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

CHINA CDC WEEKLY

Vol. 6 No. 26 Jun. 28, 2024

CCDC weekly

中国疾病预防控制中心周报

INFECTIOUS DISEASE ISSUE

ISSN 2096-7071

China CDC Weekly

Cover Image: credited from this week's author Tingting Yan, China CDC staff.

Preplanned Studies

The Epidemiological Characteristics of Mpox Cases — China, 2023

Ruiqi Ren[1,2]; Chao Li[1,2]; Wenqing Bai[1,2]; Yali Wang[1,2]; Dan Li[1,2]; Fan Ding[1,2]; Niu Yan[1,2]; Qi Wang[1,2]; Nijuan Xiang[1,2]; Lianmei Jin[1,2]; Lei Zhou[1,2]; Wenling Wang[1,3]; Fan Lyu[1,4]; Peilong Li[1,4]; Dapeng Zhang[1,4]; Yan Li[1,5]; Quanyi Wang[1,6]; Qun Li[1,2]; Guoqing Shi[1,2,#]

Summary

What is already known about this topic?

Since May 2022, a global outbreak of mpox has emerged in more than 100 non-endemic countries. As of December 2023, over 90,000 cases had been reported. The outbreak has predominantly affected men who have sex with men (MSM), with sexual contact identified as the principal mode of transmission.

What is added by this report?

Since June 2023, China has faced an occurrence of mpox, predominantly affecting the MSM population. Approximately 90% of those affected reported engaging in homosexual behavior within 21 days prior to symptom onset, a trend that aligns with the global outbreak pattern. The prompt identification of cases, diligent tracing of close contacts, and the implementation of appropriate management strategies have successfully mitigated the spread of mpox virus in China.

What are the implications for public health practice?

We propose that mpox is transmitted locally within China. Drawing from our experiences in controlling the virus spread, it is crucial to investigate and formulate effective surveillance and educational strategies. Importantly, we must encourage high-risk populations to promptly seek medical care upon the onset of symptoms.

In May 2023, the WHO declared the termination of the Public Health Emergency of International Concern (PHEIC) initiated in July 2022 due to the mpox outbreak (*1*). During the PHEIC, China reported only one imported case in September 2022 (*2*). Subsequent to this, in June 2023, China reported its first local mpox cases. By December 31, 2023, there were 1,712 confirmed cases across 29 provincial-level administrative divisions (PLADs). Our study used national surveillance data, including all confirmed mpox cases reported following a standardized protocol and unified epidemiological form, to delineate the key epidemiological characteristics of the mpox cases reported in China from June to December 2023, thereby enhancing the understanding of the outbreak's initial local transmission dynamics.

Case surveillance, diagnosis, contact tracing, and management were conducted in adherence to the Mpox Prevention and Control Protocol. Local CDCs supplied detailed epidemiological data through comprehensive investigations following the identification of suspected or confirmed mpox cases. All case information was collected in line with the "Law of the People's Republic of China on Prevention and Treatment of Infectious Diseases," under the provisions for emergency response, thereby exempting the study from requiring ethics approval and participant consent. Additionally, individual data were de-identified to ensure patient privacy and confidentiality.

Descriptive statistics were used to summarize the epidemiological characteristics of the mpox cases. An epidemic curve, constructed from the dates of illness onset and diagnosis, depicted the trend of the epidemic. Demographic and epidemiological attributes of confirmed cases were presented on a monthly basis, using both absolute and relative frequencies. It should be noted that variations in case numbers across different categories may arise from incomplete data.

Among the 29 PLADs that reported confirmed cases of mpox, the highest numbers were observed in Guangdong, Beijing, Zhejiang, Sichuan, and Jiangsu, with counts of 342, 258, 183, 142, and 123 cases, respectively. Together, these regions accounted for 61.21% of all reported cases. Additionally, 24 PLADs (approximately 80%) reported fewer than 100 cases each, and 8 reported fewer than 10 cases.

According to the onset date curve depicted in Figure 1A, the initial case of mpox likely occurred in late May, succeeded by a steady rise in case numbers through the first 20 days of June. From late June

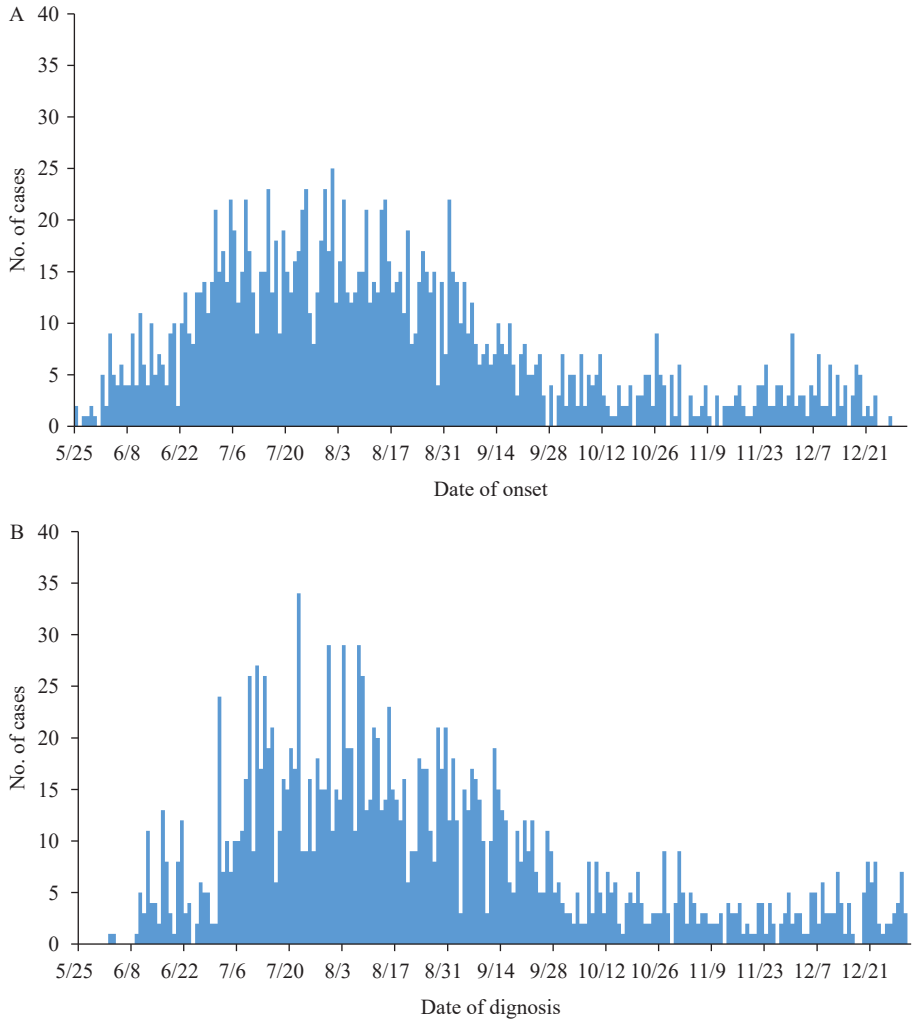


FIGURE 1. Epidemiological curves of confirmed mpox cases in China, June to December 2023. (A) Number of cases by date of onset; (B) Number of cases by date of diagnosis.

through the first week of July, there was a sharp increase in the number of incidences. Concurrently, the geographical spread of the cases widened. The incidence began to diminish in September, stabilizing at relatively low levels through November and December, with an average daily incidence of fewer than three cases. The diagnosis date curve, shown in Figure 1B, mirrored this trend.

Among the confirmed cases, 1,702 (99.42%) were male, while 10 (0.58%) were female. The median age of the affected individuals was 31 years, with a range from 15 to 71 years. Notably, 112 (6.54%) cases were individuals born before 1980, who, likely in accordance with the vaccination policies of China at that time, may have received the smallpox vaccine. The predominant demographic, representing 84.17% (1,441/1,712), comprised males aged between 18 and 39 years. Based on their current residential districts, most cases were inferred to reside in urban areas, with only approximately 72 cases residing in towns or rural locations. Regarding occupation, the most frequently reported was "unemployed," accounting for 39.54% of all cases (677/1,712). This was followed by positions in commercial services (23.01%, 394/1,712), office workers (8.53%, 146/1,712), and laborers (8.29%, 142/1,712). Additionally, there were 56 cases who were students; this group included 6 individuals under the age of 18, composed of 5 males and 1 female (Table 1).

Among the 1,654 male cases for which relevant information was provided, 94.68% (1,566/1,654) were identified as men who have had sex with men (MSM). Of these, 8.81% (138/1,566) reported being married to women. All 10 female cases reported being heterosexual.

Among the cases for which epidemiological data

TABLE 1. Characteristics of confirmed mpox cases in China, 2023.

Characteristics	June	July	August	September	October	November	December	Total
Sex (*N*)	106	491	501	305	127	80	102	1,712
Male	106	491	496	303	125	80	101	1,702
Female			5	2	2		1	10
Age (years), % (*n*)	106	491	501	305	127	80	102	1,712
15–17	0	0.2 (1)	1 (5)	0	0	1.25 (1)	0.98 (1)	0.47 (8)
18–29	39.62 (42)	38.7 (190)	41.52 (208)	40 (122)	47.24 (60)	30 (24)	40.2 (41)	40.13 (687)
30–39	51.89 (55)	45.82 (225)	44.11 (221)	42.3 (129)	40.94 (52)	47.5 (38)	39.22 (40)	44.39 (760)
40–49	6.6 (7)	12.83 (63)	11.38 (57)	15.08 (46)	8.66 (11)	21.25 (17)	18.63 (19)	12.85 (220)
50–59	1.89 (2)	2.24 (11)	1.6 (8)	2.62 (8)	1.57 (2)	1.25 (1)	0.98 (1)	1.93 (33)
≥60	0	0.2 (1)	0.4 (2)	0	0.79 (1)	0	0	0.23 (4)
Sex orientation in men, % (*n*)	106	490	490	294	111	71	92	1,654
MSM	95.28 (101)	96.53 (473)	93.88 (460)	93.88 (276)	92.79 (103)	94.37 (67)	93.48 (86)	94.68 (1,566)
Self-denial MSM	4.72 (5)	3.47 (17)	6.12 (30)	6.12 (18)	7.21 (8)	5.63 (4)	6.52 (6)	5.32 (88)
Self-reported HIV-status (*n*)	106	491	495	296	113	71	94	1,666
HIV-positive	45.28 (48)	47.25 (232)	38.79 (192)	42.23 (125)	37.17 (42)	46.48 (33)	39.36 (37)	42.56 (709)
Case-relationship available (*n*)	106	491	495	296	113	71	94	1,666
No. of clusters	13	38	29	13	3	1	2	99
Cases included in clusters	26.42 (28/106)	16.50 (81/491)	12.12 (60/495)	10.81 (32/296)	6.19 (7/113)	2.82 (2/71)	4.26 (4/94)	12.85 (214/1,666)
Hospital visit history available (*n*)	92	428	447	273	108	66	82	1,496
1 visit before diagnosis	43.48 (40)	49.07 (210)	51.68 (231)	52.01 (142)	50 (54)	50 (33)	48.78 (40)	50.13 (750)
2 visits before diagnosis	34.78 (32)	32.48 (139)	29.98 (134)	30.77 (84)	29.63 (32)	31.82 (21)	19.51 (16)	30.61 (458)
At least 3 visits before diagnosis	21.74 (20)	18.46 (79)	18.34 (82)	17.22 (47)	20.37 (22)	18.18 (12)	31.71 (26)	19.25 (288)
Time interval available (*n*)	105	479	499	303	126	80		1,537
Median time interval between onset and report (days, IQR)	7 (5.25–9)	7 (5–9)	7 (5–9)	7 (4–11)	8 (5–11.25)	7 (5–9.5)	7 (5–9)	7 (5–10)
Median time interval between onset and diagnosis (days, IQR)	8 (6–10)	8 (5–12)	7 (5–10)	7 (5–11)	8 (6–12)	7 (5–10)	7 (5–10)	8 (5–11)

Abbreivation: MSM=men who have sex with men; IQR=interquartile range.

were available, only 4.02% (67/1,666) had traveled outside China within three weeks prior to the onset of illness. Additionally, 42.56% (709/1,666) tested positive for human immunodeficiency virus (HIV). None of the cases reported a history of blood transfusion within the 21 days preceding the onset of their symptoms.

Among the 1,566 cases identified as MSM, 1,419 (90.61%) confirmed engaging in homosexual activities, with each case involving an average of 1.5 partners (as reported by 1,242 cases with available data) in the 21 days preceding symptom onset. The majority of these sexual encounters involved partners who met through social media apps or other online platforms (74.88%, 450/601) or were random encounters in public venues such as bars or bathhouses (8.49%, 51/601). Among the 88 male patients who did not identify as MSM, 17 reported sexual contact with women, and 5 with men, within 21 days before becoming ill. The remaining individuals declined to disclose their sexual activity.

Among the 10 female mpox cases examined, four reported having sexual contact with their male partners, all of whom were confirmed cases and had recently engaged in homosexual activities. Three other cases involved women who had sexual contact with their male partners; among these, two partners developed rashes that, while suggestive of mpox, had not been confirmed by laboratory tests. The third partner denied exhibiting any symptoms associated with mpox. Additionally, two cases occurred in women who were family members of confirmed mpox cases, likely acquiring the infection via general household contact. The final case involved a nurse who contracted the infection through direct exposure while providing medical care to a confirmed mpox patient, representing a probable instance of occupational transmission among healthcare workers.

Among the 1,666 cases for which epidemiological information was available, 99 clusters were identified across 22 PLADs, accounting for 12.85% (214/1,666) of the total cases. These clusters included 85 clusters with two cases, 12 clusters with three cases, and two clusters with four cases. Notably, no instances of third-generation transmission were observed within these 99 clusters. In five of these clusters, it is suspected that the initial case contracted the infection while traveling abroad, subsequently leading to local transmission. However, in the remaining 94 clusters, the definitive sources of infection could not be identified. From June to December, the proportion of cases included in these clusters displayed a decreasing trend, falling from 26.4% in June to 2.8% in November and 4.3% in December.

Among the 1,655 cases for which data were available, 92.93% (1,538/1,655) were diagnosed with mpox upon seeking medical care for their symptoms. An additional 5.26% (87/1,655) were diagnosed as close contacts of confirmed cases during testing initiatives. Moreover, 23 individuals self-reported as potential mpox infections, three cases were identified through active screening surveillance targeting high-risk populations, three cases emerged from health declarations at customs upon entry, and one case was detected during routine physical examinations. Among the 1,537 cases with detailed timelines, the median interval between the onset of symptoms and reporting was 7 days, while the interval between symptom onset and diagnosis was 8 days. These intervals remained relatively consistent from June to December.

DISCUSSION

Our research indicates that the majority of mpox cases in China occurred among middle-aged males, predominantly identifying as MSM. The primary transmission route identified was contact between cases, mainly through sexual activities, aligning with findings from prior research (*3*–*5*). Based on the data, we hypothesize that the virus may have been circulating undetected within the MSM community since late May 2023. In contrast to the outbreak dynamics observed in Europe and the United States from May to August 2022, the spread in China did not exhibit a rapid or substantial increase in cases (*6*). This disparity may stem from cultural differences influencing sexual behavior patterns among MSM in various regions. For instance, the World Health Organization has noted that common exposure settings for mpox involve gatherings at parties that include sexual contact. However, in China, a significant number of cases were associated with smaller, more intimate interactions facilitated through social media platforms and other online methods, rather than large public gatherings. This behavioral pattern might mitigate the peak of the cases but extend its duration. Furthermore, our findings suggest that mpox transmission has largely been confined to MSM and their immediate contacts in China. Nonetheless, given the high number of married MSM identified in this study, there exists a potential risk of transmission to women — a scenario observed in other countries (*7*). This underlines the importance of continued surveillance and targeted public health interventions to prevent the broader spread of the infection.

Despite significant efforts by local CDCs to investigate and identify close contacts, analyze exposure histories, and establish epidemiological links, fewer than 15% of total cases were found to have connections to other confirmed cases, consistent with a previous study in Beijing (*5*). This outcome is anticipated, as individuals may be reluctant to disclose sensitive information about their sexual partners. Consequently, there is a potential bias in self-reporting, which suggests that the extent of clustering identified in this study may be underestimated. The challenges in accurately tracing the sources of infection and managing close contacts emerged as significant obstacles in controlling the mpox virus spread in China, posing a considerable risk of ongoing, possibly undetected transmission within the MSM group.

To expedite the control of the virus, various health departments have implemented an extensive range of strategies. These strategies include enhanced

surveillance through multiple channels, rigorous contact tracing, training programs for CDC personnel and healthcare providers, and communication and educational campaigns aimed at high-risk populations. The implementation of these measures likely contributed to the observed decline in cases in China post-September. Nonetheless, from October to December, the number of mpox cases exhibited a persistently low-level trend, suggesting a need for more precise and effective strategies. Research (*8–9*) has highlighted that behavioral changes play a crucial role in mitigating the transmission of diseases such as mpox. It is therefore essential to further explore and develop targeted educational strategies that encourage high-risk groups to alter risky behaviors. Moreover, it is vital to promote prompt medical consultation among high-risk individuals upon symptom onset, enhancing the likelihood of early case detection.

The study is subject to some limitations. First, the reliance on self-reported data to investigate mpox cases raises the potential for information bias. Second, the study's exclusive focus on cases identified through the surveillance system may exclude undiagnosed and unreported cases, potentially resulting in selection bias.

Conflicts of interest: No conflicts of interest.

Acknowledgements: All domestic CDC personnel for their invaluable contributions to mpox control efforts. Their meticulous work in conducting investigations, gathering data, and compiling reports was crucial to the successful completion of this analysis.

doi: 10.46234/ccdcw2024.118

[#] Corresponding author: Guoqing Shi, shigq@chinacdc.cn.

[1] National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, Chinese Center for Disease Control and Prevention, Beijing, China; [2] Chinese Center for Disease Control and Prevention, Beijing, China; [3] National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China; [4] National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China; [5] Guangdong Provincial Center for Disease Control and Prevention, Guangzhou City, Guangdong Province, China; [6] Beijing Municipal Center for Disease Prevention and Control, Beijing, China.

