Contents

Original Papers

Pilot Influenza Syndromic Surveillance System Based on Absenteeism and Temperature in China: Development and Usability Study (e37177)
Zhen Yang, Chenghua Jiang .................................................. 2

Spatiotemporal Distribution of HIV Self-testing Kits Purchased on the Web and Implications for HIV Prevention in China: Population-Based Study (e35272)
Ganfeng Luo, Lingyun Su, Anping Feng, Yi-Fan Lin, Yiguo Zhou, Tanwei Yuan, Yuqing Hu, Song Fan, Yong Lu, Yingsi Lai, Qian Shi, Jun Li, Mengjie Han, Huachun Zou .............................................................. 16

COVID-19 Cases Among Congregate Care Facility Staff by Neighborhood of Residence and Social and Structural Determinants: Observational Study (e34927)
Huiting Ma, Kristy Yiu, Stefan Baral, Christine Fahim, Gary Moloney, Dariya Darwin, David Landsman, Adrienne Chan, Sharon Straus, Sharmistha Mishra .......................................................... 47

Corrigenda and Addendas

Correction: Cost-Effectiveness of Lung Cancer Screening Using Low-Dose Computed Tomography Based on Start Age and Interval in China: Modeling Study (e43025)
Zixuan Zhao, Lingbin Du, Yuanjuan Li, Le Wang, Youqing Wang, Yi Yang, Hengjin Dong .................................................. 42

Correction: Google Trends on Human Papillomavirus Vaccine Searches in the United States From 2010 to 2021: Infodemiology Study (e42812)
Akshaya Bhagavathula, Philip Massey .................................................. 44
Pilot Influenza Syndromic Surveillance System Based on Absenteeism and Temperature in China: Development and Usability Study

Zhen Yang1, PhD; Chenghua Jiang2, PhD

1School of Medicine, Tongji University, Shanghai, China
2Dongfang Hospital, Tongji University, Shanghai, China

Corresponding Author:
Chenghua Jiang, PhD
Dongfang Hospital
Tongji University
No 1239 Siping Road, Yangpu District
Shanghai, 200092
China
Phone: 86 18917266778
Email: jchtongji@163.com

Abstract

Background: Shortcomings of the current school-based infectious disease syndromic surveillance system (SSS) in China include relying on school physicians to collect data manually and ignoring the health information of students in attendance.

Objective: This study aimed to design and implement an influenza SSS based on the absenteeism (collected by face recognition) and temperature of attending students (measured by thermal imaging).

Methods: An SSS was implemented by extending the functionality of an existing application. The system was implemented in 2 primary schools and 1 junior high school in the Yangtze River Delta, with a total of 3535 students. The examination period was from March 1, 2021, to January 14, 2022, with 174 effective days. The daily and weekly absenteeism and fever rates reported by the system (DAR1 and DFR; WAR1 and WFR) were calculated. The daily and weekly absenteeism rates reported by school physicians (DAR2 and WAR2) and the weekly positive rate of influenza virus (WPRIV, released by the Chinese National Influenza Center) were used as standards to evaluate the quality of the data reported by the system.

Results: Absenteeism reported by school physicians (completeness 86.7%) was 36.5% of that reported by this system (completeness 100%), and a significant positive correlation between them was detected ($r=0.372, P=.002$). When the influenza activity level was moderate, DAR1s were significantly positively correlated among schools ($r_{ab}=0.508, P=.004$; $r_{bc}=0.427, P=.02$; $r_{ac}=0.447, P=.01$). During the influenza breakout, the gap of DAR1s widened. WAR1 peaked 2 weeks earlier in schools A and B than in school C. Variables significantly positively correlated with the WPRIV were the WAR1 of school A, WAR1 and WFR were calculated. The daily and weekly absenteeism rates reported by school physicians (DAR2 and WAR2) and the weekly positive rate of influenza virus (WPRIV, released by the Chinese National Influenza Center) were used as standards to evaluate the quality of the data reported by the system.

Conclusions: Data demonstrated that absenteeism calculation based on face recognition was reliable, but the accuracy of the temperature recorded by the infrared thermometer should be enhanced. Compared with similar SSSs, this system has superior simplicity, cost-effectiveness, data quality, sensitivity, and timeliness.

(JMIR Public Health Surveill 2022;8(10):e37177) doi:10.2196/37177

KEYWORDS

influenza; syndromic surveillance system; face recognition; infrared thermometer; absenteeism; temperature
Introduction

School-aged children are vulnerable to influenza. The incidence of influenza among children aged 5-14 years during the influenza season ranges from 17.31% to 46.61% [1]. School-aged children also contribute to the amplification of virus transmission. Epidemiologic evidence suggests that influenza first occurs among school-aged children and that once infected, this population transmits the virus among family members and subsequently to the general population [2,3]. An infected student spreads the virus to an estimated 2.4 (95% CI 1.8-3.2) other children within a school [4]. Incidence rates in adults who reside with school-aged children may be 2 to 3 times higher than the incidence rates in similar adults who do not reside with school-aged children [5].

Absenteism is an influenza surveillance indicator recommended by the World Health Organization and an important determinant adopted in the school-based syndromic surveillance system (SSS). In 1979, Peterson et al [6] demonstrated the effectiveness of school absenteeism for influenza surveillance. Since then, several studies have discussed the value of absenteeism surveillance from different perspectives. Compared to other methods, the main advantages of absenteeism surveillance include noninvasiveness, no requirement of clinical testing, low cost, simple operation, and good representation. Additionally, absenteeism surveillance can be used to accurately estimate the economic burden of influenza and its impact on education and promote effective cooperation between the health and education departments [6-12]. However, absenteeism itself is not a direct manifestation of the clinical symptoms of the disease, but only an approximate estimate of the illness. Therefore, it is sensitive but lacks specificity [10]. Baer et al [11] suggested that the greatest value of absenteeism surveillance may lie in “situational awareness” rather than “early detection.”

Absenteism can be divided into three types: all-cause absence, illness absence, and syndrome-specific absence [13]. These 3 indicators have high specificity, but they increase the workload of the school. The SSS must balance specificity and the burden on the school [12]. Otherwise, the schools will be uncooperative, and the system will be eventually rejected [11,14,15]. Reducing the burden to maintain and improve the compliance of schools is an important challenge for the effective operation of such systems. Researchers have continuously been focusing on three main technical approaches to address this challenge: (1) improving the automation of data collection and changing the data collection method from manual statistics to fingerprint or smart card methods [16]; (2) improving the convenience of data transmission, with data transmission changing from postcards, telephone, and fax [6] to email [10], and then to widely used network platforms [7-9,11,14-16]; (3) diversifying data reporters, with the reporter usually being a school physician or teacher and some systems requiring parents to report information [15,17] or encouraging students to participate in reporting the information [18].

Public health emergencies of infectious diseases emerging from schools account for 85.64% of the total number of all annual national public health emergencies [19]. Therefore, school-based SSSs are particularly significant in China. The SARS outbreak in 2003 and the H1N1 pandemic in 2009 prompted many places in mainland China to build systems for school-based infectious disease symptom surveillance [15,20-22], but the continuous operation of such systems is challenging. First, although the government formulated the school infectious disease reporting standard as early as 2012 [23] and issued several documents emphasizing the reporting of epidemic information in schools during the COVID-19 pandemic [24,25], absenteeism statistics are still not mandatory for schools. Second, the severe shortage of school physicians exacerbates the problem. Only 33.1% of primary and secondary schools are equipped with school physicians, with 1 school physician serving 2800 students on average; the proportions of school physicians in central and western regions, rural areas, and low-grade schools are even lower [26]. Therefore, the system based on manual information reports struggles to stimulate enthusiasm in schools [16]. Finally, Chinese people attach significant importance to studying, and parents worry that absence will delay the learning progress of their children. Consequently, they will send their children to school even if they are ill [27], which increases the possibility of misjudging the epidemic situation of infectious diseases just by calculating absenteeism.

Using artificial intelligence and information technology instead of manual methods to collect absenteeism data while considering the health information of students in attendance is highly important to solve the current development dilemma of SSS in China. To this end, an SSS was designed and trial-operated, which realized synchronous acquisition of identity, absenteeism, and temperature data based on face recognition and thermal imaging. The system has been in operation at 3 sentinel schools in the Yangtze River Delta region since November 2020. Data from these 3 schools collected by the system between March 2021 and January 2022 were exported to investigate 2 aspects. First, the alternative method of manually collecting absenteeism data explored in previous studies has produced a large amount of erroneous data [14]. Simultaneously, although infrared temperature measurement technology is widely used in the screening of suspected cases of infectious diseases in the population [28], instruments as well as environmental and individual factors can easily interfere with its accuracy [29-31]. Therefore, the completeness and accuracy of the collected data need to be verified. Second, body temperature is the first clinical symptom of most infectious diseases, and its advantage in influenza SSS has been confirmed by Miller et al [32]. Theoretically, multisource data can improve the accuracy of the perception of infectious disease information by the surveillance system [33]. The monitoring effectiveness of the combination of nonclinical symptoms (absenteism) and clinical symptoms (fever) is another issue that requires investigation.

Methods

Reporting System

The proposed SSS was based on an app called “Xiao Lian Xing” (XLL) [34], which can be downloaded for free on Alipay (Alibaba Group Holding Co, Ltd). The app integrates face recognition technology and infrared temperature measurement
technology to realize intelligent management of school public health. Its workflow is shown in Figure 1.

The operating company signed service agreements with the school and its superior management organization. According to the agreement, the school organized for the parents to register an account in Alipay with a smartphone (registration was completely voluntary), and then input information about the child, such as the name, gender, ID, school name, class, and face image into the account. Parents can also use the account to check the attendance and temperature of their children. The system collected student attendance and temperature data through terminal devices, which were composed of 2 modules, namely a face recognition system (identification accuracy ≥99.99%, Sunmi Technology Group Co, Ltd) and an infrared thermometer (measuring accuracy of ±0.5°C, Hikvision Digital Technology Co, Ltd). The 2 modules were connected using a self-designed software for synchronous input and upload of the identity, attendance, and temperature. The data processing center was responsible for processing the data uploaded by each terminal and relaying the analysis results of absenteeism and fever at different levels to family users (ie, parents or other guardians), school users (ie, teachers, school health workers, and administrators), and school district management department users.

Terminal devices were generally arranged at the school gate. After students arrived at school every day, they went to the device for testing. Face recognition of the students was performed, and their identity information, attendance information, and facial temperature were automatically recorded by the instrument. Only students who were identified and had a normal temperature were admitted directly, whereas students with fever (temperature ≥37.3°C) were processed separately. To ensure that every student was subjected to instrument testing, the school arranged for staff on duty to supervise the children. Daily information on student attendance and temperature was transmitted to the cloud data center in real time. The data processing center generated a daily statistical report of the check-in and body temperature (with infrared images) of students in real time (Figure 2), summarized and analyzed the data, and gave feedback on the results at different levels (individual, classes, schools, school districts, etc).

Figure 1. Workflow of the Xiao Lian Xing system.
Study Population and Data Collection

The study population comprised 3 schools in the Yangtze River Delta region. Schools A and B were primary schools. School A, located in Xiaoshan District, Hangzhou city, Zhejiang Province, began to use version 1.0 of the system (only face recognition) in November 2020 and version 2.0 (face recognition and infrared thermometer) in March 2021. School B, located in Binjiang District of Hangzhou city, began to pilot version 2.0 of the system in March 2021; it allowed students to use smart cards to check-in and began to fully use version 2.0 in September 2021. School C is a junior high school located in Pudong District, Shanghai, and it started to use version 2.0 of this system in October 2021.

The surveillance period was divided into 2 phases, namely, phase I, from March 1, 2021, to June 25, 2021, and phase II, from September 1, 2021, to January 14, 2022. The effective surveillance times of these phases were 83 and 91 days in total, respectively. In phase I, the numbers of students in school A and B were 1861 and 1100, respectively. In phase II, the numbers of students in schools A, B, and C were 1954, 1154, and 427, respectively, with a total of 3535 students. The system collected the following data for each school: total number of enrolled students, number of daily absentees, and number of daily students with fever. By default, students who failed to check-in within 1 hour after the specified arrival time were counted as absenteeism cases, where students whose body temperature was \( \geq 37.3^\circ C \) were counted as fever cases.

The absenteeism reported by physicians in schools A and B (school C has no record of the school physician) was collected as the reference standard to evaluate the quality of absenteeism reported by the system. Absenteeism reported by the school physicians was defined as “the student is not in school that day.” The data reported by school physicians were collected from September 1, 2021, to January 14, 2022. To verify the reliability and feasibility of the surveillance system for infectious diseases, influenza was selected as the target disease. The reference criteria for influenza activities were obtained from the weekly influenza surveillance report released by Chinese National Influenza Center [35]. These weekly report statistics show the weekly positive rate of influenza virus (WPRIV) tests in southern and northern China. The WPRIV was calculated as the ratio of the number of virus-positive samples to the total number of samples submitted [36]. As both Hangzhou and Shanghai are in southern China, only the data of the southern region in the weekly report were used. The corresponding surveillance weeks ranged from the ninth week of 2021 to the fifth week of 2022. During this period, the influenza virus strain circulating in southern China was mainly type B.

Ethics Approval

Data used in this study were anonymized, so the Tongji University Review Board classified this study as nonhuman subject research, and it was exempted from approval.

Data Analysis

The daily absence rate (DAR) of the 3 schools was calculated as follows:

\[
\text{DAR} = \frac{\text{Number of absentees}}{\text{Total number of enrolled students}} \times 100\%
\]

The DAR is divided into two categories: system-reported (DAR1) and school physician–reported (DAR2). The correlation coefficient of the 2 variables was then calculated. According to relevant literature [9,12], the data were considered abnormal if DAR1 exceeded 10%. The abnormal causes relating to the school were investigated, and if they were noninfectious factors, the abnormal data were statistically treated in an appropriate manner. Then, the DAR time series diagrams of the 3 schools were drawn (Figure 3), the Pearson correlation of the DARs among the 3 schools was calculated, and the morphological differences among their DAR curves were compared. We
compared the correlation and trend of DAR1 and DAR2 (Figure 4).

