaccines [15], it is important to know the rate of severe COVID-19 disease leading to hospitalization or death among individuals who are fully vaccinated [16], as this will help policy makers.

While the risk of breakthrough SARS-CoV-2 infection has recently been reported by type of vaccine [17], little information exists regarding the risk of *hospitalization* or *mortality* by vaccine type for breakthrough infections [1]. In addition, while a prior SARS-CoV-2 infection is associated with a lower risk of breakthrough infection, it is unknown how large an effect a prior infection has on the severity of breakthrough COVID-19 infections, should one occur [18]. There has been a growing need for retrospective studies on severe breakthrough infections to address the misinformation and vaccine hesitancy in social and public spheres [19].

In this paper, we used de-identified US medical claims records from Change Healthcare to estimate the risk of hospitalization and death, by vaccine type, age, sex, comorbidity factors and previous SARS-CoV2 infection, among SARS-CoV-2 breakthrough infections that occurred between March 10, 2021, and October 14, 2021.

Methods

Ethics Approval and Consent to Participate

This study does not constitute as human subjects research due to the use and reporting of only deidentified observational data as determined by the human subjects committee of the University of Washington and thus does not require the review and approval by the institutional review board at the University of Washington.

Data Source

Our study uses de-identified US medical claims records from Change Healthcare collected over a period from March 1, 2020, to October 14, 2021, encompassing over 100 million records from over 8 million patients. Medical claims data contain details about a patient's interaction with the medical system, which are needed for the accurate billing of the transactions. Each claims record contains patient demographic information, International Classification of Diseases, 10th Revision (ICD-10) codes indicating primary diagnosis and secondary diagnoses, place of diagnosis, ICD-10 codes of procedures performed, patient status at the end of the visit, dates pertaining to the event (where different "from" and "to" dates indicate longer visits whereas the same "from" and "to" dates are for outpatient visits).

Our claims data set includes primarily open claims and a subset of closed payer claims that are normalized for analytics purposes. The open claims are derived from broad-based health care sources and consist of all medical claims that Change Healthcare processes and for which it has the usage rights. The closed claims are derived directly from the payer (ie, health insurance provider) and capture nearly all events that occur during the patient's enrollment period. Roughly 95% of the claims used for this study are commercial, and 5% are Medicare Advantage or other types of plans.

Study Population

Our data set of 8.18 million individuals contains only COVID-19 positive patients, defined as patients with at least one claims record with the ICD-10 diagnosis codes of "U07.1" or "U07.2" in any diagnosis field. We limited our analysis to individuals who had a primary diagnosis of "U07.1" (this is indicated by the principal diagnosis code, which encodes the primary diagnosis rendered by the medical facility or the primary cause of the visit). This ICD-10 diagnosis code indicates a COVID-19 diagnosis where the virus was identified in a lab-confirmed report. We exclude patients for whom the code of U07.1 appears in the "other diagnosis" fields, which contain the list of diagnoses made in addition to the primary cause of visit, which can be any other medical condition such as cancer. We also exclude patients with the code U07.2, which indicates a non-lab–confirmed COVID-19 diagnosis.

Subsequently, fully vaccinated individuals are identified by looking for procedure codes encoding the second doses of Pfizer (0002A) and Moderna (0012A) vaccines and the first dose of the Janssen (0031A) vaccine. We do not exclude patients with missing first dose claims records (~5% of the final study cohort), because patients who went to vaccination camps and were not required to provide insurance information would have missing first dose claims records. Some of these ~5% patients with missing first dose claims records may have had mixed vaccines (eg, Pfizer for the first dose and Moderna for the second dose). Since we did not believe that this will be a significant fraction of the vaccinated population, we do not exclude them. Breakthrough patients were defined as those who had a primary, lab-confirmed COVID-19 diagnosis at least 14 days after the date of vaccination. Please see Figure S1 in Multimedia Appendix 1 for a flow diagram showing the criteria used for cohort selection.