Submitted: March 25, 2024; Accepted: June 14, 2024

REFERENCES

1. WHO. Fifth meeting of the international health regulations (2005) (IHR) emergency committee on the multi-country outbreak of mpox (monkeypox). Geneva: WHO; 2023 May. https://www.who.int/news/item/11-05-2023-fifth-meeting-of-the-international-health-regulations-(2005)-(ihr)-emergency-committee-on-the-multi-country-outbreak-of-monkeypox-(Mpox).
2. Zhao H, Wang WL, Zhao L, Ye S, Song JD, Lu RJ, et al. The first imported case of monkeypox in the mainland of China – Chongqing municipality, China, September 16, 2022. China CDC Wkly 2022;4(38):853 – 4. https://doi.org/10.46234/ccdcw2022.175.
3. Laurenson-Schafer H, Sklenovská N, Hoxha A, Kerr SM, Ndumbi P, Fitzner J, et al. Description of the first global outbreak of Mpox: an analysis of global surveillance data. Lancet Global Health 2023;11(7):e1012 – 23. https://doi.org/10.1016/S2214-109X(23)00198-5.
4. Thornhill JP, Barkati S, Walmsley S, Rockstroh J, Antinori A, Harrison LB, et al. Monkeypox virus infection in humans across 16 countries – April-June 2022. N Engl J Med 2022;387(8):679 – 91. https://doi.org/10.1056/NEJMoa2207323.
5. Dou XF, Li F, Ren ZY, Zhang DT, Li J, Li D, et al. Clinical, epidemiological, and virological features of Mpox in Beijing, China - May 31-June 21, 2023. Emerg Microbes Infect 2023;12(2):2254407. https://doi.org/10.1080/22221751.2023.2254407.
6. World Health Organization. 2022-24 Mpox (Monkeypox) outbreak: global trends. Geneva: WHO; 2024 May. https://worldhealthorg.shinyapps.io/mpx_global.
7. Vallejo-Plaza A, Rodríguez-Cabrera F, Sebastián VH, Herrador BRG, Balader PS, Rodríguez-Alarcón LGSM, et al. Mpox (formerly monkeypox) in women: epidemiological features and clinical characteristics of mpox cases in Spain, April to November 2022. Euro Surveill 2022;27(48):2200867. https://doi.org/10.2807/1560-7917.ES.2022.27.48.2200867.
8. Clay PA, Asher JM, Carnes N, Copen CE, Delaney KP, Payne DC, et al. Modelling the impact of vaccination and sexual behaviour adaptations on mpox cases in the USA during the 2022 outbreak. Sex Transm Infect 2024; 100(2), 70 – 76. https://doi.org/10.1136/sextrans-2023-055922
9. Zhang XS, Mandal S, Mohammed H, Turner C, Florence I, Walker J, et al. Transmission dynamics and effect of control measures on the 2022 outbreak of Mpox among gay, bisexual, and other men who have sex with men in England: a mathematical modelling study. Lancet Infect Dis 2024;24(1):65 – 74. https://doi.org/10.1016/S1473-3099(23)00451-6.

Preplanned Studies

Protection of Omicron Sub-Lineage Infection Against Reinfection with Another Omicron Sub-Lineage: Systematic Review, Meta-Analysis, and Meta-Regression — Worldwide, 2022–2023

Xu Guo[1]; Zuyao Cheng[1]; Junhong Li[1]; Yudan Song[1]; Hui Zheng[1]; Yamin Wang[1]; Chao Ma[1,#]; Zijian Feng[2,#]

Summary

What is already known about this topic?

Both the decline in immunity over time and the evolution of the virus play a role in the level of protection offered by a prior infection.

What is added by this report?

Point estimates indicated variations in protection levels based on the initial infecting variant and the reinfecting variant. There was a consistent correlation between real-world protection, antigenic distance, and humoral immunity levels. Specifically, shorter antigenic distances and higher humoral immunity levels corresponded to enhanced real-world protection.

What are the implications for public health practice?

Our findings suggest that virological and immunological studies could help identify and assess the epidemic risk posed by new variants before they become dominant. Prompt incorporation of the latest variants into the antigen components of the coronavirus disease 2019 (COVID-19) vaccines can significantly contribute to effective epidemic prevention and control measures.

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Omicron variant has infected over 90% of the global population at least once (*1*). The protection conferred by previous infections is gradually becoming a crucial factor in controlling the pandemic (*2*). Our research used systematic reviews and meta-analyses to estimate the degree and longevity of protection against reinfection by another Omicron sub-lineage, relative to uninfected individuals, under a similar vaccination status. Out of 14,105 publications, we selected 10 studies that had either a cohort, test-negative design, or case-control approach, and utilized their data for a statistical analysis. Our findings indicate that the immunity provided against reinfection tends to vary based on the previous variant encountered and the variant causing reinfection. Moreover, protection offered by Omicron sub-lineage infection against reinfection with another Omicron sub-lineage tends to decrease over time. The degree of protection from a prior infection increases with more closely related antigenic distance and higher humoral immunity levels.

We employed three-level meta-analytic models using the 'metagen' function of the 'meta' package (version 6.3) in R (version 4.3.1, R Foundation for Statistical Computing, Vienna, Austria) for consolidating protection data. We extracted multiple protection data points from a single study, incorporating all in the meta-analyses. Three-level meta-analyses permit the explicit modeling of nested data structures, such as when individual studies provide multiple estimates for varying subgroups or time points. These models yield more valid and reliable estimates than traditional fixed and random-effect models under such conditions (*3*). For research data specifying time from initial infection, we applied a meta-regression of the log odds to approximate the waning of protection over time, assessing at 1-month intervals. We performed meta-analysis and meta-regression only on groups comprising more than two articles with verifiable extracted data. Database searches covered PubMed, the World Health Organization (WHO) coronavirus disease 2019 (COVID-19) database, SSRN, MedRxiv, Embase, and the WanFang Database. We searched for cohort, test-negative design, and case-control studies published on or before October 24, 2023, using keywords related to reinfection, prior infection, and Omicron (Supplementary Table S1, available at https://weekly.chinacdc.cn/). Included studies were those that considered the protective effect of prior Omicron infection in individuals against those who were infection-naïve and had comparable vaccination status. We evaluated the risk of bias using the ROBINS-I tool (Supplementary Table S4, available at https://weekly.chinacdc.cn/). Our study, which

complies with PRISMA, was registered with PROSPERO (CRD42023466200). Supplementary Materials (available at https://weekly.chinacdc.cn/) provide detailed methodology.

We reviewed the titles and abstracts of 14,105 articles, of which 491 passed our screening to undergo a thorough full-text review (Supplementary Figure S1, available at https://weekly.chinacdc.cn/). From this process, we identified 10 relevant studies providing 81 data sets, sourced from eight nations: Qatar, Canada, China, Denmark, Japan, the Republic of Korea, Singapore, and Portugal. These studies encompassed a combined sample size of 17,214,915. For our meta-analysis and meta-regression, we included 12 data sets from 2 studies in the BA.1 to BA.2 group, 10 data sets from 3 studies in the BA.1 to BA.4/5 group, 12 data sets from 3 studies in the BA.2 to BA.4/5 group, and 15 data sets from 4 studies in the BA.1/2 to BA.4/5 group (Supplementary Table S2, available at https://weekly.chinacdc.cn/).

Compared to a non-infected cohort, individuals previously infected with the BA.1 variant showed 87.5% protection (47.9–97.0) against reinfection with the BA.2 variant (Supplementary Figure S2, available at https://weekly.chinacdc.cn/). This protection, however, waned from 89.8% at 1 month (64.6–97.1) to 81.1% at 5 months (31.9–94.8) (Tables 1–2). Between 30 and 60 days post-infection, an 82.0% protection rate (49.0–94.0) was observed amongst those unvaccinated, and a protection rate of 94.2% (89.2–96.9) amongst those vaccinated (Supplementary Table S2). In vaccinated individuals, the effectiveness of protection against reinfection with BA.4/5 variant after an initial BA.1 infection was 75.2% (42.1–89.4) (Supplementary Figure S2). Notably, this protection waned from 77.2% at 5 months post-infection (47.6–90.1) to 40.9% at 12 months (–32.8–73.7) (Tables 1–2).

Research conducted in Singapore (*4*) tracked the infection history of cohorts unexposed to COVID-19 who later contracted the BA.1 variant. Observations indicated that the protective effect of prior BA.1 infection against clinically attended symptomatic XBB variant reinfection diminished from a span of 40.0% (32.0–47.0) between 3–8 months post-infection to 27.0% (24.0–30.0) subsequent to 8 months (Table 1, Supplementary Table S2).

The protective effect of a BA.2 infection against subsequent reinfection with BA.4/5 variants was 88.9% (76.6–94.8) (Supplementary Figure S2), and this waned from 91.6% (80.9–96.3) at 4 months post-infection to 80.4% (56.7–91.1) at 8 months post-infection (Tables 1–2). Comparable findings were reported by a Singaporean cohort study (*4*), which demonstrated that the protective effect of a primary BA.2 infection against symptomatic reinfection by the XBB variant also declined over time, from 74.0% (72.0–75.0) during the 3–6 months post-infection period to 37.0% (32.0–43.0) after 8 months (Table 1, Supplementary Table S2). It is important to note that these studies were conducted exclusively among vaccinated populations due to a lack of data from unvaccinated groups.

An analysis of protection against reinfection with the

TABLE 1. Protection (%) against reinfection by different Omicron sub-lineages.

Type of variant for prior infection	Type of variant for reinfection				
	Time since primary infection	BA.2 (Meta-analysis)	BA.4/5 (Meta-analysis)	BA.2.75 (Systematic review)	XBB (Systematic review)
BA.1	5 months	81.1 (31.9–94.8)	77.2 (47.6–90.1)	NA	40.0 (32.0–47.0)[†]
	8 months	NA	65.7 (26.5–84.0)	NA	27.0 (24.0–30.0)[§]
	Total[*]	87.5 (47.9–97.0)	75.2 (42.1–89.4)	NA	NA
BA.2	5 months	NA	89.6 (76.9–95.4)	NA	74.0 (72.0 to 75.0)[¶]
	8 months	NA	80.4 (56.7–91.1)	NA	37.0 (32.0 to 43.0)[**]
	Total[*]	NA	88.9 (76.6–94.8)	NA	NA
BA.1/2	Total[*]	NA	86.2 (73.6–92.8)	49.9 (47.6 to 52.1)	NA
BA.4/5	Total[*]	NA	NA	80.6 (71.2 to 87.0)	NA

Note: NA means no data available.
[*] Total protection, regardless of time since primary infection.
[†] Time since primary infection: 3–8 months.
[§] Time since primary infection: ≥8 months.
[¶] Time since primary infection: 3–6 months.
[**] Time since primary infection: ≥8 months.

TABLE 2. Estimates of protection (%) against various Omicron variants based on the time elapsed since primary infection.

Time since primary infection (months)	BA.1 to BA.2	BA.1 to BA.4/5	BA.2 to BA.4/5
1	89.8 (64.6–97.1)	NA	NA
2	88.1 (60.6–96.4)	NA	NA
3	86.1 (54.5–95.8)	NA	NA
4	83.8 (45.3–95.2)	NA	91.6 (80.9–96.3)
5	81.1 (31.9–94.8)	77.2 (47.6–90.1)	89.6 (76.9–95.4)
6	NA	73.9 (41.9–88.3)	87.2 (71.8–94.2)
7	NA	70.1 (35.0–86.2)	84.1 (65.2–92.8)
8	NA	65.7 (26.5–84.0)	80.4 (56.7–91.1)
9	NA	60.1 (16.1–81.6)	NA
10	NA	55.0 (3.2–79.1)	NA
11	NA	48.5 (−12.8–76.4)	NA
12	NA	40.9 (−32.8–73.7)	NA

Note: NA means no data available.

Omicron BA.4/5 variants following a BA.1/2 infection incorporated studies from four nations: Denmark, Japan, Portugal, and Qatar. Due to a gradual shift from the dominance of the Omicron BA.1 subvariant to the BA.2 subvariant in these countries, the two infection peaks of BA.1 and BA.2 combined, making it difficult to distinguish between their timelines. Our meta-analysis findings indicated that the conferred protection against reinfection with BA.4/5 after a BA.1/2 infection was 86.2% (73.6–92.8) in comparison to an uninfected population (Supplementary Figure S2). Notably, a test-negative design study conducted in Qatar (*5*) reported protection rates of 49.9% (47.6–52.1) against a BA.2.75 infection after a primary BA.1/2 infection; this protection rate decreased to 32.2% (25.5–38.3) for an unvaccinated population. The study also reported protection rates of 50.2% (43.1–56.4) against symptomatic infection. Furthermore, a primary BA.4/5 infection offered a protection rate of 80.6% (71.2–87.0) against a BA.2.75 variant infection; this dropped to 44.4% (-4.0–70.3), however, for the unvaccinated population. The reported protection against symptomatic infection was 91.4% (35.8–98.8) (Table 1, Supplementary Table S2).

According to meta-regression analyses of studies noting the duration since the initial infection, we discerned a decline in immunity against reinfection over time. However, due to limitations in available data, these estimated protection rates yielded wide confidence intervals (*CI*s), preventing any statistically significant differences from being determined within the meta-regression results. In spite of overlapping *CI*s within the meta-regression, at the same time since primary infection, BA.2 variant showed a higher protection estimate than BA.5 against BA.1 variant infection [at 5 months: 89.6% (76.9, 95.4) versus 77.2% (47.6, 90.1); at 8 months: 80.4% (56.7, 91.1) versus 65.7% (26.5, 84.0)]. A similar trend was observed when comparing immunity from BA.1 against BA.2, and BA.1 against BA.4/5 [at 5 months: 81.1% (31.9–94.8) versus 77.2% (47.6–90.1)] (Table 1-2, Figure 1, Supplementary Table S3).

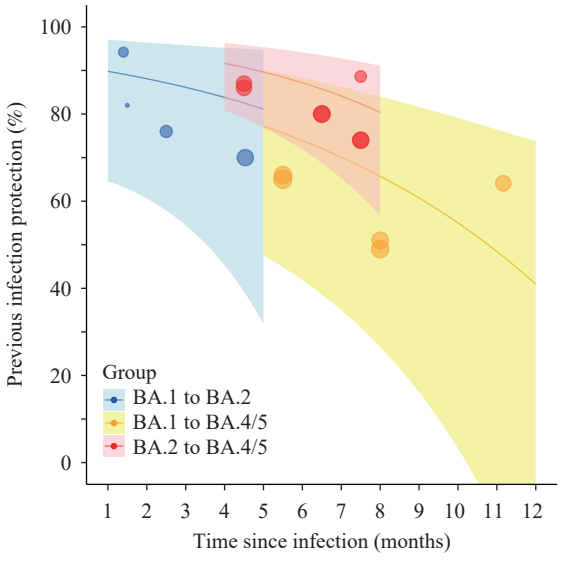


FIGURE 1. Estimated levels of protection against various Omicron variants based on the time elapsed since primary infection.
Note: The shaded areas represent the 95% confidence intervals (*CI*s). The size of each bubble is proportional to the reciprocal of the standard error (SE).

DISCUSSION

The likelihood of reinfection with the same Omicron variant is exceedingly low in instances where a previous infection has occurred. In two respective studies, there was an impressive 94%–97% immunity rate observed amongst individuals previously infected with the BA.4/5 strain, which remained effective for up to 100 days post primary infection (*6*–*7*). Our comprehensive review and meta-analysis revealed elevated protection rates against reinfection in individuals previously exposed to later, evolutionarily similar variants, as compared to those exposed to earlier strains. This was independent of the time elapsed since the initial infection. This conclusion is congruent with studies conducted in the Netherlands (*8*), Qatar (*5*), Singapore (*4*), and the United Kingdom (*9*). This could potentially elucidate the cyclical nature of the COVID-19 pandemic, characterized by repeated bouts of infections and reinfections, triggered by distinct variants in succession.

The variants BA.4 and BA.5 are derivatives of BA.2. In terms of antigenic distance, BA.4/5 shares a closer resemblance with BA.2 relative to BA.1. Serological research has revealed that after infection by the BA.2 variant, the convalescent sera possess a superior amount of neutralizing antibodies against BA.5 than the sera derived from BA.1 variant infection (*10*). This collective evidence suggests that the effectiveness of protection from prior COVID-19 infection against further infections is not solely dependent on declining immunity but is also influenced by viral evolution. Higher humoral immunity levels and closer antigenic distances contribute to enhanced protection provided by previous infections.

In assessing the future risks associated with COVID-19, it is crucial to consider not only the time elapsed since the peak of the previous wave but also the antigenic difference and evasion ability of humoral immunity between any potential new variant and the formerly dominant ones. Ensuring a timely update of the antigenic components within the COVID-19 vaccine, coupled with inoculating individuals not recently infected, stands as a vital strategy in combating this disease. While this approach may not fully synchronize with the evolution of SARS-CoV-2, potentially leading to a mismatch between subsequent infection strains and the vaccine strain, a narrower antigenic distance can ostensibly offer improved protection over a match with a more antigenically distant strain. Given that both infection- and vaccine-induced protection diminish over time, the duration since either the infection or vaccination must be factored into vaccination policy considerations.

This study was subject to at least two limitations. One limitation of our research is that the time since primary infection provided by some studies is a time range. In order to conduct meta-regression, we used the median of this range. Another limitation lies in the limited number of studies incorporated. Upon retrieval and examination, only ten studies pertaining to the protection against Omicron variant infection satisfied the inclusion criteria. To offset potential overemphasis arising from the inclusion of numerous data points from a single study, we employed a three-tier meta-analysis approach. Nonetheless, as the available data pool was relatively insufficient, stratified analyses could not be conducted. Some specialists assert that a minimum of ten studies is required to facilitate valid meta-regression, which is greater than what we currently have. The scarcity of data broadens *CI*s, hampers the extraction of useful statistical inferences, and adds uncertainty to the stability of our final findings.

The global population has previously encountered pandemics involving the BA.1, BA.2, and BA.4/5 variants, with the infectious strain now transitioning to the XBB sub-lineage. While these past variant pandemics have subsided, further exploration of existing literature can deepen our understanding of the immune mechanisms underlying COVID-19. This will assist in securing epidemiological parameters of the disease, comprehending the mechanisms behind COVID-19 outbreaks, and providing essential evidence for infectious disease model research and assessments of reinfection risk.

Our comprehension of COVID-19 continues to evolve, highlighting the need for critical research into topics such as antigenic separation, immune evasion, and the extent of cross-protection. While predicting the trajectory of SARS-CoV-2 mutation remains a formidable challenge, establishing a link between pathogenesis and immunology through empirical research might expedite and enhance the precision of risk assessments for new variants. Understanding the degree of protection provided by previous COVID-19 infections against new variants can further inform and guide national response strategies.

Conflicts of interest: No conflicts of interest.

Acknowledgements: Kathy Leung from the University of Hong Kong for her insightful suggestions on epidemiological concepts and statistical methods.

Lance Rodewald from China CDC for his assistance with English language editing.

Funding: Supported by the National Key Research and Development Program of China (2021YFC2301600), the National Natural Science Foundation of China (82341034), and the Chinese Preventive Medicine Association's COVID-19 Prevention and Control Modeling Research Project.

doi: 10.46234/ccdcw2024.103

[#] Corresponding authors: Chao Ma, machao@chinacdc.cn; Zijian Feng, fengzj@chinacdc.cn.

[1] National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Disease, National Immunization Program, Chinese Center for Disease Control and Prevention, Beijing, China; [2] Chinese Preventive Medicine Association, Beijing, China.