Second, according to the time series diagram of DAR1 (Figure 3), the starting date was determined as the date when the DAR1 of each school began to stabilize, and then the daily fever rate (DFR) of each school was calculated after this date. When the DFR deviated by 3 SDs from the mean value, the infrared images of students in the system were examined. If the problem was operational, the corresponding DFR was used as the missing value and was replaced with the mean of the DFR for the previous and next days of that date. Absenteeism and fever were assumed to represent different severity degrees of influenza symptoms, with absenteeism representing severe symptoms and attendance with fever representing mild symptoms. Therefore, the denominators of both the DAR and DFR were set as the number of total enrollments. The calculation formula of the DFR was as follows:

The DFR time series diagrams of the 3 schools are shown in Figure 5; the morphological differences among their DFR curves were compared, and the Pearson correlation of the DFRs among the 3 schools was calculated.

Third, based on the DAR and DFR, the weekly absenteeism rate (WAR) and weekly fever rate (WFR) for the three schools were calculated as follows:

The time series diagrams of the WAR and WFR of the 3 schools were produced, and their coincidence with the time series diagram trends of the WPRIV was compared (Figures 6-8).

Finally, the following statistical process was performed: (1) The sums of the WAR and WFR of each school were calculated. (2) Based on the data of school A, the data of school B were added at the beginning of week 37; then the data of school C were added again in week 45, and then the combined WAR and WFR were combined based on the weight of the total enrollment in each school. (3) The sums of the combined WAR and WFR were calculated. Simultaneously, the current time was set to \( t \); then the WPRIV sequence was advanced by 1 week (\( t-1 \)), 2 weeks (\( t-2 \)), and 3 weeks (\( t-3 \)). The correlation of the WPRIV with the aforementioned WARS and WFRs and their sum was calculated under these 4 conditions to investigate the reliability, accuracy, and timeliness of different types of data derived from this system for influenza activity surveillance.

Figure 3. Time series of daily absenteeism rates reported by the systems of schools A, B, and C.
Figure 4. Time series of daily absenteeism rates reported by the system (DAR1) and school physicians (DAR2). DAR: daily absenteeism rate.

Figure 5. Time series of daily fever rates in 3 schools.
Figure 6. Weekly positive rates of influenza virus, weekly absenteeism rates, and weekly fever rates of school A.

Figure 7. Weekly positive rates of influenza virus, weekly absenteeism rates, and weekly fever rates of school B.
Figure 8. Weekly positive rates of influenza virus, weekly absenteeism rates, and weekly fever rates of school C.

Results

Analysis of DAR1

In phase I, all 2961 students in schools A and B had registered accounts in the system. In phase II, the total number of students in the 3 schools was 3535, with 3530 students (99.86%) having registered accounts (Table 1). The DAR of school A remained relatively stable throughout the surveillance period, without surpassing the 10% critical threshold (Figure 3). The DAR of school B fluctuated sharply in phase I, but the overall trend was decreasing. In phase II, after about 2 weeks, the DAR curves of schools B and A were similar. The DAR of school C varied considerably during the first 2 weeks of system commissioning and began to stabilize about 2 weeks later. The DAR of school C exceeded 10% on 4 days, December 24, December 31, January 7, and January 10, 2022. These 4 days were weekends or holidays.

The DAR1s of the 3 schools were very similar along with the trends over the 30 monitoring days from November 8 to December 20 (Figure 3). During this time, 3 pairs of absenteeism rates, namely schools A and B ($r_a=0.508$, $P=.004$), B and C ($r_b=0.427$, $P=.017$), and A and C ($r_ac=0.447$, $P=.012$), were moderately positively correlated. After December 20, the DARs at the 3 schools showed different trends, with the differences between school C and the other 2 schools being significantly different.

Table 1. Daily absenteeism rates and daily fever rates of 3 schools reported by the system in phases I and II.

<table>
<thead>
<tr>
<th>School</th>
<th>Total number</th>
<th>Number of enrollments</th>
<th>Rate of enrollment (%)</th>
<th>Daily absenteeism rate (%)</th>
<th>Daily fever rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>Phase I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1861</td>
<td>1861</td>
<td>100</td>
<td>0.86</td>
<td>4.19</td>
</tr>
<tr>
<td>B</td>
<td>1100</td>
<td>1100</td>
<td>100</td>
<td>0.82</td>
<td>31.09</td>
</tr>
<tr>
<td>Phase II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1954</td>
<td>1954</td>
<td>100</td>
<td>1.07</td>
<td>5.63</td>
</tr>
<tr>
<td>B</td>
<td>1154</td>
<td>1153</td>
<td>99.91</td>
<td>1.21</td>
<td>15.70</td>
</tr>
<tr>
<td>C</td>
<td>427</td>
<td>423</td>
<td>99.06</td>
<td>1.42</td>
<td>16.08</td>
</tr>
</tbody>
</table>

^aN/A: not applicable.

Correlation Analysis of DAR1 and DAR2

Between September 1, 2021, and January 14, 2022, students from school A were expected to attend school for 92 days and students from school B for 90 days. Both the system and the school physician of school A recorded 92 days of data. The system recorded 3452 absences over 90 days, whereas the school physician recorded 1691 absences over 66 days (73.3% completeness), with the number reported by school physicians being 48.99% of the number reported by the system.
A significant positive correlation between DAR1 and DAR2 in school A was detected ($r=0.809$, $P<.001$), and their change trends were highly consistent. In school B, DAR1 and DAR2 were also significantly positively correlated ($r=0.766$, $P<.001$), but there was a notable gap between them in November 2021 (Figure 4). The abovementioned data show that the absence reported by school physicians (completeness 86.7%) was only 36.5% of the absence reported by the system (completeness 100%), and that a significant positive correlation between the 2 was present ($r=0.372$, $P=.002$).

**Analysis of DFR.**

The DFR calculation time for school A started from March 1, 2021 (week 9), for school B from September 13, 2021 (week 37), and for school C from November 8, 2021 (week 45). The infrared image investigation of the system confirmed that high-temperature objects were placed next to the infrared thermometer in school A on 2 days (April 30, September 18), and on 1 day (April 1), the thermometer was unable to measure temperatures owing to physical failure. In school B, there were 2 days (November 11 and December 29) when a high-temperature object was placed next to the thermometer. The DFR of these 5 days was regarded as a missing value and replaced with the mean of the DFR for the previous and next days.

Overall, the DFR curve of school A resembles a parabola with a downward opening (Figure 5). This trend was similar to the daily average temperature change trend in this area during the surveillance period. During the 30 days from September 13 to October 29, the DFR of school A was always higher than that of school B, and after October 29, this relationship was gradually reversed. The DFR curves of schools C and A were closer, and the DFRs of these 2 schools were also significantly positively correlated ($r_{ab}=0.493$, $P<.001$). Although no significant correlation was detected between schools A and B ($r_{ab}=-0.023$, $P=.84$), and schools B and C ($r_{bc}=0.091$, $P=.54$), the DFR curves of the 3 schools were consistent in some wave peaks and troughs.

**Correlation Analysis of WAR1, WFR, and WPRIV.**

The WAR1s and WFRs for the 3 schools were calculated, and their time series diagrams were plotted. The WAR1 of school A was consistent with the trend of the WPRIV (Figure 6), that is, both the WAR1 and WPRIV gradually increased over time. After week 42, influenza activity levels spiked, with the first peak of WAR1 of school A presenting in week 46 (1 week earlier than WPRIV), and the second peak in week 50 (the same time as the WPRIV). The first peak of the WFR1 in school A was at week 45 (2 weeks earlier than the WPRIV).

School B showed a peak of the WAR1 in weeks 42 and 43 (Figure 7). Investigation on this aspect showed that the school had problems with the health codes of some students; therefore, school B prudently allowed 4 classes of grade 1 to study at home. The second WAR1 peak of school B was in week 45, and the third one was in week 50. These 2 peak times were the same as those of school A, but school B had another WAR1 peak in week 53, which was different from school A. As the level of influenza activity increased, the WFR1 of school B peaked at week 45 (same as in school A), week 47, and week 52.

The surveillance weeks of school C were few, and the WAR1 of school C has been on the rise throughout the surveillance period (Figure 8), which was highly consistent with the trend of the WPRIV. The 3 peaks of the WFR1 in school C were in week 45 (same as school A and B), week 47, and week 52 (same as school B). On combining with Figures 5-7, the mismatch trend of the WAR1 and WFR became evident. This phenomenon was especially apparent in school B. For example, the WARs peaked in weeks 46, 50, and 53, whereas the corresponding WFRs were at a low point. For weeks 45, 47, and 52, the WFRs peaked, whereas the WAR1s were at their troughs.

Among the WAR1s reported by the system, only the WAR1 of schools A and C was significantly positively correlated with the WPRIV. At $t–3$, the correlation coefficient between the WAR1 of school A and WPRIV was the highest, whereas at $t$, the correlation coefficient between the WAR of school C and WPRIV was the highest (Table 2). Although the WAR of school B was not significantly correlated with the WPRIV in the 4 conditions, the variation trend of the correlation coefficients was similar to that in school A. Only the WFR of school B was significantly positively correlated with the WPRIV, and the maximum coefficient was at $t–3$. The addition of the WAR1 to the WFR produced an increased correlation with the WPRIV only in the sum of school B. The WAR1 of school A reported by the school physician was significantly positively correlated with the WPRIV, with the maximum coefficient located at $t–3$. The correlation coefficients between the WAR1 and WPRIV that were calculated based on system-reported data were higher than those calculated based on data reported by school physicians in each condition. The WAR1 of school B reported by both the school physician and the proposed system was not significantly correlated with the WPRIV in any of the 4 conditions.
Discussion

Principal Findings

In 2017, Groseclose and Buckeridge [37] proposed a surveillance system evaluation framework comprising 12 indicators, namely simplicity, acceptability, representativeness, stability, data quality, timeliness, flexibility, security, sensitivity, predictive value positive, cost-effectiveness, and standard use. Comparison of the proposed system with existing similar systems showed its advantages in in terms of simplicity, cost-effectiveness, data quality, sensitivity, and timeliness.

This system is simple to operate and has a low construction cost. The simplicity of the proposed surveillance system is reflected in four aspects: data availability and type, organization and type, data exchange and conversion, and personnel operation and training [37]. Sickness absence surveillance based on manual data collection requires the data reporter to have extensive medical expertise. Therefore, the reporter is required to be a school physician or a part-time worker with adequate medical training. On the other hand, the operation of the proposed system does not require professional personnel, but it only requires students to stand in front of the instrument in accordance with the standard for 1 second so that all symptom data can be automatically obtained. In this way, the system is highly cost-effective in terms of workforce. Moreover, contrary to the government-led SSS, which involves high investment for platform research and development, the proposed system is built based on the mature network platform and is thus cost-effective.

The absenteeism reported by the system was complete and accurate, and the completeness of the temperature was high, but the accuracy needs to be improved. Fingerprints, smart cards, or face recognition were used to guarantee that every student in attendance is checked and accurately identified. Lawpoolsri et al [14] attempted to use fingerprints instead of manual collection but failed after many students missed and checked wrongly. In this study, the proportion of students registering system accounts was close to 100%, the accuracy of face recognition technology was more than 99.99%, and the school also arranged personnel to supervise the students during check-in. Data showed that the absenteeism reported by the system was highly positively correlated with that reported by school physicians, and the completeness of the former was higher than that of the latter. The system also needs to confidently guarantee that the temperature of each identified student is accurately measured. As face recognition and infrared temperature measurement were almost synchronized, the completeness of temperature was adequate, but its accuracy needs to be improved. The accuracy of thermal imaging is affected by the instrument, environment, and individual factors [29-31]. In this study, only the temperature of school B met the monitoring requirements. The instruments in school B were installed indoors and strictly supervised by students. These measures are helpful to reduce improper measurement behavior and improve the accuracy. Future research will focus on the construction and installation of instruments and implementation of the operation standards.

The proposed system showed good sensitivity. The surveillance sensitivity includes case detection, outbreak detection, and case definition [37]. Existing systems focus on students who are absent from school [6-12,14-18,20-22], whereas others also consider students who visit the school health room [9,10]. During an influenza outbreak, about 20% of the population shows symptoms, but only about 2% require physician consultation [38]. Students do not miss school or consult school physicians unless they are seriously ill. Because of academic pressure, it is common for Chinese students to attend school.

Table 2. Correlation analysis of weekly absenteeism rate, weekly fever rate, and weekly positive rate of influenza virus.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Weekly positive rate of influenza virus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
</tr>
<tr>
<td><strong>Data reported by the system</strong></td>
<td></td>
</tr>
<tr>
<td>Weekly absence rate of school A</td>
<td>0.662</td>
</tr>
<tr>
<td>Weekly absence rate of school B</td>
<td>0.334</td>
</tr>
<tr>
<td>Weekly absence rate of school C</td>
<td>0.771</td>
</tr>
<tr>
<td>Weekly fever rate of school A</td>
<td>-0.368</td>
</tr>
<tr>
<td>Weekly fever rate of school B</td>
<td>0.492</td>
</tr>
<tr>
<td>Weekly fever rate of school C</td>
<td>0.229</td>
</tr>
<tr>
<td>Weekly absence and fever rate of school A</td>
<td>0.058</td>
</tr>
<tr>
<td>Weekly absence and fever rate of school B</td>
<td>0.6</td>
</tr>
<tr>
<td>Weekly absence and fever rate of school C</td>
<td>0.686</td>
</tr>
<tr>
<td><strong>Data reported by school physicians</strong></td>
<td></td>
</tr>
<tr>
<td>Weekly absence rate of school A</td>
<td>0.58</td>
</tr>
<tr>
<td>Weekly absence rate of school B</td>
<td>0.429</td>
</tr>
</tbody>
</table>
sick, and about 74.7% of those who took sick leave were absent for less than 1 day [26]. This is a potential explanation for the number of absences reported by school physicians being only about 36.5% of the number reported by the system in this study. The correlation between the WAR of school A reported by the system and the WPRIV was higher than that between the WAR reported by school physicians and the WPRIV, thus appearing to support this conclusion. Data showed that in the influenza outbreak season, the DAR in some schools was up to 8%. By adjusting the absenteeism criteria and combining temperature screening, the presented system is more sensitive to case and outbreak detection than existing systems.