Hospitalization and Mortality

We explicitly identify hospitalization by looking for claims where the claim type is "institutional" or "professional" and the bill type indicates an "inpatient" facility. We also look at the dates associated with the hospital stay and only consider patients whose admission duration was at least 2 days (derived from the "admission_from" and "admission_to" date fields). For mortality, we look at the patient status code and consider all codes indicating "expired." As described already, we only consider cases where the primary diagnosis was COVID-19 for both hospitalization and expiration. The patient status code is available for all hospitalized patients but only for 42% of the outpatients (who went to clinics). Among patients who had the patient status code available, we found only 17 (0.22%) deaths out of a total of 7843 outpatients; therefore, we consider outpatients with missing patient status to be alive.

Study Period

COVID-19 vaccinations began in the United States in late December 2020. By late February 2021, the Pfizer-BioNTech (Pfizer), Moderna, and Johnson & Johnson (J&J/Janssen) vaccines were all approved for emergency use authorization. The Pfizer and Moderna vaccination drives started much earlier, in late December (Figures S2 and S3 in Multimedia Appendix 1), as compared to that of the Janssen vaccines, which also saw a stall in vaccine rates in mid-April (Figure S4 in Multimedia Appendix 1). To keep the *COVID-19 exposure* of the individuals taking any of the 3 vaccines consistent, we use the same study window, though we have data for Pfizer and Moderna from late December. We construct our cohort to consist of individuals who were fully vaccinated between March 10, 2021, and April 27, 2021, the period during which all 3 vaccines were being widely administered. Every individual in this cohort was followed from the date of vaccination of each individual up to the end of the study period, October 14, 2021. The study period over the entire cohort is thus March 10, 2021, to October 14, 2021. In Figure S6 in Multimedia Appendix 1, we show statistics showing the number of days of follow-up after full vaccination by vaccine type in our study cohort.

Comorbidities

Preexisting comorbidities were defined based on ICD-10 codes assigned to medical encounters, which contain pointers to previously diagnosed conditions, using claims records during the 6-month period from March 2020 to September 2020. This period does not overlap with the study period, so events during the study period will not also be counted as comorbidities. The Elixhauser comorbidity index [20] was used to define comorbid conditions. This index has a series of codes that define comorbidities with each code mapping to one or several ICD-10 diagnosis codes. For example, the Elixhauser code "BLDLOSS" (blood loss) includes the following four ICD-10 diagnosis codes: D50.0, O90.81, O99.02, and O99.03. We provide the index that we used and the corresponding ICD-10 codes in Multimedia Appendix 2. We also show the relative abundance of comorbidities in our cohort, by vaccine type, in Table S2 and Figure S5 in Multimedia Appendix 1.

Previous COVID-19 Infection

Since some of the individuals in our cohort may have had a COVID-19 infection during the year 2020, we introduce an additional feature to encode the effect of already being infected with COVID-19. This feature is "yes" if we see a claim involving a COVID-19 diagnosis in any diagnosis field, in the period from March 1, 2020, to the beginning of the study period, March 10, 2021.

Statistical Methods

Date of full vaccination was defined as 14 days after (1) a single Janssen vaccine, (2) the second Moderna vaccine dose, or (3) the second Pfizer vaccine dose. Cox proportional hazards regression was used to estimate univariate hazard ratios (HRs) and multivariable HRs in a model including the following features: age (categorized), sex (male and female), vaccine type, Elixhauser comorbidities (encoded as independent binary variables), and SARS-CoV-2 infection prior to the first dose of vaccination (yes or no). We remove the negligible number of individuals with sex=unknown. We also model interactions between vaccine type and all other covariates as well as previous infection and all other covariates but find that none were statistically significant. Further, the interaction terms had a negligible impact on the hazard ratios of the other terms and were thus removed for greater clarity in the results. All analyses were performed using the "coxph" function from the R package "survival" (R Foundation for Statistical Computing) [21].

Results

Our study includes 19,815 fully vaccinated patients with breakthrough SARS-CoV-2 infections between March 10, 2021, and October 14, 2021. Of those patients, 11,339 (57.22%) received the Pfizer vaccine, 5480 (27.66%) received the Moderna vaccine, and 2996 (15.12%) received the Janssen vaccine. Breakthrough cases receiving Janssen were younger than those receiving Pfizer or Moderna and had a slightly greater proportion of male patients (Table 1). Breakthrough cases receiving Moderna had a greater proportion of patients with COVID-19 prior to vaccination.