Submitted: December 13, 2023; Accepted: May 27, 2024

REFERENCES

1. Leung K, Lau EHY, Wong CKH, Leung GM, Wu JT. Estimating the transmission dynamics of SARS-CoV-2 Omicron BF.7 in Beijing after adjustment of the zero-COVID policy in November-December 2022. Nat Med 2023;29(3):579 – 82. http://dx.doi.org/10.1038/s41591-023-02212-y.
2. Diani S, Leonardi E, Cavezzi A, Ferrari S, Iacono O, Limoli A, et al. SARS-CoV-2-the role of natural immunity: a narrative review. J Clin Med 2022;11(21):6272. https://doi.org/10.3390/jcm11216272.
3. Wu NN, Joyal-Desmarais K, Ribeiro PAB, Vieira AM, Stojanovic J, Sanuade C, et al. Long-term effectiveness of COVID-19 vaccines against infections, hospitalisations, and mortality in adults: findings from a rapid living systematic evidence synthesis and meta-analysis up to December, 2022. Lancet Respir Med 2023;11(5):439 – 52. https://doi.org/10.1016/S2213-2600(23)00015-2.
4. Tan CY, Chiew CJ, Pang D, Lee VJ, Ong B, Lye DC, et al. Protective immunity of SARS-CoV-2 infection and vaccines against medically attended symptomatic Omicron BA. 4, BA.5, and XBB reinfections in Singapore: a national cohort study. Lancet Infect Dis 2023;23(7):799 – 805. https://doi.org/10.1016/S1473-3099(23)00060-9.
5. Chemaitelly H, Tang P, Coyle P, Yassine HM, Al-Khatib HA, Smatti MK, et al. Protection against reinfection with the Omicron BA. 2.75 subvariant. N Engl J Med 2023;388(7):665 – 7. https://doi.org/10.1056/NEJMc2214114.
6. Chen XS, Xu YL, Xie Y, Song WL, Hu Y, Yisimayi A, et al. Protective effect of plasma neutralization from prior SARS-CoV-2 Omicron infection against BA. 5 subvariant symptomatic reinfection. Lancet Reg Health West Pac 2023;33:100758. https://doi.org/10.1016/j.lanwpc.2023.100758.
7. Tamandjou C, Auvigne V, Schaeffer J, Vaux S, du Châtelet IP. Effectiveness of second booster compared to first booster and protection conferred by previous SARS-CoV-2 infection against symptomatic Omicron BA. 2 and BA.4/5 in France. Vaccine 2023;41(17):2754 – 60. https://doi.org/10.1016/j.vaccine.2023.03.031.
8. Andeweg SP, de Gier B, Vennema H, van Walle I, van Maarseveen N, Kusters NE, et al. Higher risk of SARS-CoV-2 Omicron BA.4/5 infection than of BA.2 infection after previous BA.1 infection, the Netherlands, 2 May to 24 July 2022. Euro Surveill 2023;28(7):2200724. http://dx.doi.org/10.2807/1560-7917.ES.2023.28.7.2200724.
9. Wei J, Stoesser N, Matthews PC, Khera T, Gethings O, Diamond I, et al. Risk of SARS-CoV-2 reinfection during multiple Omicron variant waves in the UK general population. Nat Commun 2024;15(1):1008. https://doi.org/10.1038/s41467-024-44973-1.
10. Kimura I, Yamasoba D, Tamura T, Nao N, Suzuki T, Oda Y, et al. Virological characteristics of the SARS-CoV-2 Omicron BA. 2 subvariants, including BA.4 and BA.5. Cell 2022;185(21):3992 – 4007. e16. https://doi.org/10.1016/j.cell.2022.09.018.

SUPPLEMENTARY MATERIAL

METHODS

Search Strategy and Selection Criteria

We conducted a thorough search of databases, including PubMed, WHO COVID-19, SSRN, MedRxiv, Embase, and WanFang Database. The search, undertaken prior to October 24, 2023, focused on cohort studies, test-negative designs, and case-control studies with keywords related to reinfection, prior infection, and Omicron. This study adheres to the PRISMA guidelines and is officially registered with PROSPERO (CRD42023466200).

Inclusion and Exclusion Criteria

To remove the potential bias introduced by vaccination status, we included studies examining the protection conferred by a prior Omicron infection in individuals who had been infected once, compared to infection-naive individuals with a similar vaccination status. We excluded studies that failed to differentiate between an initial infection and a reinfection with a variant strain.

Outcomes

Reinfection was characterized as two separate outbreaks caused by two distinct predominant strains, confirmed through either two positive results from polymerase-chain-reaction (PCR), rapid antigen test, or self-reported infections. The studies included did not need to differentiate between symptomatic and asymptomatic infections.

Study Selection and Data Extraction

Upon reviewing titles and abstracts, we pinpointed studies and reports pertaining to immunity from COVID-19 infection. Relevant studies and reports had their main texts and supplementary materials scrutinized by two independent reviewers to ascertain if they met the inclusion criteria. One reviewer manually undertook the extraction process, which a second reviewer independently confirmed. In cases of disagreement, the input of a third reviewer was sought.

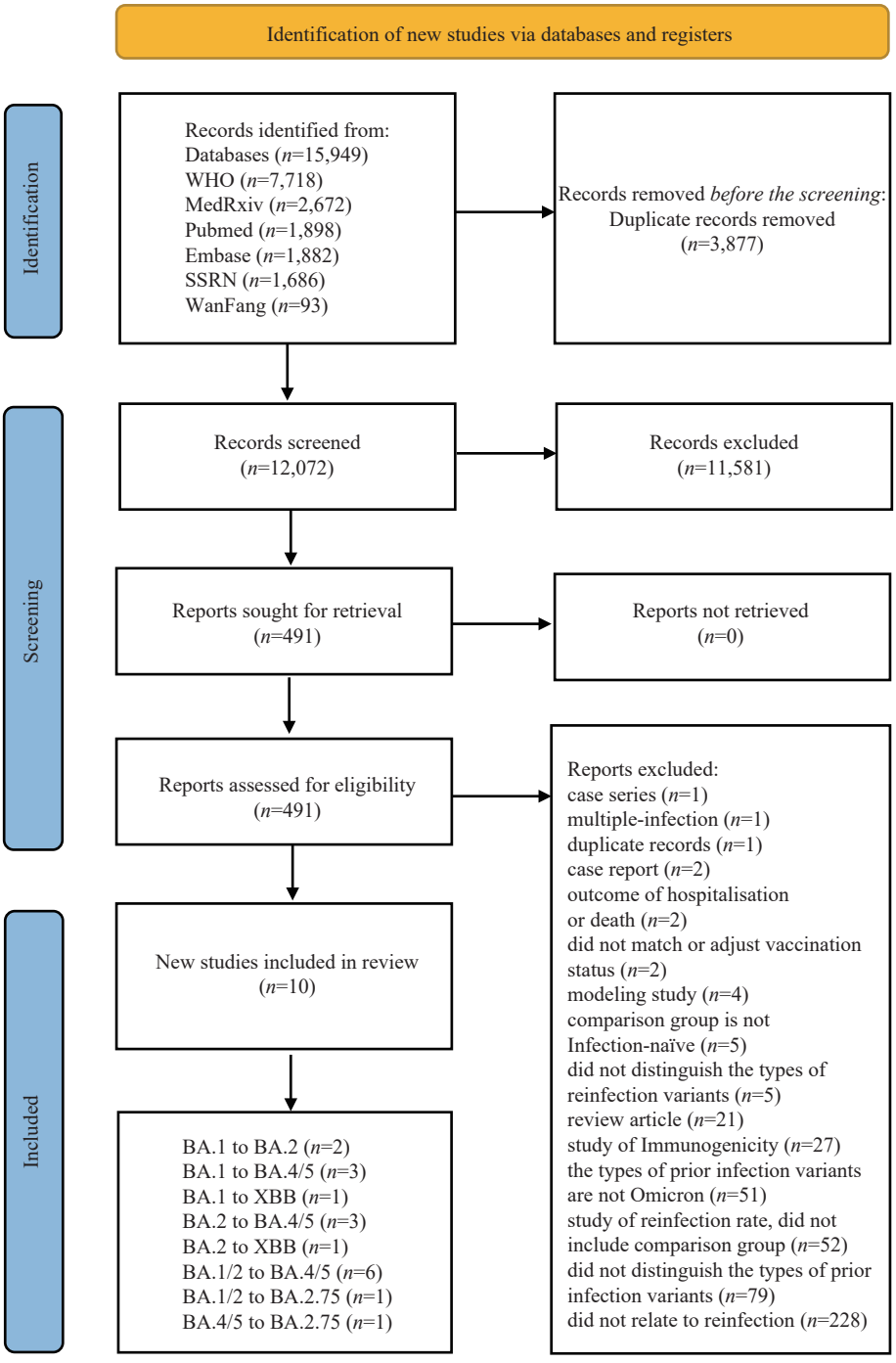
The data extracted encompassed the author's name, country of research, study design, COVID-19 vaccination status, reinfection outcome, variant type during initial infection and reinfection, time elapsed since primary infection, sample size, and the effectiveness of protection (expressed as 1-OR/HR/RRHR) along with its 95% confidence interval (*CI*). The effect value was adjusted for covariates in a multivariate analysis, and this adjusted value was favored and utilized when available.

Risk-of-bias Assessment

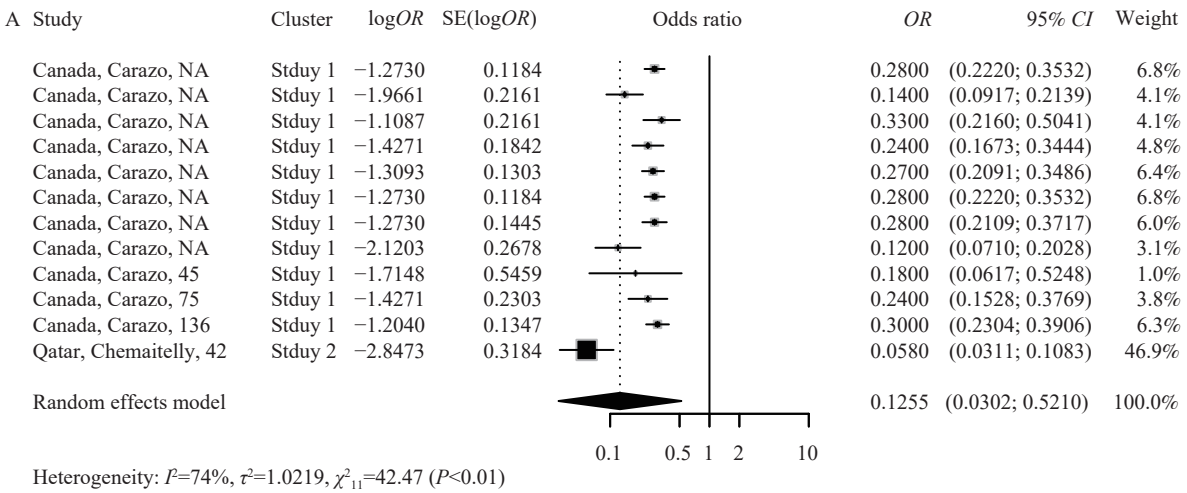
The risk of bias in the studies was evaluated using the ROBINS-I tool. This assessment was independently carried out by two reviewers for each documented outcome. Any discrepancies between the reviewers were reconciled by a third party. All studies received equal treatment in the primary analysis, irrespective of their quality rating.

Data Analysis

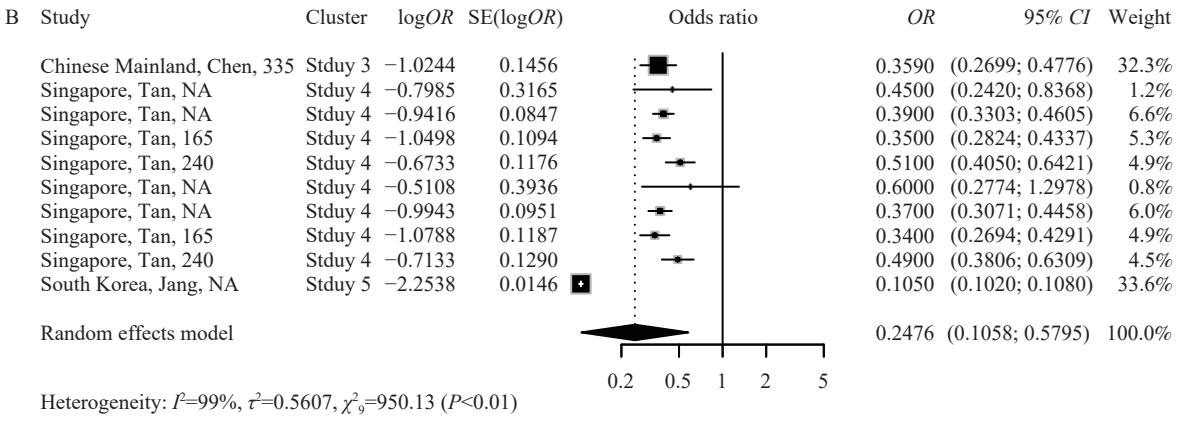
We employed three-level meta-analytic models using the 'metagen' function of the 'meta' package (version 6.3) in R (version 4.3.1, R Foundation for Statistical Computing, Vienna, Austria) for consolidating protection data. The restricted maximum likelihood estimator was leveraged for determining within-study and between-study heterogeneity variance, denoted by τ^2. In cases where data incorporated time since primary infection, we implemented a meta-regression of log odds to estimate the attenuation of protection over time, evaluated in one-month increments. For studies that did not provide time since primary infection, we estimated this parameter employing GISAID data, which was then incorporated into our meta-regression analyses. Both meta-analysis and meta-regression were restricted to groups containing more than two articles with valid, extracted data.



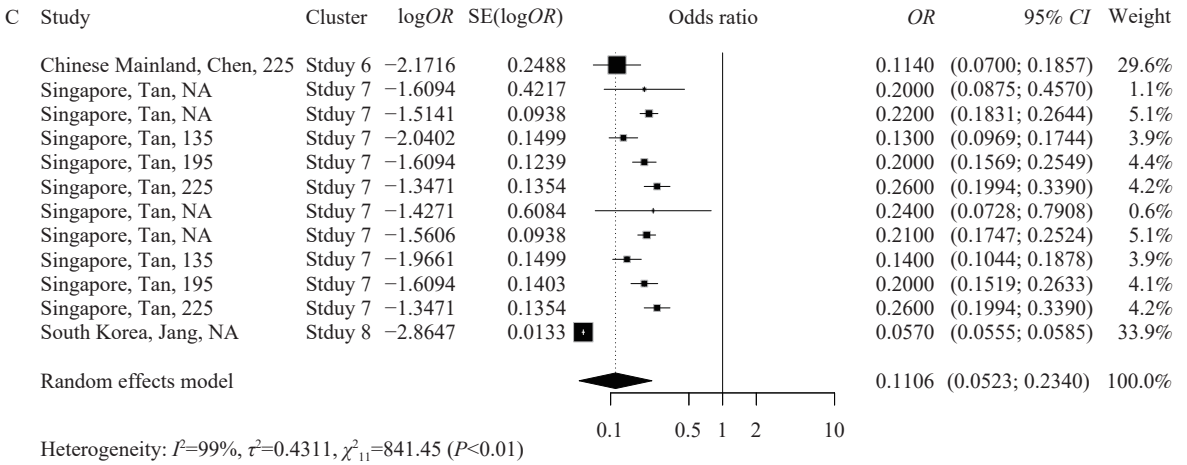
SUPPLEMENTARY FIGURE S1. Study selection.
Abbreviation: WHO=World Health Organization; COVID-19=coronavirus disease 2019; SSRN=social science research network.



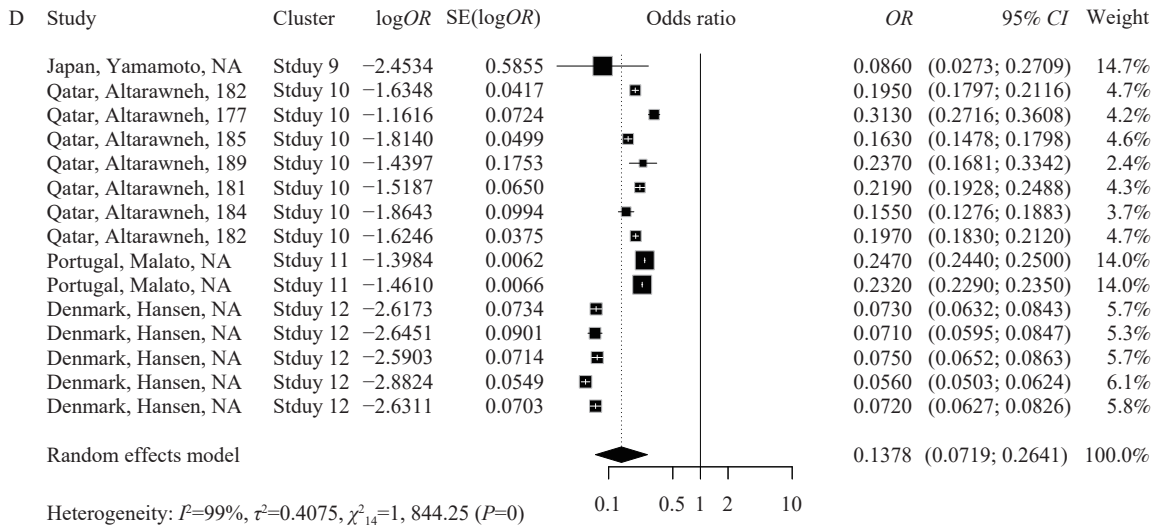
Heterogeneity: I^2=74%, τ^2=1.0219, χ^2_{11}=42.47 (P<0.01)



Heterogeneity: I^2=99%, τ^2=0.5607, χ^2_9=950.13 (P<0.01)



Heterogeneity: I^2=99%, τ^2=0.4311, χ^2_{11}=841.45 (P<0.01)



SUPPLEMENTARY FIGURE S2. Forest plots of protection against reinfection by different Omicron sub-lineages. (A) Protection of BA.1 infection against reinfection with BA.2. (B) Protection of BA.1 infection against reinfection with BA.4/5. (C) Protection of BA.2 infection against reinfection with BA.4/5. (D) Protection of BA.1/2 infection against reinfection with BA.4/5. Abbreviation: *OR*=odds ratio; SE=standard error; *CI*=confidence interval.

SUPPLEMENTARY TABLE S1. Search strategy.

Keyword	Platform	Search Results
(re-infection) OR (reinfection) OR (repeated infection) OR (recurrent Infection) OR (previously infected) OR (previous infection) OR (prior infection) AND ("Omicron" OR "B.1.1.529") AND ("2020/1/1"[Date - Publication] : "2023/10/24"[Date - Publication])	Pubmed	1,898
(Reinfection OR repeated infection OR recurrent infection OR previously infected OR previous infection OR prior infection) AND ("Omicron" OR "B.1.1.529")	WHO COVID-19 database	7,718
Omicron	SSRN	1,686
(Reinfection OR repeated infection OR recurrent infection OR previous infection OR prior infection) AND (Omicron OR B.1.1.529)" and posted between "01 Jan, 2020 and 24 Oct, 2023"	MedRxiv	2,672
('re infection' OR reinfection OR (repeated AND infection) OR (recurrent AND infection) OR (previously AND infected) OR (previous AND infection) OR (prior AND infection)) AND ('omicron' OR 'b.1.1.529') AND [01-01-2020]/sd NOT [24-10-2023]/sd	Embase	1,882
(再感染 OR 既往感染 OR 重复感染) AND ("奥密克戎" OR "Omicron")	WanFang Database	93

SUPPLEMENTARY TABLE S2. Characteristics of the included studies.

