ISSN 2096-7071 (Print)
ISSN 2096-3101 (Online)
CN 10-1629/R1

CHINA CDC WEEKLY

Vol. 6 No. 37 Sept. 13, 2024

中国疾病预防控制中心周报

ISSN 2096-7071

China CDC Weekly

Cover Image: adapted from Canva, https://www.canva.cn/.

Methods and Applications

Mapping the Characteristics of Respiratory Infectious Disease Epidemics in China Based on the Baidu Index from November 2022 to January 2023

Dazhu Huo[1,&]; Ting Zhang[2,&]; Xuan Han[2]; Liuyang Yang[2]; Lei Wang[3]; Ziliang Fan[4]; Xiaoli Wang[5]; Jiao Yang[2]; Qiangru Huang[2]; Ge Zhang[6]; Ye Wang[2]; Jie Qian[2]; Yanxia Sun[2]; Yimin Qu[2]; Yugang Li[2]; Chuchu Ye[7]; Luzhao Feng[2]; Zhongjie Li[2]; Weizhong Yang[2,#]; Chen Wang[1,#]

ABSTRACT

Introduction: Infectious diseases pose a significant global health and economic burden, underscoring the critical need for precise predictive models. The Baidu index provides enhanced real-time surveillance capabilities that augment traditional systems.

Methods: Baidu search engine data on the keyword "fever" were extracted from 255 cities in China from November 2022 to January 2023. Onset and peak dates for influenza epidemics were identified by testing various criteria that combined thresholds and consecutive days.

Results: The most effective scenario for indicating epidemic commencement involved a 90th percentile threshold exceeded for seven consecutive days, minimizing false starts. Peak detection was optimized using a 7-day moving average, balancing stability and precision.

Discussion: The use of internet search data, such as the Baidu index, significantly improves the timeliness and accuracy of disease surveillance models. This innovative approach supports faster public health interventions and demonstrates its potential for enhancing epidemic monitoring and response efforts.

Infectious diseases are a leading cause of death and disability worldwide, imposing a significant burden on public health and economic stability (*1*). The recent increase in emerging infectious diseases highlights the urgent need for accurate disease propagation predictions (*2*–*3*). The widespread use of internet data, particularly from platforms like Baidu, provides complementary real-time insights that enhance traditional infectious disease surveillance mechanisms (*4*–*5*). The Baidu index, distinguished by its superior forecasting accuracy and stability (*6*), has emerged as an invaluable asset for enriching existing surveillance systems.

This study aimed to develop a surveillance model for epidemiological trends using the Baidu index as a cornerstone. Precise trend detection promises to facilitate prompt and effective public health interventions (*7*).

METHODS

Data Source

This study utilized data from the Baidu search engine, extracted from the publicly accessible Baidu Index website, covering trend analyses across 31 provincial-level administrative divisions (PLADs) in the Chinese mainland. The research focused on tracking the keywords "fever," including "fa re" and "fa shao" in Chinese, to analyze coronavirus disease 2019 (COVID-19)-related data trends and compiled data from 255 cities. Notably, China enforced a dynamic COVID-zero strategy between 2019 and 2022 and initiated a pivotal policy transition on November 11, 2022, ultimately abandoning the COVID-zero strategy on December 7, 2022 (*8*).

Baseline data were established using Baidu Index data from August to October 2020–2022 for cities without pandemic activity. The year with the lowest average index was selected. If outbreaks occurred during these months, data from May to July were used.

This study investigated the 2023 influenza outbreak using data collected from November 1, 2022, to January 2, 2023. Data from the northern city of Weifang and the southern city of Yichang were used to identify regional variations. To avoid the influence of the COVID-19 pandemic, pre-2019 influenza-related internet search data were used. Data from October to December 2018 served as the baseline for adjusting changes in internet usage over time.

Onset Indicators

Various thresholds, paired with criteria for consecutive days, were used to pinpoint the precise start of the influenza epidemic. First, thresholds based on the 70th, 80th, and 90th percentiles of days with non-zero values on the Baidu index were established. These percentiles spanned from August 1, 2020, to November 1, 2020. The onset criteria were defined as the Baidu index exceeding the thresholds for at least three consecutive days or a longer sequence of seven consecutive days. This criterion was applied universally, resulting in six scenarios. To mitigate the influence of extraneous variables on the search index, a three-day moving average was incorporated into the analysis.

Peak Indicators

The real-time Baidu index from November 1, 2022, to January 2, 2023, was compared to a historical baseline calculated as the moving average of the Baidu index from 2020 to 2021. The Baidu Excess Search index, which measures excess search activity, was calculated using the following formula:

$$\text{Baidu Excess Search index} = \frac{\sum\left(\int x - \int n\right)}{\int m}$$

Here, x denotes the real-time Baidu index, n represents the n-day moving average of the Baidu index for the historical period, and m denotes the m-day moving average for the same period.

To determine the peak date reflecting the highest surge in search volume attributable to new cases in the current outbreak, we adjusted for the influence of other factors and diseases on the keyword search index. The optimal "m" value for the moving average was determined by evaluating three scenarios: m=3, m=7, and m=3l (assuming "l" is defined in the context). The "m" value was selected by comparing the outcomes of these scenarios. To minimize the impact of extraneous factors, we applied the "m"-day moving average for the historical Baidu index.

Subsequently, a city's peak was identified based on two criteria, with outcomes compared to render a judgment: the Baidu Engine Excess Search index demonstrated an abnormal rise from the start date followed by a decline for three to five consecutive days (*1*).

Study Structure

Baidu index data from 331 Chinese cities underwent a stringent quality assessment. This process excluded 76 cities based on two criteria. First, cities with a Baidu index below the 30th percentile were excluded. Second, cities reporting a zero index for any week between August 1 and November 31, 2022, were removed (Figure 1). After establishing the methodology and confirming its efficacy using the testing data, influenza search data were used to corroborate its versatility.

RESULTS

Evaluation and Comparison of Onset Criteria

Baseline threshold assessment. A comparative analysis of the average Baidu index and standard deviation was conducted across cities without pandemics during August–October 2020, 2021, and 2022. The year 2020, demonstrating the lowest average, was subsequently chosen as the reference for establishing the threshold (Supplementary Figure S1, available at https://weekly.chinacdc.cn/).

The criterion combining the 90th percentile threshold with seven consecutive days exceeding this threshold identified the fewest epidemic onsets among the six scenarios. Therefore, this combination was adopted to define the commencement of an outbreak.

Onset criteria evaluation revealed that the epidemic began on November 9 in more northern cities and PLADs, while southern cities and PLADs tended to see an onset date of December 28 (Figure 2A).

Criteria Evaluation for Peak Identification

Daily Baidu engine excess search values and peak dates were calculated for m-values of 3, 7, and 31. An m-value of 3 exhibited data fluctuations rather than a consistent trend (Supplementary Table S1). Conversely, an m-value of 31 blurred details and risked distortion. An m-value of 7 provided stable results and was chosen for its reliable measurement, minimizing noise and avoiding the loss of significant data variations.

Peak criteria evaluation. A criterion of three consecutive days of decline adequately signaled the peak in Sanya, Kashi, and Baishan, but five consecutive days did not. Therefore, a three-day decline was designated as the criterion. Applying this criterion to 255 cities, eight cities had not reached their peak by January 2, 2023 (Figure 2B).

Northern cities experienced a median ILI onset date of December 9 (interquartile range: December 4–10),

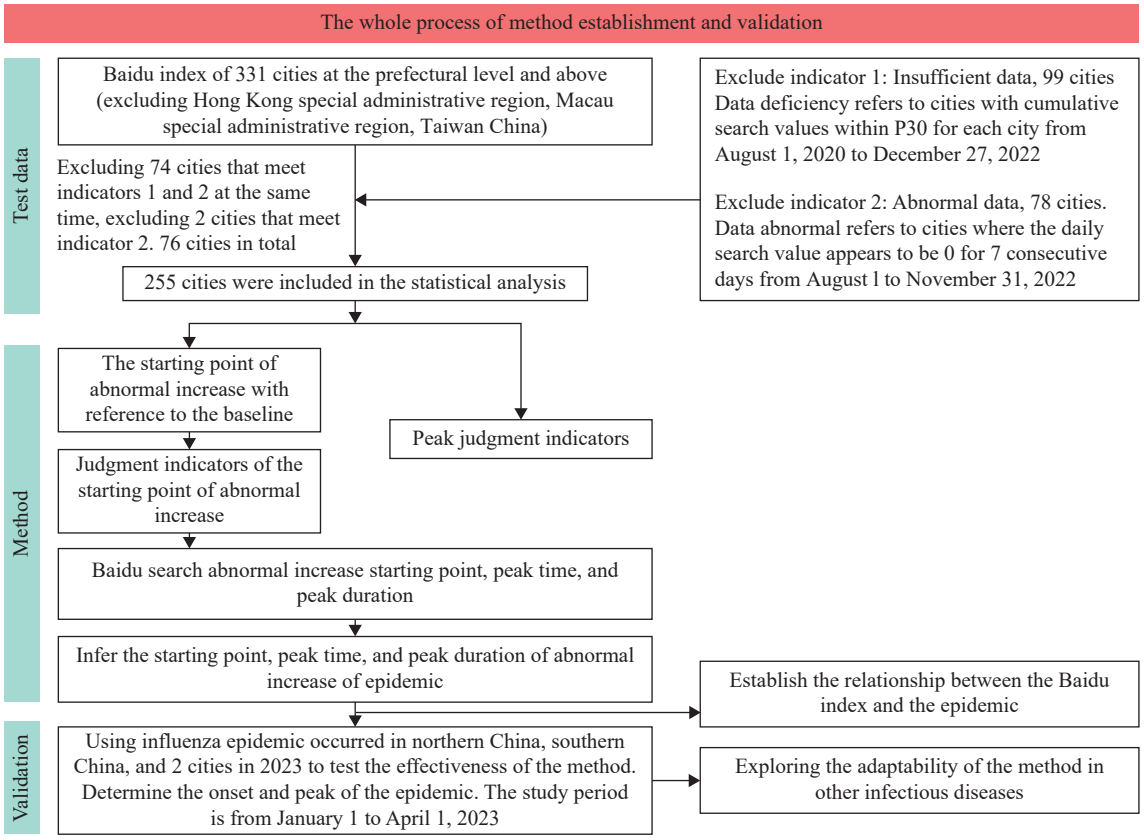


FIGURE 1. Study structure.

which was slightly earlier than the median onset date of December 11 (interquartile range: December 9–13) observed in southern cities. These data suggest that the influenza pandemic was initiated earlier in northern cities than southern cities.

Northern cities peaked around December 18 (interquartile range: December 17–20), whereas southern cities peaked later on December 20 (interquartile range: December 17–23), indicating earlier pandemic intensification in the north.

Comparison with official reports. The research findings indicated a pandemic onset date of December 8, consistent with results from the China CDC that over 96% of cities passed their peak ILI activity before January 2023 (*9*). The results of this study were highly consistent with nucleic acid assay results. Additionally, the peak dates for the Beijing (December 13) and Tianjin (December 17) municipalities aligned with the China CDC reports in December 14 and 19, respectively.

Furthermore, the study observed a sharp increase in search volume starting on December 9 in northern cities and December 11 in southern cities. This increase corresponds with the rise in ILI% reported by the China CDC from 824 sentinel hospitals during the 50th week (December 12–18) (*10*).

Fever clinic visit data, as reported by the China CDC, peaked on December 23, 2022 (*11*). Our study found that the peak occurred from December 18 to 20, 2022, suggesting that the Baidu Index may be a leading indicator of healthcare utilization during an outbreak.

Overall, the China CDC data on ILI percentage and fever clinic visits support the study's findings, validating the methodology and the use of search engine data as a reliable, real-time indicator of pandemic spread and intensity.

Validation

To test our model's effectiveness, we analyzed the 2023 influenza outbreak (January 1 to April 1, 2023) in northern and southern PLADs, Yichang, and Weifang to demonstrate its utility across different regions. We used October to December 2018 as our baseline to account for changes in internet usage. Outbreak onset was determined using an 80th percentile threshold of search volume for seven consecutive days, a method that proved effective across all regions. For peak detection, we used a three-day

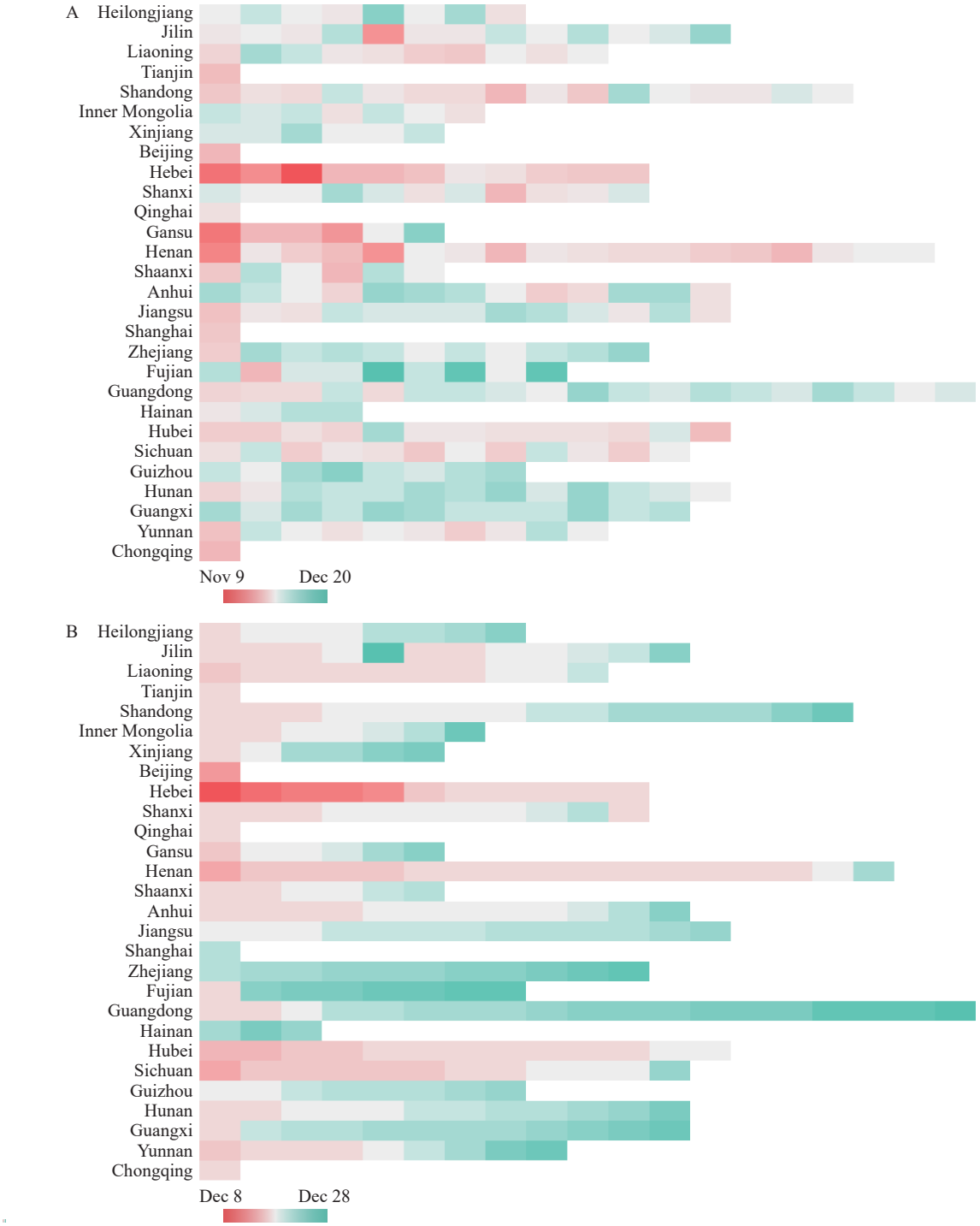


FIGURE 2. Onset dates and peak dates across 255 cities in various PLADs of China. (A) Onset dates evaluation across 255 cities in various PLADs of China; (B) Peak dates evaluation across 255 cities in various PLADs of China.
Note: Each row corresponds to a PLAD, and each colored block represents a city. The color gradient indicates the timing: shades closer to red represent earlier times, while shades closer to green indicate later times.
Abbreviation: PLAD=provincial-level administrative division.

continuous decline in search activity after an initial increase, with calculations based on a three-day moving average (m-value). This approach allowed for rapid, minimal data analysis while accurately capturing the outbreak peak.

Validation Results

The onset dates for Northern and Southern China

were February 20 and February 26, respectively, and February 25 and February 27 for Weifang and Yichang, respectively. Table 1 presents the peak dates for Northern China, Southern China, Yichang, and Weifang.

Comparison with official reports

ILI percentages reported by the China CDC rose in Northern and Southern China during week 9 of 2023 (February 27–March 5), remained high in week 10 (March 6–12), and declined in week 11 (March 13–19). The onset and peak dates for ILI activity in these regions aligned with these results.

Peak dates in Yichang (March 16) and Weifang (March 10) preceded the dates for the highest ILI% (March 29 and March 20, respectively). The median dates for the three highest ILI% (March 26 and March 16) were over 10 and 7 days later, respectively (Figure 3). Additionally, the epidemic outset period reported by the Yichang CDC and Weifang CDC was week 9 in 2023, consistent with our outset dates of February 27 and February 25.

DISCUSSION

This research underscores the vital role of search engine analytics in bolstering public health surveillance and early warning systems. By leveraging internet search data, this study demonstrates the potential for a more nuanced and immediate understanding of disease dynamics, facilitating the early identification of both pandemic outbreaks and seasonal epidemic patterns. Using varied threshold levels, this approach discerned the preliminary and peak phases of disease spread more rapidly and accurately than traditional methods.

As a potential supplement to traditional surveillance systems, internet search data has shown promise in identifying trends and peak timing before official reports (December 12–18, December 23, 2022) (*10*). This earlier detection is attributable to the immediacy of internet data, which circumvents lengthy processing and validation steps required for official reporting, and its ability to reflect real-time shifts in public concern and interest.

Validation in two cities under seasonal influenza scenarios in Northern and Southern China has confirmed that this procedure is extrapolatable for identifying the onset and peak of respiratory infectious diseases. The emergence of novel respiratory pathogens is unpredictable, and traditional surveillance systems often struggle to adapt quickly when a new pathogen spreads rapidly and causes a pandemic. Due to the ready availability of data and the simplicity of the method, this procedure can serve as an alternative option. Moreover, it remains timely and effective in detecting patterns even when epidemiological trends of seasonal respiratory diseases change.

In application, different thresholds may need to be adopted based on the actual conditions in areas with varying population sizes, search behaviors, climates, and epidemiological characteristics of diseases. This study analyzed only 255 cities, excluding those with a continuous Baidu index of zero. This exclusion could be related to the scale of internet users and their online habits. The timeliness of detection in northern cities was earlier than in southern cities, potentially due to the stable seasonal epidemic trends historically observed in the north, typically characterized by a single peak in cases. Additionally, population size might play a role in the timeliness of detection. Weifang has a population of 9.4 million compared to Yichang's 4 million, suggesting that the earlier detection in Weifang might also be related to its larger population size.

This study is subject to some limitations. First, due to the lack of referential data from surveillance systems during the pandemic, this study was unable to validate the results for the 255 cities. Second, as the study aimed to provide a scalable and simple tool, it did not account for other factors that could affect Baidu searches, which may impact the accuracy of the results.

TABLE 1. Evaluation and comparison of criteria for reaching the peak.

Region	m=3		m=7		m=31	
	Excess search index	Peak date	Excess search index	Peak date	Excess search index	Peak date
Northern PLADs	36.69	March 10	39.37	March 16	42.11	March 13
Southern PLADs	24.00	March 14	30.70	March 09	24.07	March 14
Yichang	39.87	March 16	36.75	March 10	39.04	March 16
Weifang	32.46	March 10	24.09	March 14	30.66	March 9

Abbreviation: PLADs=provincial-level administrative divisions.

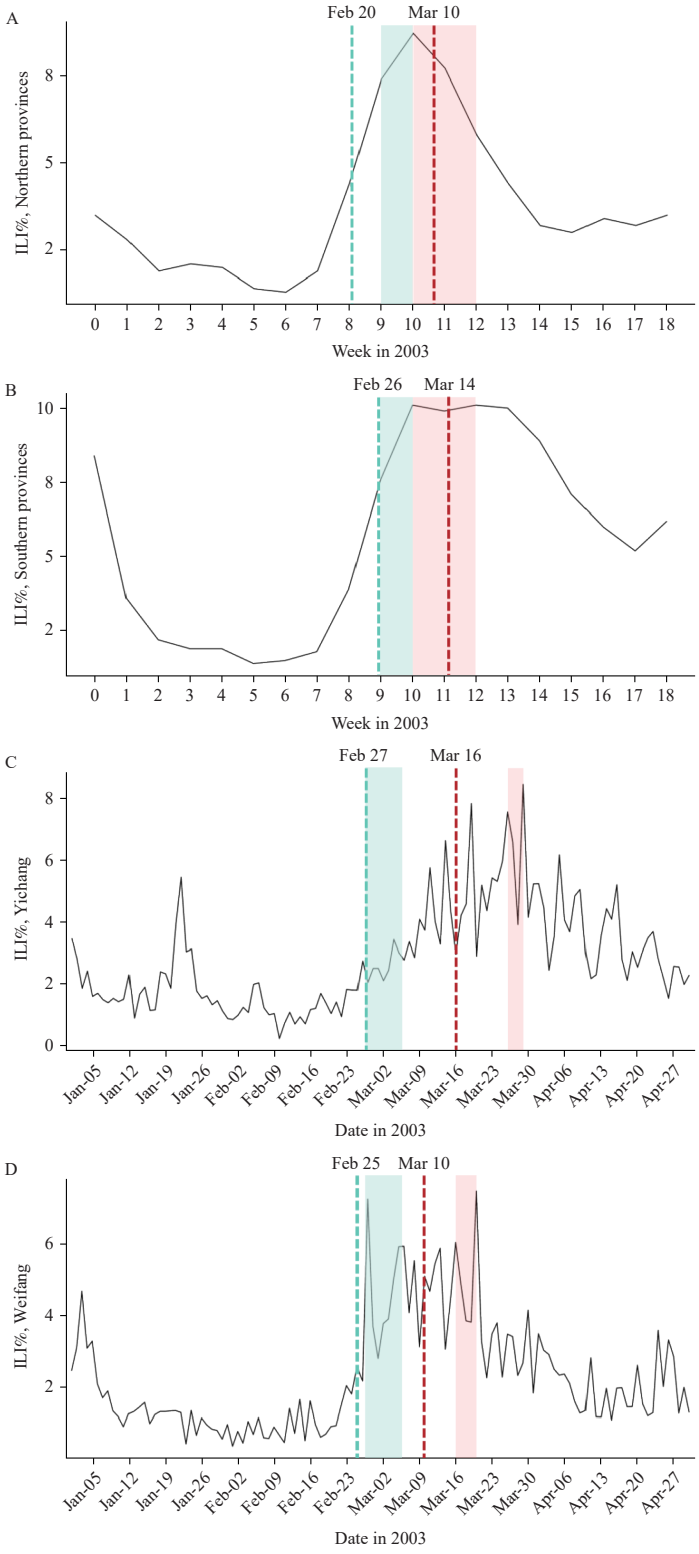


FIGURE 3. Onset and peak dates of the influenza epidemic in 2023. (A) Onset and peak dates evaluation in northern PLADs. (B) Onset and peak dates evaluation in southern PLADs. (C) Onset and peak dates evaluation in Yichang. (D) Onset and peak dates evaluation in Yichang.