The results also showed good timeliness rates. The data of this system were transmitted in real time and analyzed automatically. Parents, teachers, and administrators were able to instantly access the attendance and temperature status of an individual, class, and school. Baer et al [11] proposed that the surveillance value of absenteeism lies more in situational awareness than in early detection, which may need to be revised when applied to China. This study revealed that absenteeism in the 3 schools was significantly positively correlated when influenza activity levels were low, which was not consistent with the findings of Schmidt et al [8] stating that absenteeism is very similar across age groups. With the increase in influenza prevalence, absenteeism in primary schools was the first to peak, and absenteeism in middle schools reached the first peak 2 weeks later, which is consistent with the findings of Mook et al [39]. Subsequently, the gap in the absenteeism timing among schools increased, even between 2 primary schools; at this point, the claim of Schmidt et al [8] does apply. The addition of temperature data increased the timeliness of the system. The results showed that temperature could predict influenza up to 3 weeks earlier when the temperatures of students were accurately measured. Combined with the study of Miller et al [32], the timeliness advantage of temperature in influenza surveillance may indeed exist. Therefore, it is extremely critical to standardize instrument placement and operation procedures in the future to further improve the reliability and accuracy of temperature data.

Limitations
First, although thermal infrared imaging is simple, its accuracy is easily influenced by environmental factors [29]. Because of the need for COVID-19 prevention, some schools introduced the proposed system. Nonetheless, these schools attach different importance levels to epidemic prevention and control, which in turn leads to different levels of hardware support and operation supervision. These disparities led to differences in data quality, especially temperature data, which hindered further analysis of these data. Second, COVID-19 was a prevailing issue during the surveillance period. Although it facilitated the roll-out of the system to a certain extent, the number of COVID-19 cases in the surveillance area was too small to be the target disease. However, because influenza was selected as the target infection, the status data for the influenza epidemic in general are distorted by the pressure of COVID-19 prevention. Consequently, some of the presented results need to be re-evaluated without considering the impact of COVID-19. Finally, although the system has been authorized by the government to collect sensitive personal information, possible changes to the national personal information and privacy protection policy in the future will substantially affect the operational stability of the system.

Conclusions
Comparison of the proposed system with existing similar systems showed its advantages in terms of simplicity, cost-effectiveness, data quality, sensitivity, and timeliness. This study showed that the absenteeism recorded using face recognition technology was reliable, but the accuracy of the temperature recorded by infrared thermometers should be enhanced. The implementation of the influenza SSS based on absenteeism and temperature data was feasible. When influenza activity levels were moderate, a significant positive correlation between the DARs was detected; however, as the levels increased, the gap among the DARs gradually increased and peaked about 2 weeks earlier in primary schools than in junior high schools. The introduction of temperature measurement substantially strengthened the surveillance timeliness, allowing detection of influenza outbreaks up to 3 weeks earlier than traditional systems. This study demonstrated a feasible way to solve the challenge of developing a surveillance system and promote the automation of symptom data acquisition in the surveillance system.

Acknowledgments
This study was supported by Ji’an Science and Technology Bureau, Jiangxi Province, China (grant Ji’an Science and Technology Bureau Project 2020-05). We thank Hangzhou Huibei Technology Co, Ltd for cooperation in developing and implementing the system. We would also like to thank all the parents, teachers, and principals of the 3 schools that participated in the study. We also thank the China National Influenza Center for providing influenza surveillance data.

Conflicts of Interest
None declared.

References


34. Xiaolianxing. URL: https://k12.xiaolianxing.com/#/login?redirect=%2F [accessed 2019-12-10]


Abbreviations

DAR: Daily absence rate
DFR: Daily fever rate
SSS: Syndromic surveillance system
WAR: Weekly absence rate
WFR: Weekly fever rate
WPRIV: Weekly positive rate of influenza virus
Spatiotemporal Distribution of HIV Self-testing Kits Purchased on the Web and Implications for HIV Prevention in China: Population-Based Study

Ganfeng Luo¹*, PhD; Lingyun Su²*, MS; Anping Feng¹*, MSc; Yi-Fan Lin¹*, PhD; Yiguo Zhou¹, MSc; Tanwei Yuan¹, MSc; Yuqing Hu³, MPH; Song Fan³, PhD; Yong Lu⁴, PhD; Yingsi Lai⁵, PhD; Qian Shi⁶, PhD; Jun Li⁷, PhD; Mengjie Han⁸, MPH; Huachun Zou¹, PhD

¹School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, China
²US-China Health Summit, Beijing, China
³School of Public Health, Southwest Medical University, Luzhou, China
⁴School of Public Health, the Key Laboratory of Environmental Pollution Monitoring and Disease Control, Ministry of Education, Guizhou Medical University, Guiyang, China
⁵School of Public Health, Sun Yat-sen University, Guangzhou, China
⁶School of Geography and Planning, Sun Yat-sen University, Guangzhou, China
⁷School of Computer Science, China University of Geosciences, Wuhan, China
⁸National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

*these authors contributed equally

Corresponding Author:
Huachun Zou, PhD
School of Public Health (Shenzhen)
Sun Yat-sen University
No. 66, Gongchang Road, Guangming District
Shenzhen, 518107
China
Phone: 86 2087335651
Email: zouhuachun@mail.sysu.edu.cn

Abstract

Background: HIV self-testing (HIVST) holds great promise for expanding HIV testing. Nonetheless, large-scale data on HIVST behavior are scant. Millions of HIVST kits are sold through e-commerce platforms each year.

Objective: This study aims to analyze the spatiotemporal distribution of the HIVST kit–purchasing population (HIVSTKPP) in China.

Methods: Deidentified transaction data were retrieved from a leading e-commerce platform in China. A jointpoint regression model was used to examine annual trends of the HIVSTKPP rates by calculating average annual percentage change. Bayesian spatiotemporal analysis was performed to locate hot spots with HIVSTKPP rates. Spatial autocorrelation analysis and space-time cluster analysis were conducted to identify clusters of HIVSTKPP. High-high clusters of HIVSTKPP can be identified by spatial autocorrelation analysis, and high-high clusters indicate that a region and its surrounding region jointly had a higher-than-average HIVSTKPP rate. Spatial regression analysis was used to elucidate the association between the number of HIV testing facilities, urbanization ratio (the proportion of urban population in the total population), and gross domestic product per capita and the HIVSTKPP.

Results: Between January 1, 2016, and December 31, 2019, a total of 2.18 million anonymous persons in China placed 4.15 million orders and purchased 4.51 million HIVST kits on the web. In each of these 4 years, the observed monthly size of the HIVSTKPP peaked in December, the month of World AIDS Day. HIVSTKPP rates per 100,000 population significantly increased from 20.62 in 2016 to 64.82 in 2019 (average annual percentage change=48.2%; P<.001). Hot spots were mainly located in municipalities, provincial capitals, and large cities, whereas high-high clusters and high-demand clusters were predominantly detected in cities along the southeast coast. We found positive correlations between a region’s number of HIV testing facilities, urbanization ratio, and gross domestic product per capita and the HIVSTKPP.
Conclusions: Our study identified key areas with larger demand for HIVST kits for public health policy makers to reallocate resources and optimize the HIV care continuum. Further research combining spatiotemporal patterns of HIVST with HIV surveillance data is urgently needed to identify potential gaps in current HIV-monitoring practices.

(JMIR Public Health Surveill 2022;8(10):e35272) doi:10.2196/35272

KEYWORDS
HIV self-testing; spatiotemporal patterns; China; e-commerce platform

Introduction

Background

HIV is a growing public health challenge in China, with the annual rate of persons with newly diagnosed HIV rising from 3.66 per 100,000 in 2007 to 9.35 per 100,000 in 2020 [1,2]. At the end of 2020, there were 1.053 million people living with HIV and 351,000 cumulative reported HIV-related deaths in China [2]. Although HIV testing services are widely available in China [3,4], only an estimated 68.9% of the people living with HIV were aware of their serostatus [5], which is far below the Joint United Nations Program on HIV and AIDS target of 90% of the people living with HIV knowing their HIV status [6]. Many high-risk groups, including men who have sex with men (MSM), sex workers, and people who use drugs, are reluctant to seek venue-based HIV testing services because of concerns about confidentiality, stigma, discrimination, and inconvenience [7]. Determining how to improve the HIV detection rate and promote the identification of people living with HIV has always been the focus and difficulty of AIDS epidemic prevention and control efforts.

HIV self-testing (HIVST) holds great promise for expanding HIV testing [8]. China’s 13th Five-Year Plan for HIV Prevention and Control, a national policy framework for HIV response adopted in 2017, encouraged innovative strategies to expand HIV testing, including distribution of HIVST kits through web-based platforms [9]. E-commerce platforms such as Amazon and eBay as well as Taobao and Jingdong in China are central to how people shop and purchase goods, including HIVST kits [10]. These platforms reach billions of registered users across high-, middle-, and low-income countries [10]. HIVST kits are easily available through e-commerce platforms in China [11], with >1 million kits sold on the web in 2018 [12]. Web-based purchasing of HIVST kits is common among some populations at high risk of HIV acquisition. One-third of MSM in China have purchased HIVST kits on the web [7]. In addition, a recent study showed that 591 individuals who ever purchased HIVST kits on the web in China, 64.7% (220/340) of the heterosexual respondents, 69.9% (112/161) of the homosexual respondents, and 72% (65/90) of the bisexual respondents had engaged in unprotected sex in the last 6 months [12]. In other words, the web-based HIVST kit–purchasing behavior largely indicates that recent high-risk sexual behavior is a factor in HIV acquisition.

Spatiotemporal Analyses

In recent decades, spatiotemporal analyses have been applied to HIV surveillance and outbreak investigation among different populations [13-16]. Applying these same principles to analyze the spatiotemporal evolution of HIVST kit–purchasing patterns can provide insight into the distribution of those at potential high risk for HIV and in need of HIVST and HIV care [17]. This information can be used to identify gaps in HIV prevention and optimize allocation of resources [17]. Identifying associations between HIVST kit purchasing and macroscopic factors such as number of HIV testing facilities, urbanization ratio (the proportion of urban population in the total population), and gross domestic product (GDP) may help in the development of interventional strategies or policy responses contextualized to different settings.

No previous publication has analyzed the spatiotemporal distribution of HIVST, and the effect of expanding HIVST based on China’s 13th Five-Year Plan for HIV Prevention and Control remains unclear. In this study, we used transaction data collected from a leading e-commerce platform to analyze the spatiotemporal patterns of the HIVST kit–purchasing population (HIVSTKPP) to uncover clusters of HIVST kit purchasing and evaluate associations with macroscopic factors in China.

Methods

Data Collection

Records of the sales of HIVST kits between January 1, 2016, and December 31, 2019, were retrieved from a leading e-commerce platform in China [17]. To protect the consumers’ privacy, the name of the e-commerce platform is not being disclosed. The extracted variables included (1) anonymized ID details; (2) purchase date; (3) shipping province, city, and provincial-controlled county; and (4) purchase quantity. In addition, IP addresses were not included in the original data set. The size of the resident population, urbanization ratio, and GDP per capita (CN¥ 10,000 [US $1410]) for each shipping province, city, and provincial-controlled county were drawn from the Statistical Yearbook of China [18]. The number of HIV testing facilities in each area was collected from the Chinese Center for Disease Control and Prevention (CCDC) [19]. Maps were obtained from the National Catalogue Service for Geographic Information [20].

Data Management

All personal identifiable information was removed or deidentified. All data were maintained entirely within a sandbox environment of the e-commerce platform.

Inclusion and Exclusion Criteria

To minimize the impact of bulk or proxy purchasing of HIVST kits, we only included those persons who had purchased ≤48 kits cumulatively over the entire study period. We only included...
records with shipping addresses within mainland China. Orders shipped to Hong Kong, Macao, Taiwan, and overseas were excluded.

**Definition of HIVSTKPP**
In our study, HIVSTKPP refers to anonymous persons who purchased HIVST kits from a leading e-commerce platform for their own use for HIV testing in China between January 1, 2016, and December 31, 2019.

**Statistical Analysis**
The rate of HIVST kit purchasing per 100,000 population was calculated by dividing the number of anonymous persons who bought the kits by the total population in an area. When calculating the HIVSTKPP rate, we only included each anonymous person’s latest recorded purchase to avoid duplication. In this study, anonymous person was defined as a unique individual by anonymized user ID of the e-commerce platform. Data were analyzed at the regional, provincial, city, and provincial-controlled–county levels.

In China, 31 provinces are aggregated into 7 regions (North China, Northeast China, East China, Central China, South China, Southwest China, and Northwest China) based on geography, climate, economy, history, and ethnicity by the Chinese government (Textbox 1).

The geographical distribution of these 7 regions and 31 provinces is presented in Figure 1. In addition, 4 municipalities (Beijing, Shanghai, Tianjin, and Chongqing cities) belong to provincial administrative units. The 31 provinces and their capitals are presented in Textbox 2.