Group	Country	Study Design	Vaccination Status	Outcome of reinfection	Protection (95% *CI*)	Time since primary infection (reported)	Time since primary infection (predicted)*	sample size
BA.1 to BA.2								
Chemaitelly et al. (2022) (*1*)	Qatar	TND	Vaccinated (adjusted)	Infection	94.2 (89.2 to 96.9)	39 to 45 days	NA	41,988
Carazo et al. (2023) (*2*)	Canada	Cohort	Unvaccinated	Infection	82.0 (49.0 to 94.0)	30 to 59 days	NA	2,567
Carazo et al. (2023) (*2*)	Canada	Cohort	Unvaccinated	Infection	76.0 (63.0 to 85.0)	60 to 89 days	NA	2,567
Carazo et al. (2023) (*2*)	Canada	Cohort	Unvaccinated	Infection	70.0 (61.0 to 77.0)	90 to 182 days	NA	2,567
Carazo et al. (2023) (*2*)	Canada	Cohort	Unvaccinated	Infection	72.0 (65.0 to 78.0)§	NA	81	2,567
Carazo et al. (2023) (*2*)	Canada	Cohort	Unvaccinated	Symptomatic infection	86.0 (79.0 to 91.0)§	NA	81	630
Carazo et al. (2023) (*2*)	Canada	Cohort	Unvaccinated	Infection	67.0 (79.0 to 91.0)§	NA	81	1,167
Carazo et al. (2023) (*2*)	Canada	Cohort	Unvaccinated	Infection	76.0 (65.0 to 83.0)§	NA	81	1,258
Carazo et al. (2023) (*2*)	Canada	Cohort	Unvaccinated	Infection	73.0 (65.0 to 79.0)§	NA	81	2,416
Carazo et al. (2023) (*2*)	Canada	Cohort	Unvaccinated	Infection	72.0 (65.0 to 78.0)§	NA	81	2,567
Carazo et al. (2023) (*2*)	Canada	Cohort	Unvaccinated	Infection	72.0 (63.0 to 79.0)§	NA	81	2,040
Carazo et al. (2023) (*2*)	Canada	Cohort	Unvaccinated	Symptomatic infection	88.0 (80.0 to 93.0)§	NA	81	433
BA.1 to BA.4/5								
Chen et al. (2023) (*3*)	China	Cohort	Vaccinated (matched)	Symptomatic infection	64.1 (52.4 to 73.1)	329 to 341 days	NA	386
Tan et al. (2023) (*4*)	Singapore	Cohort	Completed primary series	Medically attended symptomatic infection	55.0 (17.0 to 76.0)	NA	265	1,733,535
Tan et al. (2023) (*4*)	Singapore	Cohort	Boosted	Medically attended symptomatic infection	61.0 (54.0 to 67.0)	NA	265	1,902,581
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	65.0 (57.0 to 72.0)	3 to <8 months	NA	1,779,968
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	49.0 (35.0 to 59.0)	≥8 months	NA	1,858,209
Tan et al. (2023) (*4*)	Singapore	Cohort	Completed primary series	Medically attended symptomatic infection	40.0 (-31.0 to 72.0)§	NA	265	1,351,636
Tan et al. (2023) (*4*)	Singapore	Cohort	Boosted	Medically attended symptomatic infection	63.0 (55.0 to 69.0)§	NA	265	1,514,582
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	66.0 (57.0 to 73.0)§	3 to <8 months	NA	1,402,809
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	51.0 (37.0 to 62.0)§	≥8 months	NA	1,465,449
Jang et al. (2023) (*5*)	South Korea	Case-control	Vaccinated (adjusted)	Infection	89.5 (89.2 to 89.8)	NA	190	5,085,535
BA.1 to XBB								
Tan et al. (2023) (*4*)	Singapore	Cohort	Completed primary series	Medically attended symptomatic infection	3.00 (−13.0 to 16.0)	NA	NA	800,221
Tan et al. (2023) (*4*)	Singapore	Cohort	Boosted	Medically attended symptomatic infection	30.0 (27.0 to 32.0)	NA	NA	878,615
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	40.0 (32.0 to 47.0)	3 to <8 months	NA	796,534
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	27.0 (24.0 to 30.0)	≥8 months	NA	883,260
Tan et al. (2023) (*4*)	Singapore	Cohort	Completed primary series	Medically attended symptomatic infection	16.0 (−6.0 to 33.0)§	NA	NA	630,473
Tan et al. (2023) (*4*)	Singapore	Cohort	Boosted	Medically attended symptomatic infection	30.0 (27.0 to 33.0)§	NA	NA	706,028
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	41.0 (32.0 to 49.0)§	3 to <8 months	NA	631,880
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	28.0 (25.0 to 31.0)§	≥8 months	NA	705,569

Continued

Group	Country	Study Design	Vaccination Status	Outcome of reinfection	Protection (95% *CI*)	Time since primary infection (reported)	Time since primary infection (predicted)*	sample size
BA.2 to BA.4/5								
Chen et al. (2023) (*3*)	China	Cohort	Vaccinated (matched)	Symptomatic infection	88.6 (81.7 to 93.1)	210 to 231 days	NA	346
Tan et al. (2023) (*4*)	Singapore	Cohort	Completed primary series	Medically attended symptomatic infection	80.0 (53.0 to 91.0)	NA	181	1,737,378
Tan et al. (2023) (*4*)	Singapore	Cohort	Boosted	Medically attended symptomatic infection	78.0 (74.0 to 82.0)	NA	181	2,124,162
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	87.0 (82.0 to 90.0)	3 to <6 months	NA	1,866,720
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	80.0 (74.0 to 84.0)	6 to <7 months	NA	1,805,491
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	74.0 (66.0 to 80.0)	7 to <8 months	NA	1,885,447
Tan et al. (2023) (*4*)	Singapore	Cohort	Completed primary series	Medically attended symptomatic infection	76.0 (24.0 to 93.0)§	NA	181	1,352,984
Tan et al. (2023) (*4*)	Singapore	Cohort	Boosted	Medically attended symptomatic infection	79.0 (74.0 to 82.0)§	NA	181	1,743,385
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	86.0 (82.0 to 90.0)§	3 to <6 months	NA	1,492,493
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	80.0 (74.0 to 85.0)§	6 to <7 months	NA	1,430,362
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	74.0 (66.0 to 80.0)§	7 to <8 months	NA	1,501,919
Jang et al. (2023) (*5*)	South Korea	Case-control	Vaccinated (adjusted)	Infection	94.3 (94.1 to 94.4)	NA	106	5,002,210
BA.2 to XBB								
Tan et al. (2023) (*4*)	Singapore	Cohort	Completed primary series	Medically attended symptomatic infection	22.0 (9.0 to 33.0)	NA	NA	802,046
Tan et al. (2023) (*4*)	Singapore	Cohort	Boosted	Medically attended symptomatic infection	51.0 (49.0 to 53.0)	NA	NA	982,831
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	74.0 (72.0 to 75.0)	3 to <6 months	NA	855,858
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	58.0 (55.0 to 61.0)	6 to <7 months	NA	820,235
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	49.0 (47.0 to 52.0)	7 to <8 months	NA	881,230
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	37.0 (32.0 to 43.0)	≥8 months	NA	809,865
Tan et al. (2023) (*4*)	Singapore	Cohort	Completed primary series	Medically attended symptomatic infection	22.0 (−10.0 to 39.0) §	NA	NA	631,104
Tan et al. (2023) (*4*)	Singapore	Cohort	Boosted	Medically attended symptomatic infection	51.0 (49.0 to 53.0) §	NA	NA	810,325
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	74.0 (72.0 to 76.0) §	3 to <6 months	NA	689,979
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	59.0 (55.0 to 62.0) §	6 to <7 months	NA	654,656
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	50.0 (47.0 to 52.0) §	7 to <8 months	NA	710,310
Tan et al. (2023) (*4*)	Singapore	Cohort	Vaccinated (adjusted)	Medically attended symptomatic infection	37.0 (31.0 to 43.0) §	≥8 months	NA	643,953
BA.1/2 to BA.4/5								
Yamamoto et al. (2023) (*6*)	Japan	Cohort	All vaccinated (92% 3 doses)	Infection	91.4 (73.2 to 97.3)	NA	NA	2,368
Altarawneh et al. (2022) (*7*)†	Qatar	TND	Vaccinated (matched)	Infection	80.5 (78.8 to82.0)§	168 to 193 days	NA	65,853
Altarawneh et al. (2022) (*7*)†	Qatar	TND	Unvaccinated	Infection	68.7 (64.0 to 72.9)§	167 to 190 days	NA	22,850
Altarawneh et al. (2022) (*7*)†	Qatar	TND	All vaccinated	Infection	83.7 (82.0 to 85.2)§	170 to 194 days	NA	43,003
Altarawneh et al. (2022) (*7*)†	Qatar	TND	Vaccinated (adjusted)	Symptomatic infection	76.3 (66.6 to 83.2)§	175 to 196 days	NA	2,838
Altarawneh et al. (2022) (*7*)†	Qatar	TND	Vaccinated (adjusted)	Infection	78.1 (75.1 to 80.7)§	169 to 193 days	NA	23,125

Continued

Group	Country	Study Design	Vaccination Status	Outcome of reinfection	Protection (95% *CI*)	Time since primary infection (reported)	Time since primary infection (predicted)*	sample size
Altarawneh et al. (2022) (*7*)†	Qatar	TND	Vaccinated (adjusted)	Symptomatic infection	84.5 (81.1 to 87.2)§	168 to 195 days	NA	12,363
Altarawneh et al. (2022) (*7*)†	Qatar	TND	Vaccinated (adjusted)	Infection	80.3 (78.8 to 81.7)§	168 to 193 days	NA	77,399
Malato et al. (2022) (*8*)	Portugal	Cohort	All vaccinated (98% 2 doses)	Infection	75.3 (75.0 to 75.6)	NA	NA	6,885,922
Malato et al. (2022) (*8*)	Portugal	Cohort	All vaccinated (98% 2 doses)	Infection	76.8 (76.5 to 77.1)§	NA	NA	6,279,978
Hansen et al. (2023) (*9*)	Denmark	Case-control	Three mRNA doses	Infection	92.7 (91.6 to 93.7)	NA	NA	187,347
Hansen et al. (2023) (*9*)	Denmark	Case-control	Three mRNA doses	Infection	92.9 (91.6 to 93.7)§	NA	NA	17,238
Hansen et al. (2023) (*9*)	Denmark	Case-control	Three mRNA doses	Infection	92.5 (91.4 to 93.5)§	NA	NA	104,339
Hansen et al. (2023) (*9*)	Denmark	Case-control	Three mRNA doses	Infection	94.4 (93.8 to 95.0)§	NA	NA	219,643
Hansen et al. (2023) (*9*)	Denmark	Case-control	Three mRNA doses	Infection	92.8 (91.7 to 93.7)§	NA	NA	187,347
BA.1/2 to BA.2.75								
Chemaitelly et al. (2023) (*10*)	Qatar	TND	Vaccinated (matched)	Infection	49.9 (47.6 to 52.1)	NA	NA	105,431
Chemaitelly et al. (2023) (*10*)	Qatar	TND	Vaccinated (matched)	Symptomatic infection	50.2 (43.1 to 56.4)	NA	NA	13,099
Chemaitelly et al. (2023) (*10*)	Qatar	TND	Vaccinated (matched)	Infection	47.4 (44.8 to 49.8)§	NA	NA	105,431
Chemaitelly et al. (2023) (*10*)	Qatar	TND	Vaccinated (matched)	Symptomatic infection	49.7 (42.3 to 56.1)§	NA	NA	13,099
Chemaitelly et al. (2023) (*10*)	Qatar	TND	Unvaccinated	Infection	32.2 (25.5 to 38.3)§	NA	NA	35,577
Chemaitelly et al. (2023) (*10*)	Qatar	TND	All vaccinated	Infection	53.9 (51.5 to 56.3)§	NA	NA	69,854
BA.4/5 to BA.2.75								
Chemaitelly et al. (2023) (*10*)	Qatar	TND	Vaccinated (matched)	Infection	80.6 (71.2 to 87.0)	NA	NA	102,271
Chemaitelly et al. (2023) (*10*)	Qatar	TND	Vaccinated (matched)	Symptomatic infection	91.4 (35.8 to 98.8)	NA	NA	12,680
Chemaitelly et al. (2023) (*10*)	Qatar	TND	Vaccinated (matched)	Infection	79.4 (69.4 to 86.2)§	NA	NA	102,271
Chemaitelly et al. (2023) (*10*)	Qatar	TND	Vaccinated (matched)	Symptomatic infection	91.3 (35.0 to 98.8)§	NA	NA	12,680
Chemaitelly et al. (2023) (*10*)	Qatar	TND	Unvaccinated	Infection	44.4 (−4.0 to 70.3)§	NA	NA	34,862
Chemaitelly et al. (2023) (*10*)	Qatar	TND	All vaccinated	Infection	87.4 (78.7 to 92.5)§	NA	NA	67,409

Note: "NA" means not applicable.

Abbreviation: TND=Test-negative design.

* Time since primary infection was determined based on GISAID data, with 50% serving as the judgment standard for epidemic strain. A variant was considered dominant if it exceeds 50%.

† Based on the research start date and the definition of interval of reinfection, if the most recent infection occurred during the BA.1/2 dominant period, the variant for prior infection was considered to belong to BA.1/2.

§ Sensitivity analysis results of the original literature.

SUPPLEMENTARY TABLE S3. Results of meta-regression.

Group	Estimate	Se	*P*
Result of meta-regression (BA.1 to BA.2):			
Intercept	−2.4363	0.6758	0.0003
Time since infection(days)	0.0051	0.0037	0.1622*
Result of meta-regression (BA.1 to BA.4/5):			
Intercept	−2.1593	0.5595	0.0001
Time since infection(days)	0.0045	0.0015	0.0029*
Result of meta-regression (BA.2 to BA.4/5):			
Intercept	−3.3310	0.5131	<0.0001
Time since infection(days)	0.0071	0.0016	<0.0001*

Note: Even though the *P* value for the BA.1 to BA.2 group surpassed 0.05, meta-regression was conducted due to the substantial impact of time on the analysis.
Abbreviation: SE=Standard error.
* *P* values represent results of test of moderators.

SUPPLEMENTARY TABLE S4. Results of the bias assessment (ROBINS-I).

First author (Year)	Bias due to confounding	Bias in selection of participants into the study	Bias in classification of interventions	Bias due to deviations from intended interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall bias
Chemaitelly et al. (2022) (*1*)	Moderate	Low	Low	Low	Low	Low	Moderate	Moderate
Carazo et al. (2023) (*2*)	Moderate	Serious	Low	Low	Serious	Low	Low	Serious
Chen et al. (2023) (*3*)	Serious	Serious	Low	Low	Moderate	Serious	Moderate	Serious
Tan et al. (2023) (*4*)	Moderate	Low	Low	Low	Serious	Low	Low	Serious
Jang et al. (2023) (*5*)	Moderate	Serious	Low	Low	Low	Low	Moderate	Moderate
Yamamoto et al. (2023) (*6*)	Moderate	Low	Low	Low	Moderate	Moderate	Low	Moderate
Altarawneh et al. (2022) (*7*)	Moderate	Low	Low	Low	Low	Low	Low	Moderate
Malato et al. (2022) (*8*)	Serious	Low	Low	Low	Serious	Low	Moderate	Serious
Hansen et al. (2023) (*9*)	Moderate	Serious	Low	Low	Moderate	Low	Low	Serious
Chemaitelly et al. (2023) (*10*)	Moderate	Low	Low	Low	Low	Low	Low	Moderate

REFERENCES

1. Chemaitelly H, Ayoub HH, Coyle P, Tang P, Yassine HM, Al-Khatib HA, et al. Protection of Omicron sub-lineage infection against reinfection with another Omicron sub-lineage. Nat Commun 2022;13(1):4675. https://doi.org/10.1038/s41467-022-32363-4.
2. Carazo S, Skowronski DM, Brisson M, Barkati S, Sauvageau C, Brousseau N, et al. Protection against Omicron (B.1.1.529) BA.2 reinfection conferred by primary Omicron BA.1 or pre-Omicron SARS-CoV-2 infection among health-care workers with and without mRNA vaccination: a test-negative case-control study. Lancet Infect Dis 2023;23(1):45 – 55. http://dx.doi.org/10.1016/S1473-3099(22)00578-3.
3. Chen XS, Xu YL, Xie Y, Song WL, Hu Y, Yisimayi A, et al. Protective effect of plasma neutralization from prior SARS-CoV-2 Omicron infection against BA. 5 subvariant symptomatic reinfection. Lancet Reg Health West Pac 2023;33:100758. https://doi.org/10.1016/j.lanwpc.2023.100758.
4. Tan CY, Chiew CJ, Pang D, Lee VJ, Ong B, Lye DC, et al. Protective immunity of SARS-CoV-2 infection and vaccines against medically attended symptomatic Omicron BA. 4, BA.5, and XBB reinfections in Singapore: a national cohort study. Lancet Infect Dis 2023;23(7):799 – 805. https://doi.org/10.1016/S1473-3099(23)00060-9.
5. Jang EJ, Choe YJ, Kim RK, Lee S, Park SK, Park YJ. Estimated effectiveness of prior SARS-CoV-2 BA. 1 or BA.2 infection and booster vaccination against Omicron BA.5 subvariant infection. JAMA Netw Open 2023;6(3):e232578. https://doi.org/10.1001/jamanetworkopen.2023.2578.
6. Yamamoto S, Mizoue T, Ohmagari N. Analysis of previous infection, vaccinations, and anti-SARS-CoV-2 antibody titers and protection against infection with the SARS-CoV-2 Omicron BA. 5 variant. JAMA Netw Open 2023;6(3):e233370. https://doi.org/10.1001/jamanetworkopen.2023.3370.
7. Altarawneh HN, Chemaitelly H, Ayoub HH, Hasan MR, Coyle P, Yassine HM, et al. Protective effect of previous SARS-CoV-2 infection against

Omicron BA. 4 and BA.5 subvariants. N Engl J Med 2022;387(17):1620 – 2. https://doi.org/10.1056/NEJMc2209306.
8. Malato J, Ribeiro RM, Leite PP, Casaca P, Fernandes E, Antunes C, et al. Risk of BA. 5 infection among persons exposed to previous SARS-CoV-2 variants. N Engl J Med 2022;387(10):953 – 4. https://doi.org/10.1056/NEJMc2209479.
9. Hansen CH, Friis NU, Bager P, Stegger M, Fonager J, Fomsgaard A, et al. Risk of reinfection, vaccine protection, and severity of infection with the BA. 5 Omicron subvariant: a nation-wide population-based study in Denmark. Lancet Infect Dis 2023;23(2):167 – 76. https://doi.org/10.1016/S1473-3099(22)00595-3.
10. Chemaitelly H, Tang P, Coyle P, Yassine HM, Al-Khatib HA, Smatti MK, et al. Protection against reinfection with the Omicron BA. 2.75 subvariant. N Engl J Med 2023;388(7):665 – 7. https://doi.org/10.1056/NEJMc2214114.