Note: The green dotted line represents the outset date identified in this study, while the green block indicates the corresponding officially reported week. The red dotted line signifies the estimated peak time of the epidemic based on this study, and the red block represents the officially reported week for the peak.

Abbreviation: PLAD=provincial-level administrative division.

Conflicts of interest: No conflicts of interest.

Acknowledgements: Baidu for the data publication and Sinosoft Company Limited for technical support.

Funding: Supported by grants from the CAMS Innovation Fund for Medical Sciences (2021-I2M-1-044, 2023-I2M-3-011) and National Key Research and Development Program of China (2023YFC2308701).

doi: 10.46234/ccdcw2024.195

[#] Corresponding authors: Weizhong Yang, yangweizhong@cams.cn; Chen Wang, cyh-birm@263.net.

[1] School of Health Policy and Management, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China; [2] School of Population Medicine and Public Health, Chinese Academy of Medical Sciences (CAMS) & Peking Union Medical College, Beijing, China; [3] Yichang Center for Disease Prevention and Control, Yichang City, Hubei Province, China; [4] Weifang Center for Disease Prevention and Control, Weifang City, Shandong Province, China; [5] Beijing Center for Disease Prevention and Control, Beijing, China; [6] School of Public Health, Dali University, Dali City, Yunnan Province, China; [7] Shanghai Pudong New Area Center for Disease Control and Prevention, Shanghai, China.

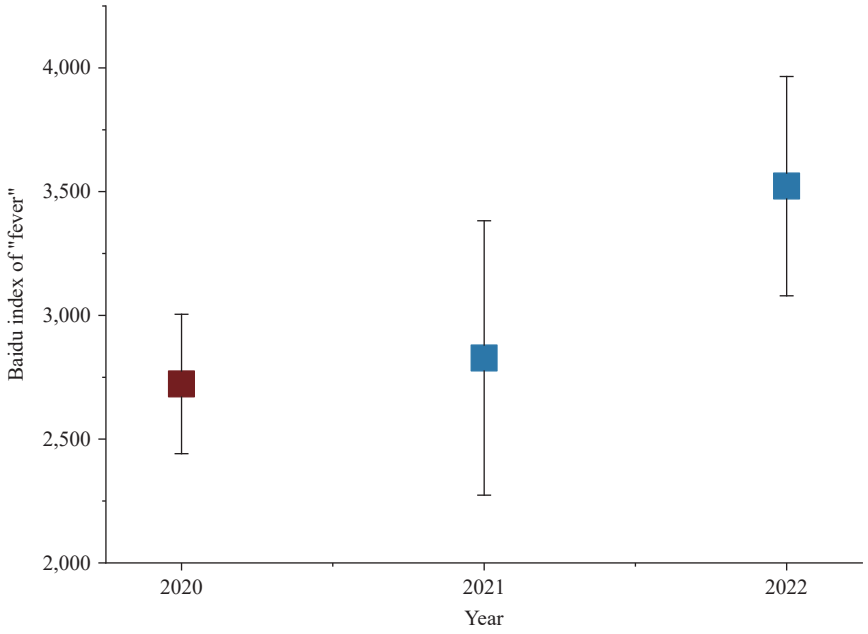
[&] Joint first authors.

Submitted: February 05, 2024; Accepted: September 06, 2024

REFERENCES

1. Nii-Trebi NI. Emerging and neglected infectious diseases: insights, advances, and challenges. BioMed Res Int 2017;2017:5245021. https://doi.org/10.1155/2017/5245021.
2. Ellwanger JH, Kaminski VDL, Chies JAB. Emerging infectious disease prevention: where should we invest our resources and efforts? J Infect Public Health 2019;12(3):313-6. http://dx.doi.org/10.1016/j.jiph.2019.03.010.
3. Ma SM, Yang SH. COVID-19 forecasts using Internet search information in the United States. Sci Rep 2022;12(1):11539. https://doi.org/10.1038/s41598-022-15478-y.
4. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009;457(7232):1012 – 4. https://doi.org/10.1038/nature07634.
5. Liang F, Guan P, Wu W, Huang DS. Forecasting influenza epidemics by integrating internet search queries and traditional surveillance data with the support vector machine regression model in Liaoning, from 2011 to 2015. PeerJ 2018;6:e5134. https://doi.org/10.7717/peerj.5134.
6. Gong X, Han YY, Hou MC, Guo R. Online public attention during the early days of the COVID-19 pandemic: infoveillance study based on Baidu Index. JMIR Public Health Surveill 2020;6(4):e23098. https://doi.org/10.2196/23098.
7. Barros JM, Duggan J, Rebholz-Schuhmann D. The application of internet-based sources for public health surveillance (Infoveillance): systematic review. J Med Internet Res 2020;22(3):e13680. https://doi.org/10.2196/13680.
8. Xinhua. China Focus: China releases measures to optimize COVID-19 response; https://english.news.cn/20221111/d4399114a082438eaac32d08a02bf58d/c.html. [2024-8-30].
9. China CDC Weekly. Epidemic situation of novel coronavirus infection in China.https://www.chinacdc.cn/jkzt/crb/zl/szkb_11803/jszl_13141/202302/t20230201_263576.html[2024-8-30].
10. Lu J, Lin AR, Jiang CM, Zhang AM, Yang ZZ. Influence of transportation network on transmission heterogeneity of COVID-19 in China. Transp Res Part C Emerg Technol 2021;129:103231. https://doi.org/10.1016/j.trc.2021.103231.
11. Xiang W, Chen L, Yan XD, Wang B, Liu XB. The impact of traffic control measures on the spread of COVID-19 within urban agglomerations based on a modified epidemic model. Cities 2023;135:104238. https://doi.org/10.1016/j.cities.2023.104238.

SUPPLEMENTARY MATERIAL



SUPPLEMENTARY FIGURE S1. Comparison of average Baidu index in different years.
Note: Red refers to the smallest mean and standard deviation.

SUPPLEMENTARY TABLE S1. Comparison of m-value in peak date and comparison peak date in different criteria for reaching the peak

City	Comparison of m-value in peak date			Comparison peak date in different criteria for reaching the peak	
	Peak date m=3	Peak date m=7	Peak date m=31	Peak date 3 days decrease	Peak date 5 days decrease
Beijing	2022/12/13	2022/12/13	2022/12/12	2022/12/13	2022/12/12
Tianjin	2022/12/16	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Shijiazhuang	2022/12/14	2022/12/9	2022/12/17	2022/12/9	2022/12/17
Hengshui	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Zhangjiakou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Chengde	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Qinhuangdao	NA	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Langfang	2022/12/13	2022/12/11	2022/12/12	2022/12/11	NA
Cangzhou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Baoding	2022/12/11	2022/12/10	2022/12/11	2022/12/10	2022/12/11
Tangshan	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Handan	2022/12/9	2022/12/12	2022/12/17	2022/12/12	2022/12/17
Xingtai	2022/12/10	2022/12/8	2022/12/8	2022/12/8	2022/12/8
Taiyuan	2022/12/19	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Datong	2022/12/17	2022/12/17	2022/12/17	2022/12/17	NA
Changzhi	2022/12/19	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Xinzhou	2022/12/17	2022/12/23	2022/12/23	2022/12/23	2022/12/23
Jinzhong	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Linfen	2022/12/22	2022/12/22	2022/12/22	2022/12/22	NA
Jincheng	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17

Continued

City	Comparison of m-value in peak date			Comparison peak date in different criteria for reaching the peak	
	Peak date m=3	Peak date m=7	Peak date m=31	Peak date 3 days decrease	Peak date 5 days decrease
Shuozhou	NA	2022/12/17	2022/12/17	2022/12/17	NA
Yangquan	NA	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Lvliang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Huhehaote	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Baotou	2022/12/17	2022/12/17	2022/12/18	2022/12/17	2022/12/18
Erdos	NA	2022/12/21	2022/12/17	2022/12/21	2022/12/17
Xilinguolemeng	NA	2022/12/25	2022/12/18	2022/12/25	NA
Chifeng	2022/12/17	2022/12/17	2022/12/17	2022/12/17	NA
Tongliao	NA	2022/12/18	2022/12/18	2022/12/18	2022/12/18
Hulunbeier	2022/12/26	2022/12/26	NA	2022/12/26	NA
Shenyang	2022/12/17	2022/12/16	2022/12/16	2022/12/16	2022/12/16
Dalian	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Panjin	NA	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Anshan	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Chaoyang	NA	2022/12/21	2022/12/17	2022/12/21	NA
Jinzhou	2022/12/17	2022/12/17	2022/12/16	2022/12/17	2022/12/16
Tieling	NA	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Dandong	2022/12/16	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Fushun	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Huludao	NA	2022/12/20	2022/12/17	2022/12/20	2022/12/17
Changchun	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Huainan	2022/12/22	2022/12/18	2022/12/18	2022/12/18	2022/12/18
Lu'an	2022/12/25	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Chuzhou	2022/12/18	2022/12/17	2022/12/18	2022/12/17	2022/12/18
Huabei	NA	2022/12/24	2022/12/24	2022/12/24	2022/12/24
Fuyang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Anqing	2022/12/17	2022/12/18	2022/12/17	2022/12/18	2022/12/17
Bengbu	2022/12/17	2022/12/17	2022/12/17	2022/12/17	NA
Wuhu	2022/12/18	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Bozhou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Fuzhou	2022/12/26	2022/12/25	2022/12/25	2022/12/25	2022/12/25
Putian	2022/12/22	2022/12/28	2022/12/28	2022/12/28	2022/12/28
Sanming	2022/12/27	NA	2022/12/27	NA	NA
Longyan	2022/12/24	2022/12/26	2022/12/27	2022/12/26	2022/12/27
Xiamen	2022/12/25	2022/12/26	2022/12/25	2022/12/26	2022/12/25
Quanzhou	2022/12/26	2022/12/27	2022/12/26	2022/12/27	2022/12/26
Zhangzhou	2022/12/23	2022/12/24	2022/12/28	2022/12/24	2022/12/28
Ningde	2022/12/22	2022/12/25	2022/12/25	2022/12/25	2022/12/25
Nanping	2022/12/16	2022/12/16	2022/12/16	2022/12/16	NA
Nanchang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Jiujiang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17

Continued

City	Comparison of m-value in peak date			Comparison peak date in different criteria for reaching the peak	
	Peak date m=3	Peak date m=7	Peak date m=31	Peak date 3 days decrease	Peak date 5 days decrease
Fuzhou	2022/12/21	2022/12/21	2022/12/21	2022/12/21	2022/12/21
Shangrao	2022/12/25	2022/12/23	2022/12/21	2022/12/23	2022/12/21
Ganzhou	2022/12/22	2022/12/17	2022/12/22	2022/12/17	2022/12/22
Ji'an	2022/12/19	2022/12/19	2022/12/18	2022/12/19	NA
Pingxiang	NA	NA	2022/12/18	NA	NA
Yichun	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Jinan	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Binzhou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Qingdao	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Yantai	2022/12/18	2022/12/18	2022/12/18	2022/12/18	2022/12/18
Linyi	2022/12/22	2022/12/22	2022/12/21	2022/12/22	2022/12/21
Weifang	2022/12/20	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Zibo	2022/12/21	2022/12/20	2022/12/17	2022/12/20	2022/12/17
Dongying	2022/12/22	2022/12/23	2022/12/17	2022/12/23	NA
Liaocheng	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Heze	2022/12/17	2022/12/21	2022/12/17	2022/12/21	2022/12/17
Zaozhuang	2022/12/25	2022/12/25	2022/12/25	2022/12/25	2022/12/25
Dezhou	2022/12/20	2022/12/21	2022/12/20	2022/12/21	2022/12/20
Weihai	2022/12/26	2022/12/26	2022/12/26	2022/12/26	NA
Jining	2022/12/21	2022/12/21	2022/12/21	2022/12/21	2022/12/21
Taian	2022/12/18	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Rizhao	2022/12/19	2022/12/17	2022/12/19	2022/12/17	2022/12/19
Zhengzhou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Nanyang	2022/12/17	2022/12/16	2022/12/17	2022/12/16	2022/12/17
Xinxiang	2022/12/16	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Kaifeng	2022/12/14	2022/12/14	2022/12/17	2022/12/14	2022/12/17
Jiaozuo	2022/12/18	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Pingdingshan	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Xuchang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	NA
Anyang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Zhumadian	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Xinyang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Hebi	NA	2022/12/17	2022/12/16	2022/12/17	2022/12/16
Zhoukou	2022/12/19	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Shangqiu	2022/12/16	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Luoyang	2022/12/13	2022/12/13	2022/12/17	2022/12/13	2022/12/17
Luohe	2022/12/21	2022/12/23	2022/12/17	2022/12/23	NA
Puyang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Sanmenxia	2022/12/14	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Jiyuan	NA	NA	2022/12/17	NA	2022/12/17
Wuhan	2022/12/14	2022/12/14	2022/12/16	2022/12/14	2022/12/16

Continued

City	Comparison of m-value in peak date			Comparison peak date in different criteria for reaching the peak	
	Peak date m=3	Peak date m=7	Peak date m=31	Peak date 3 days decrease	Peak date 5 days decrease
Huangshi	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Jingzhou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Xiangyang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Huanggang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Jingmen	2022/12/16	2022/12/16	2022/12/17	2022/12/16	2022/12/17
Yichang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Shiyan	2022/12/15	2022/12/16	2022/12/15	2022/12/16	2022/12/15
Suizhou	NA	2022/12/17	2022/12/17	2022/12/17	NA
Enshi	2022/12/11	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Ezhou	NA	2022/12/15	2022/12/16	2022/12/15	NA
Xianning	2022/12/20	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Xiaogan	2022/12/14	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Changsha	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Yueyang	2022/12/20	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Hengyang	2022/12/18	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Zhuzhou	NA	2022/12/21	2022/12/21	2022/12/21	NA
Xiangtan	NA	2022/12/26	2022/12/17	2022/12/26	NA
Yiyang	2022/12/20	2022/12/20	2022/12/20	2022/12/20	NA
Chenzhou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Xiangxi	NA	NA	NA	NA	NA
Loudi	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Huaihua	NA	2022/12/20	2022/12/20	2022/12/20	2022/12/20
Changde	2022/12/22	2022/12/23	2022/12/23	2022/12/23	2022/12/23
Yongzhou	2022/12/17	2022/12/17	2022/12/20	2022/12/17	NA
Shaoyang	2022/12/24	2022/12/22	2022/12/22	2022/12/22	2022/12/22
Guangzhou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Shenzhen	2022/12/21	2022/12/22	2022/12/21	2022/12/22	2022/12/21
Dongguan	2022/12/21	2022/12/22	2022/12/22	2022/12/22	2022/12/22
Yunfu	2022/12/20	2022/12/27	2022/12/27	2022/12/27	NA
Foshan	2022/12/20	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Zhanjiang	2022/12/21	2022/12/21	2022/12/21	2022/12/21	NA
Jiangmen	2022/12/25	2022/12/25	2022/12/24	2022/12/25	2022/12/24
Bazhong	NA	2022/12/18	2022/12/18	2022/12/18	2022/12/18
Neijiang	2022/12/19	2022/12/18	2022/12/17	2022/12/18	2022/12/17
Luzhou	NA	2022/12/16	2022/12/17	2022/12/16	2022/12/17
Deyang	2022/12/17	2022/12/16	2022/12/16	2022/12/16	2022/12/16
Leshan	NA	2022/12/19	2022/12/17	2022/12/19	2022/12/17
Zigong	2022/12/15	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Ya'an	NA	2022/12/23	NA	2022/12/23	NA
Meishan	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Liangshan	2022/12/16	2022/12/16	2022/12/17	2022/12/16	NA

Continued

City	Comparison of m-value in peak date			Comparison peak date in different criteria for reaching the peak	
	Peak date m=3	Peak date m=7	Peak date m=31	Peak date 3 days decrease	Peak date 5 days decrease
Guiyang	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Qiannan	2022/12/20	2022/12/20	2022/12/20	2022/12/20	NA
Zunyi	2022/12/21	2022/12/20	2022/12/17	2022/12/20	2022/12/17
Qiandongnan	2022/12/18	2022/12/18	2022/12/18	2022/12/18	2022/12/18
Tongren	2022/12/23	2022/12/23	2022/12/20	2022/12/23	2022/12/20
Anshun	2022/12/20	2022/12/19	2022/12/20	2022/12/19	2022/12/20
Bijie	2022/12/25	2022/12/23	2022/12/24	2022/12/23	NA
Qianxinan	NA	2022/12/21	2022/12/22	2022/12/21	2022/12/22
Kunming	2022/12/13	2022/12/17	2022/12/16	2022/12/17	2022/12/16
Yuxi	2022/12/19	2022/12/19	2022/12/19	2022/12/19	NA
Chuxiong	2022/12/17	2022/12/19	2022/12/17	2022/12/19	2022/12/17
Dali	2022/12/26	2022/12/26	2022/12/26	2022/12/26	2022/12/26
Zhaotong	NA	2022/12/23	2022/12/16	2022/12/23	2022/12/16
Honghe	2022/12/16	2022/12/16	2022/12/16	2022/12/16	NA
Qujing	2022/12/16	2022/12/17	2022/12/17	2022/12/17	NA
Lincang	NA	2022/12/26	2022/12/18	2022/12/26	NA
Pu'er	2022/12/27	2022/12/17	2022/12/17	2022/12/17	NA
Dehong	NA	NA	2022/12/17	NA	NA
Xi'an	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Ankang	NA	2022/12/21	2022/12/17	2022/12/21	2022/12/17
Baoji	NA	2022/12/17	2022/12/17	2022/12/17	NA
Jilin	2022/12/17	2022/12/18	2022/12/18	2022/12/18	2022/12/18
Tonghua	NA	2022/12/17	2022/12/17	2022/12/17	NA
Baishan	NA	2022/12/28	2022/12/20	2022/12/28	NA
Yanbian	2022/12/15	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Harbin	2022/12/18	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Daqing	2022/12/20	2022/12/21	2022/12/21	2022/12/21	2022/12/21
Qiqihar	2022/12/22	2022/12/22	2022/12/21	2022/12/22	2022/12/21
Jiamusi	2022/12/24	2022/12/21	2022/12/19	2022/12/21	2022/12/19
Mudanjiang	NA	2022/12/22	2022/12/22	2022/12/22	NA
Jixi	NA	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Suihua	NA	2022/12/25	2022/12/17	2022/12/25	NA
Shuangyashan	NA	2022/12/17	2022/12/20	2022/12/17	NA
Shanghai	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Nanjing	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Suzhou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Wuxi	2022/12/21	2022/12/21	2022/12/21	2022/12/21	2022/12/21
Lianyungang	2022/12/22	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Huai'an	2022/12/20	2022/12/20	2022/12/20	2022/12/20	2022/12/20
Yangzhou	2022/12/23	2022/12/22	2022/12/22	2022/12/22	2022/12/22
Taizhou	2022/12/21	2022/12/21	2022/12/21	2022/12/21	2022/12/21

Continued

City	Comparison of m-value in peak date			Comparison peak date in different criteria for reaching the peak	
	Peak date m=3	Peak date m=7	Peak date m=31	Peak date 3 days decrease	Peak date 5 days decrease
Yancheng	2022/12/21	2022/12/21	2022/12/21	2022/12/21	2022/12/21
Xuzhou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Changzhou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Nantong	2022/12/18	2022/12/23	2022/12/23	2022/12/23	2022/12/23
Zhenjiang	2022/12/22	2022/12/18	2022/12/22	2022/12/18	NA
Suqian	2022/12/18	2022/12/18	2022/12/21	2022/12/18	2022/12/21
Hangzhou	2022/12/21	2022/12/21	2022/12/21	2022/12/21	2022/12/21
Lishui	NA	2022/12/27	2022/12/27	2022/12/27	NA
Jinhua	2022/12/27	2022/12/24	2022/12/22	2022/12/24	2022/12/22
Wenzhou	2022/12/24	2022/12/22	2022/12/22	2022/12/22	2022/12/22
Taizhou	NA	2022/12/21	2022/12/24	2022/12/21	2022/12/24
Quzhou	2022/12/26	2022/12/27	2022/12/17	2022/12/27	2022/12/17
Ningbo	2022/12/23	2022/12/22	2022/12/22	2022/12/22	2022/12/22
Shaoxing	2022/12/24	2022/12/24	2022/12/24	2022/12/24	2022/12/24
Jiaxing	2022/12/21	2022/12/22	2022/12/22	2022/12/22	2022/12/22
Huzhou	2022/12/18	2022/12/25	2022/12/24	2022/12/25	NA
Zhoushan	NA	2022/12/25	2022/12/25	2022/12/25	NA
Hefei	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Tongling	NA	NA	2022/12/21	NA	2022/12/21
Huangshan	NA	2022/12/17	2022/12/17	2022/12/17	NA
Xuancheng	2022/12/19	2022/12/21	2022/12/21	2022/12/21	2022/12/21
Hanzhong	2022/12/20	2022/12/19	2022/12/20	2022/12/19	NA
Xianyang	2022/12/18	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Yan'an	2022/12/17	2022/12/17	2022/12/17	2022/12/17	NA
Lanzhou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Dingxi	NA	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Pingliang	2022/12/23	2022/12/23	2022/12/23	2022/12/23	NA
Tianshui	2022/12/23	2022/12/23	2022/12/17	2022/12/23	NA
Longnan	NA	2022/12/18	2022/12/18	2022/12/18	2022/12/18
Linxia	NA	2022/12/21	2022/12/16	2022/12/21	2022/12/16
Xining	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Yinchuan	NA	NA	2022/12/17	NA	2022/12/17
Shizuishan	NA	2022/12/24	2022/12/22	2022/12/24	2022/12/22
Urumqi	2022/12/18	2022/12/17	2022/12/18	2022/12/17	2022/12/18
Changji	2022/12/21	2022/12/22	2022/12/22	2022/12/22	2022/12/22
Aksu	NA	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Kashgar	NA	2022/12/17	2022/12/17	2022/12/17	NA
Bayin Goleng	2022/12/25	2022/12/25	2022/12/25	2022/12/25	2022/12/25
Yili	2022/12/26	2022/12/22	2022/12/17	2022/12/22	2022/12/17
Tacheng	NA	2022/12/24	2022/12/21	2022/12/24	2022/12/21
Huizhou	2022/12/23	2022/12/22	2022/12/23	2022/12/22	2022/12/23

Continued

City	Comparison of m-value in peak date			Comparison peak date in different criteria for reaching the peak	
	Peak date m=3	Peak date m=7	Peak date m=31	Peak date 3 days decrease	Peak date 5 days decrease
Zhuhai	2022/12/23	2022/12/23	2022/12/23	2022/12/23	2022/12/23
Shaoguan	2022/12/24	2022/12/24	2022/12/24	2022/12/24	NA
Yangjiang	2022/12/28	2022/12/26	2022/12/26	2022/12/26	NA
Maoming	2022/12/25	2022/12/26	2022/12/27	2022/12/26	NA
Chaozhou	2022/12/25	2022/12/27	2022/12/24	2022/12/27	2022/12/24
Zhongshan	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Qingyuan	2022/12/26	2022/12/24	2022/12/24	2022/12/24	2022/12/24
Zhaoqing	2022/12/24	2022/12/24	2022/12/24	2022/12/24	2022/12/24
Heyuan	2022/12/20	2022/12/22	2022/12/17	2022/12/22	NA
Meizhou	NA	NA	NA	NA	NA
Shantou	2022/12/25	2022/12/25	2022/12/25	2022/12/25	2022/12/25
Shanwei	2022/12/28	2022/12/28	2022/12/28	2022/12/28	NA
Nanning	2022/12/21	2022/12/21	2022/12/17	2022/12/21	2022/12/17
Liuzhou	2022/12/22	2022/12/22	2022/12/22	2022/12/22	2022/12/22
Guilin	2022/12/20	2022/12/20	2022/12/17	2022/12/20	2022/12/17
Hezhou	2022/12/26	2022/12/26	2022/12/25	2022/12/26	2022/12/25
Guigang	2022/12/17	2022/12/17	2022/12/24	2022/12/17	2022/12/24
Yulin	2022/12/21	2022/12/22	2022/12/25	2022/12/22	NA
Hechi	NA	2022/12/22	2022/12/17	2022/12/22	2022/12/17
Beihai	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Qinzhou	2022/12/22	2022/12/22	2022/12/25	2022/12/22	2022/12/25
Baise	2022/12/25	2022/12/25	2022/12/25	2022/12/25	NA
Wuzhou	NA	2022/12/23	2022/12/23	2022/12/23	2022/12/23
Chongzuo	2022/12/23	2022/12/23	2022/12/22	2022/12/23	2022/12/22
Haikou	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Sanya	2022/12/19	2022/12/22	2022/12/22	2022/12/22	NA
Chongqing	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Chengdu	2022/12/16	2022/12/16	2022/12/16	2022/12/16	2022/12/16
Yibin	2022/12/17	2022/12/17	2022/12/17	2022/12/17	2022/12/17
Mianyang	2022/12/12	2022/12/14	2022/12/15	2022/12/14	2022/12/15

Methods and Applications

Establishment of a Lateral Flow Dipstick Detection Method for Influenza A Virus Based on CRISPR/Cas12a System

Xiaoyan Zhao[1]; Ximing Zheng[1]; Xiyong Yang[1]; Qi Guo[2]; Yi Zhang[3,#]; Jun Lou[1,#]

ABSTRACT

Objective: This study aimed to develop a rapid, visual PCR-CRISPR/Cas12-LFD method for detecting influenza A by utilizing the conserved region of the matrix protein gene.