Risk of hospitalization and mortality among breakthrough cases increased with older age and was higher for male patients (Table 2). In multivariable analyses controlling for age, male sex, comorbidities, and prior SARS-CoV-2 infection, the risk of hospitalization was the lowest for breakthrough cases receiving the Moderna vaccine (adjusted hazard ratio [aHR]: 0.42, 95% CI 0.35-0.5; $P<.001$), comparably low for Pfizer vaccinated individuals (aHR: 0.45, 95% CI 0.39-0.53; $P<.001$), compared with that for the recipients of the Janssen vaccine. The comorbidities with statistically significant HRs for hospitalization or mortality from a breakthrough SARS-CoV-2 infection include severe liver disease, moderate and severe renal failures, alcohol abuse, chronic lung disease, coagulopathy, cancers, anemia, seizures, and arthritis (Table S1 in Multimedia Appendix 1).

Table 1. Characteristics of SARS-CoV-2 breakthrough infections cohort tracked from March 10, 2021, to October 14, 2021. Prevalence of comorbidities by vaccine type is shown in Figure S5 in Multimedia Appendix 1.

Variable	Pfizer (n=11,339), n (%)	Moderna (n=5480), n (%)	Janssen (n=2996), n (%)	Overall (n=19,815), n (%)
Age range (years)				
0-20	108 (0.95)	31 (0.57)	34 (1.13)	173 (0.87)
20-35	1005 (8.86)	455 (8.30)	337 (11.25)	1797 (9.07)
35-50	1801 (15.88)	795 (14.51)	722 (24.10)	3318 (16.74)
50-64	3663 (32.30)	1684 (30.73)	1224 (40.85)	6571 (33.16)
64-80	4007 (35.34)	2041 (37.24)	580 (19.36)	6628 (33.45)
>80	755 (6.66)	474 (8.65)	99 (3.30)	1328 (6.70)
Sex (male)	5032 (44.4)	2360 (43.06)	1385 (46.2)	8777 (44.23)
SARS-CoV2 infection before vaccination	1536 (13.5)	1137 (20.7)	437 (13.9)	3090 (15.6)

Table 2. Correlates of hospitalization and mortality after breakthrough SARS-CoV-2 infection, estimated from Cox proportional hazards models. We show the adjusted hazard ratio (aHR) and the 95% CI for the significant correlates (*P* values indicated via superscripts d, e, and f). An aHR of <1.0 indicates a lower risk of hospitalization or mortality as compared to the baseline population for that covariate (analogously, aHR>1.0 indicates a higher risk than the baseline). Hazard ratios (HRs) of comorbidities are shown in Table S1 in Multimedia Appendix 1.

Variable	Hospitalization (n/pys[a])	Hospitalization, univariate HR (95% CI)[b]	Hospitalization, multivariate aHR (95% CI)[b]	Mortality (n/pys)	Mortality, univariate HR (95% CI)[c]	Mortality, multivariate aHR (95% CI)[c]
Vaccine						
Pfizer	20.1	0.55 (1.8-0.46)[d]	0.45 (0.39-0.53)[d]	2.6	0.68 (1.5-0.45)	0.43 (0.28-0.65)[d]
Moderna	19.2	0.59 (1.7-0.5)[d]	0.42 (0.35-0.5)[d]	2.3	0.61 (1.6-0.37;[e] (*P*=.04)	0.38 (0.23-0.62)[d]
Janssen	26.5	1.0	1.0	3.0	1.0	1.0
Age range (years)						
0-20	1.9	0.29 (0.04-2.1)	0.30 (0.04-2.2)	1.9	7.48 (0.78-71.9)	7.8 (0.81-75)
20-35	1.9	0.27 (0.15-0.5)[d]	0.29 (0.15-0.54)[d]	0.0	0 (0)	0 (0)
35-50	6.8	1.0	1.0	0.3	1.0	1.0
50-64	16.9	2.08 (1.62-2.7)[d]	2.1 (1.6-2.7)[d]	1.8	5.82 (1.8-19.0)[f] (*P*=.004)	5.98 (1.8-20)[f] (*P*=.003)
64-80	31.7	2.96 (2.33-3.7)[d]	3.32 (2.6-4.2)[d]	3.9	12.30 (3.9-38.9)[d]	14.20 (4.5-45)[d]
>80	52.9	4.35 (3.34-5.7)[d]	4.99 (3.8-6.5)[d]	9.1	24.60 (7.6-79.7)[d]	29.10 (8.9-95)[d]
Sex						
Female	17.5	1.0	1.0	2.2	1.0	1.0
Male	25.0	1.38 (1.23-1.5)[d]	1.25 (1.1-1.4)[d]	3.0	1.26 (0.93-1.7)	1.11 (0.82-1.5)
SARS-CoV2 infection before vaccination						
No	21.9	1.0	1.0	2.7	1.0	1.0
Yes	7.5	0.56 (0.4-0.78)[d]	0.57 (0.41-0.80)[f] (*P*=.001 for above aHR)	0.5	0.21 (0.05-0.84)[f] (*P*<.01 for above aHR)	0.22 (0.05-0.91)[e] (*P*=.04 for above aHR)