Methods and Applications

Exploring the Lagged Correlation Between Baidu Index and Influenza-Like Illness — China, 2014–2019

Xuan Han[1]; Jiao Yang[1]; Yan Luo[1]; Dazhu Huo[2]; Xuya Yu[1]; Xuancheng Hu[3]; Ling Xin[1]; Liuyang Yang[1]; Hualei Xin[1]; Ting Zhang[1,#]; Zhongjie Li[1,#]; Weizhong Yang[1,#]

ABSTRACT

Introduction: This study investigated the lagged correlation between Baidu Index for influenza-related keywords and influenza-like illness percentage (ILI%) across regions in China. The aim is to establish a scientific foundation for utilizing Baidu Index as an early warning tool for influenza-like illness epidemics.

Methods: In this study, data on ILI% and Baidu Index were collected from 30 provincial-level administrative divisions (PLADs) spanning April 2014 to March 2019. The Baidu Index was categorized into Overall Index, Ordinary Index, Prevention Index, Symptom Index, and Treatment Index based on search query themes. The lagged correlation between the Baidu Index and ILI% was examined through the cross-correlation function (CCF) method.

Results: Correlating the Baidu Overall Index of 30 PLADs with ILI% revealed CCF values ranging from 0.46 to 0.86, with a median lag of 0.5 days. Subcategory analysis indicated that the Prevention Index and Symptom Index exhibited quicker responses to ILI%, with median lags of −9 and −0.5 days, respectively, compared to 0 and 3 days for the Ordinary and Treatment Indexes. The median lag days between the Baidu Index and the ILI% were earlier in the northern PLADs compared to the southern PLADs.

Discussion: The Prevention and Symptom Indexes show promising predictive capabilities for influenza-like illness epidemics.

Influenza, an acute respiratory infection caused by influenza viruses (*1*), affected approximately one billion people worldwide annually between 1999 and 2015 (*2*). Timely identification of the influenza season' s onset is crucial for preparing national and local healthcare resources (*3*), fostering vaccination uptake, strengthening public health measures (*4*), curtailing disease spread, and reducing the impact of seasonal influenza.

Surveillance plays a crucial role in the prevention and management influenza (*5*). Currently, the primary global surveillance methods for influenza include monitoring influenza-like illness and influenza virus activity. These methods aim to capture fluctuations in patient visits and the intensity of influenza viral circulation, providing insights into the onset, peak, and conclusion of seasonal influenza outbreaks (*6*). Traditional surveillance approaches, which involve weekly data reporting and case-based analysis, are prone to delays in detecting early signs of influenza epidemics (*7*). Previous research has demonstrated the utility of internet search data in identifying infectious disease outbreaks (*8*) and highlighted the potential of monitoring epidemic trends using information from social media and online activities of Internet users (*9*). However, accurate monitoring hinges on selecting and utilizing disease-specific web search keywords effectively.

This study examines the lagged correlation between influenza-like illness percentage (the ratio of the total number of influenza-like cases to the total number of outpatient emergency department visits; ILI%) and the Baidu Index of influenza-related keywords in 30 provincial-level administrative divisions (PLADs) in China. It further categorizes the Baidu Index to identify a specific subset that shows a strong correlation and timely response. These findings aim to establish a scientific foundation for enabling early detection of influenza-like disease outbreaks.

METHODS

This study utilized weekly influenza surveillance data, including the number of ILI cases and total outpatient emergency department visits, obtained from the National Influenza Center of China (*10*). The data

were sourced from outpatient emergency departments in sentinel hospitals located across 30 PLADs (excluding Xizang Autonomous Region) spanning from April 2014 to March 2019. To calculate daily ILI%, a cubic spline function was employed to interpolate weekly ILI%. The cubic spline function is a widely accepted method for curve fitting and interpolation, commonly used to convert weekly influenza data into daily estimates based on weekly reports (*11*).

The Baidu Index is a calculated total of search frequencies for specific keywords on Baidu web pages, sourced from the public Baidu Index website (*12*). The selection and refinement of keywords for the Baidu Index were based on methodologies outlined in prior research, encompassing "influenza bidding terms, Baidu Index demand mapping, expert consultation, and literature summarization." Keywords unrelated to influenza, not included in the Baidu Index database, not searched for in the past year, or pertaining to specific strains were excluded (*13*). A total of 39 influenza-related keywords were compiled and categorized under the "Overall Index." Based on search patterns, these keywords were further segmented into four groups: basic terminology, prevention, symptoms, and treatment, designated as the Ordinary Index, Prevention Index, Symptom Index, and Treatment Index, respectively (Table 1).

The lagged correlation between the Baidu Index and ILI% was examined using the Cross-Correlation Function (CCF) method. This technique was utilized to assess the cross-correlation between the two time series, specifically to investigate if a particular pattern in one series tends to follow a pattern in the other series. The method generates a judgment indicator in the form of CCF values to determine the correlation between the two time series, with the formulas detailed as per reference (Table 2) (*14*).

Initially, using sample estimates of cross-covariance in Equation 1 to measure co-variation at different time points, offering preliminary insights.

$$\hat{\gamma}_{xy}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y}) \quad (1)$$

Then the quantitative correlation strength was standardized by cross-correlation coefficient in Equation 2. This step normalized covariance, providing a clearer interpretation.

$$\rho_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\,\hat{\gamma}_y(0)}} \quad (2)$$

Equation 3 finalized the expression of the cross-correlation coefficient, offering a systematic approach for a profound understanding of the dynamic relationship between the targeted time series.

$$\rho_{xy}(s,t) = \frac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\,\gamma_y(t,t)}} \quad (3)$$

The CCF values fall within the range of [-1,1], denoting the correlation of the series at various lag orders. Correlation was categorized into three groups: CCF values >0.4 were considered correlated (*14*); values >0.6 indicated strong correlation; while values <0.4 were deemed not correlated and were therefore excluded from the analysis.

Based on findings from pertinent studies, a maximum lag period of ±14 days was determined. The day displaying the highest values of CCF was identified as the optimal lag day. Optimal lag days were categorized into three groups: <0 days, =0 days, >0 days. These categories signify that the Baidu Index may have the capacity to anticipate fluctuations in influenza epidemics before, during, or after ILI%.

Data processing and graphical representation were performed using R (version 4.2.1) software, developed by R Core Team, headquartered in Vienna, Austria.

RESULTS

The findings from the lagged correlation analysis

TABLE 1. The classification of Baidu Index keywords related to influenza.

Classification	Keywords
Overall Index	Contains all the keywords below
Ordinary Index	Influenza; Cold; Viral Influenza; Seasonal Influenza; Influenza Virus; Influenza transmission route; How influenza is transmitted
Prevention Index	Flu Vaccine; Prevention of influenza; Precautions against influenza; Prevent influenza; How to prevent influenza; Flu Vaccine Side Effects; Flu Vaccine Prices; Is the Flu Vaccine Necessary; Influenza Prevention Knowledge
Symptom Index	Febrile; Fever; Cough; Pharyngalgia; Sore throat; Runny nose; Pneumonia; Chest tightness; Symptoms of influenza; Sneezing; Lacking in strength; Muscle soreness
Treatment Index	Flu Treatment; Cold Medicine; Antipyretic; Lianhuaqingwen*; What is the Most Effective Flu Medicine; Liuganwan; Ganmaoqingre*; Banlangen*; Baijiahei*; Oseltamivir; Tamiflu

* The names of traditional Chinese medicines.

TABLE 2. Description of formula components.

Equation component	Explanation
Equation 1	
$\hat{\gamma}_{xy}$	The covariance between time series x and y at a lag of h time points.
x_{t+h}	The values of time series x and y at time points t+h and t.
$\bar{x}$ and $\bar{y}$	The means of time series x and y.
n	The length of the time series.
Equation 2	
$\rho_{xy}(h)$	The cross-correlation coefficient between time series x and y at a lag of h time points.
$\hat{\gamma}_{xy}(h)$	The covariance between time series x and y at a lag of h time points.
$\hat{\gamma}_x(0)$ and $\hat{\gamma}_y(0)$	The variances of time series x and y at time 0.
Equation 3	
$\rho_{xy}(s,t)$	The cross-correlation coefficient between time series x and y at time points s and t.
$\gamma_{xy}(s,t)$	The covariance between time series x and y at time points s and t.
$\gamma_x(s,s)$ and $\gamma_y(t,t)$	The variances of time series x at time s and y at time t.

between the Baidu Overall Index and ILI% reveal significant results. The CCF values fluctuate within the range of 0.46 to 0.86, indicating a lagged correlation pattern across 30 PLADs. Notably, 25 PLADs demonstrate a robust correlation trend (Figure 1). The median lag day between the Baidu Overall Index and ILI% is 0.5 days, with a range from −5 to 11 days. Specifically, 7 PLADs show a lag of less than 0 days, 7 PLADs display a lag of 0 days, and 16 PLADs exhibit a lag exceeding 0 days.

The Baidu Overall Index consists of four distinct subindices: Ordinary Index, Prevention Index, Symptom Index, and Treatment Index. The shortest median time discrepancy was observed in the Prevention Index at −9 days compared to the incidence of ILI%, with 18 PLADs registering a negative lag and 2 PLADs exhibiting no lag. The Symptom Index followed closely with a median lag of −0.5 days related to ILI%, where 13 PLADs displayed a negative lag and 2 PLADs recorded no lag. The Ordinary Index aligned most closely with ILI% showing a median lag of 0 days, with 9 PLADs experiencing a negative lag and 7 PLADs reporting no lag. In contrast, the Treatment Index showed the longest median lag of 3 days when compared to ILI%, with all 25 PLADs showing a lag exceeding 0 days.

When analyzed according to northern and southern regions, the median lag days for the Prevention Index, Symptom Index, and ILI% in the northern PLADs were −9 days and −5 days, while the Ordinary Index, Overall Index, and ILI% had median lag days of 0 days. The Treatment Index showed a median lag of 3 days behind ILI%. In the southern PLADs, the Prevention Index had a median lag of −6 days compared to ILI%, indicating its potential for early influenza outbreak detection. The Ordinary Index aligned closely with ILI%, with a median lag of 0 days. The Symptom, Treatment, and Overall Indexes showed median lags of 4, 3, and 3 days after ILI%, respectively (Figure 2).

DISCUSSION

This study reveals a significant correlation between the Baidu Overall Index and ILI%. Refining the classification of the Overall Index resulted in more advanced trends for the Prevention Index and Symptom Index, indicating potential value for warning of influenza disease epidemics. Notably, the Baidu Index trend advanced for a longer duration in the northern PLADs compared to the southern PLADs. This novel provincial-specific analysis of the Baidu Index deepens our insight into its intricate relationship with influenza transmission. Exploring geographic variations in the correlation between the Baidu Index and ILI% suggests regional differences in the utility of Internet search data for predicting future influenza-like illness epidemics.

The Prevention and Symptom Indexes demonstrate significant lead times, with median delays of −9 days and −0.5 days, respectively, compared to ILI%. On the other hand, the Treatment Index lags behind ILI% by a median lag of 3 days. Findings indicate that changes in the Prevention Index precede changes in the ILI% trend, possibly because individuals draw on past experiences and local outbreaks to proactively seek

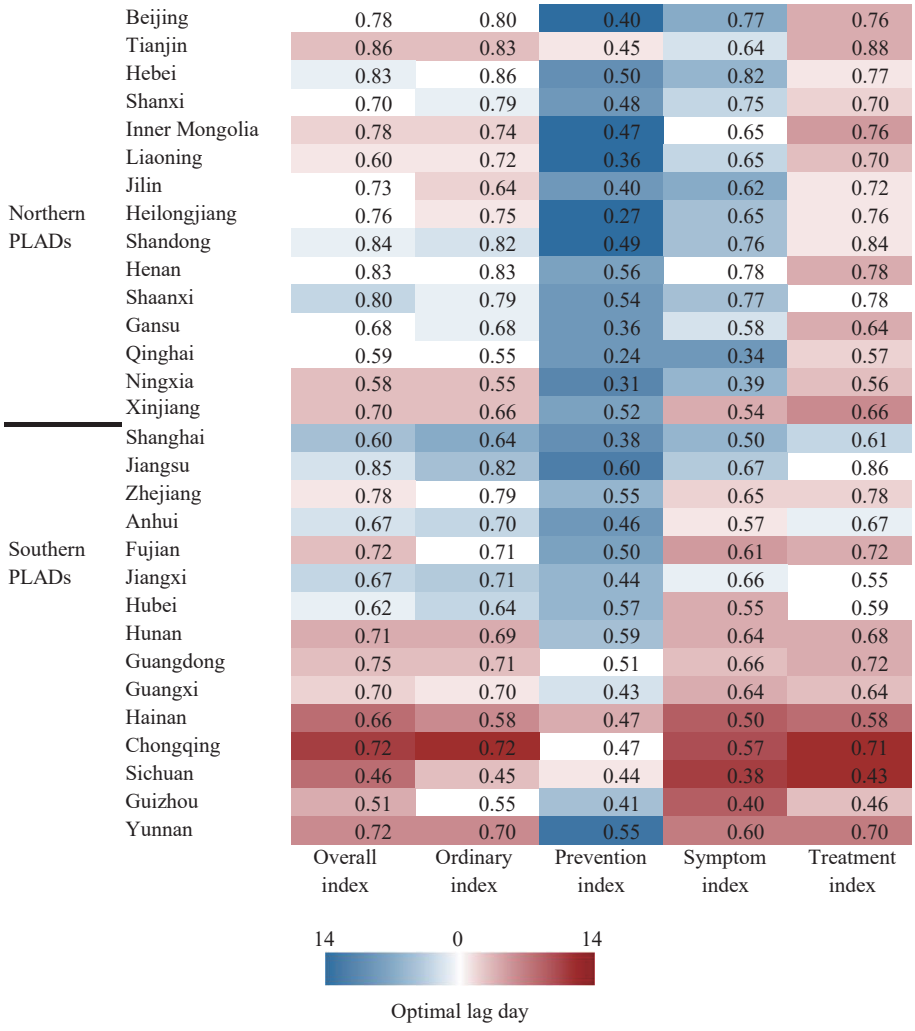


FIGURE 1. The optimal lag days and CCF values between Baidu Index and ILI% in 30 PLADs*.
Abbreviation: PLAD=provincial-level administrative division; CCF=cross-correlation function; ILI%=influenza-like illness percentage.
* The numbers on each plate represent the CCF values.

influenza prevention information before the onset of infection. The Symptom Index leads ILI% trends, potentially as individuals begin feeling unwell, search for symptom-related terms to identify potential causes, and subsequently seek care at outpatient clinics or emergency departments. In contrast, the Treatment Index trend lags behind ILI%, as individuals may search for influenza treatment-related keywords (such as medications) only when symptoms become severe or self-recovery is unsuccessful. It is important to acknowledge that fluctuations in epidemic surveillance data may introduce discrepancies in the correlation analysis results between the Baidu index and ILI%.

The predictive capability of the Baidu Index in anticipating changes in ILI% trends might be more effective in the northern region compared to the southern region. In the northern region, the median lag days between the Baidu Overall Index and ILI% were 0 days, while in the southern region they were 3 days. Upon further analysis of the subcategories of the four types of Baidu indexes, it was observed that two types of keywords (Prevention Index and Symptom Index) preceded the ILI% trend in the northern region, whereas only the Prevention Index did so in the southern region. This discrepancy could be attributed to the distinct seasonal influenza patterns in the northern region, characterized by more prominent single-peak epidemics than those in the southern region. These findings suggest that utilizing the Baidu Index for developing future influenza epidemic forecasting models may be more successful in the northern region than in the southern region.

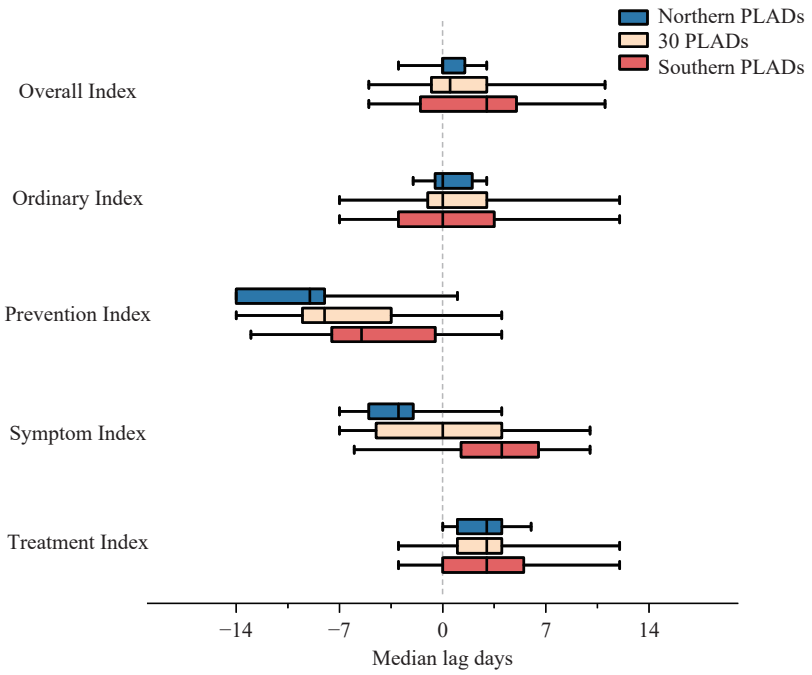


FIGURE 2. The median lag days between Baidu Index and ILI% in 30 PLADs, northern and southern PLADs.
Abbreviation: PLADs=provincial-level administrative divisions; ILI%=influenza-like illness percentage.

The Baidu Index demonstrated more prompt changes in trends compared to the ILI reporting process. According to the National Influenza Surveillance Technical Guidelines (2017 edition), surveillance sentinel sites report weekly cases every Monday (*15*), introducing a delay of 1–7 days from the case date. Conversely, the Baidu Index webpage displays data from the day before up to the following day, trailing just one day behind real-time conditions. Essentially, when the Baidu Index trend corresponds with the ILI% (with a lag of 0 days), it can provide insights into influenza epidemics 0–6 days in advance of the current reporting period, thereby circumventing delays in manual statistical reporting.