**Textbox 1.** The 7 regions and 31 provinces of mainland China.

<table>
<thead>
<tr>
<th>Regions and provinces</th>
</tr>
</thead>
<tbody>
<tr>
<td>• North China</td>
</tr>
<tr>
<td>• Beijing, Tianjin, Hebei, Shanxi, and Inner Mongolia</td>
</tr>
<tr>
<td>• Northeast China</td>
</tr>
<tr>
<td>• Heilongjiang, Jilin, and Liaoning</td>
</tr>
<tr>
<td>• East China</td>
</tr>
<tr>
<td>• Shanghai, Jiangsu, Zhejiang, Anhui, Jiangxi, Shandong, and Fujian</td>
</tr>
<tr>
<td>• Central China</td>
</tr>
<tr>
<td>• Henan, Hubei, and Hunan</td>
</tr>
<tr>
<td>• South China</td>
</tr>
<tr>
<td>• Guangdong, Guangxi, and Hainan</td>
</tr>
<tr>
<td>• Southwest China</td>
</tr>
<tr>
<td>• Chongqing, Sichuan, Guizhou, Yunnan, and Tibet</td>
</tr>
<tr>
<td>• Northwest China</td>
</tr>
<tr>
<td>• Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang</td>
</tr>
</tbody>
</table>

Time series analysis [21] and joint regression analysis [22] were performed to examine seasonal patterns of the monthly size of the HIVSTKPP and annual trends of the HIVSTKPP rates, respectively. A spatial autocorrelation model was constructed to evaluate the spatial distribution patterns (clustered, dispersed, and random) of the HIVSTKPP [23,24]. Four types of local spatial clusters can be identified by a spatial autocorrelation model: high-high, low-low, high-low, and low-high [13]. High-high and low-low clusters indicate that a city and its surrounding region jointly had higher-than-average and lower-than-average HIVSTKPP rates, respectively. A high-low cluster represents a city with an above-average HIVSTKPP rate surrounded by cities with below-average rates, whereas a low-high cluster represents a city with a below-average HIVSTKPP rate surrounded by cities with above-average rates. Temporal-spatial clustering analysis was implemented to identify high-risk (higher demand for HIVST kits) and low-risk (lower demand for HIVST kits) clusters of HIVSTKPP over space and time simultaneously [25,26], and we set a maximum spatial cluster size of 50% of the population at risk and a maximum temporal cluster size of 50% of the study period to scan for spatial clusters with high and low rates in our study. A Bayesian spatiotemporal model was constructed to detect hot spots and cold spots of the HIVSTKPP [27]. Spatial lag model or spatial error model (SEM) and geographically weighted regression model were constructed to assess the global and local spatial correlation between the size of the HIVSTKPP and 3 factors (number of HIV testing facilities, urbanization ratio, and GDP per capita [CN¥ 10,000 {US $1410}]), respectively [13,28]. Details of the statistical analysis can be found in Multimedia Appendix 1 [1-3,6-9,17,18,29-39].
Figure 1. Geographical distribution of 7 regions and 31 provinces in mainland China. The 31 provinces are as follows: 1, Beijing; 2, Tianjin; 3, Hebei; 4, Shanxi; 5, Inner Mongolia; 6, Heilongjiang; 7, Jilin; 8, Liaoning; 9, Shanghai; 10, Jiangsu; 11, Zhejiang; 12, Anhui; 13, Jiangxi; 14, Shandong; 15, Fujian; 16, Henan; 17, Hubei; 18, Hunan; 19, Guangdong; 20, Guangxi; 21, Hainan; 22, Chongqing; 23, Sichuan; 24, Guizhou; 25, Yunnan; 26, Tibet; 27, Shaanxi; 28, Gansu; 29, Qinghai; 30, Ningxia; and 31, Xinjiang.
**Textbox 2. Mainland China’s 31 provincial capitals.**

<table>
<thead>
<tr>
<th>Provinces and capitals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing: Beijing municipality</td>
</tr>
<tr>
<td>Tianjin: Tianjin municipality</td>
</tr>
<tr>
<td>Hebei: Shijiazhuang city</td>
</tr>
<tr>
<td>Shanxi: Taiyuan city</td>
</tr>
<tr>
<td>Inner Mongolia: Hohhot city</td>
</tr>
<tr>
<td>Heilongjiang: Harbin city</td>
</tr>
<tr>
<td>Jilin: Changchun city</td>
</tr>
<tr>
<td>Liaoning: Shenyang city</td>
</tr>
<tr>
<td>Shanghai: Shanghai municipality</td>
</tr>
<tr>
<td>Jiangsu: Nanjing city</td>
</tr>
<tr>
<td>Zhejiang: Hangzhou city</td>
</tr>
<tr>
<td>Anhui: Hefei city</td>
</tr>
<tr>
<td>Jiangxi: Nanchang city</td>
</tr>
<tr>
<td>Shandong: Nanchang city</td>
</tr>
<tr>
<td>Fujian: Fuzhou city</td>
</tr>
<tr>
<td>Henan: Zhengzhou city</td>
</tr>
<tr>
<td>Hubei: Wuhan city</td>
</tr>
<tr>
<td>Hunan: Wuhan city</td>
</tr>
<tr>
<td>Guangdong: Guangzhou city</td>
</tr>
<tr>
<td>Guangxi: Nanning city</td>
</tr>
<tr>
<td>Hainan: Haikou city</td>
</tr>
<tr>
<td>Chongqing: Chongqing municipality</td>
</tr>
<tr>
<td>Sichuan: Chengdu city</td>
</tr>
<tr>
<td>Guizhou: Guiyang city</td>
</tr>
<tr>
<td>Yunnan: Kunming city</td>
</tr>
<tr>
<td>Tibet: Lhasa city</td>
</tr>
<tr>
<td>Shaanxi: Xi’an city</td>
</tr>
<tr>
<td>Gansu: Lanzhou city</td>
</tr>
<tr>
<td>Qinghai: Xining city</td>
</tr>
<tr>
<td>Ningxia: Yinchuan city</td>
</tr>
<tr>
<td>Xinjiang: Urumqi city</td>
</tr>
</tbody>
</table>

**Ethics Approval**

This study was conducted with the approval of the institutional review board and ethics committee of Sun Yat-sen University (SYSU-SPH2021026). We did not identify or reidentify any individual, purposefully or inadvertently, in our analysis.

**Results**

**HIVST Kit–Purchasing Patterns**

Between January 1, 2016, and December 31, 2019, a total of 2.18 million anonymous persons in China placed 4.15 million orders and purchased 4.51 million HIVST kits on the web. HIVST kits were delivered to 366 cities and provincial-controlled counties (Table 1). The mean number of HIVST kits sold per month was 94.01 thousand, and the mean number of HIVST kits sold per day was 3.09 thousand. Of the 2.18 million anonymous persons who purchased HIVST kits, 1.39 million (63.79%), 0.41 million (18.73%), and 0.16 million (7.39%) anonymous persons purchased HIVST kits on 1, 2, and 3 occasions between January 1, 2016, and December 31, 2019, respectively; in addition, 1.33 million (60.88%), 0.43 million (19.59%), and 0.17 million (7.66%) anonymous persons purchased 1, 2, and 3 kits, respectively (Table 2). The average annual HIVSTKPP rate between January 1, 2016, and December 31, 2019, was 39.16 (SD 19.90) per 100,000 (Table 1).
Table 1. The number of purchasers, purchases, and HIV self-testing kits sold as well as the HIV self-testing kit–purchasing population rate between January 1, 2016, and December 31, 2019, in mainland China.

<table>
<thead>
<tr>
<th>Year</th>
<th>Purchasers (N=2,180,284), n (%)</th>
<th>Purchases (N=4,148,429), n (%)</th>
<th>Kits (N=4,512,353), n (%)</th>
<th>Population (thousand; N=5,568,220), n (%)</th>
<th>Rate^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>285,107 (13.08)</td>
<td>641,566 (15.47)</td>
<td>750,752 (16.64)</td>
<td>1,382,710 (24.83)</td>
<td>20.62</td>
</tr>
<tr>
<td>2017</td>
<td>369,140 (16.93)</td>
<td>832,775 (20.07)</td>
<td>924,661 (20.49)</td>
<td>1,390,080 (24.96)</td>
<td>26.55</td>
</tr>
<tr>
<td>2018</td>
<td>618,498 (28.37)</td>
<td>1,203,435 (29.01)</td>
<td>1,282,882 (28.43)</td>
<td>1,395,380 (25.06)</td>
<td>44.32</td>
</tr>
<tr>
<td>2019</td>
<td>907,539 (41.62)</td>
<td>1,470,653 (35.45)</td>
<td>1,554,058 (34.44)</td>
<td>1,400,050 (25.14)</td>
<td>64.82</td>
</tr>
</tbody>
</table>

^a^The rate of HIV self-testing kit purchasing per 100,000 population was calculated by dividing the number of purchasers by the total population in an area. The rate for the 4-year period from January 1, 2016, to December 31, 2019, was 39.16.

Table 2. The number of purchasers with ≥1 purchases and the number of purchasers of ≥1 HIV self-testing kits between January 1, 2016, and December 31, 2019, in mainland China.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Values (N=2,180,284), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purchases</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1,390,790 (63.79)</td>
</tr>
<tr>
<td>2</td>
<td>408,471 (18.73)</td>
</tr>
<tr>
<td>3</td>
<td>161,192 (7.39)</td>
</tr>
<tr>
<td>4</td>
<td>78,898 (3.62)</td>
</tr>
<tr>
<td>5</td>
<td>43,773 (2.01)</td>
</tr>
<tr>
<td>6</td>
<td>27,069 (1.24)</td>
</tr>
<tr>
<td>7</td>
<td>17,708 (0.81)</td>
</tr>
<tr>
<td>8</td>
<td>12,174 (0.56)</td>
</tr>
<tr>
<td>9</td>
<td>8,804 (0.4)</td>
</tr>
<tr>
<td>10</td>
<td>6,469 (0.3)</td>
</tr>
<tr>
<td><strong>HIV self-testing kits</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1,327,298 (60.88)</td>
</tr>
<tr>
<td>2</td>
<td>427,037 (19.59)</td>
</tr>
<tr>
<td>3</td>
<td>167,042 (7.66)</td>
</tr>
<tr>
<td>4</td>
<td>88,952 (4.08)</td>
</tr>
<tr>
<td>5</td>
<td>49,154 (2.25)</td>
</tr>
<tr>
<td>6</td>
<td>31,585 (1.45)</td>
</tr>
<tr>
<td>7</td>
<td>20,195 (0.93)</td>
</tr>
<tr>
<td>8</td>
<td>14,508 (0.67)</td>
</tr>
<tr>
<td>9</td>
<td>10,689 (0.49)</td>
</tr>
<tr>
<td>10</td>
<td>8,364 (0.38)</td>
</tr>
</tbody>
</table>

**Temporal Trends in the HIVSTKPP**

The observed annual HIVSTKPP markedly increased from 0.29 million anonymous persons in 2016 to 0.91 million anonymous persons in 2019 (Figure 2). The observed HIVSTKPP showed a clear periodic pattern, with 46.31% (1,009,704/2,180,284) of the anonymous persons purchasing HIVST kits between September and December (Figures 3-5). The maximum monthly number was observed in December for all provinces (Figure 5 and Figure 6).
Figure 2. Temporal trend of the size of the HIV self-testing kit–purchasing population (HIVSTKPP) by year and month between January 1, 2016, and December 31, 2019, in mainland China.

Figure 3. The seasonal and trend decomposition using Loess of the size of the HIV self-testing kit–purchasing population by month from January 1, 2016, to December 31, 2019, in mainland China.
**Figure 4.** The monthly plots of the size of the HIV self-testing kit-purchasing population (HIVSTKPP) from January 1, 2016, to December 31, 2019, in mainland China.

**Figure 5.** The seasonal plots of the size of the HIV self-testing kit-purchasing population (HIVSTKPP) from January 1, 2016, to December 31, 2019, in mainland China.
Temporal Trends and Spatial Distribution in the HIVSTKPP Rates per 100,000 Population

From January 1, 2016, to December 31, 2019, the HIVSTKPP rates significantly increased among all 7 regions and 31 provinces (all \( P < .001 \); Figures 8-11, and Table 3). Southwest China, with a median annual HIVSTKPP rate of 35.25 per 100,000, showed the largest increase in HIVSTKPP rates (AAPC=53.3%; \( P < .001 \); Figure 8A and Figure 9). Most provinces had low annual HIVSTKPP rates but exhibited large increases in HIVSTKPP rates (Figure 8B and Figure 11). Beijing and Shanghai municipalities had both high average annual HIVSTKPP rates and large increases in HIVSTKPP rates.

The spatial distributions in HIVSTKPP rates at city and provincial-controlled-county levels were distinctly heterogeneous (Figure 12A and Figure 13). The top 10 cities with the highest average annual HIVSTKPP rates are shown in Textbox 3, and those with the largest sizes of HIVSTKPP are shown in Textbox 4.
Table 3. Average annual percentage change (AAPC) in HIV self-testing kit–purchasing population rates per 100,000 population between January 1, 2016, and December 31, 2019, among the 7 regions and 31 provinces in mainland China.