Method: We crafted universal degradation primers and clustered regularly interspaced short palindromic repeats RNA (CRISPR RNA, crRNA) targeting the conserved matrix protein gene of the influenza virus (IFV), integrated with lateral flow dipstick (LFD) technology. This new PCR-CRISPR/Cas12-LFD approach was designed to determine its sensitivity and specificity through the analysis of various clinical samples collected in 2023.

Results: The developed nucleic acid assay for influenza A viruses (IAV) demonstrated a sensitivity of 10 copies/μL without cross-reactivity with other respiratory pathogens. Evaluation of 82 clinical samples showed high concordance with results from fluorescent Polymerase Chain Reaction (PCR), achieving a kappa value of 0.95.

Conclusion: A highly sensitive and specific PCR-CRISPR/Cas12-LFD method has been successfully established for the detection of influenza A, offering a robust tool for its diagnosis and aiding in the prevention and control of this virus.

Influenza (flu) is an acute respiratory illness caused by the influenza virus (IFV), characterized by symptoms such as fever, cough, headache, and muscle pain, posing a significant threat to public health (*1*). The virus falls within the Orthomyxoviridae family, known for single-stranded, negative-sense RNA viruses divided into four subtypes — A, B, C, and D (*2*). The influenza A virus (IAV) comprises eight distinct RNA segments and is further subclassified into 16 HA and 9 NA subtypes. The high mutation rates of HA and NA facilitate recombination and reassortment events, significantly contributing to IAV's potential to cause recurring seasonal outbreaks and pandemics (*3*–*4*). Recent surveillance data from the China CDC indicates that IAV was responsible for 86.1% of influenza cases in China in 2023 (https://ivdc.chinacdc.cn/cnic/). The ongoing COVID-19 pandemic has further exacerbated the incidence of clustered influenza A outbreaks since 2023, underscoring the urgent need for enhanced influenza prevention and control, notably through early diagnosis and continuous monitoring (*5*).

Molecular diagnosis is crucial for detecting influenza A (*6*). Isothermal methods, characterized by longer primer lengths, often suffer from non-specific amplification, leading to false-negative outcomes. Conversely, gene chips and sequencing require highly skilled personnel, limiting their utility in standard laboratory settings (*7*). Polymerase Chain Reaction (PCR) remains the preferred method due to its high sensitivity, specificity, accuracy, and ease of use, establishing it as the "gold standard" in molecular diagnostics. The CRISPR/Cas system, a defense mechanism in bacteria and archaea against phages, includes CRISPR and associated proteins (Cas). This system is categorized into two types and six subtypes based on its components and functions (*8*). Cas12, a member of the second type, subtype five, utilizes crRNA (CRISPER RNA) to locate specific 5'-TTTN-3' sites in the PAM sequence. It achieves precise binding via the complementarity between crRNA and the target sequence, activating Cas12's cleavage activity to indiscriminately cut trans-acting single-stranded DNA (ssDNA). Exploiting this mechanism, ssDNA probes, labeled with fluorescent signals and biotin, can be tailored for nucleic acid detection (*9*).

In this study, we developed a PCR-CRISPR/Cas12-LFD method for detecting IAV by designing amplification primers and crRNAs targeting the M gene, as characterized in recent GeneBank entries. This approach utilized dual-labeled ssDNAs — modified with fluorescent FAM and biotin motifs — as probes, integrated with lateral flow dipstick (LFD) technology

(Figure 1). We further analyzed the sensitivity and specificity of this method and validated it using clinical samples.

METHOD

Primer and crRNA Design and Synthesis

The M gene sequences from 126 IAVs, encompassing various subtypes such as H3N2, H1N1, H1N2, H5N1, H5N6, H5N8, H7N7, H7N9, H9N2, H10N3, among others, were retrieved from the NCBI GenBank database and analyzed individually. Sequence alignment was conducted using VectorNTI software, focusing on identifying highly homologous conserved sequences while avoiding the formation of secondary and hairpin structures. To target these conserved regions, multiple sets of primers and probes were designed using Primer 5 software and assessed with Oligo software, incorporating degenerate bases to accommodate variations among sequences. Specifically, the M gene of the H3N2 subtype (GenBank: OR865615.1) was utilized to design three sets of primers for both forward and reverse orientations, as detailed in Table 1. Additionally, for crRNA design, the reverse 20 nucleotide segment of the PAM sequence (TTTN or AAAN) was selected from the amplified targets, with a backbone sequence of UAAUUUCUACUAAGUGUAGAU, also listed in Table 1. The synthesis of all primers and crRNA sequences was performed by Shanghai Sangong Biotechnology.

PCR Reaction and Primer Selection

The M gene was cloned into the pUC57 plasmid

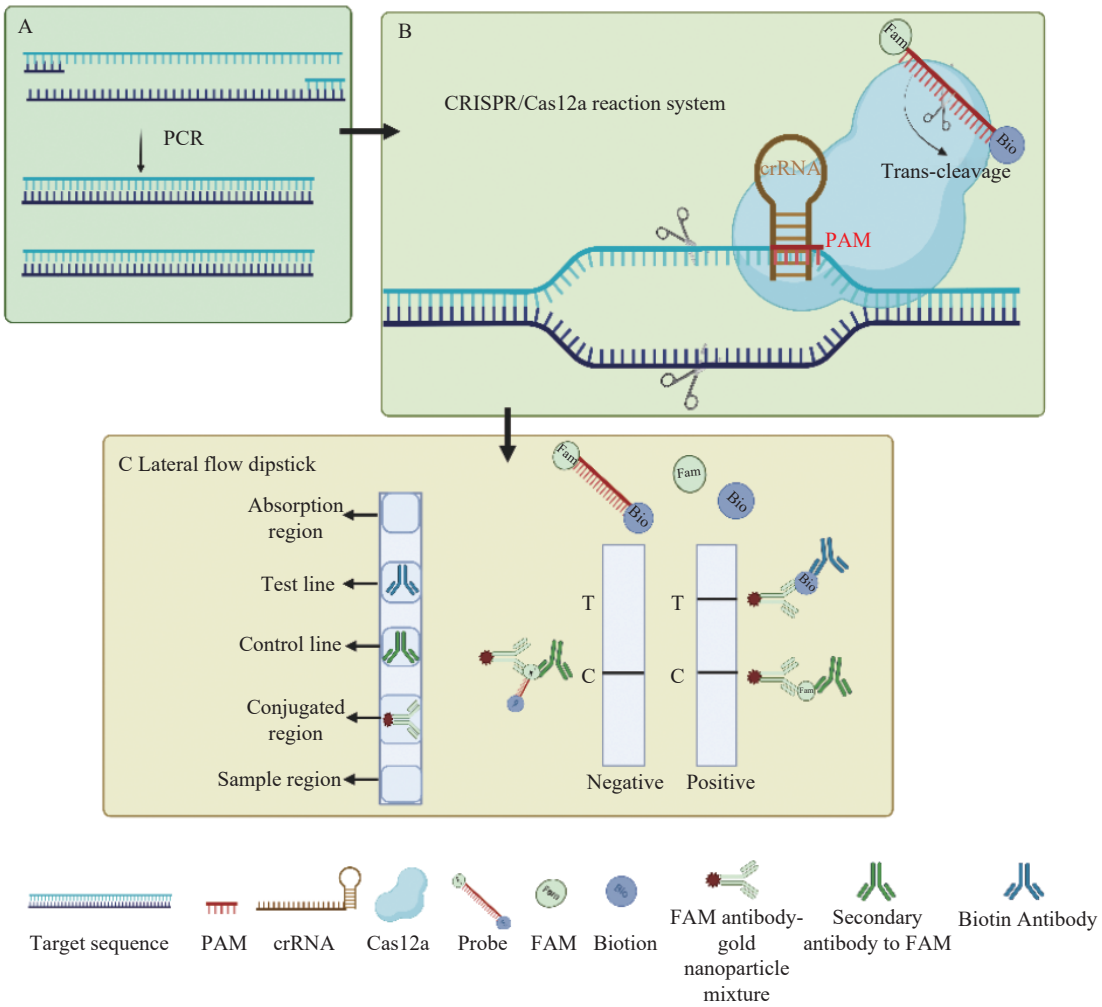


FIGURE 1. Schematic diagram of the PCR-CRISPR/Cas12a-LFD technology. (A) PCR amplification process; (B) the CRISPR/Cas12a and crRNA reaction system; (C) detection of reaction products using LFD.
Abbreviation: LFD=lateral flow dipstick; crRNA= CRISPER RNA.

TABLE 1. List of the primers and crRNA sequences.

Primer/crRNA	Sequence (5'–3')	Position
F1	GAGGTCGAAACGTATGTTCT	41–60
R1	TCACTGGGCACGGTGAGCGT	218–237
F2	ACCGAGGTCGAAACGTATGT	38–57
R2	CGTCTACGCTGCAGTCCCCG	239–258
F3	CTTCTTACCGAGGTCGAAAC	32–51
R3	CGTAGACGCTTTGTCCAAAA	251–270
crRNA-1	UAAUUUCUACUAAGUGUAGAUCUGGGAAAAACACAGAUCUU	98–117
crRNA-2	UAAUUUCUACUAAGUGUAGAUCCAGCAAAGACAUCUUCAAG	82–101

Abbreviation: crRNA=CRISPER RNA.

vector, and the synthesis was carried out by Shanghai Sangong Biotechnology. To amplify the cloned plasmid containing the target gene, three pairs of primers were utilized. The optimal primer pairs were determined through PCR. The composition of the PCR mixture was as follows: 25 μL of 2×*EasyTaq*® PCR SuperMix, 1 μL of 10 μmol/L forward primer, 1 μL of 10 μmol/L reverse primer, 2 μL of template DNA, and 21 μL of nuclease-free water (Supplementary Material). The PCR conditions included an initial denaturation at 94 °C for 30 seconds; followed by 40 cycles of 94 °C for 30 seconds, 58 °C for 15 seconds, and 72 °C for 10 seconds. The PCR products were analyzed using agarose gel electrophoresis, employing 2 μL of the PCR products for analysis; the electrophoresis was conducted at 120 V for 40 minutes.

CRISPR/Cas12a-LFD Detection and CrRNA Selection

The CRISPR/Cas12a reaction setup was as follows: the reaction mixture contained 2 μL of 10x NEB buffer, 2 μL of 500 nmol/L crRNA, 2 μL of 500 nmol/L Cas12a, and 11 μL of nuclease-free water. This mixture was incubated at 25 °C for 10 min. Subsequently, 1 μL of ssDNA-LFD probe (10 μmol/L, 5'-FAM-AAAAAAAA-Bio-3') and 2 μL of PCR products were added. The mixture was thoroughly mixed by shaking, followed by centrifugation, and then incubated in a 37 °C water bath for an additional 10 minutes. After the incubation, an LFD was inserted into the reaction tube, and results were typically observed within 5–10 minutes. The reactions were conducted separately using crRNA-1 and crRNA-2 to determine the more effective crRNA, as indicated by the intensity of the test strip's color.

PCR-CRISPR/Cas12-LFD Sensitivity Analysis

The plasmid encoding the M gene template was synthesized following the previously mentioned procedures. Nucleic acid concentrations were measured using a Qubit 2.0 fluorometer, and the copy number was calculated using the formula: Concentration (copies/μL) = $[(6.02 \times 10^{23}) \times$ Concentration (ng/μL) $\times 10^{-9}]$ / [DNA length (bp) × 660]. The plasmid was serially diluted in tenfold increments from 10^7 to 10^0 copies/μL. PCR reactions utilized these varying concentrations of plasmids as templates and were analyzed through qPCR and CRISPR/Cas12a diagnostic strips. For each dilution level, 2 μL of the sample was added to the reaction mix, with eight replicates conducted for each dilution.

PCR-CRISPR/Cas12-LFD Specificity Analysis

Nucleic acids from positive samples of common respiratory viruses, including IBV, RSV, HRV, MP, HAdV, and 2019-nCoV, were detected using the established PCR-CRISPR/Cas12-LFD method. Additionally, the specificity of this method was analyzed.

Clinical Detection

Fifty-six IAV-positive clinical samples, along with 26 negative controls, underwent nucleic acid extraction followed by reverse transcription using the specified reagents: 4 μL of 5×ES RT Buffer, 1 μL of Total RNA, 1 μL of 10 mmol/L dNTPs, 1 μL of Anchored Oligo$(dT)_{18}$ Primer, 1 μL of *EasyScript*®, and 12 μL of nuclease-free water. The reverse transcription conditions were set at 42 ℃ for 30 minutes. Subsequently, the samples were analyzed using PCR-

CRISPR/Cas12-LFD and fluorescent PCR methods. The concordance between these two methods was then assessed using kappa statistics.

RESULTS

Establishment of PCR-CRISPR/Cas12-LFD detection.

The plasmid was identified utilizing three primer pairs and two crRNAs targeting the M gene, with results in Figure 2. All primer pairs successfully amplified the target gene. The two crRNAs detected the amplification products. Notably, the F2R2 primer exhibited higher amplification efficiency compared to the other pairs (Figure 2A), and crRNA-2 produced the most robust signal (Figure 2B). Consequently, F2R2 and crRNA-2 were chosen to develop the PCR-CRISPR/Cas12-LFD method.

PCR-CRISPR/Cas12-LFD Sensitivity Analysis

The M gene of IAV underwent further sensitivity analysis and reproducibility evaluation. Eight dilution gradients from 10^7 to 10^0 were established, and detection utilized both the qPCR and the PCR-CRISPR/Cas12-LFD method. The detection limit for the PCR-CRISPR/Cas12-LFD method was identified as 10 copies/μL. Repeated experiments (eight repetitions per dilution) demonstrated consistent results, affirming the method's superior stability (Figure 3).

PCR-CRISPR/Cas12-LFD Specificity Analysis

This study established a PCR-CRISPR/Cas12-LFD method that successfully detected twenty-six common

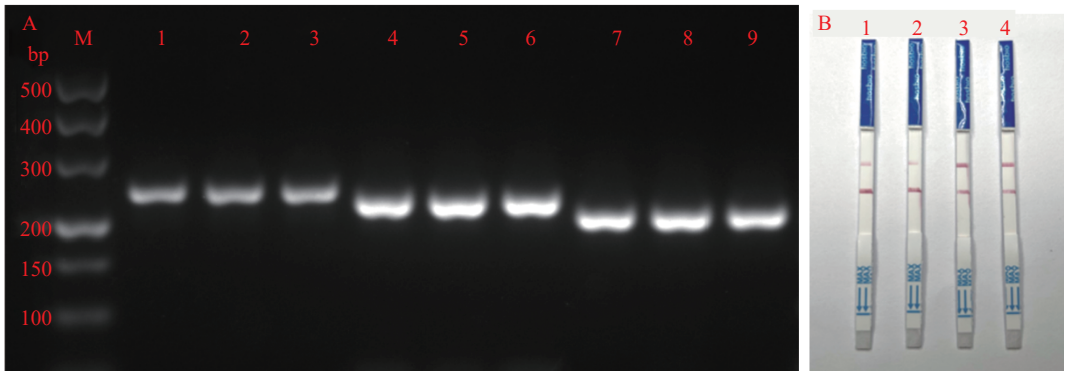


FIGURE 2. Establishment of the PCR-CRISPR/Cas12-LFD Detection Method. (A) Electrophoretic analysis of PCR products with three primer pairs; (B) CRISPR RNA (crRNA) selection using the LFD method.
Note: Panel A, 1–3: three replicated amplicons obtained with F1R1 primers; 4–6: three replicated amplicons using F2R2 primers; 7–9: three replicated amplicons generated by F3R3 primers. Panel B, 1–2: two replicates tested with crRNA-1; 3–4: two replicates tested with crRNA-2.
Abbreviation: LFD=lateral flow dipstick; crRNA= CRISPER RNA.

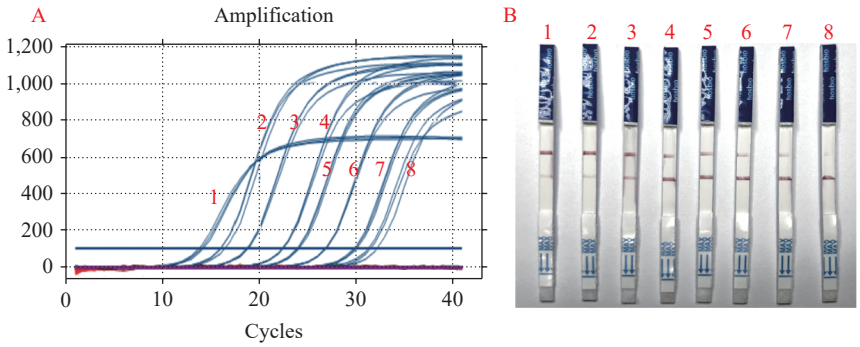


FIGURE 3. Sensitivity analysis of PCR-CRISPR/Cas12-LFD for the detection of IAV. (A) Sensitivity of qPCR in detecting IAV, with three replicates for each of eight dilutions ranging from 10^7 to 10^0; (B) Sensitivity of PCR-CRISPR/Cas12-LFD in detecting IAV, with eight dilutions from 10^7 to 10^0.
Abbreviation: LFD=lateral flow dipstick; IAV=influenza A virus.

respiratory viruses, including IBV, RSV, HRV, MP, HAdV, and 2019-nCoV, without any false positive results, demonstrating the method's high specificity.

Evaluation with Clinical Samples

The practicality of the PCR-CRISPR/Cas12-LFD method developed in this study was assessed using 82 clinical samples, including 56 IAV-positive and 26 IAV-negative samples identified by qPCR. For comparative purposes, the Universal Nucleic Acid Detection Kit for Influenza A Virus from SunSure Biotechnology was used as a control. Results indicated a kappa coefficient of 0.95 between the two methods, indicating identical performance (Table 2).

DISCUSSION

Data show that between November 15, 2023, and December 30, 2023, the incidence of IFV infection among outpatients exhibiting respiratory symptoms in North China, particularly in the Beijing area, ranged from 20% to 50%. Influenza is expected to continue its prevalence across different regions, potentially leading to localized outbreaks. It is anticipated that the IAV, specifically strains H3N2 and H1N1, will remain predominant in these pandemic occurrences (https://www.nmdc.cn/). Concurrently, other respiratory pathogens such as RSV and mycoplasma infections are also imposing significant public health challenges (*10*–*11*). Accurate and fast identification of these pathogens is crucial for effective disease prevention and control.

The CRISPR/Cas system, renowned for its high precision and ease of use, has evolved swiftly into an effective tool for pathogen detection (*12*). Initial research primarily concentrated on the CRISPR/Cas9 system, yet the inherent limitations associated with the Cas9 protein, such as its size, have restricted broader applications of this technology (*13*). Recently, the CRISPR/Cas12 system, also known as CRISPR/Cpf1, has emerged as a focal area of interest. This system falls under the V type category and includes Cas proteins ranging from Cas12a to Cas12e, with Cas12a being more extensively utilized in nucleic acid detection. This interaction activates the RuvC structural domain of the protein for DNA cleavage and initiates trans-cleavage activity, a distinct feature of the CRISPR/Cas12 system (*14*–*15*). The CRISPR/Cas12 system offers advantages such as increased reaction speed, lower off-target rates, and higher editing efficiency, making it particularly suitable for the precise identification of pathogens (*16*). Moreover, leveraging the trans-cleavage activity of Cas proteins, CRISPR-Cas has been integrated with photoelectrochemical biosensors for high-throughput detection (*17*). Additionally, the high sensitivity of the CRISPR-Cas system allows for the direct detection of pathogens using electrochemical methods without the need for nucleic acid amplification (*18*).

TABLE 2. Comparison of PCR-CRISPR/Cas12-LFD and qPCR assay results.

PCR-CRISPR/Cas12-LFD	qPCR		
	Positive	Negative	Total
Positive	54	0	54
Negative	2	26	28
Total	56	26	82

LFD has become a commonly employed technique for nucleic acid detection due to its rapidity, convenience, and cost-effectiveness (*19*). In our study, we utilized double-antibody sandwich immunochromatography for the swift identification of nucleic acid amplification products. The front end of the test strip was pre-coated with a mixture of FAM antibody and gold nanoparticles, the control line with a secondary antibody to the FAM antibody, and the test line with a biotin antibody. We labeled the ends of the probe with FAM and biotin, respectively. Upon activation of Casase's cutting activity, colorful precipitates were observed at both the detection and quality control lines. In the absence of probe cleavage, a precipitate formed only at the quality control line. This reaction with the amplified product allows for visual results within 2–10 minutes (*20*). Due to the lower cost of traditional PCR combined with LFD compared to fluorescence quantitative PCR, it is especially favored in primary-level laboratories and field detection.

The CRISPR/Cas12 system holds promise as a diagnostic tool for pathogen detection (*21*–*23*). Recent advancements have incorporated isothermal detection combined with CRISPR and test strips to identify IFA and other prevalent pathogens. However, these methods face challenges, such as the need for 30–35 bp primers in recombinase polymerase amplification (RPA) and 4–6 pairs of primers in Loop-mediated isothermal amplification (LAMP), complicating primer design. This complexity can lead to mismatches and false-positive results, thereby reducing specificity (*24*). In contrast, PCR remains a traditional method with advantages including lower costs for instruments and

reagents, along with enhanced sensitivity and specificity. The integration of PCR, CRISPR-Cas12a, and LFD offers superior sensitivity and specificity, making it particularly suitable for local primary laboratories.

Challenges to this combined detection include less sensitivity than qPCR, and requirement for integrated microfluidics for portability, in which we will develop further.