[a]Incidence per 100 person years.

[b]n=19,815; events=1140.

[c]n=19,815; events=159.

[d]*P*<.001.

[e]*P*<.05.

[f]*P*<.01.

We see a similar trend with the risk of mortality for breakthrough cases, with the risk being the lowest for those receiving the Moderna vaccines (aHR: 0.38, 95% CI 0.23-0.62; *P*<.001) and comparably lower for Pfizer recipients (aHR: 0.43, 95% CI 0.28-0.65; *P*<.001) as compared to that for Janssen recipients. Finally, as expected, the protection offered by vaccines was enhanced for breakthrough cases who already had a previous SARS-CoV-2 infection. These individuals were 40% less likely to be hospitalized due to COVID-19 (aHR: 0.57, 95% CI 0.41-0.80; *P*=.001) and four times less likely to die of COVID-19 (aHR: 0.22, 95% CI 0.05-0.91; *P*=.04), when compared to those without a prior SARS-CoV-2 infection independent of age, sex, comorbidities, and vaccine type.

We repeat this analysis by excluding the population who had a prior SARS-CoV-2 infection for completeness and show the resulting HRs Table S3 in Multimedia Appendix 1.

Discussion

Principal Findings

Using medical claims data, we found that the risk of hospitalization in SARS-CoV-2 breakthrough infections was lower for those receiving the Moderna and Pfizer vaccines compared to those receiving the Janssen vaccine. The risk of mortality was similarly low in breakthrough infections who received Pfizer and Moderna vaccines compared to those receiving the Janssen vaccine. There was no statistically significant difference between the HRs of Pfizer and Moderna for both risks. We also found older age, male sex, and certain comorbidities to be risk factors for hospitalization and mortality in breakthrough infections. Further, we found that risk of hospitalization was 40% less and risk of death was 75% less in SARS-CoV-2 breakthrough infections among individuals who already had a SARS-CoV-2 infection prior to their vaccination compared with fully vaccinated individuals without a previous SARS-CoV-2 infection. While other studies have reported lower risk of breakthrough infection with previous SARS-CoV-2 infection [18], our study analyzes both hospitalization and mortality and shows that the immunity provided by previous infection seems to increase the protection provided by vaccines, against severe COVID-19, independent of vaccine type, age, comorbidities, and sex. Since our cohort only consists of individuals who were all fully vaccinated, this is by no means a comparison of vaccine-induced immunity against acquired immunity from previous infections.

Excluding patients who had COVID-19 infection prior to vaccination from our Cox regression analysis results in a similar HR for hospitalization risk in patients who received the Pfizer (aHR=0.42) and Moderna (aHR=0.41) vaccines (Table S3 in Multimedia Appendix 1). This might be explained by the fact that 20.7% of patients who received Moderna had a prior COVID-19 infection as compared to ~13% of patients who received Pfizer. Hence, removing all patients with prior COVID-19 infection reduced the influence of the additional immunity that some of Moderna-vaccinated individuals had.