The study is subject to some limitations. Initially, it focused on categorizing and analyzing Baidu Index data relating to influenza-related keywords, overlooking an analysis of the lagged correlation between the Baidu Index of individual keywords and ILI%. Furthermore, the study did not account for discrepancies resulting from regional variations in the processes and methodologies of influenza-like illness reporting by provincial influenza surveillance sentinel sites. Future research could investigate the correlation between individual keywords and ILI% on a provincial or municipal level, leading to potential adjustments in the surveillance keywords used in each region based on findings. Additionally, a more in-depth exploration of keyword weighting may enhance surveillance accuracy and early warning capabilities.

In conclusion, a substantial correlation was observed between the Baidu Overall Index and ILI%. Additionally, the changes in trend for the Prevention Index and Symptom Index preceded those of the ILI%. Future research could explore the development of a predictive model using the Prevention Index and Symptom Index to anticipate and provide early warnings for influenza epidemics. The predictive nature of these indices could aid health authorities in identifying patient care requirements earlier in the influenza season, facilitating the prompt implementation of preventive measures. Furthermore, initiating early public health campaigns focused on influenza prevention could enhance public awareness.

Conflicts of interest: No conflicts of interest.

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

doi: 10.46234/ccdcw2024.084

[#] Corresponding authors: Ting Zhang, zt0416@126.com; Zhongjie Li, lizhongjiecdc@163.com; Weizhong Yang, yangweizhong@cams.cn.

[1] School of Population Medicine and Public Health, Chinese Academy of Medical Sciences (CAMS) & Peking Union Medical College, Beijing, China; [2] School of Health Policy and Management, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China; [3] Department of management science and information

system, Faculty of Management and Economics, Kunming University of Science and Technology, Kunming City, Yunnan Province, China.

Submitted: December 26, 2023; Accepted: March 26, 2024

REFERENCES

1. National Immunization Advisory Committee (NIAC) Technical Working Group (TWG), Influenza Vaccination TWG. Technical guidelines for seasonal influenza vaccination in China (2022-2023). Chin J Epidemiol 2022;43(10):1515 – 44. https://doi.org/10.3760/cma.j.cn112338-20220825-00734.
2. Iuliano AD, Roguski KM, Chang HH, Muscatello DJ, Palekar R, Tempia S, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet 2018;391(10127):1285 – 300. https://doi.org/10.1016/S0140-6736(17)33293-2.
3. Lee VJ, Ho ZJM, Goh EH, Campbell H, Cohen C, Cozza V, et al. Advances in measuring influenza burden of disease. Influenza Other Respir Viruses 2018;12(1):3 – 9. https://doi.org/10.1111/irv.12533.
4. The Lancet. Preparing for seasonal influenza. Lancet 2018;391(10117):180. https://doi.org/10.1016/S0140-6736(18)30087-4.
5. WHO. Influenza (seasonal). 2018. http://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal). [2023-12-2].
6. Pearce N, Vandenbroucke JP, VanderWeele TJ, Greenland S. Accurate statistics on COVID-19 are essential for policy guidance and decisions. Am J Public Health 2020;110(7):949 – 51. https://doi.org/10.2105/AJPH.2020.305708.
7. Paul R, Han D, DeDoncker E, Prieto D. Dynamic downscaling and daily nowcasting from influenza surveillance data. Stat Med 2022;41(21):4159 – 75. https://doi.org/10.1002/sim.9502.
8. 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.
9. Lai SJ, Feng LZ, Leng ZW, Lyu X, Li RY, Yin L, et al. Summary and prospect of early warning models and systems for infectious disease outbreaks. Chin J Epidemiol 2021;42(8):1330 – 5. https://doi.org/10.3760/cma.j.cn112338-20210512-00391.
10. Chinese National Influenza Center. Influenza weekly report of the national influenza center of China. https://ivdc.chinacdc.cn/cnic/zyzx/lgzb/. [2023-10-2]. (In Chinese).
11. Ali ST, Cowling BJ, Wong JY, Chen DX, Shan SW, Lau EHY, et al. Influenza seasonality and its environmental driving factors in mainland China and Hong Kong. Sci Total Environ 2022;818:151724. https://doi.org/10.1016/j.scitotenv.2021.151724.
12. Baidu Index Public Website. https://index.baidu.com. (In Chinese).
13. Yang LY, Zhang T, Han X, Yang J, Sun YX, Ma LB, et al. Influenza epidemic trend surveillance and prediction based on search engine data: deep learning model study. J Med Internet Res 2023;25:e45085. https://doi.org/10.2196/45085.
14. Shumway RH, Stoffer DS. Time series analysis and its applications. New York: Springer. 2000. http://dx.doi.org/10.1007/978-1-4757-3261-0.
15. Chinese National Influenza Center. National influenza surveillance technical guidelines (2017 edition). https://ivdc.chinacdc.cn/cnic/zyzx/jcfa/201709/t20170930_153976.htm. (In Chinese).

Methods and Applications

Comparison Between Threshold Method and Artificial Intelligence Approaches for Early Warning of Respiratory Infectious Diseases — Weifang City, Shandong Province, China, 2020–2023

Ting Zhang[1,2,3,&]; Liuyang Yang[4,5,&]; Ziliang Fan[6,&]; Xuancheng Hu[1,2,3]; Jiao Yang[1,2,3]; Yan Luo[1,2,3]; Dazhu Huo[7]; Xuya Yu[1,2,3]; Ling Xin[1,2,3]; Xuan Han[1,2,3]; Jie Shan[6]; Zhongjie Li[1,2,3]; Weizhong Yang[1,2,3,#]

ABSTRACT

Introduction: Respiratory infectious diseases, such as influenza and coronavirus disease 2019 (COVID-19), present significant global public health challenges. The emergence of artificial intelligence (AI) and big data offers opportunities to improve traditional disease surveillance and early warning systems.

Methods: The study analyzed data from January 2020 to May 2023, comprising influenza-like illness (ILI) statistics, Baidu index, and clinical data from Weifang. Three methodologies were evaluated: the adaptive dynamic threshold method (ADTM) for dynamic threshold adjustments, the machine learning supervised method (MLSM), and the machine learning unsupervised method (MLUM) utilizing anomaly detection. The comparison focused on sensitivity, specificity, timeliness, and warning consistency.

Results: ADTM issued 37 warnings with a sensitivity of 71% and a specificity of 85%. MLSM generated 35 warnings, with a sensitivity of 82% and a specificity of 87%. MLUM produced 63 warnings with a sensitivity of 100% and specificity of 80%. The initial warnings from ADTM and MLUM preceded those from MLSM by five days. The Kappa coefficient indicated moderate agreement between the methods, with values ranging from 0.52 to 0.62 (P<0.05).

Discussion: The study explores the comparison between traditional methods and two machine learning approaches for early warning systems. It emphasizes the validation of machine learning's reliability and underscores the unique advantages of each method. Furthermore, it stresses the significance of integrating machine learning models with various data sources to enhance public health preparedness and response, alongside acknowledging limitations and the need for broader validation.

Respiratory infectious diseases like seasonal influenza and coronavirus disease 2019 (COVID-19) have the potential to escalate into pandemics or epidemics, rapidly spreading and endangering global public health (*1*). The World Health Organization (WHO) estimates that influenza results in around 1 billion infections, 3–5 million instances of severe illness, and 290,000–650,000 deaths each year (*2*). Timely detection and swift responses to these diseases are crucial in averting outbreaks and controlling the public health threats they bring (*3*).

Threshold-based approaches have traditionally been utilized to promote vigilance regarding respiratory diseases. Models such as the moving percentile method, cumulative sum control chart, and exponentially weighted moving average control chart (*4–5*) evaluate the dynamic nature of time-series data in infectious disease early warning systems. These models issue alerts when reported case numbers meet or exceed predefined thresholds (*6*). With advancements in information technology, there has been a significant shift from reliance on single-source data to incorporating multiple sources. This shift introduces complex analytical processes and the challenge of mitigating noise from large datasets. In the context of COVID-19 management, the application of artificial intelligence (AI) has proven to be exceptionally promising in overcoming these obstacles within surveillance and early warning frameworks (*7*). As a result, the development of robust and dependable AI-driven methods has become crucial in the realm of infectious disease epidemiology.

This study compares the outcomes of traditional methods, specifically the process-credible threshold approach, with two machine learning techniques to assess the suitability and reliability of machine learning methods for early warning systems in infectious disease detection.

METHODS

In this study, "infectious disease early warning" refers to identifying outbreak signs before or during its initial phases through the analysis of infectious disease data from various surveillance sources. Data from January 2020 to May 2023, including influenza-like illness (ILI) statistics, the Baidu index, and clinical data, were analyzed. All methodologies used in this study relied on a uniform and collective data origin.

ILI data from the National Influenza Surveillance Network in China were segmented into China Northern ILI%, Shandong Province ILI%, and Weifang City ILI%. The ILI definition matched the criteria established by the Department of Disease Control and Prevention of the National Health Commission of China, identifying ILI as fever (body temperature ≥38 °C) with cough or sore throat, as referenced (*8*). ILI% represented the ratio of cases among individuals seeking medical care.

The Baidu index, sourced from the publicly accessible Baidu index website, represents the aggregate search frequency of specified keywords on Baidu web pages, with each keyword assigned a particular weight. In the context of "treatment," the index takes into account search terms including "Flu Treatment," "Cold Medicine," "Antipyretic," "Lianhuaqingwen," "What is the most effective flu medicine," "Liuganwan," "Ganmaoqingre," "Banlangen," "Baijiahei," "Oseltamivir," and "Tamiflu." Conversely, the Baidu index for "non-fever symptoms" comprises phrases such as "Fever," "Cough," "Pharyngalgia," "Sore throat," "Runny nose," "Pneumonia," "Chest tightness," "Symptoms of influenza," "Sneezing," "Lacking in strength," and "Muscle soreness" (*9*).

Clinical data from primary and tertiary medical institutions in Weifang City included 21,584,148 chief complaints, 23,128,256 initial diagnoses, 39,486,100 pharmaceutical sales, and 426,171 instances of emergency call data (120). This study focused on respiratory symptoms data, incorporating chief complaints, diagnoses, pharmaceutical sales, and emergency call data (120) with proportional representation.

This study performed a comparative analysis of three early warning methods used in Weifang City, China. The first method improves upon the conventional threshold approach by autonomously determining an optimal threshold, enhancing its practical usability. The second method utilizes supervised machine learning models, whereas the third employs unsupervised machine learning models. Specific details of these models are provided below.

Method 1: Adaptive Dynamic Threshold Method (ADTM)

The ADTM method integrates automatic adjustments into conventional fixed-threshold methods to improve sensitivity and specificity. It consists of five comprehensive phases.

Phase 1. Modeling and parameter setting: Establish models for three distinct scenarios: the beginning of an epidemic season, sudden increases in case numbers, and outliers surpassing historical levels. Each scenario had specific thresholds set through various techniques (Table 1). A total of 1,620 thresholds were determined based on the three warning signal scenarios and different criteria (Figure 1). This process aimed to ensure the model's accuracy in accommodating the dynamic and changing patterns in epidemiological data.

Figure 1 illustrates the criteria for activating alerts in various scenarios. In the epidemic season, an alert is triggered by either "Abrupt Growth" or "Outliers

TABLE 1. Scenarios and criteria for setting early warning thresholds for infectious diseases.

Warning signal scenarios	Criteria
A. Outliers over historical levels	**A1.** Exceeds standard deviations (0.5x, 1x, 1.5x, 2x, 3x) compared to the same period over the last three years, calculated for the past 3 or 7 days. **A2.** Exceeds the 50th to 90th percentiles of case numbers compared to the same period in the past three years, calculated for the past 2 days or weeks.
B. Abrupt Growth	Retrospective time: Two intervals of three days, for a total of six days. **B1.** Absolute change, calculated as the percentage difference between the mean case numbers of two 3-day intervals, with thresholds at 10%, 20%, and 30%. **B2.** Acceleration of absolute change, defined as the difference between absolute changes at adjacent intervals. Acceleration thresholds established at 0.005, 0.01, and 0.015.
C. Epidemic season	Criteria based on exceeding historical data thresholds over 3 or 7 consecutive days. **C1.** Set at 0.5 times the historical mean. **C2.** Standard deviation thresholds at 0.8x, 1x, 1.2x, and 1.5x of the historical average. **C3.** Percentile thresholds at the 50th, 70th, 80th, and 90th percentiles based on historical data.

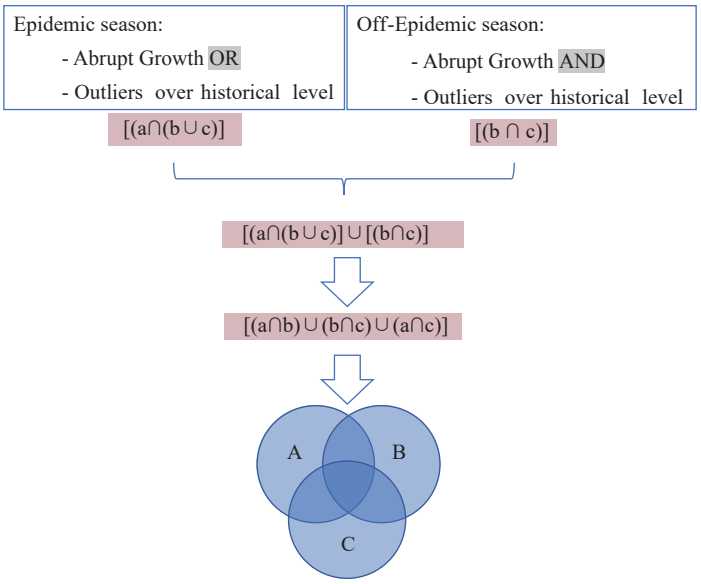


FIGURE 1. Mapping of scenarios and criteria for infectious disease early warning thresholds.

exceeding historical levels." Conversely, during the off-epidemic season, an alert necessitates the concurrent presence of both "Abrupt Growth" and "Outliers exceeding historical levels."

Phase 2. Threshold determination using SoftMax function: This method employs a Softmax function to determine the warning thresholds. The aim is to optimize the balance between timeliness, sensitivity, and specificity for threshold determination, which is crucial for accurate and timely epidemic detection.

Phase 3. Optimal warning strategy for single-source data: Optimally calibrated warning thresholds are applied to single-source data indicators, such as ILI. Warning signals are generated whenever such data surpasses the established threshold, signaling potential health risks that warrant immediate attention.

Phase 4. Integration of multi-source warning signals: Warning signals are synthesized from multi-source warning signals. A comprehensive assessment of warning probability is achieved by calculating a weighted ensemble probability, where each data source is assigned a specific weight. This integration enhances the reliability and accuracy of the warning systems.

Phase 5. Threshold setting for warning probability: A definitive threshold for the warning probability is established to evaluate integrated warning signals. Exceeding this threshold prompts the issuance of an alert, signaling the potential emergence of a public health threat or the initiation of an epidemic.

During these specified phases, the ADTM provides a comprehensive strategy for epidemic surveillance. It incorporates single-source and multi-source data while adjusting thresholds dynamically based on critical epidemiological parameters.

Comparative study of early warning methods: The timeliness of an early warning method was determined by the date of the first warning signal, positioned within the timeline of an outbreak period. The volume of warnings is reflected in the count of days with issued warning signals, as dictated by the warning rules. Consistency was assessed using the Kappa coefficient, which accounts for the probability of random agreement. Statistical significance was attributed to findings with a $P<0.05$.

Method 2: Machine Learning Supervised Method (MLSM)

This approach employs fully supervised learning to reframe the warning issue as a classification task. It accomplishes the categorization of warning levels through the acquisition of multi-source time-series characteristics. The efficacy of early warning for the target metric (Weifang ILI%) is attained by constructing a dataset suitable for supervised learning and utilizing the eXtreme Gradient Boosting (XGBoost) machine learning model (*10*). The XGBoost model, which leverages decision trees and gradient boosting, serves as the underlying framework, which we detail in the Supplementary Materials (available at https://weekly.chinacdc.cn/). Initially, aligning the multi-source time series with the warning labels of the target metric establishes a correspondence between features and labels. Subsequently, these

features and labels are fed into the XGBoost model for training, effectively addressing the supervised learning issue as illustrated in Figure 2.

In the MLSM study, we utilized a training set spanning from January 1, 2020 to November 30, 2022, comprising 1065 days. The test set ranged from December 1, 2022 to May 31, 2023, totaling 182 days. The training set to test set ratio is approximately 6∶1 requirements for dataset partitioning.

Method 3: Machine Learning Unsupervised Method (MLUM)

This approach reconceptualizes the challenge of early warning into a task of anomaly detection. Utilizing unsupervised learning, the model analyzes characteristics of multi-source time series data to identify atypical signals indicative of early warnings. We employ the Isolation Forest algorithm, a machine learning model notably used for its efficacy in anomaly detection (*11*), to thoroughly examine the intrinsic properties of the provided multi-source time series data. The fundamental principle of the Isolation Forest method is that normal and anomalous data points manifest distinct traits; by evaluating and segregating the outliers, the model successfully pinpoints potential anomalies (Supplementary Figure S1, available athttps://weekly.chinacdc.cn/). An advantage of this technique over fully supervised learning is that it eschews the necessity for data labeling, thereby simplifying the implementation of the early warning system (Figure 2). Both the training and test sets in the MLUM were identical to those in the MLSM.

The traditional adaptive dynamic threshold method's effectiveness was rigorously compared with two other methods by assessing their sensitivity and specificity. To establish a reliable benchmark for this evaluation, we used an expert-based consensus. Professionals from the Weifang CDC and senior medical experts reviewed case timelines, labeling moments requiring early warning with a "1" and all other instances with a "0." The application of our technique and the subsequent analyses were performed using Python (version 3.6.13; Python Software Foundation, Fredericksburg, VA, US), aided by the scikit-learn library (version 0.24.2), and the R (version 4.3.1; The R Foundation for Statistical Computing, Vienna, Austria). For Python analyses, the utilized packages included Pandas (1.2.0), Numpy (1.19.5), Xgboost (2.0.3), and Scikit-learn (1.0). For R, the employed packages were ggplot2 (3.4.4), patchwork (1.1.3), scales (1.2.1), dplyr (1.1.4), tidyverse (2.0.0), and readxl (1.4.3).