In this study, we developed a rapid diagnostic method for detecting IAV that integrates the benefits of PCR, CRISPR/Cas12, and LFD techniques. This method exhibits high sensitivity and specificity while eliminating the need for fluorescent PCR instruments and allowing for visual interpretation of results. When testing a positive plasmid containing the M gene, the method achieved a sensitivity of 10 copies/μL. We evaluated the method on 82 clinical samples, and it demonstrated a high degree of consistency with real-time fluorescence PCR, except for two samples not detected with a qPCR CT value exceeding 35. Tests on various common respiratory samples indicated no cross-reactivity, confirming the method's high specificity. Furthermore, this study innovatively detects different subtypes of IAV by including degenerate bases in the primer design and conducts optimized screening of crRNA to enhance both sensitivity and specificity of the detection process.

Conflicts of interest: No conflicts of interest

doi: 10.46234/ccdcw2024.198

[#] Corresponding authors: Yi Zhang, zhangyicdc@126.com; Jun Lou, 13783378825@163.com.

[1] Department of Clinical Laboratory, Zhumadian Central Hospital, Zhumadian City, Henan Province, China; [2] Laboratory of Virology, Beijing Key Laboratory of Etiology of Viral Diseases in Children, Capital Institute of Pediatrics, Beijing, China; [3] National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China.

Submitted: February 01, 2024; Accepted: June 01, 2024

REFERENCES

1. Vahey MD, Fletcher DA. Low-fidelity assembly of influenza a virus promotes escape from host cells. Cell 2020;180(1):205. https://doi.org/10.1016/j.cell.2019.12.028.
2. Stadlbauer D, Zhu XY, McMahon M, Turner JS, Wohlbold TJ, Schmitz AJ, et al. Broadly protective human antibodies that target the active site of influenza virus neuraminidase. Science 2019;366(6464):499 – 504. https://doi.org/10.1126/science.aay0678.
3. Javanian M, Barary M, Ghebrehewet S, Koppolu V, Vasigala V, Ebrahimpour S. A brief review of influenza virus infection. J Med Virol 2021;93(8):4638 – 46. https://doi.org/10.1002/jmv.26990.
4. Cowling BJ, Ip DKM, Fang VJ, Suntarattiwong P, Olsen SJ, Levy J, et al. Aerosol transmission is an important mode of influenza A virus spread. Nat Commun 2013;4(1):1935. https://doi.org/10.1038/ncomms2922.
5. Nittayananta W, Lerdsamran H, Chutiwitoonchai N, Promsong A, Srichana T, Netsomboon K, et al. A novel film spray containing curcumin inhibits SARS-CoV-2 and influenza virus infection and enhances mucosal immunity. Virol J 2024;21(1):26. https://doi.org/10.1186/s12985-023-02282-x.
6. Kim DK, Poudel B. Tools to detect influenza virus. Yonsei Med J 2013;54(3):560 – 6. https://doi.org/10.3349/ymj.2013.54.3.560.
7. Srivastava P, Prasad D. Isothermal nucleic acid amplification and its uses in modern diagnostic technologies. 3 Biotech 2023;13(6):200. https://doi.org/10.1007/s13205-023-03628-6.
8. Zhou J, Ren XM, Wang X, Li Z, Xian CJ. Recent advances and challenges of the use of the CRISPR/Cas system as a non-nucleic acid molecular diagnostic. Heliyon 2023;9(12):e22767. https://doi.org/10.1016/j.heliyon.2023.e22767.
9. Hillary VE, Ceasar SA. A review on the mechanism and applications of CRISPR/Cas9/Cas12/Cas13/Cas14 proteins utilized for genome engineering. Mol Biotechnol 2023;65(3):311 – 25. https://doi.org/10.1007/s12033-022-00567-0.
10. Vivaldi G, Pfeffer PE, Talaei M, Basera TJ, Shaheen SO, Martineau AR. Long-term symptom profiles after COVID-19 *vs* other acute respiratory infections: an analysis of data from the COVIDENCE UK study. eClinicalMedicine 2023;65:102251. https://doi.org/10.1016/j.eclinm.2023.102251.
11. Su PP, Jiang CS, Zhang YM. The implication of infection with respiratory syncytial virus in pediatric recurrent wheezing and asthma: knowledge expanded post-COVID-19 era. Eur J Clin Microbiol Infect Dis 2024;43(3):403 – 16. https://doi.org/10.1007/s10096-023-04744-0.
12. Kong HM, Yi K, Mintz RL, Wang B, Xu YT, Lao YH, et al. CRISPR/Cas detection with nanodevices: moving deeper into liquid biopsy. Chem Commun (Camb) 2024;60(17):2301 – 19. https://doi.org/10.1039/d3cc05375j.
13. Asmamaw Mengstie M, Teshome Azezew M, Asmamaw Dejenie T, Teshome AA, Tadele Admasu F, Behaile Teklemariam A, et al. Recent advancements in reducing the off-target effect of CRISPR-Cas9 genome editing. Biologics 2024;18:21 – 8. https://doi.org/10.2147/BTT.S429411.
14. Watters KE, Fellmann C, Bai HB, Ren SM, Doudna JA. Systematic discovery of natural CRISPR-Cas12a inhibitors. Science 2018;362(6411):236 – 9. https://doi.org/10.1126/science.aau5138.
15. Li J, Zhang K, Lin GG, Li JM. CRISPR-Cas system: a promising tool for rapid detection of SARS-CoV-2 variants. J Med Virol 2024;96(1):e29356. https://doi.org/10.1002/jmv.29356.
16. Zhao F, Hu YM, Fan ZL, Huang BY, Wei L, Xie Y, et al. Rapid and sensitive one-tube detection of mpox virus using RPA-coupled CRISPR-Cas12 assay. Cell Rep Methods 2023;3(10):100620. https://doi.org/10.1016/j.crmeth.2023.100620.
17. Zeng RJ, Gong HX, Li YL, Li YX, Lin W, Tang DP, et al. CRISPR-Cas12a-derived photoelectrochemical biosensor for point-of-care diagnosis of nucleic acid. Anal Chem 2022;94(20):7442 – 8. https://doi.org/10.1021/acs.analchem.2c01373.
18. Wei LY, Wang ZL, Wu L, Chen YP. CRISPR/Cas12a-based magnetic relaxation switching biosensor for nucleic acid amplification-free and ultrasensitive detection of methicillin-resistant *Staphylococcus aureus*. Biosens Bioelectron 2023;222:114984. https://doi.org/10.1016/j.bios.2022.114984.
19. Zhang FY, Shang JJ, Luo J, Yin XH, Yu XM, Jiang W, et al. Development of a recombinase-aided amplification combined with a lateral flow dipstick assay for rapid detection of H7 subtype avian influenza virus. Front Microbiol 2023;14:1286713. https://doi.org/10.3389/fmicb.2023.1286713.
20. Jiang T, Li G, Liu RD, Zhou J, Gao NN, Shen JL. Creating an ultra-sensitive detection platform for monkeypox virus DNA based on CRISPR technology. J Med Virol 2023;95(7):e28905. https://doi.org/10.1002/jmv.28905.

21. Shen PJ, Si ZJ, Huang D, Xu ZP, Wang ZY, Fang MJ, et al. CRISPR Cas12a-enabled biosensors coupled with commercial pregnancy test strips for the visible point-of-care testing of SARS-CoV-2. Analyst 2023;148(11):2573 – 81. https://doi.org/10.1039/D3AN00284E.
22. Cong L, Ran FA, Cox D, Lin SL, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science 2013;339 (6121):819 – 23. https://doi.org/10.1126/science.1231143.
23. Yin XY, Yang LL, Sun XY, Zheng QY, Piao Y, Hu B, et al. Development and validation of sensitive and rapid CRISPR/Cas12-based PCR method to detect hazelnut in unlabeled products. Food Chem 2024;438:137952. https://doi.org/10.1016/j.foodchem.2023.137952.
24. Zhou X, Wang SW, Ma Y, Li YB, Deng GH, Shi JZ, et al. Rapid detection of avian influenza virus based on CRISPR-Cas12a. Virol J 2023;20(1):261. https://doi.org/10.1186/s12985-023-02232-7.

SUPPLEMENTARY MATERIAL

2x EasyTaq® PCR SuperMix and EasyScript® Reverse Transcriptase were acquired from Beijing TransGen Biotech Company. The CFX96TM Real-Time System was purchased from Bio-Rad. Rapid Nucleic Acid Extraction Kit (Magnetic Bead Method) was sourced from BioPerfectus Company, while Tiosbio Cas12/13 Test Strips were obtained from Nanjing WarBio Technology Company. EnGen® Lba Cas12a was procured from New England Biolabs and AceQ qPCR SYBR Green Master Mix from Nanjing Vazyme Company. The Influenza A Virus Universal Nucleic Acid Detection Kit (PCR-fluorescent probe method) was purchased from SunSure Company. Zhumadian Central Hospital provided samples of several respiratory pathogens, including IAV, influenza B virus (IBV), human adenovirus (HAdV), respiratory syncytial virus (RSV), human rhinovirus (HRV), Mycoplasma pneumonia (MP), COVID virus (2019-nCoV), and 26 other common respiratory pathogens. Additionally, 56 IAV clinical samples were supplied by the Inspection Department of Zhumadian Central Hospital.

Methods and Applications

Development and Diagnosis Performance of IgM-Based Rapid Antigen Test for Early Detection of SARS-CoV-2 Infection in a Large Cohort of Suspected COVID-19 Cases — USA, Poland, and Sweden, 2021–2022

Yihua Huang[1,&]; Yiyi Pu[2,&]; Youhong Weng[3,4,&]; Yahan Wu[3]; Qing He[3]; Sofia Litchev[5]; Longyou Zhao[1]; Haojie Ding[3]; Yunru Lai[1]; Jie Li[1]; Xiaojun Zheng[6]; Jinshu Chen[6]; XianqinXiong [6]; Shaohong Lu[3]; Fei Gao[6,#]; Meng Gao[3,#]; Qingming Kong[2,#]

ABSTRACT

Introduction: Antigen testing has been crucial in effectively managing the coronavirus disease 2019 (COVID-19) pandemic. This study evaluated the clinical performance of a nasopharyngeal swab (NPS)-based antigen rapid diagnostic test (Ag-RDT) compared to the gold standard real-time reverse transcription-polymerase chain reaction (RT-PCR) for early detection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

Methods: We developed an IgM-based rapid antigen test for early detection of SARS-CoV-2 infection. Between July 2021 and January 2022, we analyzed 1,030 NPS samples from participants at three centers in different countries, using both antigen rapid diagnostic tests (Ag-RDT) and RT-PCR.

Results: The Ag-RDT demonstrated minimal detection limits as low as 0.1 ng/mL for recombinant N antigen and 100 TCID50/mL for heat-inactivated SARS-CoV-2 virus. Specificity assessments involving four human coronaviruses and 13 other respiratory viruses showed no cross-reactivity. The Ag-RDT assay (ALLtest) exhibited high sensitivity (93.18%–100%) and specificity (99.67%–100%) across all centers. Factors such as cycle threshold (Ct) values and the timing of symptoms since onset were influential, with sensitivity increasing at lower Ct values (<30) and within the first week of symptoms.

Conclusion: The ALLtest Ag-RDT demonstrated high reliability and significant potential for diagnosing suspected COVID-19 cases.

Coronavirus disease 2019 (COVID-19), a contagious illness caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has imposed an unprecedented burden on public health systems worldwide. As of mid-July 2023, there have been over 768 million confirmed cases of SARS-CoV-2 infection, resulting in more than 6.9 million deaths globally (*1*). The rapid increase in new cases continues to be fueled by various genomic variants of the virus. Notably, the recent emergence of the BA.1 through BA.5 and XBB (Omicron) variants, classified as the sixth variant of concern, demonstrates increased transmissibility and resistance to immunity (*2*).

SARS-CoV-2 transmits among humans via respiratory droplets and aerosols during actions such as sneezing, talking, or coughing (*3*). The rapid and widespread dissemination of SARS-CoV-2 across geographical and species boundaries highlights the critical need for adopting the One Health approach to effectively tackle public health emergencies. Central to managing the COVID-19 pandemic is the development and implementation of swift and accessible diagnostic tools. Currently, real-time reverse transcription-polymerase chain reaction (RT-PCR) is considered the benchmark for COVID-19 diagnosis; however, challenges persist in its practical use. Notably, during the collection of nasopharyngeal secretions, the recommended one-meter distance between the healthcare provider and the patient is often not maintained, potentially heightening the risk of COVID-19 transmission. Furthermore, RT-PCR can take several hours to yield results and may result in false negatives (*4*). Consequently, developing a rapid detection method is imperative for timely isolation of suspected SARS-CoV-2 cases and minimizing the transmission risk.

The SARS-CoV-2 antigen rapid diagnostic test (Ag-RDT) has been crucial in the early detection of cases, containment strategies, and reducing transmission (*5*–*7*). Multiple studies have assessed Ag-RDTs'

diagnostic accuracy across different populations and clinical environments, revealing varied results (*8*–*10*). Ag-RDTs demonstrate high sensitivity during the initial week following symptom onset and exhibit excellent specificity when viral loads are elevated (*5*). Currently, Ag-RDTs are accessible to the general public, delivering results within 15–30 minutes. Consequently, they serve as a supplementary method to RT-PCR for the quick, economical, and straightforward identification of SARS-CoV-2 infections. Furthermore, Ag-RDTs are effective as point-of-care tests (POCT) for diagnosis in the disease's early stages, providing a significant advantage in curbing SARS-CoV-2 spread.

In the conceptual design phase, our ALLtest distinguishes itself from other Ag-RDTs through its innovative utilization of IgM antibodies targeting the nucleocapsid (N) protein. The N protein is integral to various stages of the replication cycle and serves as a serological marker for SARS-CoV-2 infection (*11*). This technique enhances early-stage infection detection and provides deeper insights into transmission dynamics. By incorporating this novel element into antigen diagnosis, the ALLtest enables a more comprehensive evaluation of SARS-CoV-2 infection, thereby facilitating improved public health interventions and mitigation strategies. In this study, nasopharyngeal swabs (NPS) were collected from suspected COVID-19 cases and subjected to diagnostic RT-PCR in the United States (US), Sweden (SE), and Poland (PL) from July 2021 to January 2022. All samples were subsequently analyzed using the ALLtest to conduct a detailed correlation analysis among various diagnostic assays. Furthermore, we explore the challenges and limitations of IgM-based antigen diagnosis, underscoring the necessity for versatile diagnostic approaches to effectively address the global health crisis. By elucidating the details of this innovative methodology, we not only enhance our comprehension of COVID-projected diagnostics but also contribute to establishing a more robust and proactive global health response.

METHODS

Preparation and Characterization of Monoclonal Antibodies (mAbs) Against SARS-CoV-2 N Protein

The nucleocapsid protein (NCBI: NC_045512.2) was expressed under identical conditions as described in our previous study (*12*). BALB/c mice, ranging in age from 4 to 8 weeks, were intraperitoneally injected with 50 μg of N antigen per mouse for the initial immunization, followed by 100 μg per mouse for the second and third immunizations at biweekly intervals. Spleen cells from mice with high antibody titers were fused with myeloma cells to produce hybridoma cells. These cells were then injected into BALB/c mice to generate mAbs specific for the SARS-CoV-2 N protein. The titers of these mAbs were evaluated using indirect ELISA on plates coated with 5 μg/well of N antigen. We tested 12 distinct mAbs, initially diluted from 1∶10 to $1:10^8$ starting from a concentration of 1 mg/mL. Pair screening of the mAbs was conducted using the ForteBio Octet® Red96 system, where the response value (nm) indicative of specific binding to the SARS-CoV-2 N antigen was recorded for each mAb pair. High-titer mAbs demonstrating effective pairing were further characterized for their isotypes using the multiple antibody isotypes kit from Proteintech Group Inc., China (Cat No. PK20003), which includes IgG1, IgG2a, IgG2b, IgG3, IgA, IgM, kappa light chain, and lambda light chain.

The Design of Ag-RDT Assay for COVID-19

The SARS-CoV-2 Ag-RDT was developed using the immunochromatographic method to detect the SARS-CoV-2 N antigen in NPS samples. The structural design of the Ag-RDT cassette is principally similar to that used in the ALLtest for antibody detection (*13*). The process begins by placing the sample and buffer on the sample pad. The fluid then migrates from left to right, sequentially interacting with the conjugation pad that contains labeled colloidal gold-mAb17, the NC membrane where specific antibodies are immobilized at distinct lines [test (T) line with mAb16 at 1 mg/mL, and control (C) line with goat anti-mouse IgG at 1 mg/mL], and finally the absorption pad.

Determination of the ALLtest Ag-RDT Assay Sensitivity and Specificity

To assess the sensitivity of the ALLtest Ag-RDT Assay, preparations were made using 1.0×10^6 median tissue culture infectious doses (TCID50)/mL of the SARS-CoV-2 Shenzhen/02/2020 strain and 1 mg/mL of full-length nucleocapsid protein (NCBI: NC_045512.2). For specificity validation of the same assay, samples from four human coronaviruses (HCoV-229E, HCoV-HKU1, CoV-OC43, HCoV-NL63), parainfluenza viruses 2 and 3, influenza A and B,

adenovirus types 3 and 7, human rhinovirus 2, 14, and 16, measles virus, mumps virus, and respiratory syncytial virus were prepared. Furthermore, to evaluate potential interference, substances, including budesonide nasal spray, mucin, dexamethasone, flunisolide, mupirocin, oxymetazoline, phenylephrine, rebetol, relenza, tamiflu, tobramycin, and whole blood, were spiked with negative and weak positive SARS-CoV-2 samples. Swabs spiked with virus tissue cultures, nucleocapsid protein, various virus strains, and interfering substances at specified concentrations were tested under defined reaction conditions.

Evaluation of the Application of the SARS-CoV-2 Variant Strains

To assess the sensitivity of the ALLtest Ag-RDT Assay across various SARS-CoV-2 variants, including Alpha, Beta, VUI-21ARP-03, Gamma, Delta, and Omicron, these strains were prepared in dilutions using 0.5% BSA-PBS to create different concentration solutions. A volume of 30 µL from each solution was applied to the tip of a nasal swab, subsequently immersed in extraction buffer. Three to four drops of this mixture were then placed into the sample well of the test cassette. Results were visually read from the cassette after 15 minutes, according to the protocol specified by the manufacturer.

Validation of Clinical Samples

Antigen tests were conducted on 1,030 individuals suspected of having COVID-19 across three international centers: Phamatech, Inc. in America (PI-US), iLab in Sweden (I-SE), and Uniwersyteckie Centrum Kliniczne in Poland (UCK-PL) from July 2021 to January 2022. Participant data from these centers is compiled in Table 1. Each participant provided two NPS specimens. One swab was analyzed immediately using the ALLtest Ag-RDT Assay or processed using sample lysates without protein denaturant and stored at either 4 ℃ for up to 24 hours or below –70 ℃ for extended storage. The second swab was preserved in universal transport medium and tested for ORF1a/b through quantitative reverse transcription PCR (qRT-PCR) within 24 hours. The Ct values from the qRT-PCR were documented for subsequent analysis. This research received ethical approval from the Ethical Committee of Phamatech, Inc. in America (Approval No. 21-HANG-101) and Uniwersyteckie Centrum Kliniczne in Poland (Approval No. NKBBN/710/2021). The requirement for informed consent was waived for the clinical trial conducted by iLab in Sweden.

Statistical Analysis

The sensitivity and specificity served as the primary metrics for assessing our antigen test. Sensitivity, or the true positive rate, was determined by dividing the number of positive cases identified by the ALLtest by the number of true positive cases confirmed by the RNA test. Specificity, or the true negative rate, was calculated by dividing the number of negative cases

TABLE 1. Summary of participant information from three centers.

Parameter	PI-US		I-SE		UCK-PL	
	RT-PCR positive (*n*=44)	RT-PCR negative (*n*=247)	RT-PCR positive (*n*=100)	RT-PCR negative (*n*=114)	RT-PCR positive (*n*=223)	RT-PCR negative (*n*=302)
Age in years, mean (SD)	34.6 (18.9)	33.3 (17.5)	–	–	47.8 (21.7)	46.2 (24.5)
Gender (%)						
Male	20.0 (45.5)	99.0 (40.1)	–	–	115.0 (51.6)	144.0 (47.8)
Female	24.0 (54.5)	148.0 (59.9)	–	–	108.0 (48.4)	158.0 (52.3)
Cycle threshold (%)						
<15	0 (0)		1.0 (1.0)		0 (0)	
15 to <20	5.0 (11.4)		6.0 (6.0)		178.0 (79.8)	
20 to <25	12.0 (27.3)		8.0 (8.0)		15.0 (6.7)	
25 to <30	20.0 (45.5)		29.0 (29.0)		10.0 (4.5)	
30 to <35	7.0 (15.9)		56.0 (56.0)		20.0 (9.0)	
≥35	0 (0)		0 (0)		0 (0)	

Note: "–" means information not recorded.

Abbreviation: PI-US=Phamatech, Inc. in USA; I-SE=iLab in Sweden; UCK-PL=Uniwersyteckie Centrum Kliniczne in Poland; RT-PCR= real-time reverse transcription-polymerase chain reaction; SD=standard deviation.

identified by the ALLtest by the total number of true negative cases. The 95% confidence intervals (*CI*s) were derived using exact binomial tests.

To determine factors strongly correlated with antigen test outcomes, participants were divided into two categories according to their test results. Factors in each category were then compared for significance. For categorical variables such as sex and Ct category, Pearson's chi-square test was utilized. For continuous variables including age, symptom duration, and Ct value — depending on their distribution — either the T-test or the Wilcoxon test was employed. All statistical analyses were performed using R software (version 3.6.0, R Core Team, Vienna, Austria).

RESULTS

Preparation and Characterization of mAbs Against N Protein

The PCR amplification of the SARS-CoV-2 N gene resulted in a 1,200 bp band, as observed through 1% agarose gel electrophoresis (Figure 1A). Following immunization and cell fusion, 12 mAbs were produced against the SARS-CoV-2 N protein (Figure 1B). The titers for five of these mAbs were notably high, approximately 1∶102,400, whereas the titers for the remaining mAbs were around 1∶25,600 (Figure 1C). Subsequent to pair-wise screening (Figure 1D), mAbs N1, N12, N15, N16, and N17 were chosen for further analysis to identify antibody subtypes and develop the Ag-RDT Assay, due to their effective pairing performance and elevated titers. Notably, N16 and N17 were classified as IgM, while the other mAbs were identified as IgG1 subtype (Figure 1E).

The Assessment of the Sensitivity and Specificity of the Ag-RDT Assay

The Ag-RDT assay was constructed using paired mAbs based on the standard schematic for colloidal gold test strip assembly, as illustrated in Figure 2A. One mAb was conjugated with colloidal gold while its counterpart served as the capture line (T-line) antibody. Sensitivity varied across different antibody pairs when assessing both positive and negative samples. The assay achieved maximal detection sensitivity and specificity when colloidal gold was conjugated with N17 and the T-line utilized N16, as shown in Figure 2B. Testing with the optimized Ag-RDT assay on virus tissue culture and recombinant SARS-CoV-2 N protein antigen at varying dilutions identified the lowest detection limits: 100 TCID50/mL for heat-inactivated SARS-CoV-2 virus (Figure 2C) and 0.1 ng/mL for the recombinant N antigen.

To verify the specificity of the Ag-RDT assay, we tested it against four other human coronaviruses and 13 viruses that cause respiratory illnesses. The results, which showed no cross-reactions (Supplementary Table S1, available at https://weekly.chinacdc.cn/), confirmed the assay's specificity for detecting the SARS-CoV-2 antigen. Additionally, various substances including budesonide nasal spray, mucin, dexamethasone, flunisolide, mupirocin, oxymetazoline, phenylephrine, rebetol, relenza, tamiflu, tobramycin, and whole blood were demonstrated not to interfere with the Ag-RDT results at the concentrations used (Supplementary Table S2, available at https://weekly.chinacdc.cn/).