A number of studies have found that age has a direct effect on the risk of severe COVID-19 disease [22,23]. We find that the proportion of the elderly cohort in our data set who were hospitalized is much higher than the proportion of the younger cohort (Figure S7 in multimedia Appendix 1). In addition, we find a higher HR for the elderly subset of our study cohort (aHR=2.1 for age>50, aHR=3.3 for age>65, and aHR=5.0 for age>80, as compared to the baseline age group of 35-50 years).

Our findings comparing vaccine types are similar to those reported by the Centers for Disease Control and Prevention for mortality but provide additional information by vaccine type [1,16]. There have been several studies on individual risk factors such as age [22], specific comorbidities [7], focused populations such as Veterans [24], or large-scale projects such as OpenSAFELY, which involved 17 million unvaccinated patients [10]. Our work advances this body of literature by analyzing vaccine type, age, sex, and 39 different comorbidities in a large cohort of breakthrough patients from the general US population. Some of the risk factors that we find for severe breakthrough SARS-CoV-2 infections, such as age, male gender, and certain comorbidities (eg, chronic lung infection, kidney disease, and cancers) are similar to what has been reported in prior studies of SARS-CoV-2 infections among unvaccinated individuals [8,23,25]. However, we find that some risk factors found by initial studies such as hypertension are not a risk factor for breakthrough COVID-19 hospitalization or death (aHRs of 0.75 and 0.59, respectively), neither are diabetes or congestive heart failure. We instead find that both moderate and severe renal failure are significant risk factors, independent of age or other factors, which agrees with other large-scale studies such as OpenSAFELY [10] and the Global Burden of Disease collaboration [26] which identified that worldwide chronic kidney disease is the most prevalent risk factor for severe COVID-19. Even mild impairment of renal function has been found to be an independent risk factor for COVID-19 infection, hospitalization, and mortality [27].

Lastly, to understand the association between outcomes and the *time of vaccination*, we incorporate a variable indicating the number of days between full vaccination and the onset of the surge in infections caused by the delta variant. However, our population-based data set is inadequate to derive any significant conclusions vis-à-vis the best time for vaccination in anticipation of a surge.

Limitations

Limitations of our study include, first, a lack of access to data on unvaccinated individuals or those who had a negative SARS-CoV-2 test result. The former is due to the lack of a medical claims record for vaccinations that were done in vaccination drives and camps; the absence of a vaccination-related claim in our data set therefore does not imply an unvaccinated individual. Second, our medical claims source consists of mostly privately insured individuals and can thus miss people who may be susceptible to the most adverse outcomes. Another caveat of our data set is that most of the claims are open claims, which have the benefit of capturing a patient's activities over a longer time frame regardless of their insurance provider, but do not necessarily track all medical encounters of patients.

Conclusions

Our findings add to the growing literature regarding the risk factors for severe breakthrough SARS-CoV-2 infections in fully vaccinated individuals, where we identify the influence of age, sex, and comorbidities that are risk factors; importantly, we found that previous SARS-CoV-2 infections can provide additional protection over that offered by vaccines against severe disease. Our results also necessitate further studies on the optimal number of vaccine doses to protect from the most severe breakthrough SARS-CoV-2 infections. An important strength of our study is that we consider US-wide breakthrough hospitalizations covering a broad demographic and compare all 3 vaccines, whereas most previous studies lack specific data on Janssen.

Acknowledgments

MK worked on the study design, experiments, analysis, and writing the paper. MN, SM, NB, RD, WBW, JLF, and BR worked on the study design and writing the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental data.
[DOCX File , 838 KB - publichealth_v8i11e38898_app1.docx]

Multimedia Appendix 2
Elixhauser comorbidities table.
[XLSX File (Microsoft Excel File), 571 KB - publichealth_v8i11e38898_app2.xlsx]