RESULTS

This study evaluated the performance of the traditional ADTM in comparison with two machine-learning-based methods, MLSM and MLUM, over a period of 182 days from December 1, 2022 to May 31, 2023. ADTM issued 37 warnings with a sensitivity of 71% and a specificity of 85%. MLSM generated 35 warnings with a sensitivity of 82% and a specificity of 87%, while MLUM produced 63 warnings with a sensitivity of 100% and a specificity of 80%. ADTM

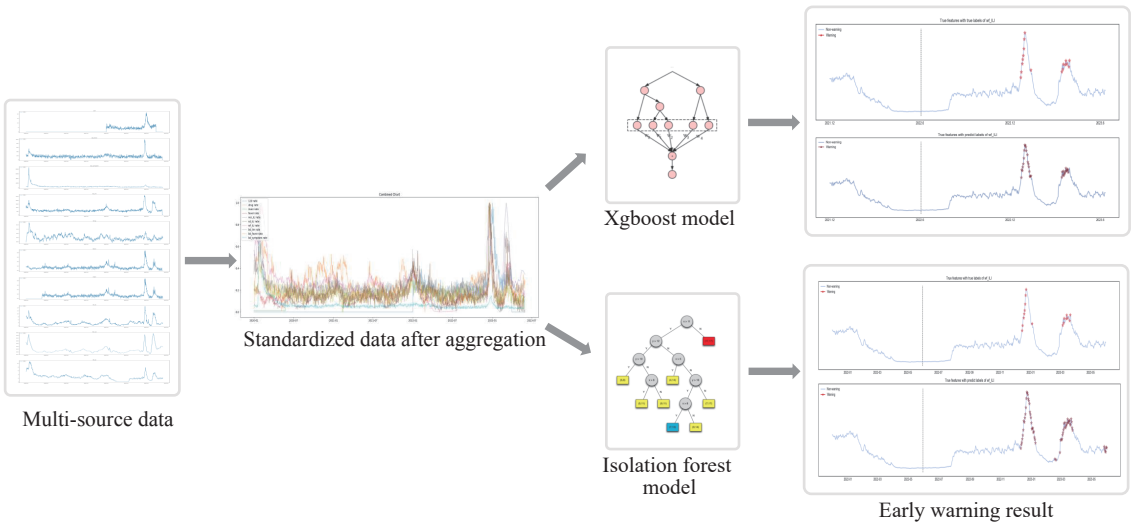


FIGURE 2. Schematic diagram of the MLSM and MLUM models
Abbreviation: MLSM=machine learning supervised method; MLUM=machine learning unsupervised method.

and MLUM issued initial warnings on December 11, with MLSM following on December 16. Pairwise Kappa coefficient analysis indicated significant consistency among these methods (P<0.05) (Figure 3, Table 2).

Panel A illustrates the warning signals derived from the ADTM method for the Weifang ILI% data. Panel B shows the results using the MLSM method, and Panel C depicts the outcomes from the MLUM method. The blue line represents the ILI percentage

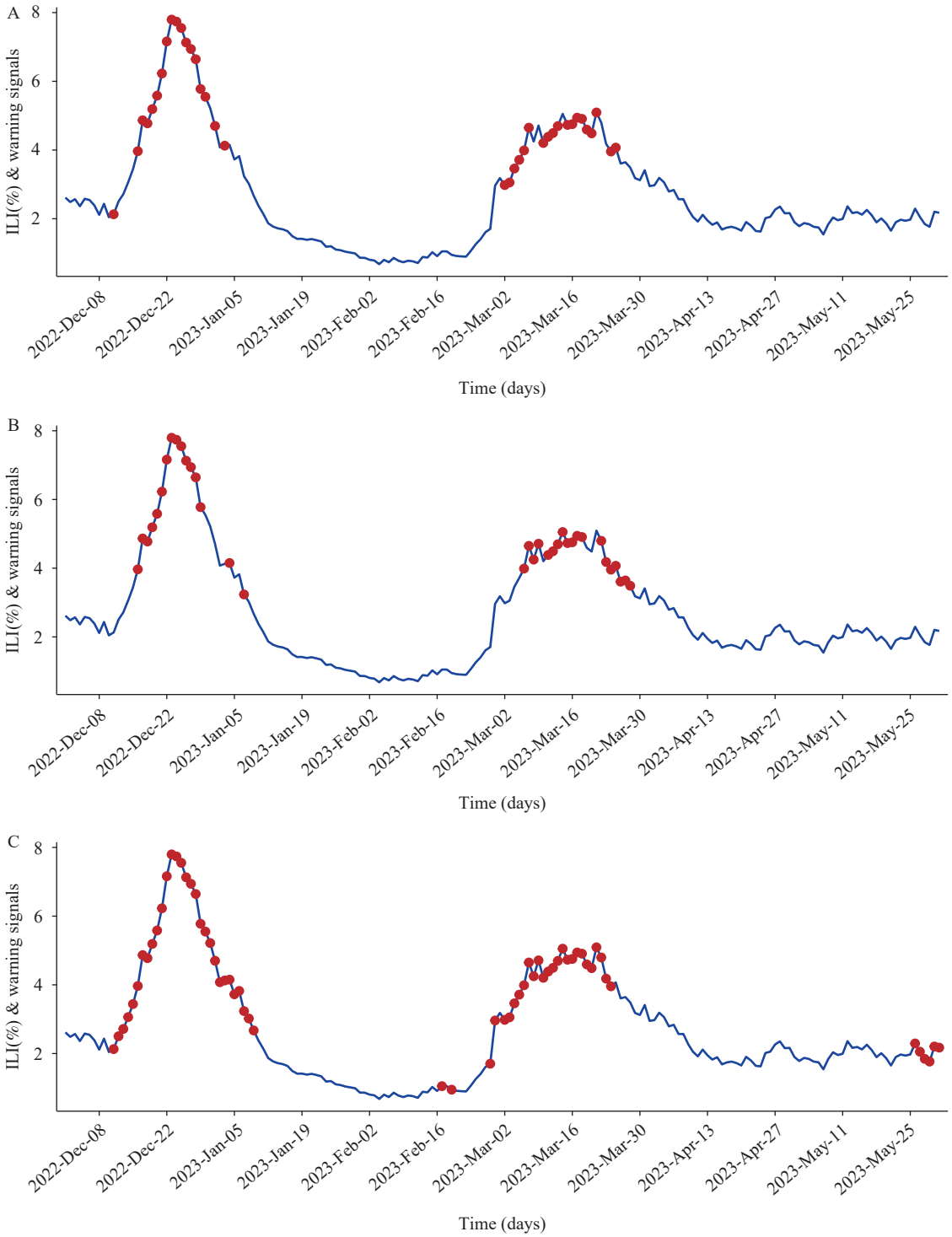


FIGURE 3. Comparative early warning models using three different approaches. (A) ADTM; (B) MLSM; (C) MLUM. Abbreviation: ADTM=adaptive dynamic threshold method; MLSM=machine learning supervised method; MLUM=machine learning unsupervised method.

TABLE 2. Comparative analysis of warning signal detection methods: ADTM, MLSM, and MLUM.

Methods	Warning signals	Timeliness*	Sensitivity (%)	Specificity (%)	Consistency† (Kappa)		
					ADTM	MLSM	MLUM
ADTM	37	Dec 11, 2022	71	85	1.00	§	§
MLSM	35	Dec 16, 2022	82	87	0.62	1.00	§
MLUM	63	Dec 11, 2022	100	80	0.62	0.52	1.00

Note: The values denote the Kappa coefficient.
Abbreviation: ADTM=adaptive dynamic threshold method; MLSM=machine learning supervised method; MLUM=machine learning unsupervised method.
* indicates the date when the warning signals were first initiated.
† represents the consistency, measured by the pairwise Kappa coefficient.
§ signifies a statistical difference with a *P* value of less than 0.05.

curve, while the red points indicate the warning signals.

In addition, in this study, we utilized three evaluation metrics, precision, recall, and F1-score, to evaluate the warning results of the test set, as shown in Supplementary Table S1 (available at https://weekly.chinacdc.cn/).

DISCUSSION

Early warning systems have advanced by utilizing diverse data sources, incorporating big data and machine learning to improve surveillance. This research compares the ADTM approach with MLSM and MLUM, assessing their effectiveness in early warning scenarios. Through establishing ADTM as the reference point, we evaluate the consistency of results from MLSM and MLUM, emphasizing the impact of machine learning on enhancing public health readiness.

This study introduces an enhanced method for infectious disease surveillance, integrating an automatic threshold selection system to enhance adaptability and scalability across various regions. The China Infectious Diseases Automated-Alert and Response System (CIDARS) implements a spatiotemporal early warning model for Type 1 diseases, covering nine infectious diseases, and Type 2 diseases involving 19 infectious diseases, utilizing Fixed-threshold, Temporal, and Spatial detection methods (*10*). However, challenges arose in determining precise thresholds for different regions, times, populations, policies (*11*), and behaviors, hindering rapid adjustments to these factors. To tackle this issue, a dynamic threshold selection function has been developed in this study, enabling real-time adaptation of thresholds, thereby increasing the method's versatility and facilitating its application across diverse geographical areas.

The three methodologies, ADTM, MLSM, and MLUM, exhibit varying effectiveness and suitability in early warning systems. MLSM and MLUM represent the fundamental paradigms of machine learning, each offering unique approaches to problem solving. ADTM and MLUM are particularly relevant for timely anomaly detection crucial for outbreak response. ADTM excels in specificity, reducing false alarms and saving resources, but may lack sensitivity in detecting certain anomalies. MLSM strikes a balance between sensitivity and specificity, albeit with less straightforward interpretability. MLUM stands out for its high sensitivity, benefiting disease detection at the expense of specificity, making it valuable for conditions with significant clinical impacts. Validating the reliability of machine learning methods in infectious disease early warning, the study uses ADTM as a benchmark. Machine learning's computational strength, combined with independence from traditional benchmarks, bears promise for future applications. However, caution is advised as predictive models, with moderate Kappa coefficient agreement, are not infallible and should not be the sole determinants of public health decisions. A more robust approach involves integrating diverse data sources and surveillance methods with predictive models to enhance early warning system reliability and effectiveness, mitigating the limitations of single-model predictions and fortifying public health strategies.

The methodology of the study has limitations, particularly in finding a dependable benchmark for early warning models, notably with machine learning. The study aimed to compare models under similar conditions without in-depth exploration of their intricacies. The MLUM model prioritized timeliness and sensitivity, albeit with a trade-off in specificity due to its parameter settings. Future research may consider more sophisticated models to enhance accuracy. Furthermore, the study was confined to the Weifang area, suggesting the necessity for broader validation in

other regions in subsequent work.

Conflicts of interest: No conflicts of interest.

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

doi: 10.46234/ccdcw2024.119

[#] Corresponding author: Weizhong Yang, yangweizhong@cams.cn.

[1] School of Population Medicine and Public Health, Chinese Academy of Medical Sciences (CAMS) & Peking Union Medical College (PUMC), Beijing, China; [2] State Key Laboratory of Respiratory Health and Multimorbidity, Beijing, China; [3] Key Laboratory of Pathogen Infection Prevention and Control (Peking Union Medical College), Ministry of Education, Beijing, China; [4] The Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, Kunming City, Yunnan Province, China; [5] School of Data Science, Fudan University, Shanghai, China; [6] Weifang Center for Disease Control and Prevention, Weifang City, Shandong Province, China; [7] School of Health Policy and Management, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China.

[&] Joint first authors.

Submitted: March 05, 2024; Accepted: June 24, 2024

REFERENCES

1. Van Kerkhove MD, Ryan MJ, Ghebreyesus TA. Preparing for "Disease X". Science 2021;374(6566):377. https://doi.org/10.1126/science.abm7796.
2. Krammer F, Smith GJD, Fouchier RAM, Peiris M, Kedzierska K, Doherty PC, et al. Influenza. Nat Rev Dis Primers 2018;4(1):3. https://doi.org/10.1038/s41572-018-0002-y.
3. Hamalaw SA, Bayati AH, Babakir-Mina M, Benvenuto D, Fabris S, Guarino M, et al. Assessment of core and support functions of the communicable disease surveillance system in the Kurdistan Region of Iraq. J Med Virol 2022;94(2):469 – 79. https://doi.org/10.1002/jmv.27288.
4. Hutwagner L, Thompson W, Seeman GM, Treadwell T. The bioterrorism preparedness and response Early Aberration Reporting System (EARS). J Urban Health 2003;80(2 Suppl 1):i89-96. http://dx.doi.org/10.1007/pl00022319.
5. Li ZJ, Lai SJ, Zhang HL, Wang LP, Zhou DL, Liu JZ, et al. Hand, foot and mouth disease in China: evaluating an automated system for the detection of outbreaks. Bull World Health Organ 2014;92(9):656 – 63. https://doi.org/10.2471/BLT.13.130666.
6. Unkel S, Farrington CP, Garthwaite PH, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. J R Stat Soc Ser A: Stat Soc 2012;175(1):49 – 82. https://doi.org/10.1111/j.1467-985X.2011.00714.x.
7. MacIntyre CR, Chen X, Kunasekaran M, Quigley A, Lim S, Stone H, et al. Artificial intelligence in public health: the potential of epidemic early warning systems. J Int Med Res 2023;51(3):3000605231159335. http://dx.doi.org/10.1177/03000605231159335.
8. Feng LZ, Zhang T, Wang Q, Xie YR, Peng ZB, Zheng JD, et al. Impact of COVID-19 outbreaks and interventions on influenza in China and the United States. Nat Commun 2021;12(1):3249. https://doi.org/10.1038/s41467-021-23440-1.
9. Yang LY, Zhang T, Han X, Yang J, Sun YX, Ma LB, et al. Influenza epidemic trend surveillance and prediction based on search engine data: deep learning model study. J Med Internet Res 2023;25:e45085. https://doi.org/10.2196/45085.
10. Yang WZ, Li ZJ, Lan YJ, Wang JF, Ma JQ, Jin LM, et al. A nationwide web-based automated system for outbreak early detection and rapid response in China. Western Pac Surveill Response J 2011;2(1):10 – 5. https://doi.org/10.5365/WPSAR.2010.1.1.009.
11. Han SS, Zhang T, Lyu Y, Lai SJ, Dai PX, Zheng JD, et al. The incoming influenza season - China, the United Kingdom, and the United States, 2021-2022. China CDC Wkly 2021;3(49):1039 – 45. https://doi.org/10.46234/ccdcw2021.253.

SUPPLEMENTARY MATERIAL

Model Introduction

XGBoost is a popular ensemble learning algorithm called eXtreme Gradient Boosting, commonly applied for classification and regression purposes. It utilizes decision trees and the gradient boosting technique to enhance model performance through iteratively training new decision trees.

XGBoost relies on decision trees as its fundamental components, with each tree functioning as a weak learner. Decision trees encompass nodes, branches, and leaves, where nodes split based on features and leaf nodes correspond to output values. The model's goal is to minimize an objective function consisting of a loss function and regularization term. This objective function evaluates the model's performance, aiming to minimize it by seeking new decision trees in each iteration. This optimization process can be described by Equation 1, which also represents the objective function. In Equation 1, y_t denotes the objective function, l represents the loss function, y_i is the actual label, $\hat{y}_i$ stands for the model's predicted label, K denotes the number of trees, and $\Omega(f_k)$ is the regularization term.

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

XGBoost employs the gradient boosting strategy to minimize the objective function gradient at each iteration.

The process of generating a new tree includes fitting the negative gradient of the current model. This guides the creation of new trees to prioritize poorly-performing samples from the previous model.

The negative gradient can be represented by Equation 2.

$$G_n = -\frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i} \quad (2)$$

To address overfitting, XGBoost incorporates regularization methods such as weight decay (L2 regularization) and minimum split loss to manage tree depth and leaf node weights. The calculation for the regularization term is detailed in Equation 3.

$$\Omega(f_k) = \frac{1}{2}\sum_{j=1}^{L} w_j^2 \quad (3)$$

The ultimate prediction is calculated by summing the predicted values of all generated trees, with each tree's impact adjusted by the learning rate. This methodology enables XGBoost to boost performance by amalgamating diverse decision trees.

XGBoost is a potent machine learning algorithm, well-suited for medium to large-scale datasets and intricate classification tasks. Optimal performance is attainable via meticulous parameter adjustment. Leveraging multithreading and parallel computing, the model demonstrates efficient performance, enabling swift training on extensive datasets. Incorporating regularization terms aids in averting overfitting and enhancing the model's generalization capabilities. XGBoost offers insights on feature importance, facilitating comprehension of the model's sensitivity to specific features. It adeptly manages missing values without necessitating supplementary processing. Additionally, the algorithm accommodates diverse loss functions and evaluation metrics, rendering it versatile across various problem types.

In this study, the model successfully meets the demands of the research task by providing scientifically precise predictions for influenza-like case activity levels.

The Isolation Forest

Forest algorithm model was first introduced by Fei Tony Liu et al. It focuses on anomaly detection by differentiating typical data from anomalies, enabling efficient classification by isolating the anomalous data points.

The primary operational concept of the Isolation Forest model involves randomly selecting a feature from the dataset and choosing a separation value within its range. Samples are then split into branches based on this value. This recursive process continues for each branch until only one sample is left or the defined recursion depth is met.

Anomalous data is distinguished from normal data by its unique features, requiring fewer partitioning steps for isolation. Conversely, normal data necessitates more steps for isolation, leading to a longer path to the endpoint. The model assesses the path length for each data point to reach isolation, with a shorter path indicating early isolation and suggesting a higher likelihood of anomaly.

To provide a clearer understanding of the Isolation Forest model's training methodology, a straightforward example is discussed. Refer to Supplementary Figure S1, which presents a set of data points plotted along a number line, with their values arranged in ascending order. The objective is to identify any anomalies within these data points. The initial step involves determining the median of all values, which lies between the maximum and minimum data points, to serve as the initial split value. Following this initial split, data point Z becomes separated. Subsequently, the median value amongst the remaining data points, A through I, is selected to define the second split, consequently isolating data points F through I and leaving data points A through E. This bifurcation proceeds recursively until each data point stands alone. Ultimately, data point Z is sequestered in just one split, whereas data point E requires four splits for isolation. The fewer partitions required to isolate a data point signal a higher likelihood of it being anomalous, hence data point Z is deemed more likely to be an outlier compared to data point E.

Evaluation of Model Warning Results

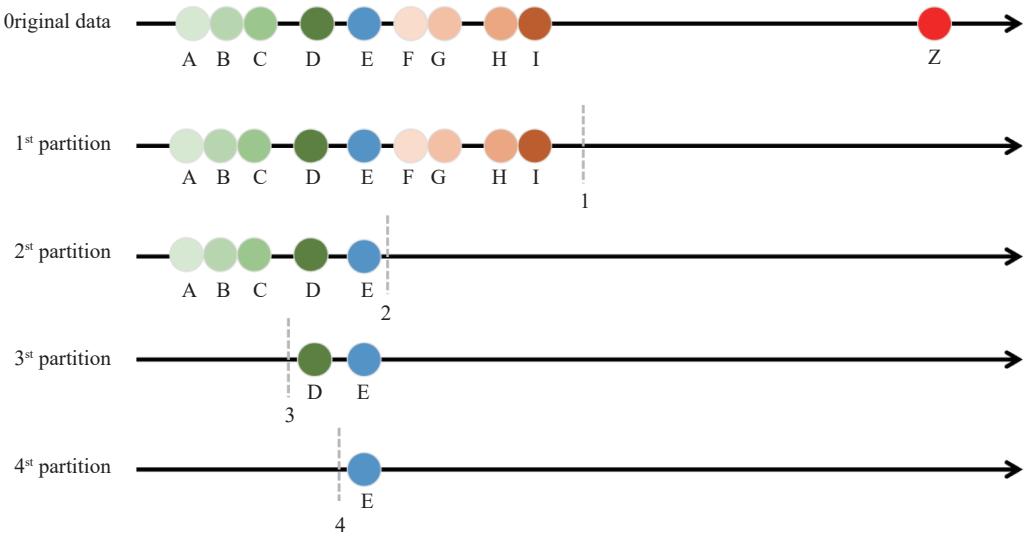
In this study, three evaluation metrics, Precision, Recall, and F1-Score, were utilized to assess the warning outcomes of the test set. The calculation equations for these metrics are presented below:

$$\text{Precision (Pre)} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Recall (Rec)} = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = 2 \times \frac{Pre \times Rec}{Pre + Rec} \tag{6}$$

True positive (TP) refers to the count of normal events correctly classified as normal, while true negative (TN) denotes the count of abnormal events accurately identified as abnormal. False positive (FP) indicates the number of abnormal events erroneously detected as normal, and false negative (FN) represents the count of normal events mistakenly presented as abnormal. The detailed evaluation results are presented in Supplementary Table S1.