Performance Evaluation of the Ag-RDT Assay with SARS-CoV-2 Variants

To assess the detection limit of the Ag-RDT assay for various SARS-CoV-2 variants, including Alpha, Beta, Gamma, Delta, and Omicron, each variant was subjected to testing. Results presented in Supplementary Table S3 (available at https://weekly.chinacdc.cn/) demonstrate that the Ag-RDT assay can effectively detect SARS-CoV-2 variants such as B.1.1.7, B.1.351, B.1.617.3, P.1.2, B.1.617.2, B.1.1.529, BA.2, BA.4, and BA.5, down to specific dilution rates and concentrations.

The Clinical Validation of the Ag-RDT Assay

The ALLtest antigen test was assessed across three centers, as detailed in Table 2. Within the PI-US cohort of 291 participants, the ALLtest correctly identified 41 out of 44 confirmed cases and showed no false positives among excluded cases, achieving a sensitivity of 93.18% and a specificity of 100%. Similarly, the test demonstrated high clinical efficacy in the I-SE and UCK-PL cohorts, with sensitivities of 100% and 97.31%, and specificities of 100% and 99.67%, respectively. Notably in the PI-US center, the sensitivity increased to 97.30% (95% *CI*: 85.84%, 99.93%) for samples with Ct values below 30, and to 96.77% (95% *CI*: 83.30%, 99.92%) for patients displaying symptoms within the first week. Additionally, the UCK-PL center reported a higher

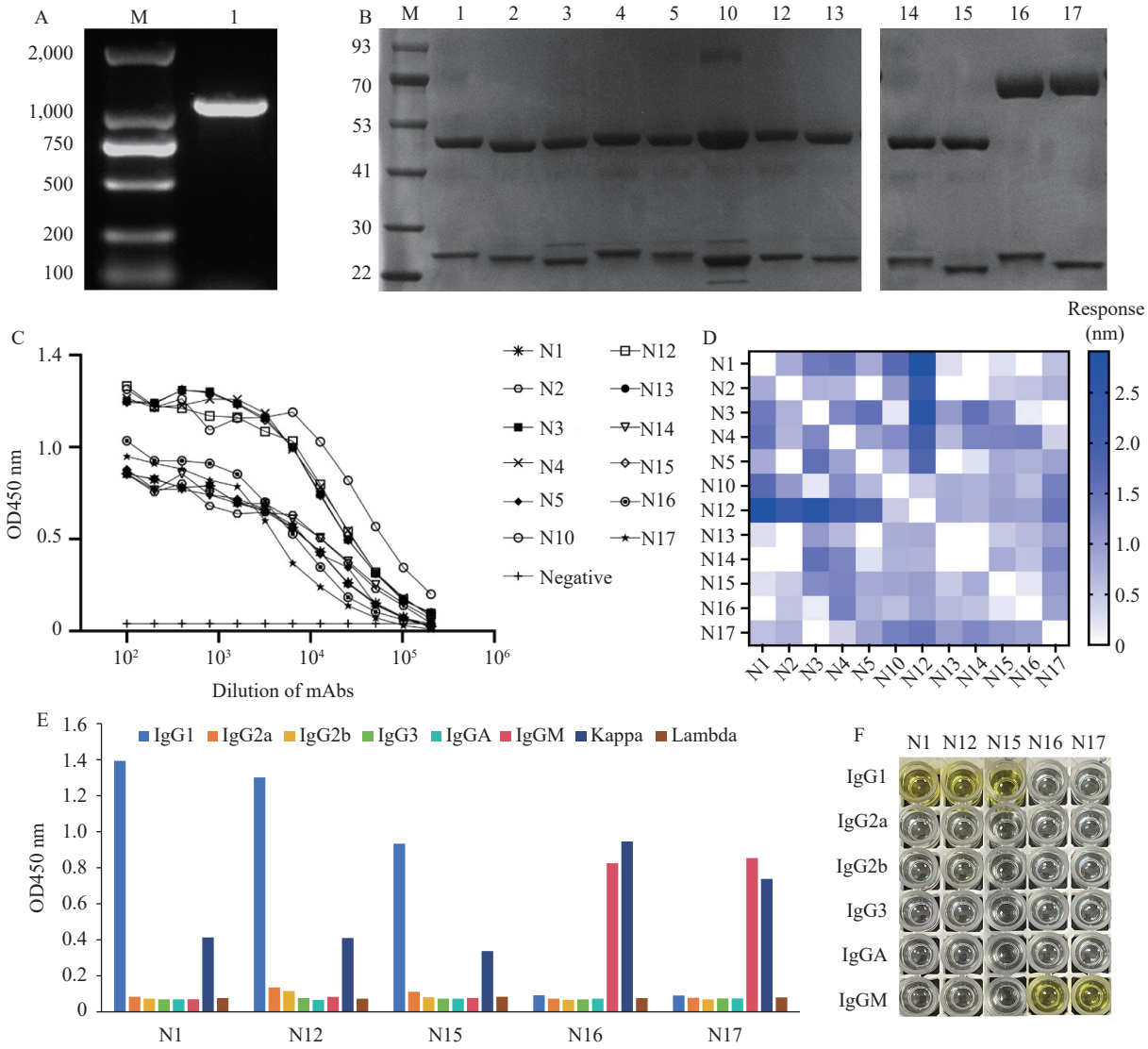
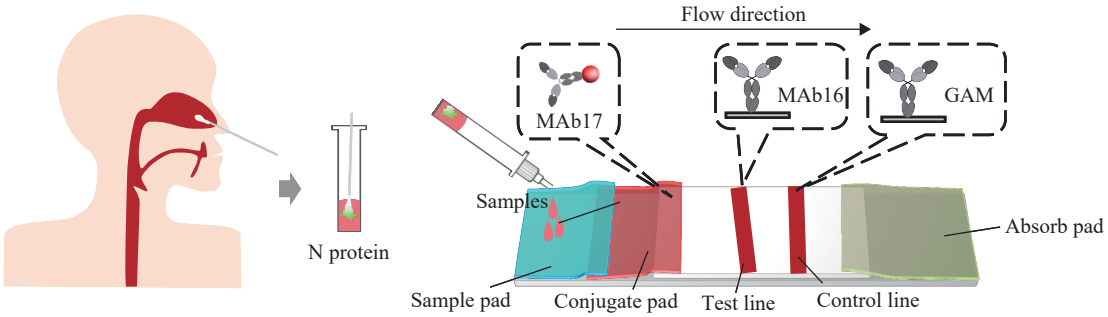


FIGURE 1. Characterization and paired screening of mAbs against the N protein. (A) Electrophoresis analysis of PCR products from pET-28a (+)-N. M: marker, L1: SARS-CoV-2 N gene; (B) SDS-PAGE analysis of SARS-CoV-2 N mAbs. M: marker, L1–L17: purified mAbs of SARS-CoV-2 N; (C) Titer detection of mAbs; (D) Heat map presenting paired results of mAbs; (E) Histogram of ELISA identification of mAbs subtypes N1, N12, N15, N16, and N17; (F) Identification of SARS-CoV-2 N mAbs subtypes using a monoclonal antibody subtyping kit.
Abbreviation: PCR=polymerase chain reaction; SARS-CoV-2=severe acute respiratory syndrome coronavirus 2; SDS-PAGE=sodium dodecyl sulfate polyacrylamide gel electrophoresis; ELISA=enzyme linked immunosorbent assay.

sensitivity of 99.51% (95% *CI*: 97.29%, 99.99%) for samples with Ct values below 30.

Factors Correlated with Our Antigen Results

Clinical factors including age, sex, duration of symptoms, and Ct values of participants were assessed for correlations with results from the ALLtest antigen assay, as shown in Supplementary Table S4 (available at https://weekly.chinacdc.cn/). Across three centers, results were consistent. A significant number of participants exhibited low viral loads that precluded RNA detection by PCR, regardless of the number of amplification cycles performed. Therefore, artificial cutoff values were employed to categorize Ct values. In PI-US and UCK-PL settings, both actual Ct values (Figure 3A and 3B; P=0.026, P=9.74×10^{-5}) and categorized Ct values (Ct<35 versus Ct≥35; P<2.2×10^{-16}) demonstrated a significant correlation with antigen test outcomes. The timing of symptom onset correlated with antigen results exclusively in UCK-PL (P=0.0018; Figure 3C). No significant correlations between age or sex and antigen test results were identified.



Colloidal gold antibody conjugates	Samples	P1	P2	P3	P4	P5	P6	Neg1	Neg2
	Capture antibody (N1)								
	N12	+	+	+	+	+	+	+	+
	N17	+	+	++	−	++	++	−	−
	Capture antibody (N12)								
	N1	−	−	−	−	−	−	−	−
	N17	−	−	−	−	−	−	−	−
	Capture antibody (N16)								
	N12	+++	++++	++++	+++	++++	++++	+	+
	N17	+++++	++++	++++	++++	++++	++++	−	−
	Capture antibody (N17)								
	N12	+	−	+	−	−	−	−	−
	N16	++++	+++++	+++	+++++	+++	+++++	−	−

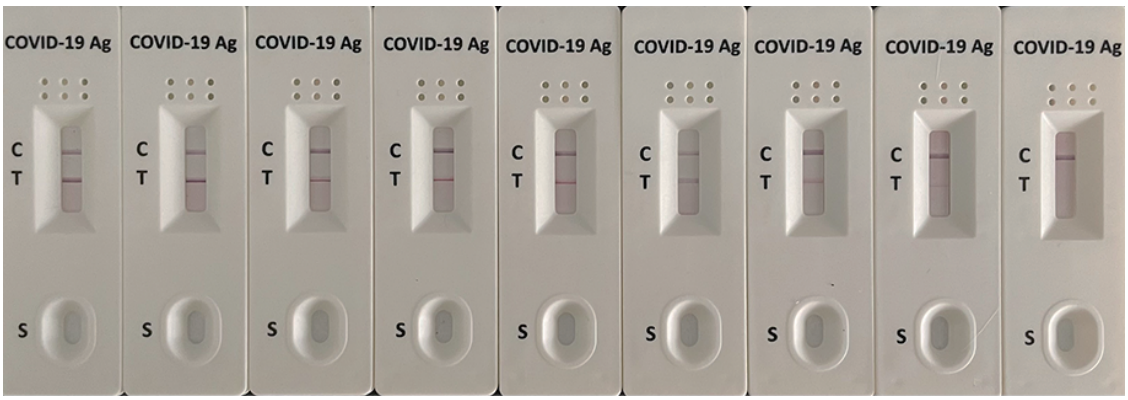


FIGURE 2. Sensitivity detection of the Ag-RDT assay for COVID-19. (A) The principal diagram of the colloidal gold test strip assembly and sample detection process; (B) Detection sensitivity and specificity using colloidal gold labeling and T-line labeling with various paired monoclonal antibodies (mAbs); (C) Sensitivity detection of the Ag-RDT assay for the heat-inactivated SARS-CoV-2 virus.
Note: In panel C, Lines 1–9 indicate concentrations ranging from 10^9 to 10 TCID50/mL; Line 8 indicates that the minimal limit of detection is 100 TCID50/mL.
Abbreviation: Ag-RDT=antigen early rapid diagnostic test; COVID-19=coronavirus disease 2019; SARS-CoV-2=severe acute respiratory syndrome coronavirus 2; TCID=tissue culture infectious doses.

DISCUSSION

The SARS-CoV-2 pandemic continues to pose significant challenges, necessitating effective strategies to curb its spread. Antigen testing has been crucial in early detection, containment, and the reduction of transmission rates. Furthermore, its integration with the One Health approach emphasizes the interconnection of human, animal, and environmental health, highlighting the need for a robust and unified global health framework. Recent literature indicates varying findings regarding the application of antigen

TABLE 2. Sensitivity and specificity of ALLtest.

Antigen test results	PI-US			I-SE			UCK-PL		
	+	−	Total	+	−	Total	+	−	Total
Positive	41	0	41	100	0	100	217	1	218
Negative	3	247	250	0	114	114	6	301	307
Total	44	247	291	100	114	214	223	302	525
Test evaluation, % (95% *CI*)									
Sensitivity	93.2 (81.3, 98.6)			100.0 (96.4, 100.0)			97.3 (94.2, 99.0)		
Specificity	100.0 (98.5, 100.0)			100.0 (96.8, 100.0)			99.7 (98.2, 100.0)		

Note: "−" means negative. "+" means positive.
Abbreviation: PI-US=Phamatech, Inc. in USA; I-SE=iLab in Sweden; UCK-PL=Uniwersyteckie Centrum Kliniczne in Poland; RT-PCR=real-time reverse transcription-polymerase chain reaction; *CI*=confidence interval.

tests in diagnosing COVID-19. A meta-analysis of antigen tests developed before September 30, 2020, showed that under specific conditions, despite varied sensitivity, rapid antigen tests could serve as alternatives to RT-PCR (*13*). Conversely, Greub et al. reported mixed performances among 30 antigen tests, advocating for comprehensive validation before official adoption (*8*). Similarly, while Navero-Castillejos et al. recommended using antigen tests solely for identifying highly infectious individuals (*10*), Regev-Yochay et al. observed high sensitivity across symptomatic and asymptomatic patients alike (*14*). Kohmer et al. also noted a significant correlation between antigen test outcomes and cell culture infectivity, underscoring their potential utility in managing COVID-19 (*9*).

Adding to this body of work, we employed our Ag-RDT (ALLtest) in a substantial cohort across three international centers. Our findings revealed high clinical sensitivities (93.18%–100%) and specificities (99.67%–100%) in suspected COVID-19 cases when compared to RT-PCR. This outperformed six other commercial rapid antigen tests, which showed sensitivities ranging from 65% to 79% (*10*). According to WHO standards — a sensitivity of ≥80% and specificity of ≥97% — our ALLtest demonstrated excellent efficacy in all three centers, meeting the criteria for SARS-CoV-2 detection. However, historical data revealed that ALLtest was grouped among those with low sensitivity (16.7%) in a review of 122 Ag-RDTs by Nubling et al. (*15*). Nonetheless, recent studies, such as those by Moons et al. and Sivro et al., show more promising results, with higher sensitivities observed in nasal versus saliva samples (83.9%) and reasonable detection (about 74%) of Omicron BA.4 and BA.5 variants (*16*–*17*). Overall, the ALLtest exhibits generally qualified performance amid evolving epidemiological landscapes.

The Ct value, indicative of the quantity of viral RNA, has been recommended for inclusion alongside binary RT-PCR outcomes to enhance their clinical utility (*18*). Bullard et al. observed that no viable viral cultures were produced from samples with a Ct value greater than 24 (*19*). Additionally, studies have shown that samples with a Ct value exceeding 30 (2.17×10^5 E gene copies/mL) are typically not cultivable (*20*). The sensitivity of the ALLtest correlates with viral loads, consistent with previous findings (*21*–*22*). Similar to other research (*14*), the sensitivity of ALLtest was higher in samples with a Ct value of 30 or less (99.30%, 95% *CI*: 97.48%, 99.91%) and increased to 100% (95% *CI*: 98.38%, 100%) in samples with a Ct value of 25 or less. Significantly lower Ct values were also observed in ALLtest-positive cases among participants with detectable Ct values across all three centers (Figure 3A, 3B, and Supplementary Table S4). According to Kucirka et al., RT-PCR sensitivity and viral load typically increase within 5–7 days post-infection, stabilize for 1–2 weeks, and subsequently decrease (*4*). Furthermore, a significant difference in the timing of symptoms was found between ALLtest positive and negative groups, with those displaying symptoms earlier (2.25 days *vs.* 2.64 days) more likely to test positive with ALLtest (P=0.0018).

In this study, we exclusively utilized NPS for both Ag-RDT and RT-PCR. This limitation may affect the generalizability of the results, as the performance of the Ag-RDT could vary with other sample types, such as oropharyngeal swabs or saliva. And due to the difficulty of collecting enough number of participants infected by the recent coronavirus strain, we only validated the detection effect of ALLtest on those strains in lab. In the future, multiple commonly used Ag-RDT will be tested compared with ALLtest on a larger COVID-19-infected population.

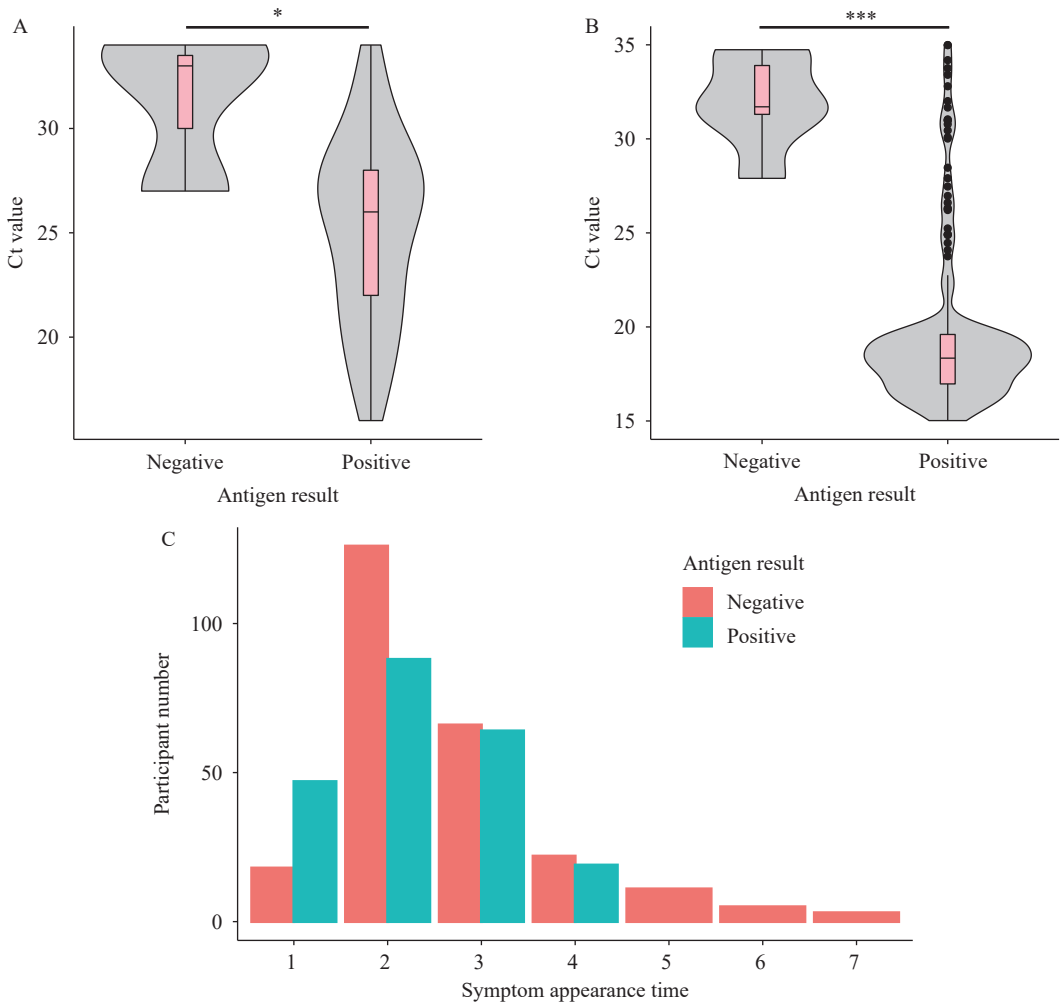


FIGURE 3. Correlation factors associated with ALLtest results. (A) Ct value distributions among participants from PI-US categorized by differing antigen test results (*P*=0.026); (B) Ct value distributions among participants from UCK-PL categorized by differing antigen test results ($P=9.74\times10^{-5}$); (C) Distributions of symptom onset among participants from UCK-PL with varying antigen test outcomes (*P*=0.0018).
Abbreviation: Ct=cycle threshold; UCK-PL=Uniwersyteckie Centrum Kliniczne in Poland; PI-US=Phamatech, Inc. in USA.

We developed and validated the ALLtest antigen test for the early diagnosis of COVID-19 across three international centers. Given its high sensitivity and specificity, we propose that the ALLtest can serve as a reliable adjunct to the RT-PCR test.

Conflicts of interest: No conflicts of interest.

Funding: Supported by the Zhejiang Provincial Program for the Cultivation of High-level Innovative Health Talents (WJW2021002); Science Foundation of National Health Commission of the PRC (WJW-ZJ-2203); the Key R&D projects of Zhejiang Province (2022C03109); the S&T Major Project of Lishui (2023GYX03) and the National Health Commission Research fund (WKJ-ZJ-2421).

doi: 10.46234/ccdcw2024.199

[#] Corresponding authors: Fei Gao, soar.gao@alltests.com.cn; Meng Gao, yourgm@hotmail.com; Qingming Kong, qmkong_1025@163.com.

[1] Department of Laboratory Medicine, Lishui Second People's Hospital Affiliated to Wenzhou Medical University, Lishui City, Zhejiang Province, China; [2] Key Laboratory of Biomarkers and In Vitro Diagnosis Translation of Zhejiang province, School of Laboratory Medicine and Bioengineering, Hangzhou Medical College, Hangzhou City, Zhejiang Province, China; [3] School of Basic Medicine and Forensics, Key Laboratory of Bio-tech Vaccine of Zhejiang Province, Engineering Research Center of Novel Vaccine of Zhejiang Province, Hangzhou Medical College, Hangzhou City, Zhejiang Province, China; [4] The Fifth Affiliated Hospital of Wenzhou Medical University, Lishui City, Zhejiang Province, China; [5] Baylor College of Medicine, Houston, TX, USA; [6] Department of Research and Development, Hangzhou AllTest Biotech Co., Ltd, Hangzhou City, Zhejiang Province, China.

[&] Joint first authors.