<table>
<thead>
<tr>
<th>Region</th>
<th>2016 to 2019</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
<th>AAPC (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>39.16</td>
<td>20.62</td>
<td>26.56</td>
<td>44.32</td>
<td>64.82</td>
<td>48.2</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>North China</td>
<td>38.73</td>
<td>19.75</td>
<td>25.92</td>
<td>45.58</td>
<td>63.43</td>
<td>50.1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Beijing</td>
<td>106.20</td>
<td>58.92</td>
<td>73.68</td>
<td>119.28</td>
<td>173.59</td>
<td>45.1</td>
<td></td>
</tr>
<tr>
<td>Tianjin</td>
<td>50.27</td>
<td>26.08</td>
<td>33.13</td>
<td>59.15</td>
<td>82.68</td>
<td>49.8</td>
<td></td>
</tr>
<tr>
<td>Hebei</td>
<td>27.23</td>
<td>12.67</td>
<td>17.23</td>
<td>33.90</td>
<td>44.83</td>
<td>56.3</td>
<td></td>
</tr>
<tr>
<td>Shanxi</td>
<td>27.42</td>
<td>13.06</td>
<td>18.21</td>
<td>32.62</td>
<td>45.56</td>
<td>54.2</td>
<td></td>
</tr>
<tr>
<td>Inner Mongolia</td>
<td>24.75</td>
<td>12.84</td>
<td>17.64</td>
<td>28.39</td>
<td>40.02</td>
<td>47.5</td>
<td></td>
</tr>
<tr>
<td>Northeast China</td>
<td>30.23</td>
<td>15.95</td>
<td>19.98</td>
<td>34.36</td>
<td>50.84</td>
<td>49.5</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Heilongjiang</td>
<td>24.69</td>
<td>13.57</td>
<td>16.15</td>
<td>28.27</td>
<td>40.98</td>
<td>47.3</td>
<td></td>
</tr>
<tr>
<td>Jilin</td>
<td>29.14</td>
<td>14.79</td>
<td>19.86</td>
<td>33.87</td>
<td>48.32</td>
<td>50.5</td>
<td></td>
</tr>
<tr>
<td>Liaoning</td>
<td>35.70</td>
<td>18.75</td>
<td>23.38</td>
<td>39.92</td>
<td>60.90</td>
<td>50.2</td>
<td></td>
</tr>
<tr>
<td>East China</td>
<td>47.44</td>
<td>25.78</td>
<td>32.63</td>
<td>53.15</td>
<td>77.62</td>
<td>46.1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Shanghai</td>
<td>103.32</td>
<td>56.09</td>
<td>70.14</td>
<td>116.35</td>
<td>170.43</td>
<td>46.8</td>
<td></td>
</tr>
<tr>
<td>Jiangsu</td>
<td>56.48</td>
<td>30.07</td>
<td>39.37</td>
<td>63.74</td>
<td>92.43</td>
<td>47.0</td>
<td></td>
</tr>
<tr>
<td>Zhejiang</td>
<td>67.79</td>
<td>38.87</td>
<td>48.80</td>
<td>73.54</td>
<td>108.15</td>
<td>41.6</td>
<td></td>
</tr>
<tr>
<td>Anhui</td>
<td>27.69</td>
<td>14.55</td>
<td>18.84</td>
<td>30.66</td>
<td>46.22</td>
<td>48.5</td>
<td></td>
</tr>
<tr>
<td>Jiangxi</td>
<td>26.02</td>
<td>14.03</td>
<td>17.79</td>
<td>29.22</td>
<td>42.80</td>
<td>46.8</td>
<td></td>
</tr>
<tr>
<td>Shandong</td>
<td>35.80</td>
<td>18.08</td>
<td>23.59</td>
<td>42.03</td>
<td>59.24</td>
<td>51.3</td>
<td></td>
</tr>
<tr>
<td>Fujian</td>
<td>51.42</td>
<td>30.76</td>
<td>34.98</td>
<td>55.58</td>
<td>83.63</td>
<td>41.4</td>
<td></td>
</tr>
<tr>
<td>Central China</td>
<td>29.71</td>
<td>14.97</td>
<td>20.37</td>
<td>33.5</td>
<td>49.77</td>
<td>50.7</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Henan</td>
<td>27.07</td>
<td>13.23</td>
<td>17.87</td>
<td>31.55</td>
<td>45.40</td>
<td>53.2</td>
<td></td>
</tr>
<tr>
<td>Hubei</td>
<td>35.28</td>
<td>19.34</td>
<td>24.28</td>
<td>39.18</td>
<td>58.17</td>
<td>46.0</td>
<td></td>
</tr>
<tr>
<td>Hunan</td>
<td>28.60</td>
<td>13.63</td>
<td>20.49</td>
<td>31.35</td>
<td>48.64</td>
<td>52.8</td>
<td></td>
</tr>
<tr>
<td>South China</td>
<td>50.85</td>
<td>28.45</td>
<td>34.04</td>
<td>56.75</td>
<td>82.97</td>
<td>45.1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Guangdong</td>
<td>60.10</td>
<td>34.28</td>
<td>39.62</td>
<td>66.73</td>
<td>98.07</td>
<td>44.4</td>
<td></td>
</tr>
<tr>
<td>Guangxi</td>
<td>31.81</td>
<td>16.97</td>
<td>22.75</td>
<td>35.49</td>
<td>51.55</td>
<td>45.9</td>
<td></td>
</tr>
<tr>
<td>Hainan</td>
<td>39.37</td>
<td>19.04</td>
<td>26.34</td>
<td>47.60</td>
<td>63.75</td>
<td>52.5</td>
<td></td>
</tr>
<tr>
<td>Southwest China</td>
<td>35.25</td>
<td>17.23</td>
<td>23.45</td>
<td>39.84</td>
<td>60.05</td>
<td>53.3</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Chongqing</td>
<td>44.75</td>
<td>22.05</td>
<td>29.40</td>
<td>49.10</td>
<td>77.67</td>
<td>53.6</td>
<td></td>
</tr>
<tr>
<td>Sichuan</td>
<td>39.32</td>
<td>19.04</td>
<td>25.18</td>
<td>44.69</td>
<td>68.01</td>
<td>55.2</td>
<td></td>
</tr>
<tr>
<td>Guizhou</td>
<td>25.04</td>
<td>11.66</td>
<td>17.25</td>
<td>28.77</td>
<td>42.17</td>
<td>54.8</td>
<td></td>
</tr>
<tr>
<td>Yunnan</td>
<td>30.35</td>
<td>15.51</td>
<td>21.51</td>
<td>34.57</td>
<td>49.46</td>
<td>48.5</td>
<td></td>
</tr>
<tr>
<td>Tibet</td>
<td>26.51</td>
<td>12.07</td>
<td>20.14</td>
<td>28.36</td>
<td>44.44</td>
<td>53.0</td>
<td></td>
</tr>
<tr>
<td>Northwest China</td>
<td>24.91</td>
<td>13.15</td>
<td>17.9</td>
<td>28.51</td>
<td>39.71</td>
<td>46.0</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Shaanxi</td>
<td>32.25</td>
<td>15.98</td>
<td>21.83</td>
<td>37.84</td>
<td>52.99</td>
<td>51.4</td>
<td></td>
</tr>
<tr>
<td>Ningxia</td>
<td>24.94</td>
<td>12.40</td>
<td>17.75</td>
<td>29.57</td>
<td>39.59</td>
<td>49.1</td>
<td></td>
</tr>
<tr>
<td>Qinghai</td>
<td>20.97</td>
<td>9.76</td>
<td>16.19</td>
<td>25.00</td>
<td>32.62</td>
<td>50.0</td>
<td></td>
</tr>
<tr>
<td>Xinjiang</td>
<td>20.13</td>
<td>13.42</td>
<td>16.35</td>
<td>20.19</td>
<td>30.11</td>
<td>30.2</td>
<td></td>
</tr>
<tr>
<td>Gansu</td>
<td>19.56</td>
<td>9.75</td>
<td>14.02</td>
<td>23.22</td>
<td>31.09</td>
<td>49.0</td>
<td></td>
</tr>
</tbody>
</table>

*aHIVSTKPP: HIV self-testing kit–purchasing population.*
The average annual percentage change from January 1, 2016, to December 31, 2019, was calculated by joinpoint regression model.

**Figure 7.** Average annual rate per 100,000 population of the HIV self-testing kit–purchasing population (HIVSTKPP) between January 1, 2016, and December 31, 2019, among the 31 provinces in mainland China.

**Figure 8.** Average annual percentage change (AAPC) in the HIV self-testing kit–purchasing population (HIVSTKPP) rate per 100,000 population between January 1, 2016, and December 31, 2019, by average annual rate per 100,000 population of HIVSTKPP between January 1, 2016, and December 31, 2019, among the 7 regions (A) and 31 provinces (B) in mainland China.
**Figure 9.** Temporal trend of the HIV self-testing kit–purchasing population (HIVSTKPP) rate per 100,000 population from January 1, 2016, to December 31, 2019, by region (refer to Textbox 1, which shows the number of regions and provinces).

**Figure 10.** The HIV self-testing kit–purchasing population (HIVSTKPP) rate per 100,000 population from January 1, 2016, to December 31, 2019, for 31 provinces in mainland China.
Figure 11. Temporal trend of the HIV self-testing kit–purchasing population (HIVSTKPP) rate per 100,000 population from January 1, 2016, to December 31, 2019, by province (there are 31 provinces in mainland China).

Figure 12. (A) Geographical distribution of the average annual rate per 100,000 population of the HIV self-testing kit–purchasing population (HIVSTKPP), (B) local clusters of the average annual rate of the HIVSTKPP identified by spatial autocorrelation analysis, (C) spatiotemporal clusters of the HIVSTKPP identified by temporal-spatial clustering analysis, and (D) hot spots and cold spots identified by the best-fitting Bayesian spatiotemporal model between January 1, 2016, and December 31, 2019, at city and provincial-controlled–county levels in mainland China. LISA: local indicators of spatial association; RR: relative risk.
Figure 13. Geographical distribution of the HIV self-testing kit–purchasing population (HIVSTKPP) rate per 100,000 population from January 1, 2016, to December 31, 2019, at city and provincial-controlled–county levels in mainland China.

Textbox 3. The top 10 cities with the highest average annual HIV self-testing kit–purchasing population (HIVSTKPP) rates.

<table>
<thead>
<tr>
<th>Top 10 cities with the highest average HIVSTKPP rates (per 100,000 population)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shenzhen in Guangdong province: 107.92</td>
</tr>
<tr>
<td>Hangzhou in Zhejiang province: 97.17</td>
</tr>
<tr>
<td>Zhuhai in Guangdong province: 93.98</td>
</tr>
<tr>
<td>Nanjing in Jiangsu province: 86.75</td>
</tr>
<tr>
<td>Guangzhou in Guangdong province: 85.74</td>
</tr>
<tr>
<td>Beijing: 83.86</td>
</tr>
<tr>
<td>Shanghai: 80.88</td>
</tr>
<tr>
<td>Chengdu in Sichuan province: 80.62</td>
</tr>
<tr>
<td>Xiamen in Fujian province: 80.49</td>
</tr>
<tr>
<td>Wuhan in Hubei province: 79.37</td>
</tr>
</tbody>
</table>
**Textbox 4.** The top 10 cities with the largest sizes of HIV self-testing kit–purchasing population (HIVSTKPP).

<table>
<thead>
<tr>
<th>City</th>
<th>HIVSTKPP Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shanghai</td>
<td>0.10 million</td>
</tr>
<tr>
<td>Beijing</td>
<td>0.09 million</td>
</tr>
<tr>
<td>Shenzhen in Guangdong province</td>
<td>0.07 million</td>
</tr>
<tr>
<td>Chengdu in Sichuan province</td>
<td>0.07 million</td>
</tr>
<tr>
<td>Guangzhou in Guangdong province</td>
<td>0.07 million</td>
</tr>
<tr>
<td>Chongqing</td>
<td>0.06 million</td>
</tr>
<tr>
<td>Hangzhou in Zhejiang province</td>
<td>0.05 million</td>
</tr>
<tr>
<td>Wuhan in Hubei province</td>
<td>0.05 million</td>
</tr>
<tr>
<td>Suzhou in Jiangsu province</td>
<td>0.04 million</td>
</tr>
<tr>
<td>Nanjing in Jiangsu province</td>
<td>0.04 million</td>
</tr>
</tbody>
</table>

**Spatial Autocorrelation Patterns**

The global Moran I showed significant positive spatial clustering of average annual HIVSTKPP rates at the city and provincial-controlled–county levels during the entire study period (Moran I=0.252; \( P<.001 \); Table S1 in Multimedia Appendix 2). The global Moran I decreased from 0.317 in 2016 to 0.234 in 2019. Significant local indicators of spatial association were mainly detected in high-high and low-low clusters (Table 4).

Between January 1, 2016, and December 31, 2019, of the 366 cities and provincial-controlled counties, 26 (7.1%) and 27 (7.4%) were identified as high-high and low-low clusters, respectively (Table 4). The average annual HIVSTKPP rate in high-high clusters (79.63 per 100,000) was 5.29-fold higher than that in low-low clusters (15.06 per 100,000). The high-high clusters, where the HIVSTKPP accounted for 26.16% (570,261/2,180,284) of all anonymous persons, were mainly located in East and South China, including Zhejiang, Guangdong, and Jiangsu provinces (Figure 12B and Table S2 in Multimedia Appendix 2). From January 1, 2016, to December 31, 2019, the local indicators of spatial association clusters remained relatively stable (Figure 14). The number of cities in high-high clusters decreased from 31 in 2016 to 24 in 2019, and the proportion of HIVSTKPP (high-high clusters vs all anonymous persons) decreased from 28.16% (80,297/285,107) in 2016 to 23.89% (216,848/907,539) in 2019, whereas the HIVSTKPP rates significantly increased from 41.61 per 100,000 in 2016 to 124.93 per 100,000 in 2019 (AAPC=45.6%; \( P<.001 \); Table 4).
Table 4. Descriptive statistics of 4 types of spatial clusters as defined by local indicators of spatial association analysis between January 1, 2016, and December 31, 2019, in mainland China.