References

1. COVID data tracker. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker/#rates-by-vaccine-status [accessed 2022-11-01]
2. Uschner D, Bott M, Santacatterina M, Gunaratne MP, Fette L, Burke BK, et al. Breakthrough SARS-CoV-2 infections after vaccination in North Carolina. medRxiv Preprint posted online October 13, 2021. [doi: 10.1101/2021.10.10.21264812]
3. Christensen PA, Olsen RJ, Long SW, Subedi S, Davis JJ, Hodjat P, et al. Delta variants of SARS-CoV-2 cause significantly increased vaccine breakthrough COVID-19 cases in Houston, Texas. Am J Pathol 2022 Feb;192(2):320-331 [FREE Full text] [doi: 10.1016/j.ajpath.2021.10.019] [Medline: 34774517]
4. SARS-CoV-2 vaccine breakthrough surveillance and case information resource. Washington State Department of Health. 2022. URL: https://www.doh.wa.gov/Portals/1/Documents/1600/coronavirus/data-tables/420-339-VaccineBreakthroughReport.pdf [accessed 2022-11-01]
5. Abu-Raddad L, Chemaitelly H, Ayoub HH, Tang P, Coyle P, Hasan MR, et al. Relative infectiousness of SARS-CoV-2 vaccine breakthrough infections, reinfections, and primary infections. Nat Commun 2022 Jan 27;13(1):532 [FREE Full text] [doi: 10.1038/s41467-022-28199-7] [Medline: 35087035]
6. Bergwerk M, Gonen T, Lustig Y, Amit S, Lipsitch M, Cohen C, et al. Covid-19 breakthrough infections in vaccinated health care workers. N Engl J Med 2021 Oct 14;385(16):1474-1484. [doi: 10.1056/nejmoa2109072]
7. Klang E, Kassim G, Soffer S, Freeman R, Levin MA, Reich DL. Severe obesity as an independent risk factor for COVID-19 mortality in hospitalized patients younger than 50. Obesity (Silver Spring) 2020 Sep 02;28(9):1595-1599 [FREE Full text] [doi: 10.1002/oby.22913] [Medline: 32445512]
8. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. Kidney Int 2020 May;97(5):829-838 [FREE Full text] [doi: 10.1016/j.kint.2020.03.005] [Medline: 32247631]
9. ERA-EDTA Council, ERACODA Working Group. Chronic kidney disease is a key risk factor for severe COVID-19: a call to action by the ERA-EDTA. Nephrol Dial Transplant 2021 Jan 01;36(1):87-94 [FREE Full text] [doi: 10.1093/ndt/gfaa314] [Medline: 33340043]
10. Williamson E, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020 Aug;584(7821):430-436 [FREE Full text] [doi: 10.1038/s41586-020-2521-4] [Medline: 32640463]
11. Lipsitch M, Krammer F, Regev-Yochay G, Lustig Y, Balicer RD. SARS-CoV-2 breakthrough infections in vaccinated individuals: measurement, causes and impact. Nat Rev Immunol 2022 Jan 07;22(1):57-65 [FREE Full text] [doi: 10.1038/s41577-021-00662-4] [Medline: 34876702]