Submitted: August 27, 2023; Accepted: September 08, 2024

REFERENCES

1. World Health Organization. Weekly epidemiological update on covid-19 - 16 March 2023. 2023. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---16-march-2023. [2024-3-18].
2. Shrestha LB, Foster C, Rawlinson W, Tedla N, Bull RA. Evolution of the SARS-CoV-2 omicron variants BA. 1 to BA.5: implications for immune escape and transmission. Rev Med Virol 2022;32(5):e2381. https://doi.org/10.1002/rmv.2381.
3. Asadi S, Bouvier N, Wexler AS, Ristenpart WD. The coronavirus pandemic and aerosols: does COVID-19 transmit via expiratory particles?. Aerosol Sci Technol 2020;54(6):635 – 8. https://doi.org/10.1080/02786826.2020.1749229.
4. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. Ann Intern Med 2020;173(4):262 – 7. https://doi.org/10.7326/M20-1495.
5. Brummer LE, Katzenschlager S, Gaeddert M, Erdmann C, Schmitz S, Bota M, et al. Accuracy of novel antigen rapid diagnostics for SARS-CoV-2: a living systematic review and meta-analysis. PLoS Med 2021;18(8):e1003735. https://doi.org/10.1371/journal.pmed.1003735.
6. Khalid MF, Selvam K, Jeffry AJN, Salmi MF, Najib MA, Norhayati MN, et al. Performance of rapid antigen tests for COVID-19 diagnosis: a systematic review and meta-analysis. Diagnostics (Basel) 2022;12(1):110. https://doi.org/10.3390/diagnostics12010110.
7. Ye Q, Shao WX, Meng HY. Performance and application evaluation of SARS-CoV-2 antigen assay. J Med Virol 2022;94(8):3548 – 53. https://doi.org/10.1002/jmv.27798.
8. Greub G, Caruana G, Schweitzer M, Imperiali M, Muigg V, Risch M, et al. Multicenter technical validation of 30 rapid antigen tests for the detection of SARS-CoV-2 (VALIDATE). Microorganisms 2021;9(12):2589. https://doi.org/10.3390/microorganisms9122589.
9. Kohmer N, Toptan T, Pallas C, Karaca O, Pfeiffer A, Westhaus S, et al. The comparative clinical performance of four SARS-CoV-2 rapid antigen tests and their correlation to infectivity in vitro. J Clin Med 2021;10(2):328. https://doi.org/10.3390/jcm10020328.
10. Navero-Castillejos J, Casals-Pascual C, Narváez S, Cuesta G, Hurtado JC, Fernandez M, et al. Diagnostic performance of six rapid antigen tests for SARS-CoV-2. Microbiol Spectr 2022;10(2):e02351 – 21. https://doi.org/10.1128/spectrum.02351-21.
11. Wu C, Qavi AJ, Hachim A, Kavian N, Cole AR, Moyle AB, et al. Characterization of SARS-CoV-2 nucleocapsid protein reveals multiple functional consequences of the C-terminal domain. iScience 2021;24(6):102681. https://doi.org/10.1016/j.isci.2021.102681.
12. Pu YY, Weng YH, Wu YH, Gao F, Zheng XJ, Xiong XQ, et al. Antibody response to SARS-CoV-2 in the first batch of COVID-19 patients in China by a self-developed rapid IgM-IgG test. Front Cell Infect Microbiol 2022;12:915751. https://doi.org/10.3389/fcimb.2022.915751.
13. Dinnes J, Deeks JJ, Adriano A, Berhane S, Davenport C, Dittrich S, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. Cochrane Database Syst Rev 2020;8(8):CD013705. https://doi.org/10.1002/14651858.CD013705.
14. Regev-Yochay G, Kriger O, Mina MJ, Beni S, Rubin C, Mechnik B, et al. Real world performance of SARS-CoV-2 antigen rapid diagnostic tests in various clinical settings. Infect Control Hosp Epidemiol 2022:1-20. http://dx.doi.org/10.1017/ice.2022.3.
15. Scheiblauer H, Filomena A, Nitsche A, Puyskens A, Corman VM, Drosten C, et al. Comparative sensitivity evaluation for 122 CE-marked rapid diagnostic tests for SARS-CoV-2 antigen, Germany, September 2020 to April 2021. Euro Surveill 2021;26(44):2100441. https://doi.org/10.2807/1560-7917.ES.2021.26.44.2100441.
16. Schuit E, Venekamp RP, Veldhuijzen IK, van den Bijllaardt W, Pas SD, Stohr JJJM, et al. Head-to-head comparison of the accuracy of saliva and nasal rapid antigen SARS-CoV-2 self-testing: cross-sectional study. BMC Med 2022;20(1):406. https://doi.org/10.1186/s12916-022-02603-x.
17. Samsunder N, Lustig G, de Vos M, Ngcapu S, Giandhari J, Tshiabuila D, et al. Performance of rapid antigen tests in identifying Omicron BA.4 and BA.5 infections in South Africa. J Clin Virol 2023;165:105498. https://doi.org/10.1016/j.jcv.2023.105498.
18. Tom MR, Mina MJ. To interpret the SARS-CoV-2 test, consider the cycle threshold value. Clin Infect Dis 2020;71(16):2252 – 4. https://doi.org/10.1093/cid/ciaa619.
19. Bullard J, Dust K, Funk D, Strong JE, Alexander D, Garnett L, et al. Predicting infectious severe acute respiratory syndrome coronavirus 2 from diagnostic samples. Clin Infect Dis 2020;71(10):2663 – 6. https://doi.org/10.1093/cid/ciaa638.
20. Igloi Z, Velzing J, van Beek J, van de Vijver D, Aron G, Ensing R, et al. Clinical evaluation of Roche SD biosensor rapid antigen test for SARS-CoV-2 in municipal health service testing site, the Netherlands. Emerg Infect Dis 2021;27(5):1323 – 9. https://doi.org/10.3201/eid2705.204688.
21. Jääskeläinen AE, Ahava MJ, Jokela P, Szirovicza L, Pohjala S, Vapalahti O, et al. Evaluation of three rapid lateral flow antigen detection tests for the diagnosis of SARS-CoV-2 infection. J Clin Virol 2021;137:104785. https://doi.org/10.1016/j.jcv.2021.104785.
22. Smith RD, Johnson JK, Clay C, Girio-Herrera L, Stevens D, Abraham M, et al. Clinical evaluation of Sofia rapid antigen assay for detection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) among emergency department to hospital admissions. Infect Control Hosp Epidemiol 2022;43(8):968 – 73. https://doi.org/10.1017/ice.2021.281.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLE S1. Verification of specificity for viruses associated with respiratory infections.

No.	Virus strains	Concentration (TCID50/mL)	Result
1	HCoV-229E	5×10^5	−
2	HCoV-HKU1	1×10^6	−
3	HCoV-OC43	1×10^6	−
4	HCoV-NL63	1×10^6	−
5	Parainfluenza viruses 2	1.58×10^7	−
6	Parainfluenza viruses 3	1.58×10^8	−
7	Influenza A (H3N2)	1×10^5	−
8	Influenza A (H1N1)	3.16×10^5	−
9	Influenza B	3.16×10^6	−
10	Adenovirus type 3	3.16×10^4	−
11	Adenovirus type 7	1.58×10^5	−
12	Human Rhinovirus 2	2.81×10^4	−
13	Human Rhinovirus 14	1.58×10^6	−
14	Human Rhinovirus 16	8.89×10^6	−
15	Measles virus	1.58×10^4	−
16	Mumps virus	1.58×10^4	−
17	Respiratory syncytial virus	8.89×10^4	−

Note: "−" means negative.

SUPPLEMENTARY TABLE S2. Results of the interfering action by the Ag-RDT assay.

No.	Analytes	Concentration	Negative	SARS-CoV-2 positive
1	budesonide nasal spray	200 μL/mL	−	+
2	whole blood	20 μL/mL	−	+
3	relenza	282 ng/mL	−	+
4	flunisolide	6.8 ng/mL	−	+
5	mucin	50 μg/mL	−	+
6	rebetol	4.5 μg/mL	−	+
7	tamiflu	1.1 μg/mL	−	+
8	mupirocin	12 mg/mL	−	+
9	phenylephrine	12 mg/mL	−	+
10	tobryamycin	2.43 mg/mL	−	+
11	dexamethasone	0.8 mg/mL	−	+
12	oxymetazoline	0.6 mg/mL	−	+

Note: "−" means negative. "+" means positive.
Abbreviation: Ag-RDT=antigen early rapid diagnostic test; SARS-CoV-2=severe acute respiratory syndrome coronavirus 2.

SUPPLEMENTARY TABLE S3. Detection of SARS-CoV-2 variants using the Ag-RDT assay.

WHO label	Pango lineages	N protein mutations	LOD
Alpha	B.1.1.7	D3L, R203K, G204R, S235F	$1{:}2\times10^{6}$
Beta	B.1.351	T205I	$1{:}2\times10^{6}$
VUI-21ARP-03	B.1.617.3	P67S, R203M, D377Y	$1{:}2\times10^{6}$
Gamma	P.1.2	P80R	$1{:}2\times10^{6}$
Delta	B.1.617.2	D63G, R203M, D377Y	$1{:}2\times10^{6}$
Omicron	B.1.1.529	P13L, del31/33, R203K, G204R	$1{:}1\times10^{6}$ (1.61 ng/mL)
Omicron	BA.2	P13L, del31/33, R203K, G204R	$1{:}1\times10^{7}$ (0.368 ng/mL)
Omicron	BA.4	P13L, del31/33, P151S, R203K, G204R, S413R	$1{:}1\times10^{7}$ (0.29 ng/mL)
Omicron	BA.5	P13L, del31/33, R203K, G204R, S413R	$1{:}1\times10^{7}$ (0.368 ng/mL)

Note: LOD indicated by the highest dilution rate of the virus that can be detected.
Abbreviation: Ag-RDT=antigen early rapid diagnostic test; SARS-CoV-2=severe acute respiratory syndrome coronavirus 2; LOD=limit of detection; WHO=World Health Organzation.

SUPPLEMENTARY TABLE S4. Correlation significance of factors across three centers.

P	PI-US	I-SE	UCK-PL
Sex	0.55	–	0.29
Age	0.78	–	0.42
Symptom time	0.47	0.11	$0.0018^{\dagger}$
Ct value	$0.026^{\dagger}$	–	$9.74\times10^{-5\dagger}$
Ct category*	$<2.2\times10^{-16\dagger}$	$<2.2\times10^{-16\dagger}$	$<2.2\times10^{-16\dagger}$

"–" means significance cannot be calculated on missing data.
Abbreviation: PI-US=Phamatech, Inc. in USA; I-SE=iLab in Sweden; UCK-PL=Uniwersyteckie Centrum Kliniczne in Poland; Ct=cycle threshold.
* Ct category include: 1) Ct<35; 2) Ct≥35.
†: $P<0.05$.

Methods and Applications

Development and Comparison of Time Series Models in Predicting Severe Fever with Thrombocytopenia Syndrome Cases — Hubei Province, China, 2013–2020

Zixu Wang[1,2,&]; Jinwei Zhang[3,&]; Wenyi Zhang[4,&]; Nianhong Lu[1]; Qiong Chen[1]; Junhu Wang[1]; Yingqing Mao[1]; Haiming Yi[1]; Yixin Ge[1]; Hongming Wang[1]; Chao Chen[1]; Wei Guo[1]; Xin Qi[5]; Yuexi Li[7,#]; Ming Yue[6,#]; Yong Qi[1,#]

ABSTRACT

Introduction: Severe fever with thrombocytopenia syndrome (SFTS) is an emerging infectious disease caused by the SFTS virus, which has a high mortality rate. Predicting the number of SFTS cases is essential for early outbreak warning and can offer valuable insights for establishing prevention and control measures.

Methods: In this study, data on monthly SFTS cases in Hubei Province, China, from 2013 to 2020 were collected. Various time series models based on seasonal auto-regressive integrated moving average (SARIMA), Prophet, eXtreme Gradient Boosting (XGBoost), and long short-term memory (LSTM) were developed using these historical data to predict SFTS cases. The established models were evaluated and compared using mean absolute error (MAE) and root mean squared error (RMSE).

Results: Four models were developed and performed well in predicting the trend of SFTS cases. The XGBoost model outperformed the others, yielding the closest fit to the actual case numbers and exhibiting the smallest MAE (2.54) and RMSE (2.89) in capturing the seasonal trend and predicting the monthly number of SFTS cases in Hubei Province.

Conclusion: The developed XGBoost model represents a promising and valuable tool for SFTS prediction and early warning in Hubei Province, China.

Severe fever with thrombocytopenia syndrome (SFTS) is an emerging infectious disease caused by the SFTS virus. Since the first confirmed case was reported in 2009 (*1*), most cases have been reported in northern and central China (*2*–*3*). The number of reported SFTS cases continues to rise, and the areas affected by the disease are expanding (*4*–*5*). Due to its high case-fatality rate and the possibility of pandemic spread, the World Health Organization included SFTS on its list of the top 10 infectious diseases needing immediate research attention (*6*). Although China has established a valuable infectious disease surveillance system to monitor and assess disease burden, the system cannot predict future trends or provide early warnings of outbreaks. Furthermore, the monitoring data obtained are often delayed. Consequently, there is an urgent need for a model to predict the number of SFTS cases in endemic regions.

As a tick-borne disease, the incidence of SFTS exhibits distinct time-series characteristics, referring to data points collected and recorded chronologically, typically at regular intervals. Specialized time-series analysis techniques are likely suitable for effectively modeling and forecasting SFTS incidence.

In this study, we utilized various time series algorithms based on historical data to predict the occurrence of SFTS in Hubei Province, one of the first provinces to report SFTS cases and a province with a high incidence of the disease in China (*7*). Predicting the number of SFTS cases in this region will provide important insights for developing prevention and control interventions.

METHODS

Data Collection

The monthly number of SFTS cases in Hubei Province was obtained from the Public Health Science Data Center (https://www.phsciencedata.cn/Share/). Data reported between January 2013 and December 2019 (84 data points total) were used for model training and development, while the remaining data from January to December 2020 (12 data points total)

were used for external validity assessment.

Model Constructions

SARIMA model: Seasonal autoregressive integrated moving average (SARIMA) is an extension of autoregressive integrated moving average (ARIMA) that requires selecting hyperparameters for both the trend and seasonal elements of the time series. The formula for SARIMA is as follows:

$$(1-B)^d\left(1-B^s\right)^D Y_t = \theta_0 + \frac{\theta(B)\theta_s(B^s)}{\phi(B)\phi_s(B^s)}\varepsilon_t \quad (1)$$

where Y_t refers to the value of the time series at time t, θ_0 is constant, ε_t is the white noise value at period t, and the parameters d and D represent the difference number and seasonal difference number, respectively. B is the backshift operator, $\phi(B)$ is the autoregressive operator, and $\theta(B)$ is the moving average operator. $\phi_s(B^s)$ and $\theta_s(B^s)$are the seasonal operators.

Prophet model: The Prophet model provides a versatile treatment of trends, seasonality, and holiday effects. The trend component, g(t), is engineered to capture non-periodic changes in the time series. The foundational equation of the Prophet model is expressed as:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t \quad (2)$$

Where y_t denotes the predicted value at time t, $s(t)$ is the seasonality component, $h(t)$ represents the impact of holidays or specific events on the time series, and ε_t is the error term accounting for aspects of the data not explained by the model.

XGBoost: eXtreme Gradient Boosting (XGBoost) iteratively constructs a series of short, basic decision trees. For a dataset with n examples and m features, a tree ensemble model in XGBoost predicts the output using K additive functions:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F \quad (3)$$

Here, $\hat{y}$ represents the predicted value for the ith sample, f_k is a function corresponding to the kth tree, and F denotes the space of regression tree functions, with x_i being the feature vector for the ith sample.

To learn the set of functions used in the model, XGBoost minimizes the following regularized objective:

$$L(\varphi) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega\left(f_k\right) \quad (4)$$

In this equation, $l\left(y_i, \hat{y}_i\right)$ is the loss function that quantifies the error between the observed and predicted data, and $\Omega(f_k)$ is the regularization term that helps smooth the learned weights. This smoothing prevents overfitting and encourages the model to select simpler, more predictive functions. The regularization term is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \quad (5)$$

Where γ and λ are regularization parameters, T is the number of leaves in the tree, and w_j represents the score on each leaf.

Bayesian optimization was used to select the optimal hyperparameters, with the objective function defined to maximize R^2.

LSTM Networks: Long short-term memory (LSTM) networks incorporate a cell state that acts as a form of memory. The key feature of LSTM networks lies in their gating mechanism, which comprises three types of gates.

The input gate regulates the flow of new information into the cell state through a two-step process. First, a sigmoid function determines the necessary update values, represented by the equation:

$$i_t = \sigma\left(W_{x_i}x_t + W_h h_{t-1} + b_i\right) \quad (6)$$

The second step employs a tanh function to generate a vector of new candidate values that may be added to the state, given by:

$$C_t = tanh\left(W_{x_C}x_t + W_{h_C}h_{t-1} + b_C\right) \quad (7)$$

Here, it is the activation of the input gate, and C_t is the candidate vector for the cell state update.

The forget gate determines which information from the cell state to retain or discard. It uses a sigmoid function to evaluate the importance of existing information in the cell state, defined by:

$$f_t = \sigma\left(W_{x_f}x_t + W_{h_f}h_{t-1} + b_f\right) \quad (8)$$

The activation vector f_t indicates the extent to which past information should be forgotten or retained.

The output gate regulates the information sent to the subsequent layer. This gate functions in two stages. First, a sigmoid function determines which parts of the cell state are outputted, as shown by:

$$o_t = \sigma\left(W_{x_0}x_t + W_{h_g}h_{t-1} + b_0\right) \quad (9)$$

Then, the final output is calculated by multiplying this activation o_t with the tanh of the cell state, resulting in:

$$h_t = o_t \times \tanh(C_t) \quad (10)$$

The output vector h_t represents the information transmitted to subsequent layers or units in the network.

In the models described above, clipping, a data post-processing technique, was used to address unrealistic negative values in the results. A detailed explanation of each model is provided in Supplementary Material (available at https://weekly.chinacdc.cn/).

Performance Evaluations

The predictive performance of the models was assessed using two indices: mean absolute error (MAE) and root mean squared error (RMSE), defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (12)$$

Software

Descriptive statistics and time series modeling were conducted using Python (version 3.7; Python Software Foundation, Beaverton, OR, USA). The SARIMA, Prophet, XGBoost, and LSTM models were implemented using the statsmodels, fbprophet, scikit-learn, and Keras packages, respectively. A $P<0.05$ was considered statistically significant.

RESULTS

General Analysis

A total of 1,695 SFTS cases were reported in Hubei Province from January 2013 to December 2020, exhibiting clear seasonal characteristics. More cases were reported from April to August each year and fewer from December to February of the following year. Interestingly, a prominent peak occurred in June and a smaller peak in October (Figure 1).

Models

In the SARIMA model construction, the augmented Dickey-Fuller (ADF) test indicated that the time series data were unstable with a $P>0.05$ (Dickey-Fuller=−1.339, $P=0.611$). After the first difference, the original sequence tended to become stationary. The parameters p and q were determined from the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots (Figure 2), and the final model parameters were determined as SARIMA (1,1,1), (0,1,1)12 based on the minimum Akaike information criterion (AIC) (AIC=543.302). All parameters were significant with $P<0.01$ (Supplementary Table S1, available at https://weekly.chinacdc.cn/). The residual autocorrelation test (Ljung-box test) indicated that the residual was not significantly different from a white noise series (Q-statistic=0.32, $P=0.57$), suggesting that the model was acceptable.

The optimized parameters of the other three models are summarized in Table 1.

Model Evaluation and Comparison

The trained SARIMA, Prophet, XGBoost, and LSTM models were used to predict the number of reported SFTS cases in 2020 and were compared with real external validation data (Figure 3). All four models performed well in predicting the trends of SFTS cases;

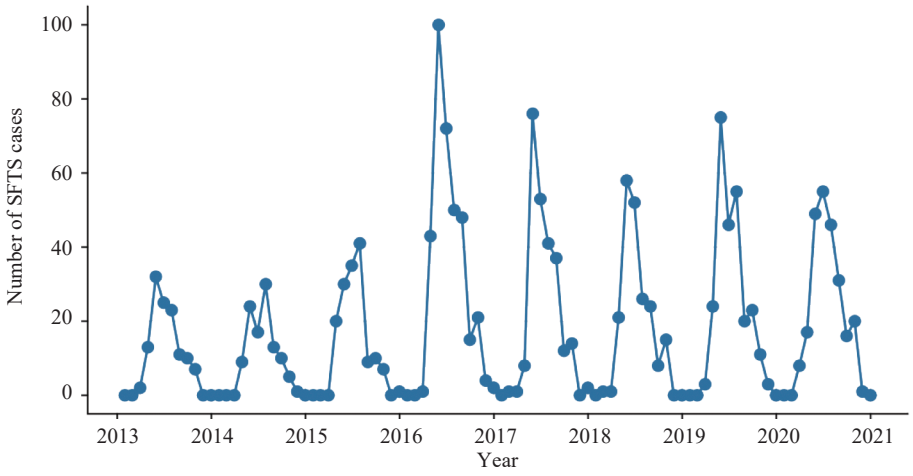


FIGURE 1. Trends of the actual number of SFTS cases from January 2013 to December 2020 in Hubei Province, China.
Abbreviation: SFTS=Severe Fever with Thrombocytopenia Syndrome.

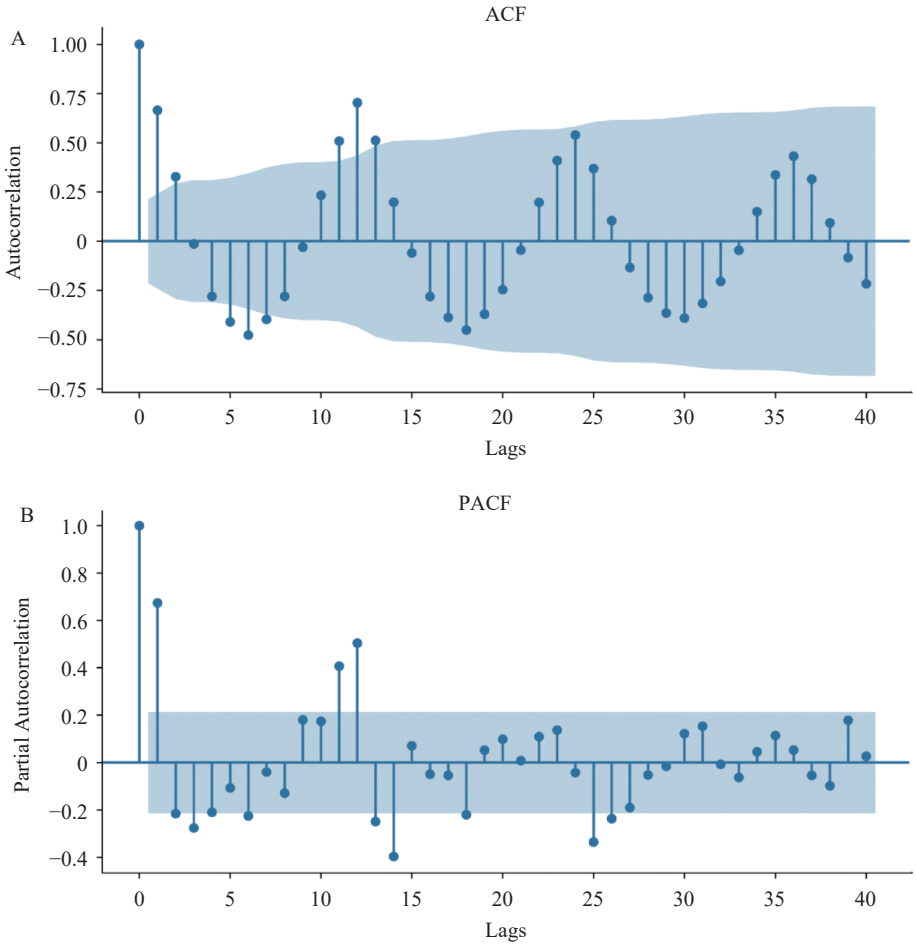


FIGURE 2. ACF and PACF charts after the first-order difference. (A) ACF; (B) PACF.
Abbreviation: ACF=autocorrelation function; PACF=partial autocorrelation function

however, the XGBoost model yielded the closest fit to the actual case numbers (Figure 3). The prediction performances of the models were then compared using error indices, including MAE and RMSE. As shown in Supplementary Table S2 (available at https://weekly.chinacdc.cn/), the MAE and RMSE of XGBoost were lower than those of the other three models, indicating that XGBoost performed best in predicting SFTS cases, followed by Prophet, LSTM, and SARIMA, respectively.

DISCUSSION

Previous studies have conducted multivariate modeling analyses to examine the risk factors associated with SFTS incidence in Hubei Province (*7*). However, to our knowledge, this study is the first to construct predictive models for the number of SFTS cases in Hubei Province.

In this study, we developed four models based on different algorithms to predict SFTS cases in Hubei Province. Each algorithm has advantages and disadvantages. SARIMA models are relatively simple, linear models capable of uncovering dynamic relationships between historical and predicted data. However, they require the original sequence to be stable before modeling and struggle to capture nonlinear relationships in the data. This limitation becomes evident when abrupt changes or nonlinear trends are present in the data, as SARIMA is less flexible in adapting to these complexities.

In contrast, the Prophet model does not require consideration of time series data stationarity and offers greater parameter adjustability, enhancing its flexibility. This model can automatically detect and handle outliers in the data, making it suitable for noisy or irregular datasets. It demonstrates rapid computation, making it appropriate for large datasets and real-time forecasting applications. Prophet has shown excellent performance in predicting various

TABLE 1. Parameters of the optimized Prophet, XGBoost, and LSTM models.