<table>
<thead>
<tr>
<th>Period and patterns</th>
<th>Rate&lt;sup&gt;a&lt;/sup&gt;</th>
<th>HIV self-testing kit–purchasing population, n (%)</th>
<th>Cities (N=366), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>January 1, 2016, to December 31, 2019 (N=2,180,284)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-high</td>
<td>79.63</td>
<td>570,261 (26.16)</td>
<td>26 (7.10)</td>
</tr>
<tr>
<td>High-low</td>
<td>52.01</td>
<td>100,477 (4.61)</td>
<td>7 (1.91)</td>
</tr>
<tr>
<td>Low-high</td>
<td>22.53</td>
<td>6062 (0.28)</td>
<td>2 (0.55)</td>
</tr>
<tr>
<td>Low-low</td>
<td>15.06</td>
<td>34,220 (1.57)</td>
<td>27 (7.38)</td>
</tr>
<tr>
<td><strong>2016 (N=285,107)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-high</td>
<td>41.61</td>
<td>80,297 (28.16)</td>
<td>31 (8.47)</td>
</tr>
<tr>
<td>High-low</td>
<td>26.34</td>
<td>12,377 (4.34)</td>
<td>7 (1.91)</td>
</tr>
<tr>
<td>Low-high</td>
<td>13.00</td>
<td>864 (0.30)</td>
<td>2 (0.55)</td>
</tr>
<tr>
<td>Low-low</td>
<td>7.65</td>
<td>5522 (1.94)</td>
<td>34 (9.29)</td>
</tr>
<tr>
<td><strong>2017 (N=369,140)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-high</td>
<td>54.38</td>
<td>95,153 (25.78)</td>
<td>27 (7.38)</td>
</tr>
<tr>
<td>High-low</td>
<td>34.94</td>
<td>16,744 (4.54)</td>
<td>7 (1.91)</td>
</tr>
<tr>
<td>Low-high</td>
<td>15.80</td>
<td>1057 (0.29)</td>
<td>2 (0.55)</td>
</tr>
<tr>
<td>Low-low</td>
<td>10.80</td>
<td>5396 (1.46)</td>
<td>22 (6.01)</td>
</tr>
<tr>
<td><strong>2018 (N=618,498)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-high</td>
<td>86.16</td>
<td>140,666 (22.74)</td>
<td>23 (6.28)</td>
</tr>
<tr>
<td>High-low</td>
<td>60.35</td>
<td>30,046 (4.86)</td>
<td>7 (1.91)</td>
</tr>
<tr>
<td>Low-high</td>
<td>26.54</td>
<td>2381 (0.38)</td>
<td>3 (0.82)</td>
</tr>
<tr>
<td>Low-low</td>
<td>16.91</td>
<td>9840 (1.59)</td>
<td>27 (7.38)</td>
</tr>
<tr>
<td><strong>2019 (N=907,539)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-high</td>
<td>124.93</td>
<td>216,848 (23.89)</td>
<td>24 (6.56)</td>
</tr>
<tr>
<td>High-low</td>
<td>89.98</td>
<td>45,060 (4.97)</td>
<td>7 (1.91)</td>
</tr>
<tr>
<td>Low-high</td>
<td>38.68</td>
<td>3491 (0.38)</td>
<td>3 (0.82)</td>
</tr>
<tr>
<td>Low-low</td>
<td>22.16</td>
<td>11,617 (1.28)</td>
<td>29 (7.92)</td>
</tr>
</tbody>
</table>

<sup>a</sup>The rate of HIV self-testing kit purchasing per 100,000 population was calculated by dividing the number of purchasers by the total population in an area.
Spatiotemporal Clusters

Clusters with higher demand for HIVST kits available on the web were detected in 75 cities and provincial-controlled counties located in East China, South China, and Central China between January 1, 2018, and December 31, 2019, including 15 (20%) in Guangdong province, 11 (15%) in Zhejiang province, 11 (15%) in Jiangxi province, 10 (13%) in Anhui province, 9 (12%) in Fujian province, 7 (10%) in Jiangsu province, 6 (8%) in Hunan province, 5 (7%) in Hubei province, and 1 (1%) in Shanghai. Clusters with lower demand for HIVST kits available on the web were found in 181 cities and provincial-controlled counties between January 1, 2016, and December 31, 2017 (Figure 12C and Table 5). High-demand and low-demand clusters accounted for 42.2% (643,977/1,526,012) and 29.74% (194,571/654,240) of the total HIVSTKPP, respectively. The HIVSTKPP rate in high-demand clusters (165.79 per 100,000) was 5.03-fold higher than in low-demand clusters (32.94 per 100,000). Compared with neighboring cities and provincial-controlled counties, those identified in high-demand clusters had 2.55 times more HIVSTKPP (relative risk=2.55; P<.001).
Table 5. General description of the high-risk and low-risk clusters identified by temporal-spatial clustering analysis between January 1, 2016, and December 31, 2019, in mainland China.

<table>
<thead>
<tr>
<th>Cluster type</th>
<th>Time interval</th>
<th>Cluster center (latitude, longitude)</th>
<th>Radius (km)</th>
<th>Cities (N=366), n (%)</th>
<th>Provinces included (number of cities)</th>
<th>Purchasers (N=2,180,284), n (%)</th>
<th>Ratea</th>
<th>Relative risk</th>
<th>Log likelihood ratio</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-risk cluster</td>
<td>January 1, 2018, to December 31, 2019</td>
<td>26.049704 N, 119.180295 E</td>
<td>731.09</td>
<td>75 (20.49)</td>
<td>Guangdong (15), Zhejiang (11), Jiangxi (11), Anhui (10), Fujian (9), Jiangsu (7), Hunan (6), Hebei (5), Shanghai</td>
<td>643,977 (42.20)</td>
<td>165.79</td>
<td>2.55</td>
<td>171422.12</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Low-risk cluster</td>
<td>January 1, 2016, to December 31, 2017</td>
<td>36.712360 N, 94.531065 E</td>
<td>1931.67</td>
<td>181 (49.45)</td>
<td>Xinjiang (24), Sichuan (21), Henan (18), Yunnan (16), Gansu (14), Hubei (12), Shanxi (11), Shaanxi (10), Guizhou (9), Qinghai (8), Hunan (7), Inner Mongolia (7), Tibet (7), Hebei (6), Ningxia (5), Guangxi (3), Shandong (2), and Chongqing</td>
<td>194,571 (29.74)</td>
<td>32.94</td>
<td>0.36</td>
<td>119468.41</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

aThe rate of HIV self-testing kit purchasing per 100,000 population was calculated by dividing the number of purchasers by the total population in an area.

Spatiotemporal Patterns

We found that the spatiotemporal model with type IV interaction had the lowest deviance information criterion value. Therefore, this was the best-fit model and was applied in our analysis (Table S3 in Multimedia Appendix 2). During the study period, the highest spatial relative risk was noted in Shenzhen city of Guangdong province, Hangzhou city of Zhejiang province, and Zhuhai city of Guangdong province (Figure 15). Precisely 37.2% (136/366) and 48.1% (176/366) of the cities and provincial-controlled counties were classified as hot spots (higher demand for HIVST kits) and cold spots (lower demand for HIVST kits), respectively (Figure 12D and Table S4 in Multimedia Appendix 2).

Spatial variations in adjusted relative risk (ARR; relative demand for HIVST kits) of spatiotemporal interaction effect from January 1, 2016, to December 31, 2019, are presented in Figure 16. The ARRs in each region between January 1, 2016, and December 31, 2019, varied slightly, and the higher ARR (higher demand for HIVST kits) continued to be applicable in municipalities, provincial capitals, and large cities.
Figure 15. Geographical distribution of relative risk of the spatial effect at city and provincial-controlled–county levels from January 1, 2016, to December 31, 2019, in mainland China through the best-fitting Bayesian spatiotemporal model.
Correlates of HIVSTKPP

The SEM was found to be the most appropriate model for interpreting associations between city and provincial-controlled county–based factors and the HIVSTKPP (data not shown). The SEM results showed that the number of HIV testing facilities, urbanization ratio, and GDP per capita were positively associated with HIVSTKPP for the whole country, with the regression coefficients being 14.106, 11.236, and 1.057, respectively (all P values met the threshold for statistical significance; Table S5 in Multimedia Appendix 2). Notably, the geographically weighted regression model revealed a strong spatial heterogeneity in relationships between the HIVSTKPP and number of HIV testing facilities as well as urbanization ratio (Figure 17 and Table S6 in Multimedia Appendix 2). The number of HIV testing facilities and urbanization ratio were found to have a positive correlation with the HIVSTKPP in most cities (306/366, 83.6%; and 289/366, 79%, respectively). Cities with large positive correlations with the number of HIV testing facilities were concentrated in Sichuan and Guizhou provinces. In terms of urbanization ratio, cities with greater positive correlations were also predominantly in Sichuan and Guizhou provinces.

Figure 16. Geographical distribution of relative risk of the spatiotemporal interaction effect at city and provincial-controlled–county levels from January 1, 2016, to December 31, 2019, in mainland China through the best-fitting Bayesian spatiotemporal model. ARR: adjusted relative risk.
Figure 17. Geographical distribution of coefficients of the (A) number of HIV testing facilities, (B) urbanization ratio, and (C) gross domestic product (GDP) in the association with the number of the HIV self-testing kit–purchasing population at city and provincial-controlled–county levels in mainland China in 2019 through a geographically weighted regression model.

Discussion

Principal Findings

In this retrospective analysis of data involving 2.18 million anonymous persons in China who ordered 4.51 million HIVST kits on the web between January 1, 2016, and December 31, 2019, we identified key regions in which there was a larger demand for HIVST kits. Between January 1, 2016, and December 31, 2019, the size of the HIVSTKPP as well as HIVSTKPP rates increased across regions. High HIVSTKPP rates and hot spots identified by Bayesian spatiotemporal analysis were mainly located in large cities, provincial capitals, and municipalities. Spatial autocorrelation analysis identified high-high clusters and space-time cluster analysis identified higher-demand clusters that were predominantly in cities along China’s southeast coast. The number of HIV testing facilities, urbanization ratio, and GDP per capita were correlated with larger HIVSTKPP for the whole country.

The Interpretation of the Temporal Distribution of the HIVSTKPP and Its Implication for HIV Prevention

We observed a steady increase in HIVST kit purchases over the 4-year study period. The growth in the size of the HIVSTKPP accelerated after 2017. This growth may have been prompted by the 13th Five-Year Plan for HIV Prevention and Control, which directed the CCDC and nongovernmental organizations to focus on promoting HIVST at the population level, including organizing a series of promotional activities for the Nationwide HIV Testing and Consultation Month program between November 20 and December 20 in both 2018 and 2019 to raise public awareness of HIV testing and promote active testing.
Annual increases in the HIVSTKPP could also be explained by the fact that people are increasingly relying on web-based purchasing for their shopping generally, and HIVST kits are no different than other products. However, the numbers of HIVST kits among newly diagnosed people living with HIV in China did not show a similar rate of growth over the same period. The total number of HIVST kit purchases conducted in China was 169 million, 200.72 million, and 240.87 million in 2016, 2017, and 2018, respectively, which corresponded to annual growth rates of 18.77% in 2017 and 20% in 2018 [5]. The total number of newly diagnosed people living with HIV in China was 124,555, 134,512, and 148,589 in 2016, 2017, and 2018 respectively, which corresponded to annual growth rates of 7.99% in 2017 and 10.46% in 2018 [5]. Both were dwarfed by the increases in the HIVSTKPP in our study, with 0.64 million, 0.83 million, and 1.20 million in 2016, 2017 and 2018, respectively, which corresponded to annual growth rates of 29.80% in 2017 and 44.51% in 2018. The ranking of the size of the HIVSTKPP among 31 provinces was also different from those of HIV tests and newly diagnosed people living with HIV [5]. Although our study is not able to tease out whether the observed widespread and rapidly growing uptake of web-based purchase of HIVST kits is purely coming from a diffusion process of innovation or compounded with a worsening underlying epidemic, these phenomena confirm the potential existence of a previously unmet need in HIV prevention and care and highlight the urgent need for a novel strategy that links the HIVSTKPP to facility-based HIV testing and care.

The maximum monthly number of web-based purchases of HIVST kits occurs in December, which might indicate the effectiveness of public education campaigns that promote HIV testing on World AIDS Day (December 1). However, efforts to contain the epidemic should be in place throughout the year. Besides World AIDS Day, prevention campaigns may take place on a more frequent and regular basis.

The Interpretation of the Spatial Distribution of the HIVSTKPP and Its Implication for HIV Prevention

High HIVSTKPP rates and hot spots were identified in large cities, provincial capitals, and municipalities. The spatiotemporal distribution of the HIVSTKPP at local levels in our study provides policy makers with rare and valuable information about the number of people who are potentially in need of testing and treatment services. This may guide public health authorities to allocate resources, develop interventions, and deliver services.

Recent studies showed that HIV epidemic clusters among MSM spread from a few large cities in eastern China to most of the municipalities and provincial capitals countrywide between 2006 and 2015 [13]. HIV epidemic clusters among young people aged 15 to 24 years spread from southwestern China to central and northeastern China between 2005 and 2012 [14]. However, in our study, spatial clusters of the HIVSTKPP were mainly found in southeastern coastal cities between 2016 and 2019. Some studies showed that the majority of the HIVSTKPP reported that they had engaged in high-risk sexual behaviors in the last 6 months [12,42]. Thus, the web-based HIVST kit-purchasing behavior could largely indicate that recent high-risk sexual behavior may have been a factor in HIV acquisition, and the high-demand spatiotemporal cluster of the HIVSTKPP identified in our study could serve as an early warning signal for new HIV epidemics. Potential gaps in web-based HIVST kit-purchasing behavior and an HIV epidemic might exist and should be further identified by connecting and comparing our findings with these patterns of HIV epidemics in the same population and study period for a timely adjustment of the HIV prevention and control strategy.

The Interpretation of the Associations Between HIVST Kit Purchasing and Macroscopic Factors and Its Implication for HIV Prevention

Our SEM identified positive correlations between a region’s number of HIV testing facilities, urbanization ratio, and GDP per capita and the HIVSTKPP. The number of HIV testing facilities was positively correlated with the size of the HIVSTKPP in 306 cities and provincial-controlled counties. Regions with larger numbers of HIV testing facilities may have higher levels of relevant health education, thus further promoting HIV testing. In addition, the urbanization ratio was positively associated with the size of the HIVSTKPP in 289 cities and provincial-controlled counties. The positive effect could be associated with the assumption that the increase in the proportion of the urban population may promote social networking and sexually risky behaviors (eg, multiple sexual partners, unprotected sex behaviors, and commercial sexual behaviors) [28], thus leading to an increase in the demand for HIVST kits. GDP per capita did not have an important role to play with regard to the size of the HIVSTKPP, and its effect varies slightly across regions. Thus, the spatial patterns of the HIVSTKPP could be largely explained by the spatial variation in the number of HIV testing facilities and urbanization ratio.