12. Sun J, Zheng Q, Madhira V, Olex AL, Anzalone AJ, Vinson A, National COVID Cohort Collaborative (N3C) Consortium. Association Between Immune Dysfunction and COVID-19 Breakthrough Infection After SARS-CoV-2 Vaccination in the US. JAMA Intern Med 2022 Feb 01;182(2):153-162. [doi: 10.1001/jamainternmed.2021.7024] [Medline: 34962505]
13. Buchan S, Chung H, Brown KA, Austin PC, Fell DB, Gubbay JB, et al. Estimated effectiveness of COVID-19 vaccines against omicron or delta symptomatic infection and severe outcomes. JAMA Netw Open 2022 Sep 01;5(9):e2232760 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.32760] [Medline: 36136332]
14. Levine-Tiefenbrun M, Yelin I, Alapi H, Katz R, Herzel E, Kuint J, et al. Viral loads of Delta-variant SARS-CoV-2 breakthrough infections after vaccination and booster with BNT162b2. Nat Med 2021 Dec 02;27(12):2108-2110. [doi: 10.1038/s41591-021-01575-4] [Medline: 34728830]
15. Han X, Xu P, Ye Q. Analysis of COVID-19 vaccines: Types, thoughts, and application. J Clin Lab Anal 2021 Sep;35(9):e23937 [FREE Full text] [doi: 10.1002/jcla.23937] [Medline: 34396586]
16. Rates of laboratory-confirmed COVID-19 hospitalizations by vaccination status. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker/#covidnet-hospitalizations-vaccination [accessed 2022-11-01]
17. Liu C, Lee J, Ta C, Soroush A, Rogers JR, Kim JH, et al. Risk factors associated with SARS-CoV-2 breakthrough infections in fully mRNA-vaccinated individuals: Retrospective analysis. JMIR Public Health Surveill 2022 May 24;8(5):e35311 [FREE Full text] [doi: 10.2196/35311] [Medline: 35486806]
18. Abu-Raddad LJ, Chemaitelly H, Ayoub HH, Yassine HM, Benslimane FM, Al Khatib HA, et al. Association of prior SARS-CoV-2 infection with risk of breakthrough infection following mRNA vaccination in Qatar. JAMA 2021 Nov 16;326(19):1930-1939 [FREE Full text] [doi: 10.1001/jama.2021.19623] [Medline: 34724027]
19. Hou Z, Tong Y, Du F, Lu L, Zhao S, Yu K, et al. Assessing COVID-19 vaccine hesitancy, confidence, and public engagement: A global social listening study. J Med Internet Res 2021 Jun 11;23(6):e27632. [doi: 10.2196/27632] [Medline: 34061757]
20. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998 Jan;36(1):8-27. [doi: 10.1097/00005650-199801000-00004] [Medline: 9431328]
21. Therneau T, Lumlet T. R survival package. R Project. 2013. URL: https://cran.r-project.org/web/packages/survival/index.html [accessed 2022-11-01]
22. Lu G, Zhang Y, Zhang H, Ai J, He L, Yuan X, et al. Geriatric risk and protective factors for serious COVID-19 outcomes among older adults in Shanghai Omicron wave. Emerg Microbes Infect 2022 Dec;11(1):2045-2054 [FREE Full text] [doi: 10.1080/22221751.2022.2109517] [Medline: 35924388]
23. 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]
24. Cardemil CV, Dahl R, Prill MM, Cates J, Brown S, Perea A, et al. COVID-19-related hospitalization rates and severe outcomes among veterans from 5 veterans affairs medical centers: Hospital-based surveillance study. JMIR Public Health Surveill 2021 Jan 22;7(1):e24502 [FREE Full text] [doi: 10.2196/24502] [Medline: 33338028]
25. Steenblock C, Schwarz PEH, Ludwig B, Linkermann A, Zimmet P, Kulebyakin K, et al. COVID-19 and metabolic disease: mechanisms and clinical management. The Lancet Diabetes & Endocrinology 2021 Nov;9(11):786-798 [FREE Full text] [doi: 10.1016/s2213-8587(21)00244-8]
26. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang H, Mercer S, et al. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. The Lancet Global Health 2020 Aug;8(8):e1003-e1017 [FREE Full text] [doi: 10.1016/S2214-109X(20)30264-3]
27. Liakopoulos V, Roumeliotis S, Papachristou S, Papanas N. COVID-19 and the kidney: time to take a closer look. Int Urol Nephrol 2022 May 12;54(5):1053-1057 [FREE Full text] [doi: 10.1007/s11255-021-02976-7] [Medline: 34383205]

Abbreviations

aHR: adjusted hazard ratio
HR: hazard ratio
ICD-10: International Classification of Diseases, 10th Revision

Edited by A Mavragani, G Eysenbach; submitted 20.04.22; peer-reviewed by W Murk, O Rahaman; comments to author 13.09.22; revised version received 06.10.22; accepted 18.10.22; published 08.11.22.

Please cite as:
Kshirsagar M, Nasir M, Mukherjee S, Becker N, Dodhia R, Weeks WB, Ferres JL, Richardson B
The Risk of Hospitalization and Mortality After Breakthrough SARS-CoV-2 Infection by Vaccine Type: Observational Study of Medical Claims Data
JMIR Public Health Surveill 2022;8(11):e38898
URL: https://publichealth.jmir.org/2022/11/e38898
doi:10.2196/38898
PMID:36265135

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org