Models	Parameters	Values
Prophet	Growth	linear
	Seasonality mode	additive
	Interval width	0.8
	Changepoints	24
	Changepoint prior scale	0.3
XGBoost	Min_child_weight	9
	Estimators	54
	Learning rate	0.407
	Max depth	6
LSTM	No. of neurons	201
	Layers	1
	Learning rate	0.003
	Activation	tanh
	Recurrent activation	sigmoid
	Dropout	0
	Loss	mse
	Optimizer	Adam
	Batch size	1
	Epochs	100

infectious diseases, including coronavirus disease 2019 (COVID-19) and hand, foot, and mouth disease (*8–10*).

XGBoost displays robustness in handling nonlinear time series data, excelling at forecasting extreme values. This is likely due to its ability to model complex relationships through boosting. LSTM features a memory unit for storing information across time steps, which is advantageous for modeling long-term dependencies. It accommodates varying input and output dimensions for both univariate and multivariate data. However, LSTM may struggle with predicting sudden changes due to its reliance on past data patterns, as seen in our study with the surge in cases from April to May 2020.

All four models performed well in predicting SFTS cases and exhibited similar trends to the actual case counts. XGBoost demonstrated the closest predictions to the actual values, with the lowest MAE and RMSE values. Notably, SARIMA, Prophet, and LSTM did not accurately predict the May case counts (Figure 3). Additionally, SARIMA and Prophet failed to predict the peak month, possibly due to the sharp increase in actual cases from April to May 2020, which may have introduced challenges in predicting such volatile data. XGBoost displayed excellent performance in

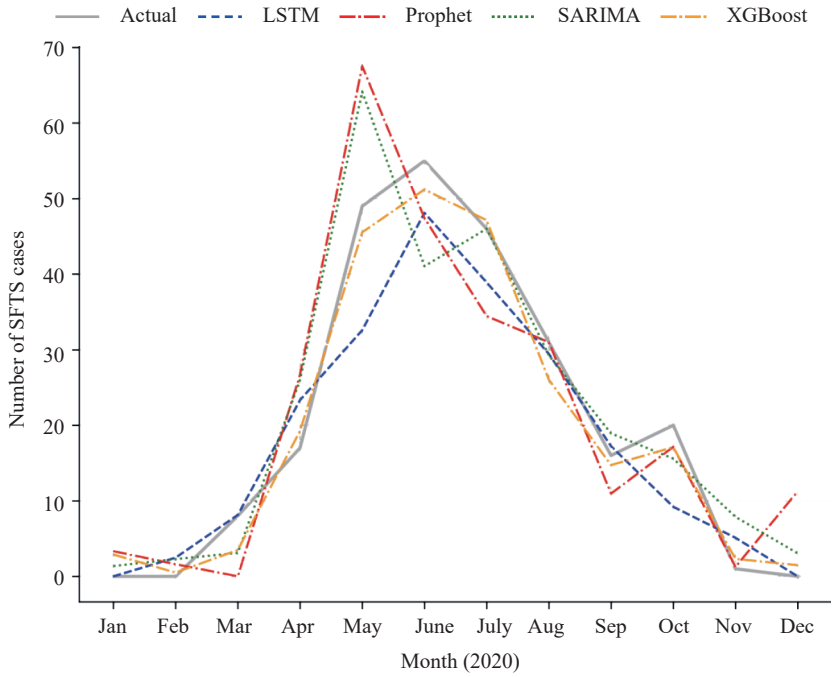


FIGURE 3. Comparison of the actual SFTS cases with the predicted cases from January to December 2020 by the four models.
Abbreviation: LSTM=long short-term memory; SFTS=severe fever with thrombocytopenia syndrome; SARIMA=seasonal auto-regressive integrated moving average; XGBoost=eXtreme Gradient Boosting.

forecasting extreme values (such as the prominent June peak and the smaller October peak) and capturing the overall trend.

Considering that meteorological, geographical, and human activity factors are considered risk factors for SFTS (*11*–*13*), incorporating additional related external variables could enhance the predictive model's performance. Furthermore, studies have indicated that combining linear and nonlinear models may yield superior predictive performance compared to single models, such as SARIMA-Prophet (*14*) and SARIMA-LSTM (*15*), representing a potential avenue for improvement.

In addition, the best model in the present study was developed based on data from Hubei Province, so it may not be suitable for other regions. This limits the model's general applicability. However, the study provides a feasible scheme for other regions to predict the disease.

In conclusion, we established and evaluated various time series models. The XGBoost model demonstrated the best predictive performance for forecasting monthly confirmed SFTS cases in Hubei Province. This model holds promise for providing valuable information and data for the early assessment of potential SFTS risks, which is crucial for developing early warning systems and formulating effective prevention and control measures.

Conflicts of interest: No conflicts of interest.

Funding: Supported by Medical Science and Technology Projects (JK2023002), National Natural Science Foundation of China (82273691), and Open Research Fund Program of the State Key Laboratory of Pathogen and Biosecurity (No. SKLPBS2137).

doi: 10.46234/ccdcw2024.200

[#] Corresponding authors: Yuexi Li, liyxi2007@126.com; Ming Yue, njym08@163.com; Yong Qi, qslark@126.com.

[1] Huadong Research Institute for Medicine and Biotechniques, Nanjing City, Jiangsu Province, China; [2] Bengbu Medical College, Bengbu City, Anhui Province, China; [3] Department of Anesthesiology, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing City, Jiangsu Province, China; [4] Chinese PLA Center for Disease Control and Prevention, Beijing, China; [5] The Second People's Hospital of Yiyuan County, Zibo City, Shandong Province, China; [6] Department of Infectious Diseases, The First Affiliated Hospital of Nanjing Medical University, Nanjing City, Jiangsu Province, China; [7] School of Public Health, Nanjing Medical University, Nanjing City, Jiangsu Province, China.
[&] Joint first authors.

Submitted: June 21, 2023; Accepted: September 06, 2024

REFERENCES

1. Yu XJ, Liang MF, Zhang SY, Liu Y, Li JD, Sun YL, et al. Fever with thrombocytopenia associated with a novel bunyavirus in China. N Engl J Med 2011;364(16):1523 – 32. https://doi.org/10.1056/NEJMoa1010095.
2. Li H, Lu QB, Xing B, Zhang SF, Liu K, Du J, et al. Epidemiological and clinical features of laboratory-diagnosed severe fever with thrombocytopenia syndrome in China, 2011-17: a prospective observational study. Lancet Infect Dis 2018;18(10):1127 – 37. https://doi.org/10.1016/S1473-3099(18)30293-7.
3. Sun JM, Lu L, Liu KK, Yang J, Wu HX, Liu QY. Forecast of severe fever with thrombocytopenia syndrome incidence with meteorological factors. Sci Total Environ 2018;626:1188 – 92. https://doi.org/10.1016/j.scitotenv.2018.01.196.
4. Sun JM, Lu L, Wu HX, Yang J, Ren JP, Liu QY. The changing epidemiological characteristics of severe fever with thrombocytopenia syndrome in China, 2011-2016. Sci Rep 2017;7(1):9236. https://doi.org/10.1038/s41598-017-08042-6.
5. Li JC, Zhao J, Li H, Fang LQ, Liu W. Epidemiology, clinical characteristics, and treatment of severe fever with thrombocytopenia syndrome. Infect Med 2022;1(1):40 – 9. https://doi.org/10.1016/j.imj.2021.10.001.
6. Mehand MS, Millett P, Al-Shorbaji F, Roth C, Kieny MP, Murgue B. World health organization methodology to prioritize emerging infectious diseases in need of research and development. Emerg Infect Dis 2018;24(9):e171427. https://doi.org/10.3201/eid2409.171427.
7. Wang T, Li XL, Liu M, Song XJ, Zhang H, Wang YB, et al. Epidemiological characteristics and environmental risk factors of severe fever with thrombocytopenia syndrome in Hubei province, China, from 2011 to 2016. Front Microbiol 2017;8:387. https://doi.org/10.3389/fmicb.2017.00387.
8. Battineni G, Chintalapudi N, Amenta F. Forecasting of COVID-19 epidemic size in four high hitting nations (USA, Brazil, India and Russia) by Fb-Prophet machine learning model. Appl Comput Inf 2020;6(1):1 – 10. https://doi.org/10.1108/ACI-09-2020-0059.
9. Lv CR, Guo WQ, Yin XY, Liu L, Huang XL, Li SM, et al. Innovative applications of artificial intelligence during the COVID-19 pandemic. Infect Med 2024;3(1):100095. https://doi.org/10.1016/j.imj.2024.100095.
10. Xie C, Wen HY, Yang WW, Cai J, Zhang P, Wu R, et al. Trend analysis and forecast of daily reported incidence of hand, foot and mouth disease in Hubei, China by Prophet model. Sci Rep 2021;11(1):1445. https://doi.org/10.1038/s41598-021-81100-2.
11. Deng B, Rui J, Liang SY, Li ZF, Li KG, Lin SN, et al. Meteorological factors and tick density affect the dynamics of SFTS in jiangsu province, China. PLoS Negl Trop Dis 2022;16(5):e0010432. https://doi.org/10.1371/journal.pntd.0010432.
12. Jiang XL, Wang YG, Zhang XM, Pang B, Yao MX, Tian XY, et al. Factors associated with severe fever with thrombocytopenia syndrome in endemic areas of China. Front Public Health 2022;10:844220. https://doi.org/10.3389/fpubh.2022.844220.
13. Wang ZJ, Yang ST, Luo L, Guo XH, Deng B, Zhao ZY, et al. Epidemiological characteristics of severe fever with thrombocytopenia syndrome and its relationship with meteorological factors in Liaoning province, China. Parasit Vectors 2022;15(1):283. https://doi.org/10.1186/s13071-022-05395-4.
14. Luo ZX, Jia XC, Bao JZ, Song ZJ, Zhu HL, Liu MY, et al. A combined model of SARIMA and prophet models in forecasting AIDS incidence in Henan province, China. Int J Environ Res Public Health 2022;19(10):5910. https://doi.org/10.3390/ijerph19105910.
15. Huang D, Grifoll M, Sanchez-Espigares JA, Zheng PJ, Feng HX. Hybrid approaches for container traffic forecasting in the context of anomalous events: the case of the Yangtze River Delta region in the COVID-19 pandemic. Transp Policy (Oxf) 2022;128:1 – 12. https://doi.org/10.1016/j.tranpol.2022.08.019.

SUPPLEMENTARY MATERIAL

Data Preprocessing

Ensuring that all timestamps in the dataset have a consistent format is crucial for accurate data processing and time series analysis. All timestamps were converted to the YYYY-MM format. This uniformity lays the foundation for applying various time series models effectively. The data were divided into training data and prediction data. Data starting from January 2020 to December 2020 were used as prediction data, while the rest as training data.

SARIMA Model

The autoregressive integrated moving average (ARIMA) and seasonal ARIMA (SARIMA) models are the most general class of models for forecasting a time series in theory (*1–2*). The ARIMA model aims to describe the autocorrelations in the data by outlining its components, including Autoregression (AR), Integrated (I), and Moving Average (MA). SARIMA is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component, adding three new hyperparameters to specify the seasonal component. The SARIMA model requires selecting hyperparameters for both the trend elements (trend autoregression order p, trend difference order d, and trend moving average order q) and seasonal elements (seasonal autoregressive order P, seasonal difference order D, and seasonal moving average order Q) of the series. The formula for SARIMA is as follows:

$$(1-B)^d\left(1-B^s\right)^D Y_t = \theta_0 + \frac{\theta(B)\theta_s(B^s)}{\phi(B)\phi_s(B^s)}\sigma_t \quad (1)$$

where Y_t refers to the value of the time series at time t, θ_0 is constant, ε_t is the white noise value at period t, and the parameters d and D represent the difference number and seasonal difference number, respectively. B is the backshift operator, $\phi(B)$ is the autoregressive operator and $\theta(B)$ is the moving average operator. $\phi_s(B^s)$ and $\theta_s(B^s)$ are the seasonal operators.

The construction process of the SARIMA model is as follows.

Grid Search: The model starts by defining possible combinations of parameters for the seasonal aspects of the time series. It uses a grid search approach where p, d, and q values (representing autoregressive, differencing, and moving average terms, respectively) are tested along with seasonal counterparts.

AIC Evaluation: For each combination, a SARIMA model is fitted, and the Akaike Information Criterion (AIC) is calculated to assess model fit. The combination with the lowest AIC is considered optimal as it suggests a model that best explains the data with minimal complexity.

Best parameters selection: Use the best parameters determined through grid search, and then fit the SARIMA model to data before the specified date (in this case, January 2020).

Diagnostic check: Perform the Ljung Box test on the residuals to check for white noise, which indicates that the model's residuals have no autocorrelation and the model has fully captured the information in the data.

Prediction: The model performs a step-by-step (single-step-ahead) forecast using the trained model, constantly updating with actual data as it becomes available. This simulates a real-world scenario where predictions are made as new data comes in.

Evaluation: As indicated in part "Performance Evaluations" below.

Prophet Model

The Prophet model is a sophisticated method for forecasting time series, particularly tailored for business data and adept at navigating through complex trends and habitual seasonal variations. It provides a versatile treatment of trends, seasonality, and holiday influences. The trend component *g(t)* is engineered to capture non-periodic changes in the time series. It can employ either a logistic growth model for data with saturation limits or a piecewise linear model for data without clear saturation points.

The foundational equation of the Prophet model is expressed as:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t \quad (2)$$

Where yt denotes the predicted value at time t, s(t) is the seasonality component, h(t) represents the impact of holidays or specific events on the time series, and ε t is the error term accounting for aspects of the data not

explained by the model. The model also incorporates an advanced feature for automatically detecting change points in trends.

The construction process for the Prophet model is as follows:

Initiation: Prophet model is initialized with a specific configuration, including Growth Trend, Seasonality Components, Seasonality Mode, and Regularization Parameters.

Model Training: The Prophet model is fitted on the training data.

Prediction: Future dates are generated (future) for 12 months (periods=12) with monthly frequency (freq='M'). Forecasting is performed [forecast=m.predict(future)], and predictions for the year 2020 are extracted (forc).

Evaluation: As indicated in part "Performance Evaluations" below.

XGBoost

eXtreme Gradient Boosting (XGBoost) is an advanced optimization technique based on Gradient-boosting decision trees (GBDT). It operates by iteratively constructing a series of short, basic decision trees, each termed as a "weak learner". The process begins with the construction of an initial tree that exhibits subpar performance. Subsequent trees are then trained to correct the errors of their predecessors. This sequence of producing weaker learners continues until a stopping condition is met, such as reaching a predetermined number of trees. This method has been demonstrated to be effective in predicting human brucellosis (*3*) and renal hemorrhagic fever syndrome (*4*).

For a dataset with n examples and m features, a tree ensemble model in XGBoost predicts the output using K additive functions:

$$\hat{y}_k = \sum_{k=1}^{K} f_k(x_t), f_k \in F \tag{3}$$

Here, $\hat{y}$represents the predicted value for the i-th sample, fk is a function corresponding to the k-th tree, and F denotes the space of regression tree functions, with xi being the feature vector for the i-th sample.

To learn the set of functions used in the model, XGBoost minimizes the following regularized objective:

$$L(\varphi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{4}$$

In this equation, $l(yi, \hat{yi})$ is the loss function quantifying the error between the observed and predicted data, and Ω(fk) is the regularization term that aids in smoothing the learned weights to prevent overfitting and encourage the model to select simpler yet predictive functions. The regularization term is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{5}$$

Where γ and λ are regularization parameters, T is the number of leaves in the tree, and wj represents the score on each leaf.

SUPPLEMENTARY TABLE S1. Parameters of the optimized SARIMA model.

Parameters	Coefficient	Standard Errors	Z	*P*
ar.L1	0.424	0.133	3.190	0.001
ma.L1	−0.923	0.101	−9.136	<0.001
ma.S.L12	−0.878	0.242	−3.621	<0.001
sigma2	118.976	25.138	4.733	<0.001

SUPPLEMENTARY TABLE S2. Comparison of four models using MAE and RMSE.

Error metrics	SARIMA	Prophet	XGBoost	LSTM
MAE	5.47	6.64	2.54	4.75
RMSE	7.21	8.50	2.89	6.77

Abbreviation: MAE=mean absolute error; RMSE=root mean square error.

The construction process of the XGBoost model is as follows:

Parameter optimization: Use methods like Bayesian optimization to optimize the parameters.

Model Training: The XGBoost model is fitted on the training data.

Prediction: The model performs a step-by-step (single-step ahead) forecast using the trained model, constantly updating with actual data as it becomes available. This simulates a real-world scenario where predictions are made as new data comes in.

Evaluation: As indicated in part "Performance Evaluations" below.

LSTM Networks

Long Short-Term Memory (LSTM) networks, a specialized type of Recurrent Neural Networks (RNNs), excel at capturing both short-term and long-term dependencies in sequential data. This is primarily due to the unique architecture of the LSTM, which incorporates a cell state acting as a form of memory. This cell state is crucial for retaining information across various time steps, addressing the limitations of traditional RNNs. Additionally, LSTMs effectively alleviate the vanishing and exploding gradient issues commonly encountered in standard RNNs, particularly in lengthy sequences. The key feature of LSTM networks lies in their gating mechanism, comprising three types of gates, each with specific roles and formulas.

The input gate in an LSTM is pivotal in regulating the influx of new information into the cell state. This gate operates in two steps. The first step entails a sigmoid function that determines the essential update values, represented by the equation:

$$i_t = \sigma\left(W_4 x_t + W_4 h_{4-1} + b_t\right) \tag{6}$$

The second step employs a tanh function to generate a vector of new candidate values that may be added to the state, given by:

$$C_t = tanh\left(W_{x_C} x_t + W_{h_C} h_{t-1} + b_C\right) \tag{7}$$

Here, it is the activation of the input gate, and C_t is the candidate vector for the cell state update.

The forget gate in an LSTM decides which information from the cell state should be retained or discarded. It operates using a sigmoid function that evaluates the importance of the existing information in the cell state, defined by:

$$f_t = \sigma\left(W_{x_f} x_t + W_{h_f} h_{t-1} + b_f\right) \tag{8}$$

The activation vector ft indicates the extent to which past information is to be forgotten or retained.

The output gate in an LSTM manages the output sent to the next layer. This gate operates in two stages. Initially, a sigmoid function determines which parts of the cell state will be outputted, as shown by:

$$o_t = \sigma\left(W_{x_t} x_t + W_{A_j} h_{t-1} + b_0\right) \tag{9}$$

Then, the final output is calculated by multiplying this activation o_t with the tanh of the cell state, resulting in:

$$h_t = o_t \times \tanh\left(C_t\right) \tag{10}$$

The output vector ht represents the information transmitted to subsequent layers or units in the network.

The construction process of the LSTM model is as follows:

Network architecture: Build a neural network containing an LSTM layer, which extracts features from the input sequence, followed by a fully connected layer that outputs prediction results.

Loss function and model training: Use mean squared error as the loss function and use Adam optimizer for model training.

Early Stopping: To avoid overfitting, the training is stopped when the loss on the validation set no longer improves.

Parameter optimization: Use methods like Bayesian optimization to optimize the parameters.

Model Training: The LSTM model is fitted on the training data.

Prediction: Rolling prediction of future time points involves using a model to gradually predict future values, and after each prediction step, the results are fed back into the input data for the next prediction step.

Evaluation: As indicated in part "Performance Evaluations" below.

In the models mentioned above, a data post-processing technique known as clipping is utilized to handle unrealistic negative values in the results. This involves adjusting all negative forecast values to zero, ensuring data consistency and interpretability, and preventing potential analytical errors stemming from impractical negative predictions.

Performance Evaluations

The predictive performance of the models was assessed using two indexes: mean absolute error (MAE) and root mean squared error (RMSE).

MAE is a metric used to measure the average absolute errors between actual and predicted values in a dataset. It is calculated by taking the average of the absolute differences between the actual values and the predicted values. MAE is often used in regression analysis to evaluate the accuracy of a regression model. A lower MAE indicates better accuracy of the model, as it means that the model's predictions are closer to the actual values.

RMSE is another metric used to measure the accuracy of a regression model by calculating the square root of the average of the squared differences between actual and predicted values in a dataset. RMSE penalizes larger errors more heavily compared to MAE because it squares the errors before taking the square root. This means that outliers or large errors have a bigger impact on the RMSE compared to the MAE. Similar to MAE, a lower RMSE indicates better accuracy of the model. RMSE is often preferred when a small number of large errors are more significant than a large number of small errors.

The formulas are defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_t - \hat{y}_j\right)^2} \tag{12}$$

Software

The descriptive statistics and time series modeling were conducted using Python 3.7. The SARIMA model, Prophet model, XGBoost model, and LSTM model were implemented using the statsmodels package, fbprophet package, scikit-learn package, and Keras package, respectively. In the analysis, a $P<0.05$ was considered significant.

REFERENCES

1. Kim KR, Park JE, Jang IT. Outpatient forecasting model in spine hospital using ARIMA and SARIMA methods. J Hosp Manage Health Policy 2020;4:20. https://doi.org/10.21037/jhmhp-20-29.
2. ArunKumar KE, Kalaga DV, Sai Kumar CM, Chilkoor G, Kawaji M, Brenza TM. Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). Appl Soft Comput 2021;103:107161. https://doi.org/10.1016/j.asoc.2021.107161.
3. Alim M, Ye GH, Guan P, Huang DS, Zhou BS, Wu W. Comparison of ARIMA model and XGBoost model for prediction of human brucellosis in mainland China: a time-series study. BMJ Open 2020;10(12):e039676. https://doi.org/10.1136/bmjopen-2020-039676.
4. Lv CX, An SY, Qiao BJ, Wu W. Time series analysis of hemorrhagic fever with renal syndrome in mainland China by using an XGBoost forecasting model. BMC Infect Dis 2021;21(1):839. https://doi.org/10.1186/s12879-021-06503-y.

Outbreak Reports

A Norovirus-Related Gastroenteritis Outbreak Stemming from a Potential Source of Infection — Pudong New Area, Shanghai Municipality, China, April 2024

Zou Chen[1,2]; Hong Zhang[1,2]; Yifeng Shen[1,2]; Chuchu Ye[1,2,#]

Summary

What is already known about this topic?

Noroviruses are highly infectious with rapid transmission capabilities, causing illness for an average duration of 12–60 hours. In China, individuals in educational agencies may return to class 72 hours after symptom resolution.

What is added by this report?

This outbreak was precipitated by a potential source of infection in a child resuming class after a 72-hour quarantine post-symptom resolution, leading to a cluster of cases within the class.

What are the implications for public health practice?

While extending the quarantine period for children may be considered from a safety perspective, it is a challenge for educational agencies. The outbreak is deemed a low-probability event; however, further investigation into the detoxification period of asymptomatic patients is warranted.

On April 27, 2024, at 13:00, the Pudong New Area Disease Prevention and Control Center (Pudong CDC), Shanghai Municipality, China, received reports from the Market Bureau of multiple children in the same class at PN Kindergarten exhibiting vomiting. Pudong CDC promptly collaborated with the community health service center to conduct an epidemiological investigation and intervention. The objectives were to identify the pathogen, delineate the outbreak's characteristics and potential transmission risks, and implement effective control strategies.

INVESTIGATION AND RESULTS

PN Kindergarten, located at 609 LS Road, Pudong New Area, Shanghai, China, is a public institution serving 219 children in four grades across nine classes. The school employs 38 faculty and staff members, including 2 health teachers. The kindergarten lacks school buses and on-site accommodations. It features a teaching building with various public classrooms, including a reading room, all of which have been in recent use. The air conditioning system was recently turned off. The kitchen and canteen provide meals for both children and staff, who dine separately in designated areas. Kitchen personnel and teachers reported good health over the previous two weeks. The kindergarten provides water cups for student use. Disinfection practices are well-maintained, with complete records.