Our study shows more HIVST kit purchases in areas with greater access to HIV testing centers and more urbanization. For areas currently lacking adequate access to HIV testing, especially areas bearing heavy HIV burdens (eg, rural provinces such as Tibet, Sichuan, Guizhou, and Guangxi) [1,5,43], our study also found that there were few web-based HIVST kit purchases. Thus, public education campaigns that promote web-based HIVST kit purchases at the population level are urgently required to expand HIV testing as well as HIV prevention and control programs in these areas.

Comparison With Prior Work

To the best of our knowledge, this study is the first to evaluate the spatiotemporal patterns of the HIVSTKPP. Monitoring HIVST kit purchasing patterns could be a new tool for public health policy makers, researchers, and program implementers to reallocate resources, promote HIVST uptake, and optimize the HIV care continuum.

Limitations

Our study includes several limitations. First, although sales records from a leading e-commerce platform in China were included in our study, we did not include records from all e-commerce platforms, HIV clinics, hospitals, offline pharmacies, and CCDC offices. The sample in our study does not represent the entire HIVSTKPP in China, and HIVSTKPP rates are underestimated in our analysis. Our findings should
be interpreted with caution, and further research that involves collecting data from all potential sources of HIVST kits is needed. Second, we could not quantitatively assess associations between the HIVSTKPP and HIV epidemiological data to identify potential gaps in current HIV-monitoring practices in China. In addition, we could not compare our findings with national CCDC data about new HIV diagnoses that were first screened using self-testing because these data were not publicly available, and we have no access to these data. Third, we could not evaluate the spatial patterns of the HIVSTKPP in finer geographic units (eg, street level). Finally, many sex workers, including male sex workers, could have purchased HIVST kits on the web in bulk at the end of the day. However, our study did not include the records of bulk purchases for analysis.

Conclusions
Our study provides an understanding of the spatiotemporal patterns of the HIVSTKPP in mainland China, which can inform the development of national and local HIVST guidelines in allocating resources and promoting HIV testing. The development of contextualized prevention and intervention strategies tailored to the HIVSTKPP in key regions identified from our study is urgently required for HIV prevention and control. Further research combining the spatiotemporal patterns of HIVST with HIV epidemiological data is needed to identify potential gaps in current HIV-monitoring practices and develop comprehensive HIV control strategies.

Acknowledgments
The authors thank Dr Thomas Fitzpatrick from the University of Washington and Dr Cong Jin from the National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, for their helpful suggestions regarding manuscript preparation. This study was supported by the Natural Science Foundation of China Excellent Young Scientists Fund (82022064), Natural Science Foundation of China International and Regional Research Collaboration Project (72061137001), Natural Science Foundation of China Young Scientist Fund (81703278), National Science and Technology Major Project of China (2018ZX10721102), National Science and Technology Major Project of China (2017ZX10201101-002), Sanming Project of Medicine in Shenzhen (SZSM201811071), High-Level Project of Medicine in Longhua, Shenzhen (HLPM201907020105), National Key Research and Development Program of China (2020YFC0840900), Shenzhen Science and Technology Innovation Commission Basic Research Program (JCYJ20190807155409373), Special Support Plan for High-Level Talents of Guangdong Province (2019TQ05Y230), and Fundamental Research Funds for the Central Universities (58000-31620005). The funding parties had no role in the design of the study or in the explanation of the data.

Authors’ Contributions
HZ and GL conceived the study design. GL participated in literature search, data cleaning, statistical analysis, creating the tables, plotting the graphics, interpreting the study findings, and manuscript writing. LS, GL, AF, and YFL participated in data collection. HZ and MH are responsible for the overall content. All authors critically reviewed and substantively revised the manuscript. All authors have approved the final version of the manuscript for publication.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Details of the statistical analysis.
[DOCX File, 46 KB - publichealth_v8i10e35272_app1.docx ]

Multimedia Appendix 2
Detailed results (Tables S1-S6) of spatiotemporal analysis of the HIV self-testing kit–purchasing population (HIVSTKPP) from January 1, 2016, to December 31, 2019 in mainland China. [DOCX File, 27 KB - publichealth_v8i10e35272_app2.docx]

References


Abbreviations

AAPC: average annual percentage change
ARR: adjusted relative risk
CCDC: Chinese Center for Disease Control and Prevention
GDP: gross domestic product
HIVST: HIV self-testing
HIVSTKPP: HIV self-testing kit–purchasing population
MSM: men who have sex with men
SEM: spatial error model

Edited by T Sanchez, A Mavragani; submitted 29.11.21; peer-reviewed by S John, J Tucker, Q Qin; comments to author 04.05.22; revised version received 19.07.22; accepted 06.09.22; published 04.10.22.

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

Zixuan Zhao\textsuperscript{1*}, PhD; Lingbin Du\textsuperscript{2*}, MPH, PhD; Yuanyuan Li\textsuperscript{3}, MD, PhD; Le Wang\textsuperscript{2}, PhD; Youqing Wang\textsuperscript{2}, MD; Yi Yang\textsuperscript{1}, PhD; Hengjin Dong\textsuperscript{1}, PhD

\textsuperscript{1}Department of Science and Education of the Fourth Affiliated Hospital, Center for Health Policy Studies, School of Public Health, Zhejiang University School of Medicine, Hangzhou, China
\textsuperscript{2}Department of Cancer Prevention, Cancer Hospital of the University of Chinese Academy of Sciences, Zhejiang Cancer Hospital, Hangzhou, China
\textsuperscript{3}Department for Science and Education, Hangzhou Ninth People’s Hospital, Hangzhou, China
\textsuperscript{*}these authors contributed equally

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

Related Article:
https://publichealth.jmir.org/2022/7/e36425

(\textit{JMIR Public Health Surveill} 2022;8(10):e43025) doi:10.2196/43025

In “Cost-Effectiveness of Lung Cancer Screening Using Low-Dose Computed Tomography Based on Start Age and Interval in China: Modeling Study” (\textit{JMIR Public Health Surveill} 2022;8(7):e36425), the authors noted an error:

In the originally published article, Affiliations 1 and 4 were inadvertently presented as two different affiliations due to erroneous presentation of organizational units within these affiliations.

In the corrected version, the two affiliations are merged into a single affiliation with the correct order of organizational units. With this update, the numbering and attribution of affiliations to the authors are updated as follows:

\begin{itemize}
  \item Zixuan Zhao\textsuperscript{1}; Lingbin Du\textsuperscript{2}; Yuanyuan Li\textsuperscript{3}; Le Wang\textsuperscript{2}; Youqing Wang\textsuperscript{2}; Yi Yang\textsuperscript{1}; Hengjin Dong\textsuperscript{1}
\end{itemize}

\textsuperscript{1}Department of Science and Education of the Fourth Affiliated Hospital, Center for Health Policy Studies, School of Public Health, Zhejiang University School of Medicine, Hangzhou, China
\textsuperscript{2}Department of Cancer Prevention, Cancer Hospital of the University of Chinese Academy of Sciences, Zhejiang Cancer Hospital, Hangzhou, China
\textsuperscript{3}Department for Science and Education, Hangzhou Ninth People’s Hospital, Hangzhou, China

Affiliation noted in the corresponding author’s contact information is also updated accordingly.

The correction will appear in the online version of the paper on the JMIR Publications website on October 17, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.
Correction: Cost-Effectiveness of Lung Cancer Screening Using Low-Dose Computed Tomography Based on Start Age and Interval in China: Modeling Study

Published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 17.10.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Corrigenda and Addenda

Correction: Google Trends on Human Papillomavirus Vaccine Searches in the United States From 2010 to 2021: Infodemiology Study

Akshaya Srikanth Bhagavathula1*, PharmD, PhD; Philip M Massey1*, MPH, PhD
Center for Public Health and Technology, Department of Health, Human Performance, and Recreation, College of Education and Health Professions, University of Arkansas, Fayetteville, AR, United States
*all authors contributed equally

Corresponding Author:
Philip M Massey, MPH, PhD
Center for Public Health and Technology
Department of Health, Human Performance, and Recreation, College of Education and Health Professions
University of Arkansas
346 West Ave.
West Avenue Annex Building, Suite 317
Fayetteville, AR, 72701
United States
Phone: 1 4795758491
Email: masseyp@uark.edu

Related Article:
Correction of: https://publichealth.jmir.org/2022/8/e37656

(JMIR Public Health Surveill 2022;8(10):e42812) doi:10.2196/42812

In “Google Trends on HPV vaccine searches in the U.S. from 2010 - 2021: An Infodemiology Study” (JMIR Public Health Surveill 2022;8(8):e37656) the authors noted one error. In the originally published article, Figure 2 appeared incorrectly (as shown in Multimedia Appendix 1). The correct Figure 2 is provided below.
Figure 2. Joinpoint regression analysis indicating trends in HPV vaccine Relative Search Volume (RSVs) on Google Trends from 2010 – 2021 in the U.S.

Multimedia Appendix 1
Originally published Figure 2.
[ PNG File , 410 KB , publichealth_v8i10e42812_app1.png ]
COVID-19 Cases Among Congregate Care Facility Staff by Neighborhood of Residence and Social and Structural Determinants: Observational Study

Huiting Ma¹, BMath, MSc; Kristy C Y Yiu¹, BHSc, MSc; Stefan D Baral², MD, MPH; Christine Fahim¹, MSc, PhD; Gary Moloney¹, BASc, MSA; Dariya Darvin¹, BSc; David Landsman¹, BSc, MScAC; Adrienne K Chan³,⁴,⁵,⁶, MD, MPH; Sharon Straus¹,⁷, MD, MSc; Sharmistha Mishra¹,³,⁴,⁶, MD, PhD

¹St Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada
²Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD, United States
³Division of Infectious Diseases, Department of Medicine, University of Toronto, Toronto, ON, Canada
⁴Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
⁵Division of Infectious Diseases, Sunnybrook Health Sciences, University of Toronto, Toronto, ON, Canada
⁶Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
⁷Department of Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Sharmistha Mishra, MD, PhD
St Michael's Hospital
Unity Health Toronto
Room 315, 209 Victoria Street
Toronto, ON, M5B 1T8
Canada
Phone: 1 416 864 6060 ext 5568
Email: sharmistha.mishra@utoronto.ca

Abstract

Background: Disproportionate risks of COVID-19 in congregate care facilities including long-term care homes, retirement homes, and shelters both affect and are affected by SARS-CoV-2 infections among facility staff. In cities across Canada, there has been a consistent trend of geographic clustering of COVID-19 cases. However, there is limited information on how COVID-19 among facility staff reflects urban neighborhood disparities, particularly when stratified by the social and structural determinants of community-level transmission.

Objective: This study aimed to compare the concentration of cumulative cases by geography and social and structural determinants across 3 mutually exclusive subgroups in the Greater Toronto Area (population: 7.1 million): community, facility staff, and health care workers (HCWs) in other settings.

Methods: We conducted a retrospective, observational study using surveillance data on laboratory-confirmed COVID-19 cases (January 23 to December 13, 2020; prior to vaccination rollout). We derived neighborhood-level social and structural determinants from census data and generated Lorenz curves, Gini coefficients, and the Hoover index to visualize and quantify inequalities in cases.

Results: The hardest-hit neighborhoods (comprising 20% of the population) accounted for 53.87% (44,937/83,419) of community cases, 48.59% (2356/4849) of facility staff cases, and 42.34% (1669/3942) of other HCW cases. Compared with other HCWs, cases among facility staff reflected the distribution of community cases more closely. Cases among facility staff reflected greater social and structural inequalities (larger Gini coefficients) than those of other HCWs across all determinants. Facility staff cases were also more likely than community cases to be concentrated in lower-income neighborhoods (Gini 0.24, 95% CI 0.15-0.38 vs 0.14, 95% CI 0.08-0.21) with a higher household density (Gini 0.23, 95% CI 0.17-0.29 vs 0.17, 95% CI 0.12-0.22) and with a greater proportion working in other essential services (Gini 0.29, 95% CI 0.21-0.40 vs 0.22, 95% CI 0.17-0.28).

Conclusions: COVID-19 cases among facility staff largely reflect neighborhood-level heterogeneity and disparities, even more so than cases among other HCWs. The findings signal the importance of interventions prioritized and tailored to the home
geographies of facility staff in addition to workplace measures, including prioritization and reach of vaccination at home (neighborhood level) and at work.

**Methods**

**Study Design, Setting, and Population**

We conducted a retrospective, observational study using provincial surveillance data on laboratory-confirmed COVID-19 cases reported between January 23, 2020, and December 13, 2020, in the Greater Toronto Area (population: 7.1 million) [17] and in accordance with the RECORD (Reporting of Studies Conducted Using Observational Routinely-Collected Data) statement [18]. We restricted the study to the period before COVID-19 vaccination was available due to differential vaccine allocation and coverage over time by each subgroup after vaccine rollout [19].

**Data Sources and Measures**

We used person-level data from Ontario’s centralized surveillance system [20], which includes information on laboratory-confirmed COVID-19 cases by reported date, demographic characteristics, exposure category, and setting-specific characteristics (eg, LTCHs), as well as data on social and structural determinant measures from the Statistics Canada 2016 Census [21]. The surveillance data classify cases as an HCW if a person works or volunteers in any health care setting (including LTCHs, retirement homes, shelters, hospitals, clinics, or homecare). We stratified HCWs into those associated with working or volunteering in an LTCH, retirement home, and/or shelter as facility staff, and all others as “other HCWs.” If an HCW fell into both categories (facility staff and other HCWs), then they were categorized as facility staff.