The first case was a 4-year-old child from junior Class II who began experiencing vomiting, abdominal pain, nausea, and fatigue at home at 17:00 on April 26. The child did not exhibit chills or urgency. After experiencing 10 vomiting episodes, the child was diagnosed with acute gastroenteritis at the hospital at 20:00 that evening and presented with elevated white blood cell counts. The parents reported no history of suspicious food intake, contact with suspected cases, or recent travel. No similar symptoms were observed in the cohabiting parents, and the child had no record of recent vaccinations.

Case definition: According to the 2015 *Norovirus Outbreak Investigation and Prevention Technical Guidelines* (*1*), a suspected case of norovirus-related gastroenteritis is defined as an individual experiencing vomiting at least twice or diarrhea at least three times within a 24-hour period. Cases were identified through classroom-based searches, teacher interviews, and review of absence records.

As of 16:00 on April 27, 11 cases met the suspect case definition through case searching, all concentrated in junior Class II. The attack rate among schoolchildren was 5.02%, affecting 11 of 219 individuals. Onset times ranged from 17:00 on April 26 to 5:00 on April 27, with all cases occurring at home. Notably, 10 cases (90.91%) emerged on April 26, and 1 case (9.09%) on April 27. Among the 11

cases, there were 7 males and 4 females, with a male-to-female ratio of 1.75∶1. Clinical presentations are detailed in Table 1, with all cases presenting mild symptoms and no severe conditions reported.

Further investigation revealed that a primary case, a potential source of infection, experienced three vomiting episodes at home on Saturday, April 20. The vomiting consisted of food and lasted for one day, accompanied by two episodes of diarrhea with sticky stools, nausea, and weakness. The child rested at home, and symptoms noticeably subsided by April 21. The child continued to rest at home for three days, from April 22 to 24. After 72 hours, the child returned to kindergarten on Thursday, April 25. According to parental feedback, the child had not exhibited symptoms such as vomiting or diarrhea since April 22. Subsequently, 11 children from the same class gradually began to experience illness after 17:00 on April 26.

On April 27, an on-site investigation was conducted, and 14 samples were collected. These included one vomit sample from the index case, one anal swab sample from the suspected source of infection, one fecal sample from one of the 11 cases, and 11 environmental samples from classrooms, toilets, and dining tables. The samples were sent to the Pudong CDC Laboratory for pathogen testing, including *Norovirus*, *Rotavirus*, *Astrovirus*, *Vibrio parahaemolyticus*, pathogenic *Escherichia coli*, *Salmonella*, and other pathogens, using real-time polymerase chain reaction or bacterial culture. Testing revealed that three case samples were positive for *Norovirus* GII nucleic acid, while all environmental samples tested negative.

During the investigation, the Pudong CDC implemented several public health responses, including: 1) strengthening morning checks and full-day observations at the kindergarten, with immediate reporting of new cases; 2) quarantining cases at home, mandating a 72-hour symptom-free period before resumption of activities; 3) setting disinfection standards for environmental surfaces at 1,000 mg/L for classrooms and 2,000 mg/L for bathrooms; 4) suspending collective activities and public classroom use, and promptly issuing parental notices to manage public perception; 5) actively promoting awareness of intestinal infectious diseases and hand hygiene; and 6) maintaining close communication with the kindergarten and Pudong CDC. As of May 6, all cases associated with this outbreak had recovered and resumed classes, with no new cases reported. All control measures have been implemented, and this outbreak is considered closed.

DISCUSSION

Norovirus is a member of the genus *Calicivirus* and is recognized as a significant pathogen causing non-bacterial acute gastroenteritis (*2*), as well as a common agent in foodborne infectious diseases (*3*). In developed countries, norovirus is responsible for over 50% of non-bacterial diarrhea outbreaks (*4*). With advancements in detection technology, norovirus has come under increased scrutiny as a key pathogen in viral diarrhea. Characterized by its potent pathogenicity, infection can occur with as few as 10–100 viral particles (*5*–*6*), leading to gastrointestinal symptoms such as nausea, vomiting, and diarrhea. These characteristics make norovirus particularly prone to outbreaks in settings with dense populations, such as kindergartens, schools, and restaurants.

Norovirus, characterized by its high infectivity and rapid transmissibility, is a predominant cause of sporadic and outbreak-related cases of acute gastroenteritis globally, posing a substantial health burden. Since 2013 in China, norovirus infection has been the primary cause of outbreaks of other infectious diarrhea diseases (infectious diarrhea diseases other than cholera, bacterial and amoebic dysentery, typhoid

TABLE 1. Distribution of clinical manifestations.

Symptoms	Number of cases (*N*=11)	Percentage (%)	Remarks
Vomiting	11	100	Occurred 3 to 20 times
Abdominal Pain	9	82	–
Leukocytosis	9	90	10 cases visited the hospital, 8 cases with white blood cells over 20×10^9/L
Nausea	6	55	–
Fatigue	3	27	–
Diarrhea	1	9	Occurred 4 times

Note: the dash indicates no remarks.

and paratyphoid fever), with a significant increase in the number of outbreaks compared to previous years (*7*–*9*).

This outbreak event was confirmed based on clinical symptoms, epidemiological investigations, and laboratory testing results. Supporting evidence includes: 1) all cases occurring within the same class; 2) the synchronized onset of symptoms within a 12-hour window after returning home, consistent with the typical norovirus incubation period; 3) no close contact between the affected class and other classes; 4) a thorough case search ensuring identification of all cases; and 5) the prompt response, early reporting, and effective preventive measures implemented by the kindergarten.

The outbreak was hypothesized to originate from a potential source of infection within the class following an individual's return after the 72-hour quarantine period post-symptom resolution. This assumption is supported by: 1) the predominant symptom of vomiting among the children, corroborated by laboratory findings confirming norovirus infection; 2) the timing of the potential source's return to class coinciding with the outbreak's start and genetic alignment with other outbreak isolates; and 3) the exclusion of other exposure possibilities, such as dining conditions and the health status of cooks, teachers, and children in other classes.

Post-infection, individuals can shed norovirus during the incubation period, peaking 2–5 days after symptom onset and persisting for approximately 2–3 weeks, with the longest reported shedding period exceeding 56 days (*10*). Standardized management protocols are crucial for controlling transmission and minimizing environmental contamination. The current guideline in China dictates a quarantine period extending from the acute phase until 72 hours after symptom resolution, as viral shedding significantly decreases thereafter. However, given norovirus's extended shedding time, transmission may still occur beyond the 72-hour mark, potentially contributing to the source of infection in this outbreak.

However, this investigation had limitations. Although the outbreak was traced to an early case (a potential source of infection), positive norovirus tests do not guarantee infectivity, and not all cases underwent testing, particularly asymptomatic carriers. Furthermore, the potential source of infection was asymptomatic upon returning to the kindergarten. This, along with the lack of positive environmental findings, suggests that the evidence linking this individual to the outbreak is inconclusive.

Norovirus currently comprises at least six genogroups, further classified into at least 30 genotypes (*11*), with GI, GII, and GIV known to infect humans. The confirmed pathogen of this outbreak is norovirus GII. Norovirus mutates rapidly, and new variants that can cause global epidemics emerge every 2–3 years.

Norovirus infection presents with fever, diarrhea, nausea, vomiting, and abdominal pain, often accompanied by secondary symptoms such as headache, discomfort, and fatigue, typically manifesting as acute gastroenteritis. Although China classifies norovirus as a Class C infectious disease for reporting and management, detection and reporting may be delayed, particularly when vomiting is the primary symptom. In this outbreak, vomiting was ubiquitous, while other symptoms were notably absent, consistent with literature indicating that children exhibit vomiting more frequently than adults (*12*).

In light of the findings, we advocate for further research into the viral shedding duration in asymptomatic norovirus gastroenteritis cases, which is essential for refining outbreak detection and containment strategies.

Conflicts of interest: No conflicts of interest.

Funding: Supported by the Key Discipline (GWVI-11.1-02 Infectious Diseases) of the three-year action plan for strengthening the construction of the public health system in Shanghai (2023–2025), Pudong New Area Science and Technology Development Innovation fund (No. PKJ2023-Y73), and Shanghai Pudong New Area Center for Disease Control and Prevention Technology Project (PDCDC-KJ-2024-05).

doi: 10.46234/ccdcw2024.197

[#] Corresponding author: Chuchu Ye, ccye@pdcdc.sh.cn.

[1] Shanghai Pudong New Area Center for Disease Control and Prevention, Shanghai, China; [2] Fudan University Pudong Institute of Preventive Medicine, Shanghai, China.

Submitted: June 30, 2024; Accepted: August 14, 2024

REFERENCES

1. Liao QH, Ran L, Jin M, Cui SH, Yuan J, Ma HL, et al. Guidelines on outbreak investigation, prevention and control of Norovirus infection (2015). Chin J Viral Dis 2015;5(6):448 – 58. https://doi.org/10.16505/j.2095-0136.2015.06.003.
2. Atmar RL, Estes MK. The epidemiologic and clinical importance of norovirus infection. Gastroenterol Clin North Am 2006;35(2):275 – 90. https://doi.org/10.1016/j.gtc.2006.03.001.
3. Netzler NE, Enosi Tuipulotu D, White PA. Norovirus antivirals: where are we now? Med Res Rev 2019;39(3):860-86. http://dx.doi.org/10.

1002/med.21545.
4. Takahashi M, Takahashi H, Kuda T, Kimura B. Viability and heat resistance of murine norovirus on bread. Int J Food Microbiol 2016;216:127 – 31. https://doi.org/10.1016/j.ijfoodmicro.2015.09.018.
5. Atmar RL, Opekun AR, Gilger MA, Estes MK, Crawford SE, Neill FH, et al. Norwalk virus shedding after experimental human infection. Emerg Infect Dis 2008;14(10):1553 – 7. https://doi.org/10.3201/eid1410.080117.
6. Teunis PFM, Moe CL, Liu PB, Miller SE, Lindesmith L, Baric RS, et al. Norwalk virus: how infectious is it? J Med Virol 2008;80(8):1468-76. http://dx.doi.org/10.1002/jmv.21237.
7. Lu J, Sun LM, Fang L, Yang F, Mo YL, Lao JQ, et al. Gastroenteritis outbreaks caused by norovirus GII.17, Guangdong Province, China, 2014-2015. Emerg Infect Dis 2015; 21(7):1240-2. http://dx.doi.org/10.3201/eid2107.150226.
8. Fu J, Ai J, Jin M, Jiang C, Zhang J, Shi C, et al. Emergence of a new GII.17 norovirus variant in patients with acute gastroenteritis in Jiangsu, China, September 2014 to March 2015. Euro Surveill 2015;20(24):21157. http://dx.doi.org/10.2807/1560-7917.es2015.20.24.21157.
9. Han JK, Ji L, Shen YH, Wu XF, Xu DS, Chen LP. Emergence and predominance of norovirus GII.17 in Huzhou, China, 2014-2015. Virol J 2015;12:139. http://dx.doi.org/10.1186/s12985-015-0370-9.
10. Teunis PFM, Sukhrie FHA, Vennema H, Bogerman J, Beersma MFC, Koopmans MPG. Shedding of norovirus in symptomatic and asymptomatic infections. Epidemiol Infect 2015;143(8):1710 – 7. https://doi.org/10.1017/S095026881400274X.
11. Vinjé J. Advances in laboratory methods for detection and typing of norovirus. J Clin Microbiol 2015;53(2):373 – 81. https://doi.org/10.1128/JCM.01535-14.
12. Götz H, Ekdahl K, Lindbäck J, de Jong B, Hedlund KO, Giesecke J. Clinical spectrum and transmission characteristics of infection with Norwalk-like virus: findings from a large community outbreak in Sweden. Clin Infect Dis 2001;33(5):622 – 8. https://doi.org/10.1086/322608.

Commentary

Intelligent Forest Hospital as a New Management System for Hospital-Acquired Infection Control

Yingxin Liu[1]; Zhousheng Lin[2]; Guanwen Lin[3]; Wanmin Lian[4]; Junzhang Tian[5]; Guowei Li[1,6,7,#]; Hongying Qu[1,5,#]

ABSTRACT

Hospital-acquired infection (HAI) is a significant global health concern, elevating the risks of morbidity and imposing a substantial socioeconomic burden. To enhance the management of HAI, particularly in the aftermath of the coronavirus disease 2019 (COVID-19) pandemic, the Guangdong Second Provincial General Hospital (GD2H) has launched a new system called Intelligent Forest Hospital (IFH). Leveraging advancements in artificial intelligence, 5G technology, and cloud networking, the IFH implements customized indoor air quality (IAQ) control strategies tailored to different medical settings. It utilizes various intelligent disinfection devices and air purification systems. The IFH features a dynamic 3D hospital model with real-time monitoring of crucial IAQ parameters and a risk assessment ranking for clinical departments, providing timely risk alerts, communication prompts, and automatic disinfection processes. The IFH aims to effectively mitigate HAI post-COVID-19 and other future pandemics, ensuring a safe and pleasant environment for patients, hospital staff, and visitors.

Hospital-acquired infection (HAI), or healthcare-associated infection, is a significant global health concern linked to increased morbidity and socioeconomic burden (*1*). The coronavirus disease 2019 (COVID-19) pandemic has highlighted the importance of addressing HAI, with reported rates ranging from 8.3% to 23.4% (*2*). HAI, including COVID-19, primarily spread through airborne and contaminated surface routes (*3*). Pathogens can persist in hospital environments for extended periods, with some resistant microorganisms surviving for years (*4*). Traditional cleaning methods often struggle to eliminate these pathogens effectively (*5*). Research has shown that even after four rounds of bleach disinfection, around 26.6% of hospital rooms remained contaminated (*6*). Therefore, adopting new technologies to maintain a safe hospital environment and improve indoor air quality (IAQ) is crucial for HAI prevention and control (*7*).

In an effort to enhance the control of HAI, particularly in the post-COVID era, Guangdong Second Provincial General Hospital (GD2H) has introduced a novel management system called the *Intelligent Forest Hospital* (IFH). Drawing on advancements in artificial intelligence (AI), 5G technology, and cloud networking that have been effectively integrated into hospital operations at GD2H (*8*–*10*), the IFH has been tailored to manage IAQ in various medical settings. This system involves the deployment of specialized disinfection equipment and air purification terminals in areas such as delivery rooms, operating rooms, intensive care units, respiratory and infection wards, and emergency departments — locations where the risk of HAI for both healthcare staff and patients is notably high.

While cleaning is crucial for preventing HAI, it is essential to consider hospital design, disinfection practices, and surface composition. GD2H utilizes various intelligent equipment categorized into disinfection [e.g., negative air ion (NAI) generator, plasma air sterilizer, ultraviolet sterilizer], surface composition (e.g., antibacterial coating, antibacterial fabric), and detection (e.g., microbial detection system). This combination effectively inhibits pathogens and controls IAQ parameters such as CO_2 levels, formaldehyde, $PM_{2.5}$, and particle mass concentration. In the outpatient hall of GD2H, where over 5,000 patients are served daily, around 60 air purification devices have been installed, including ceiling oxygen equipment, disinfection lamps, and plasma disinfection machines (Figure 1). The NAI concentration in the hospital reaches forest environment standards (>3,000/cm^3), with some areas exceeding 100,000/cm^3. Forest environments have high NAI concentrations known to benefit human health, including mental health, cognitive and

FIGURE 1. Ceiling oxygen equipment in the outpatient hall.

cardiorespiratory function (*11*–*13*), which inspired the IFH concept to create a safe and comfortable environment similar to a forest sanatorium.

In 2020, GD2H collaborated with Huawei to develop the Hospital Intelligent Data Twins, which are essentially digital twins within the framework of Industry 4.0. These twins function as an integrated system to facilitate the seamless exchange of digitalized and visualized data. At the core of the Intelligent Data Twins lies cloud services that leverage AI techniques to enhance data utilization (*10*). All data gathered by the intelligent equipment at GD2H are concurrently transmitted to a virtual cloud platform within the hospital premises for IAQ monitoring and management. The central management hub, serving as the hospital's *smart brain*, continuously monitors IAQ across all areas of GD2H round the clock. By harnessing AI, 5G technology, cloud services, and a virtual platform, the Intelligent Data Twins effectively process and receive vast amounts of data from all intelligent equipment, transmitting this information to the cognitive center for real-time IAQ visualization at GD2H (*10*). The primary interface of the cognitive center exhibits a dynamic 3D model diagram of the hospital, providing real-time monitoring of critical IAQ parameters such as pathogenic microorganisms, carbon dioxide, formaldehyde, $PM_{2.5}$, and negative ions (Figure 2). Moreover, the cognitive center conducts a comprehensive analysis of the continuously updated real-time data, evaluating HAI risks from various sources. It generates a risk ranking index for clinical departments on screen, enabling prompt risk alerts and initiating automated disinfection measures when necessary. For example, upon detecting abnormal IAQ parameters in infection wards, the cognitive center promptly assesses the situation and triggers the automatic disinfection equipment in the affected area. Simultaneously, it communicates the IAQ status to the nurse stations in the infected wards to inform healthcare providers and patients about potential contamination and the ongoing automated disinfection process.

There are various challenges that need attention in the integration of IFH using technologies like AI and 5G. These challenges include the high costs and energy consumption, the absence of medical standards and regulations, ethical concerns regarding patient privacy, societal acceptance, real-time monitoring difficulties, and the general applicability of 5G and virtual cloud platforms. The IFH requires enhancements such as regular system updates for real-time data processing, improved data privacy and security measures, and further assessments of costs and benefits. Specific training and adjustments are necessary to ensure the system's effectiveness in various contexts and to improve its generalizability. Despite these challenges, integrating IFH with AI and 5G technologies shows promise in enhancing the prevention and management of HAI, particularly in the context of post COVID-19 and future pandemics, aiming to create a safe and

FIGURE 2. The dynamic 3D hospital model diagram displaying real-time monitoring of key indoor air quality parameters on the main screen of the smart brain.

pleasant environment for patients, hospital staff, and visitors.

Conflicts of interest: No conflicts of interest.

Acknowledgements: Thank the staff at Intelligent Forest Hospital for their assistance in providing the essential materials for this manuscript.

Funding: Supported by the Science Foundation of Guangdong Second Provincial General Hospital (number: YY2018-002, recipient: GL).

doi: 10.46234/ccdcw2024.201

[#] Corresponding authors: Guowei Li, ligw@gd2h.org.cn; Hongying Qu, tggd2h@163.com.

[1] Center for Clinical Epidemiology and Methodology (CCEM), The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou City, Guangdong Province, China; [2] Medical Department, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou City, Guangdong Province, China; [3] Hospital-Acquired Infection Control Department, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou City, Guangdong Province, China; [4] Information Department, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou City, Guangdong Province, China; [5] Institute for Healthcare Artificial Intelligence Application, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou City, Guangdong Province, China; [6] Father Sean O'Sullivan Research Centre, St. Joseph's Healthcare Hamilton, Hamilton, ON, Canada; [7] Department of Health Research Methods, Evidence, and Impact (HEI), McMaster University, Hamilton, ON, Canada.

Submitted: January 07, 2024; Accepted: February 27, 2024

REFERENCES

1. Cassini A, Plachouras D, Eckmanns T, Abu Sin M, Blank HP, Ducomble T, et al. Burden of six healthcare-associated infections on European population health: estimating incidence-based disability-adjusted life years through a population prevalence-based modelling study. PLoS Med 2016;13(10):e1002150. https://doi.org/10.1371/journal.pmed.1002150.
2. Klompas M, Karan A. Preventing SARS-CoV-2 transmission in health care settings in the context of the omicron variant. JAMA 2022;327(7):619 – 20. https://doi.org/10.1001/jama.2022.0262.
3. Tang JW, Bahnfleth WP, Bluyssen PM, Buonanno G, Jimenez JL, Kurnitski J, et al. Dismantling myths on the airborne transmission of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). J Hosp Infect 2021;110:89 – 96. https://doi.org/10.1016/j.jhin.2020.12.022.
4. Suleyman G, Alangaden G, Bardossy AC. The role of environmental contamination in the transmission of nosocomial pathogens and healthcare-associated infections. Curr Infect Dis Rep 2018;20(6):12. https://doi.org/10.1007/s11908-018-0620-2.
5. Otter JA, Yezli S, French GL. The role played by contaminated surfaces in the transmission of nosocomial pathogens. Infect Control Hosp Epidemiol 2011;32(7):687 – 99. https://doi.org/10.1086/660363.
6. Manian FA, Griesenauer S, Senkel D, Setzer JM, Doll SA, Perry AM, et al. Isolation of *Acinetobacter baumannii* complex and methicillin-resistant *Staphylococcus aureus* from hospital rooms following terminal cleaning and disinfection: can we do better? Infect Control Hosp Epidemiol 2011;32(7):667-72. http://dx.doi.org/10.1086/660357.
7. Saran S, Gurjar M, Baronia A, Sivapurapu V, Ghosh PS, Raju GM, et al. Heating, ventilation and air conditioning (HVAC) in intensive care unit. Crit Care 2020;24(1):194. https://doi.org/10.1186/s13054-020-02907-5.
8. Chen XJ, Tian JZ, Li GM, Li GW. Initiation of a new infection control system for the COVID-19 outbreak. Lancet Infect Dis 2020;20(4):397 – 8. https://doi.org/10.1016/s1473-3099(20)30110-9.
9. Li GW, Lian WM, Qu HY, Li ZY, Zhou QR, Tian JZ. Improving patient care through the development of a 5G-powered smart hospital. Nat Med 2021;27(6):936 – 7. https://doi.org/10.1038/s41591-021-01376-9.
10. Cheng WB, Lian WM, Tian JZ. Building the hospital intelligent twins for all-scenario intelligence health care. Digit Health 2022;8:1–4. https://doi.org/10.1177/20552076221107894
11. Chu CH, Chen SR, Wu CH, Cheng YC, Cho YM, Chang YK. The effects of negative air ions on cognitive function: an event-related potential (ERP) study. Int J Biometeorol 2019;63(10):1309 – 17. https://doi.org/10.1007/s00484-019-01745-7.
12. Liu S, Li C, Chu MT, Zhang WL, Wang WZ, Wang YZ, et al. Associations of forest negative air ions exposure with cardiac autonomic nervous function and the related metabolic linkages: a repeated-measure panel study. Sci Total Environ 2022;850:158019. https://doi.org/10.1016/j.scitotenv.2022.158019.
13. Jiang SY, Ma AL, Ramachandran S. Negative air ions and their effects on human health and air quality improvement. Int J Mol Sci 2018;19(10):2966. https://doi.org/10.3390/ijms19102966.

Youth Editorial Board

Indexed by Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), PubMed Central (PMC), Scopus, Chinese Scientific and Technical Papers and Citations, and Chinese Science Citation Database (CSCD)

The inauguration of *China CDC Weekly* is in part supported by Project for Enhancing International Impact of China STM Journals Category D (PIIJ2-D-04-(2018)) of China Association for Science and Technology (CAST).

CCDC weekly

Vol. 6 No. 37 Sept. 13, 2024

Responsible Authority
National Disease Control and Prevention Administration

Sponsor
Chinese Center for Disease Control and Prevention

Editing and Publishing
China CDC Weekly Editorial Office
No.155 Changbai Road, Changping District, Beijing, China
Tel: 86-10-63150501, 63150701
Email: weekly@chinacdc.cn

CSSN
ISSN 2096-7071 (Print)
ISSN 2096-3101 (Online)
CN 10-1629/R1