We examined social and structural determinants at the level of the dissemination area (neighborhood) because it was the smallest geographic unit (population size ranging from 400 to 700) for which census data were available. Other geographic units include the forward sortation area and census tracts, but the dissemination area is most commonly used when examining social and structural determinants because it reflects the smallest geographic unit and is less prone to ecological fallacy than larger geographic units [22]. We conceptualized and defined the social and structural determinant measures from the Statistics Canada 2016 Census [21,23,24] and are related to socioeconomic status (per-person equivalent after tax income) and proxies for systemic racism (% visible minority, % recent immigration), or to the potential for increased contact rates (housing: % not living in high-density housing [25,26], % living in multigenerational households) and employment in other essential services (ie, excluding health care) [27] not amenable to remote work [28].

https://publichealth.jmir.org/2022/10/e34927
Analyses

We aggregated the number of confirmed COVID-19 cases at the neighborhood level during the study period into 3 mutually exclusive subgroups: community (excluding facility staff, other HCWs, congregate facility residents, and travel-related cases), facility staff (workers or volunteers in LTCHs, retirement homes, and shelters), and other HCWs. We generated Lorenz curves and Gini coefficients to quantify the magnitude of inequalities (i.e., the concentration in cases), and the Hoover index was used as an alternate measure for validation [15,16,29]. With the Gini coefficient, a value closer to zero represents greater equality [30]. The Hoover index measures the percentage of cases that would need to be redistributed to achieve equality in how cases are distributed across neighborhoods. As with the Gini coefficient, a larger Hoover index represents greater inequality [31]. We generated 95% CIs for Gini coefficients using bootstrapping [29,32].

First, we compared the magnitude of geographic concentration of cases for each subgroup (y-axis) against the distribution of total cases (x-axis, community plus travel related) at the neighborhood level. To examine the extent to which facility staff and other HCW cases mirrored community cases, we generated a separate set of Lorenz curves and Gini coefficients using community cases on the x-axis. Second, to examine the magnitude of inequalities by each social or structural determinant, we ranked the cumulative proportion of the population by each determinant (e.g., from lowest to highest income decile) on the x-axis. A detailed analytic plan can be found in Multimedia Appendix 2 [9,10,15,16,20-22,25-32]. We also generated spatial maps to describe and overlay cases among facility staff and among other HCWs using one social determinant as an example (neighborhood-level income). Analyses were conducted in R (version 4.0.2; R Core Team), and spatial maps were generated using ArcGIS (version 10.7; Esri).

Ethics Approval

The University of Toronto Health Sciences Research Ethics Board approved the study (protocol #39253).

Results

Overview

Of the 92,210 cases (excluding congregate facility residents and travel-related cases) included during our study period, there were 83,419 cases in the community, 4849 cases among facility staff, and 3942 cases among other HCWs (Table 1). Among facility staff, there were 4241, 363, and 245 cases among LTCH staff, retirement home staff, and shelter staff, respectively.

Table 1. Number of COVID-19 cases in 3 mutually exclusive subgroups (community, facility staff, and other health care workers) in the Greater Toronto Area (January 23, 2020, to December 13, 2020).

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>COVID-19 cases, n</th>
<th>Dissemination areas(^{a}) with zero cases, n ((%))(^{b})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community(^{c})</td>
<td>83,419</td>
<td>1058 (12.8)</td>
</tr>
<tr>
<td>Facility staff(^{d})</td>
<td>4849</td>
<td>5771 (69.7)</td>
</tr>
<tr>
<td>Other health care workers</td>
<td>3942</td>
<td>5879 (71.0)</td>
</tr>
</tbody>
</table>

\(^{a}\) Dissemination area refers to the geographic unit of measurement for the social and structural determinants examined in this study generated from Statistics Canada [21]. In the Greater Toronto Area (population: 7.1 million), the median population size of a dissemination area is 561 (IQR 442-800) residents.

\(^{b}\) A total of 8278 dissemination areas in the region.

\(^{c}\) Excludes residents of congregate settings and facility staff (long-term care homes, retirement homes, and shelters), other health care workers, and travel-related cases.

\(^{d}\) Includes staff and volunteers who work in long-term care homes, retirement homes, and shelters, and excludes all other health care workers.

Geographic Concentration of Cases Across Subgroups

The most affected neighborhoods (x-axis) comprising 20% of the total population accounted for 53.87% (44,937/83,419) of community cases, 48.59% (2356/4849) of congregate setting worker cases, and 42.34% (1669/3942) of other HCW cases (Figure 1). Compared with other HCWs, cases among facility staff more closely reflected the geographic distribution of community cases (Gini 0.06 vs 0.16; Hoover 0.05 vs 0.12) (Multimedia Appendix 3).
**Figure 1.** Geographic concentration of COVID-19 cases in the community population, among facility staff, and among other health care workers in the Greater Toronto Area (January 23, 2020, to December 13, 2020). The magnitude of concentration is depicted by Lorenz curves (the dashed line represents the line of equality) and the corresponding Gini coefficient for each subgroup. The x-axis represents the cumulative proportion of the population ranked by DAs from the highest to lowest number of cumulative cases per capita. “Community” excludes residents of congregate settings and facility staff (long-term care homes, retirement homes, and shelters), other health care workers, and travel-related cases. “Facility staff” includes staff and volunteers who work in long-term care homes, retirement homes, and shelters and excludes all other health care workers. DA: dissemination area.

---

**Differences in the Concentration of Cases Across Subgroups by Social and Structural Determinants**

Multimedia Appendix 4 depicts Lorenz curves and Gini coefficients by each social and structural determinant. Cases among facility staff reflected greater social and structural inequalities (larger Gini coefficients and a larger Hoover index) than other HCWs across all determinants (Figure 2, Multimedia Appendices 4 and 5). Multimedia Appendix 6 depicts how cases among facility staff and other HCWs were clustered along neighborhood-level income. Although facility-level cases mirrored that of community cases (Multimedia Appendix 3), there were greater inequalities in facility-level versus community cases with respect to income (Figure 2: Gini 0.24, 95% CI 0.15-0.38 vs 0.14, 95% CI 0.08-0.21), household density (Gini 0.23, 95% CI 0.17-0.29 vs 0.17, 95% CI 0.12-0.22), and other essential services (Gini 0.29, 95% CI 0.21-0.40 vs 0.22, 95% CI 0.17-0.28). Cases in the community, among facility staff, and among other HCWs were disproportionately concentrated in areas with lower household income and in areas with a higher proportion of visible minorities, recent immigration, high-density housing, multigenerational households, and employment in other essential services (Figure 2).
Discussion

We found that the distribution of COVID-19 cases among facility workers mirrored neighborhood heterogeneity and social and structural disparities, a pattern that was less evident with other HCWs. Facility staff cases reflected greater inequalities by social and structural determinants than cases among other HCWs, and with some determinants (income and other essential workers), greater inequalities were seen with facility staff cases compared to community cases.

Cases among facility staff more closely reflected the geographic distribution of community cases than the distribution of cases among other HCWs and in the community. The similar distributions of facility staff and community cases could occur if there was an equal distribution of facility staff living across neighborhoods and the infection rate ratio between facility staff and the community was the same across neighborhoods. The implication of this potential mechanism is that most infections among facility staff would have been acquired outside the facility or workplace. The Lorenz curve patterns may also occur if facility staff were more likely than other HCWs to live in harder-hit neighborhoods with greater social and structural disparities, irrespective of workplace exposures in congregate facilities.

Although our study was centered on the Greater Toronto Area, the findings are likely to be generalizable to other large, urban cities and metropolitan areas with similar patterns of social and structural inequalities. In Canada, most congregate facilities are concentrated in large cities [34], and previous research comparing 16 cities demonstrated a similar neighborhood clustering of COVID-19 cases by social and structural determinants [10]. Data on the neighborhood characteristics of LTCF staff in the United States [11] suggest workers tend to live in lower-income neighborhoods, and individual-level data in Canada suggest that a high proportion of LTCF staff identify as racialized women with low household income [35]. In current LTCF staffing models across Canada and the United States, approximately 60% to 90% of workers who provide direct care to residents are providing services as unregulated staff (personal support workers, care aides, orderlies, and nurse assistants) [36-40] and receive the lowest wages in the health care sector at or just above minimum wage [36], often in the context of contract or casual work without benefits [36,37,41-45].

Our study suggests that cases among facility staff may disproportionately intersect with household exposures that are connected with other essential workplaces or amplified in the context of household density [21]. Our study was limited by a lack of confirmed denominators for setting-specific workers in the region, but data from England and Wales suggest a 2-fold increased rate of COVID-19 among LTCF workers versus other HCWs [46]. Based on government reports, just over 100,000 employees serve 78,000 LTCF residents in the province [36], such that the ratio of staff to LTCF residents in Ontario is approximately 1 staff for every 0.78 LTCF resident. If we extrapolate the provincial ratio to the Greater Toronto Area, where 28,316 LTCF residents reside, the city would have approximately 36,303 LTCF staff. With 4241 cases among LTCF staff in the Greater Toronto Area during our study period and a total population in the Greater Toronto Area of 7.1 million [17], the cumulative rate of COVID-19 cases among LTCF staff (at 11,682 per 100,000) would have been 10 folds higher than that of the community (1174 cases per 100,000).
This study has several limitations. First, we derived the dissemination area–level social determinants from the 2016 Census data, which may not be representative of the population during the COVID-19 pandemic. Second, the occupation status obtained from Ontario’s centralized surveillance system could have been misclassified due to possible misinterpretation of the question in self-reporting; further, facility staff or HCWs may work in multiple settings. Third, we did not have data on the residence of all HCWs across the various congregate settings to compare neighborhood-level per-HCW rates of cases. Finally, data were not available to link cases among HCWs to specific facilities and to directly examine how cases in communities influenced outbreaks in congregate settings.

The findings have implications for COVID-19 modeling and interventions. The magnitude of inequalities can be used as calibration or validation targets for epidemic and prediction models to reproduce the observed pattern of cases in relation to the distribution of overall cases and by social and structural determinants. In doing so, detection systems (eg, neighborhood wastewater surveillance) designed to predict the potential for exposures in congregate facilities could leverage data on underlying vulnerabilities in the neighborhood of facility staff residence. A study in the United States found that neighborhood characteristics of LTCF staff’s residences were the most important predictor of LTCF outbreaks [11]. These data could then be used to implement strategies to mitigate risks. For example, proximal strategies to reduce community-level transmission risks conferred through social and structural inequalities have the potential to reduce workplace exposure risks. Examples include systematically addressing the lived realities of workers that make physical distancing challenging (eg, household density) or that remain barriers to effective isolation and quarantine (eg, precarious job security and absence of benefits such as paid sick leave), with interventions such as wraparound care including access to food, medications, and child and senior care (especially in the context of multigenerational households) to facilitate staff quarantine and isolation. Prioritizing vaccination coverage in the hardest-hit neighborhoods is another example of indirectly reducing workplace exposures in LTCF, retirement homes, and shelters. Finally, the findings highlight an urgent need for a long-term commitment and resources to comprehensively address social and structural barriers at a systems level (integration of health, education, social services, public health, and labor) given the long-standing history of infectious disease outbreaks in facilities and disparities experienced by its staff even before the COVID-19 pandemic [37].

Acknowledgments
We thank the Ontario Ministry of Health Data Analytics Branch for supporting access to data via the Ontario COVID-19 Modelling Consensus Table.

SM is supported by a Tier 2 Canada Research Chair in Mathematical Modeling and Program Science. SS is supported by a Tier 1 Canada Research Chair in Knowledge Translation and Quality of Care.

This work was funded by a Public Health Agency of Canada COVID-19 Immunology Task Force COVID-19 Hot Spots Competition Grant (grant 2021-HQ-000143) and a Canadian Institute of Health Research Grant (grant VR5-172683). The funders did not have any role in the study design; collection, analysis, and interpretation of data; writing of the paper; and decision to submit for publication.

Data Availability
Reported COVID-19 cases were obtained from Case and Contact Management Solutions via the Ontario COVID-19 Modelling Consensus Table and with approval from the University of Toronto Health Sciences Research Ethics Board. The analyses, conclusions, opinions, and statements expressed herein are solely those of the authors and do not reflect those of the funding or data sources; no endorsement is intended or should be inferred.

Authors’ Contributions
HM and SM conceived and designed the study with input from KCYY. HM, DD, and SM conducted the literature review. HM developed the analysis plan with input from SM. HM led data management, data cleaning, data linkage, and variable creation, with support from DD and DL. GM sourced and generated the census-level data and wrote the appendix related to the census variables. HM executed the analysis plan and conducted the statistical analysis. HM wrote the first draft of the manuscript. All authors provided critical input on the study design, interpretation of results, and manuscript review and editing.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Social determinants of health: variables from the Statistics Canada 2016 Census of Population. [DOCX File, 17 KB - publichealth_v8i10e34927_app1.docx]
Multimedia Appendix 2
Detailed analytic plan.
[DOCX File, 15 KB - publichealth_v8i10e34927_app2.docx ]

Multimedia Appendix 3
Geographical concentration of COVID-19 cases among facility staff and other health care workers compared with community cases in the Greater Toronto Area, from January 23, 2020, to December 13, 2020.
[DOCX File, 651 KB - publichealth_v8i10e34927_app3.docx ]

Multimedia Appendix 4
Lorenz curves and Gini coefficients of COVID-19 cases in the community, among facility staff, and among other health care workers by social determinants in the Greater Toronto Area, from January 23, 2020, to December 13, 2020.
[DOCX File, 967 KB - publichealth_v8i10e34927_app4.docx ]

Multimedia Appendix 5
Magnitude of concentration of COVID-19 cases by social and structural determinants in the community, among facility staff, and among other health care workers in the Greater Toronto Area, from January 23, 2020, to December 13, 2020.
[DOCX File, 335 KB - publichealth_v8i10e34927_app5.docx ]

Multimedia Appendix 6
Map overlay of household income deciles by dissemination area and distribution of (A) COVID-19 facility staff cases and (B) other health care worker cases in the Greater Toronto Area, from January 23, 2020, to December 13, 2020.
[DOCX File, 2870 KB - publichealth_v8i10e34927_app6.docx ]

References


Abbreviations

HCW: health care worker
LTCH: long-term care home
RECORD: Reporting of Studies Conducted Using Observational Routinely-Collected Data

©Huiting Ma, Kristy C Y Yiu, Stefan D Baral, Christine Fahim, Gary Moloney, Dariya Darvin, David Landsman, Adrienne K Chan, Sharon Straus, Sharmistha Mishra. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 04.10.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.