SUPPLEMENTARY FIGURE S1. Training process of the Isolation Forest model.

SUPPLEMENTARY TABLE S1. Evaluation results based on the XGBoost warning model.

Weighted average	Precision	Recall	F1-Score
XGBoost	0.93	0.88	0.90
Isolation forest	0.95	0.82	0.86

Recollection

The COVID-19 Vaccines Evaluation Program: Implementation, Management, and Experiences, 2021–2023

Tingting Yan[1]; Hui Zheng[1]; Mingshuang Li[1]; Chao Ma[1]; Xuanyi Wang[2]; Xiaoqi Wang[3]; Zhenjun Li[4]; Yuansheng Chen[4]; Wenshang Hu[5]; Lance Rodewald[1]; Zhijie An[1]; Zundong Yin[1]; Zijian Feng[6,#]

ABSTRACT

In 2021, China's domestically produced coronavirus disease 2019 (COVID-19) vaccines received approval from regulatory bodies and were administered worldwide. Due to a low number of infections within China during that period, it became imperative to evaluate the vaccines' real-world effectiveness through international studies. To facilitate this, China CDC launched the COVID-19 Vaccines Evaluation Program (COVEP). This program formed research collaboration agreements with health institutes across five World Health Organization regions, addressing key questions about vaccine performance through ten cooperative agreements. The findings from COVEP projects reinforced confidence, both domestically and globally, in the effectiveness of the vaccines produced in China. Moreover, the outcomes observed internationally were frequently mirrored by later studies conducted within China. COVEP thus pioneered a novel approach for fostering cross-national research collaborations, addressing significant public health issues and exemplifying a framework for international cooperation. This approach is in line with the strategic objectives and other development efforts of China CDC's national disease control and prevention initiatives.

RATIONALE

The coronavirus disease 2019 (COVID-19) pandemic has presented a significant threat to global health (*1*). In response, many nations, including China, embarked on the rapid development, testing, manufacturing, and distribution of COVID-19 vaccines using diverse technological approaches. While regulatory approvals were granted based on evidence from short-term randomized clinical trials (RCTs) demonstrating safety and efficacy within controlled demographics, these studies were not equipped to assess several critical aspects of vaccine performance. These aspects include the vaccine effectiveness (VE) in special populations, a wide range of outcomes such as protection against symptomatic infection, severe or fatal illness, the duration of protection against various outcomes, the necessity for booster doses, VE against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants, and the overall impact of vaccination at the population level.

The initial cohort of COVID-19 vaccines developed in China, including Sinopharm and Sinovac inactivated vaccines, CanSino adenovirus-vectored vaccine, and Zhifei Longcom protein subunit vaccine, underwent safety and immunogenicity testing domestically, whereas efficacy assessments were conducted internationally (*2*). Following demonstration of their safety and efficacy, these vaccines received approval both within China and globally, ultimately being approved in over one hundred countries (*3*). During the initial two years of COVID-19 vaccine distribution, international real-world studies offered invaluable policy-relevant insights into the effectiveness of China-manufactured vaccines in diverse populations, insights that were difficult to acquire domestically due to the low incidence of infections and outbreaks in China at that time (*4*).

In mid-2021, China CDC introduced the COVID-19 Vaccine Evaluation Program (COVEP). This initiative marked the first time China CDC provided competitive grants to independent international research institutions for joint research endeavors. COVEP primarily aimed to support studies in nations extensively using China-manufactured COVID-19 vaccines, regardless of whether other vaccines were also in use. The principal research priorities encompassed targeted evaluation of VE against SARS-CoV-2 variants, VE in special demographics, the longevity of protection, the impact of booster doses, the overall influence of vaccination on the pandemic, and the severity and characteristics of breakthrough infections. Special attention was given to

groups underrepresented in the phase 3 efficacy trials of these vaccines, notably the elderly, those with comorbidities, and pregnant women. The volatile progression of the COVID-19 pandemic required that the research adapt to changing conditions at international sites, necessitating both flexible study designs and research questions. The cooperative agreement framework was established to facilitate this adaptability. This article details the objectives, framework, procedures, impacts, and insights gained from COVEP.

GOALS

The objectives of COVEP were to deliver policy-relevant scientific evidence concerning the real-world effectiveness of COVID-19 vaccines produced in China; to create effective frameworks for facilitating international research during the COVID-19 pandemic; to provide technical assistance to overseas researchers in their study execution; and to serve as a potential prototype for future international research initiatives by the China CDC.

STRUCTURE

COVEP is integrated within the National Immunization Program, which offers technical guidance for vaccine policy development and played a pivotal role in the execution of China's COVID-19 vaccination strategy. Figure 1 illustrates the organizational framework and duties of COVEP, encompassing the Principal Investigator (PI), the steering committee, a project officer team, a financial management team, and an academic scientific technical support team. Members of the steering committee included representatives from the World Health Organization (WHO) China office, Gavi, the Bill and Melinda Gates Foundation China office, Fudan University, and China CDC. The committee facilitated the engagement of prospective international PIs with COVEP and offered counsel regarding scientific objectives and methodologies. Project officers and financial managers, employed by China CDC, provided support for funded research initiatives. Expertise for these funded endeavors was independently contributed by an academic team from Fudan University.

PROCEDURES

Figure 2 presents the flow diagram of the COVEP processes. The identification of relevant projects was facilitated through a two-stage application procedure, beginning with the submission of a letter of interest (LOI). COVEP released a call for LOIs on the official China CDC website on October 5th, 2021 (*5*). This call detailed eligibility criteria, research objectives and scope, application guidelines, and review procedures. PIs from public health departments, academic institutions, and international technical agencies capable of conducting studies in countries using China-manufactured COVID-19 vaccines were invited to submit their LOIs by the end of 2021. Submissions

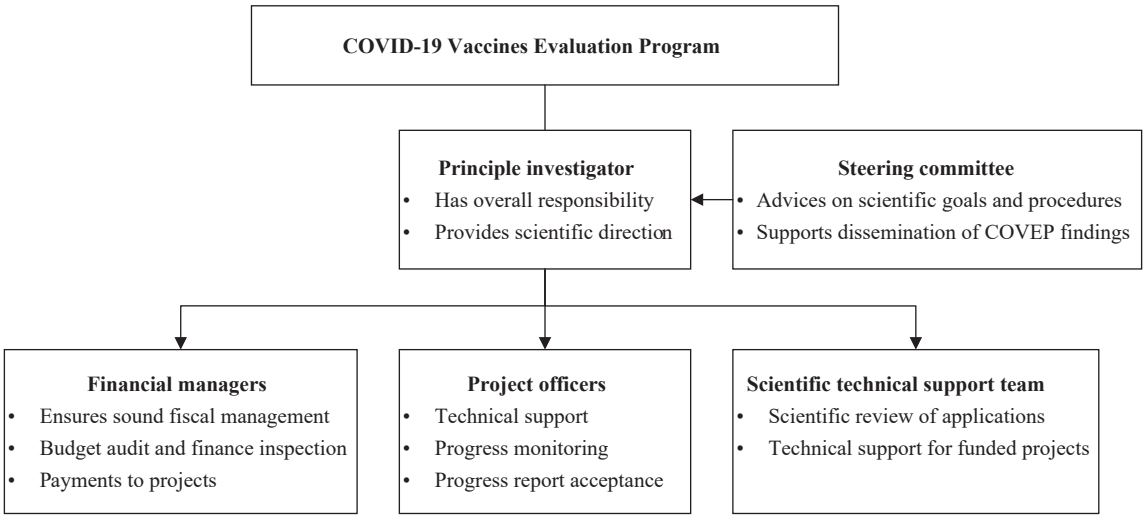


FIGURE 1. Structure and responsibilities of COVEP.
Abbreviation: COVID-19=coronavirus disease 2019; COVEP=COVID-19 Vaccines Evaluation Program.

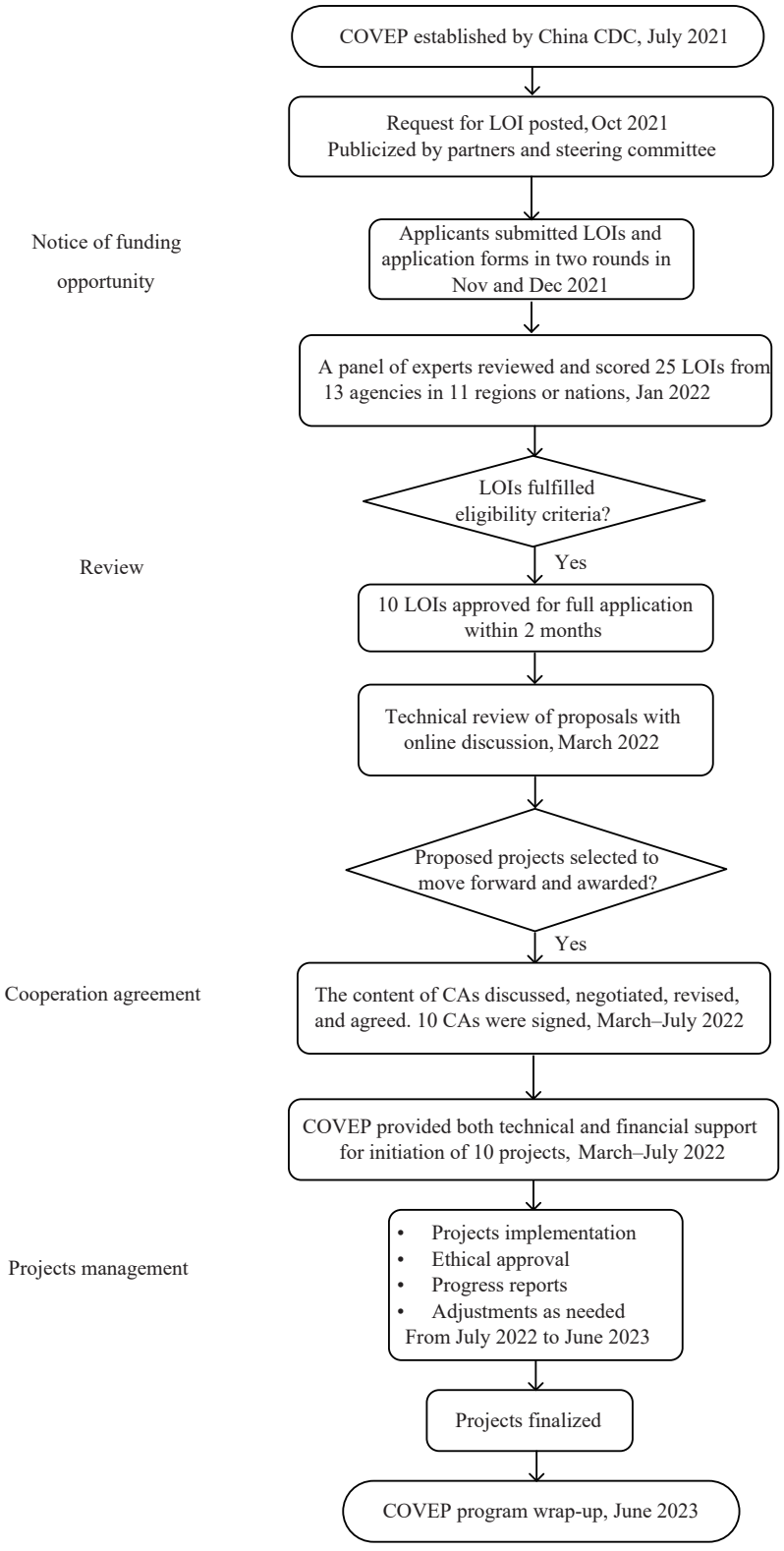


FIGURE 2. Flowchart of the COVEP program.
Abbreviation: COVEP=the COVID-19 vaccines evaluation program; COVID-19=coronavirus disease 2019; LOI=letter of interest; CAs=cooperation agreements.

required basic contact details, a concise study design outline, projections of project feasibility, a timeline (not to exceed 18 months), and an estimated budget (under one million US dollars).

Twenty-five LOIs were received from thirteen agencies across eleven regions or countries. These LOIs were reviewed by the COVEP PI, project officers, and the scientific technical support team. Ten projects, originating from eligible institutions, were identified as being within the project scope. PIs of these projects were asked to submit detailed proposals within two months following LOI approval. These proposals included budgeting and detailed protocols describing study goals, design, methodology, timelines, and PI resumes. Throughout the proposal development phase, PIs and their teams were encouraged to engage in discussions with COVEP staff regarding study objectives and methods. The review panel evaluated and scored these full proposals, ultimately awarding funding to ten projects from nine institutions.

Under the advisement of the China CDC's legal department, the COVEP team developed bilingual cooperation agreements in both Chinese and English. These agreements, which were fine-tuned through negotiations until a mutual consensus was reached between China CDC and the respective awardees, delineated specific rights and responsibilities. Key elements included research objectives and expected outcomes, standards and methodologies, financial protocols, intellectual property rights, contractual liabilities, clauses for unexpected project cessation, and provisions for force majeure that might necessitate amendments to the agreements or research protocols. Participants were offered a choice between two cooperative frameworks: financial only or combined financial and technical. Ten agreements were successfully negotiated and executed by all involved parties, with comprehensive proposals appended as implementation frameworks. These agreements explicitly stated that the project owners were responsible for manuscript drafting and publication decisions independently of China CDC. Additionally, it was mandated that manuscripts originating from COVEP-funded projects must acknowledge COVEP's financial support.

Initial funding was disbursed to the funded institution within 20 working days following the signing of the cooperative agreement. Subsequent allocations were made after the required interim reports were submitted, with the remaining funds released upon receipt of final reports. Ethical approval was mandatory before the commencement of any research activities. Progress reports facilitated ongoing awareness of research developments and hurdles among project officers. Necessary adjustments to the project plans were implemented following discussions and negotiations between the project officers and PIs, in cooperation with the Fudan University technical support team.

PROJECTS

Table 1 displays the list of the ten COVEP-supported projects, including institutions, research topics, basic study designs, and study countries/settings. COVEP funded projects across five of the six WHO regions: two in the African Region, three in the Region of the Americas, two in the European Region, one in the South-East Asia Region, and two in the Western Pacific Region. Supported entities comprised academic and government institutions as well as a health research enterprise. Studies were conducted among diverse populations including children, pregnant women, and individuals with comorbidities such as hypertension, diabetes, and HIV. All projects concluded by June 2023 with the acceptance of their final reports. Due to unexpected and unavoidable shifts in pandemic epidemiology, some project protocols or research questions required adjustments, achieved through mutual agreement between the PIs and COVEP.

IMPACT

COVEP facilitated pandemic responses by supporting independent, international research teams in generating practical, policy-relevant data on the effectiveness of COVID-19 vaccines produced in China, beneficial to nations utilizing these vaccines. The first project funded by COVEP to evaluate the real-world effectiveness of these vaccines was carried out by researchers at the University of Hong Kong, Hong Kong Special Administrative Region (SAR), China (*6*–*7*). The prompt dissemination of their findings was crucial for China's domestic COVID-19 strategy, highlighting the need for very high vaccination coverage among the elderly to optimally protect against severe or fatal COVID-19 at the population level. Notably, the results from Hong Kong SAR, China, which demonstrated an over 90% VE against severe or fatal COVID-19, were later mirrored in national studies during the Omicron transmission phase (*8*–*9*). Another study, also funded by COVEP, was conducted in the Federation of Bosnia and Herzegovina (BiH) and assessed the effectiveness of

TABLE 1. Research projects supported by COVEP.

Institute	Research topic	Study design	Setting
Biomedical Research and Training Institute	COVID-19 vaccine effectiveness in adults with co-morbidities	Prospective observational cohort study	Zimbabwe
MLI	Real-world effectiveness and determinants of effectiveness of COVID-19 vaccines	Test-negative case control study	Uganda
Institute for Immunobiology and Human Genetic, Medical Faculty in Skopje	Real-world effectiveness of COVID-19 vaccines in preventing symptomatic disease, hospitalization, and death	Retrospective test-negative case-control study	Republic of North Macedonia
BiH	Real-world effectiveness of Sinopharm COVID-19 vaccine	Retrospective/prospective test-negative case-control study	BiH
P95 Latina SAS	Real-world, brand-specific vaccine effectiveness of Chinese COVID-19 vaccines	Test-negative case-control study	Colombia
Universidad del Desarrollo	Platform for COVID-19 surveillance and evaluation of interventions in Chile	Large-scale platforms of surveillance	Chile
Universidad de Antioquia	Real-world effectiveness of SINOVAC vaccine	Population-based retrospective cohort study	Colombia
University of Oxford	Vaccination of pregnant women with CoronaVac and maternal and newborn health	Longitudinal cohort study (PregVax)	Indonesia
University of Hong Kong	COVID-19 vaccine effectiveness for the prevention of symptomatic, clinically severe and fatal COVID-19 disease	Population-based observational cohort study	Hong Kong SAR, China
University of Hong Kong	Modeling exit strategies from the COVID-19 epidemic in China	Mathematical modeling study	Hong Kong SAR, China

Abbreviation: MLI=makerere university lung institute; BiH=institute for public health of the federation of Bosnia and Herzegovina; SAR=special administrative region; COVID-19=coronavirus disease 2019; COVEP=COVID-19 Vaccines Evaluation Program; SINOVAC=Sinovac Biotech Ltd.

two doses of the inactivated vaccine in individuals aged 60 and older during a period dominated by the Delta variant (*10*). This research concluded that a primary series of inactivated COVID-19 vaccines provided significant defense against moderate to severe COVID-19 in the elderly, though it also indicated diminishing immunity that suggested the need for a booster. Additionally, a scoping review of VE study methodologies, conducted by researchers in Colombia, revealed numerous studies on inactivated vaccines, yet the diversity in methodologies posed challenges for cross-study comparisons (*11*).

Final reports from COVEP studies provided additional evidence to the China CDC before formal publication. Key findings included the following: in the Republic of North Macedonia, three doses of inactivated vaccines were shown to protect against COVID-19 hospitalization. In Colombia, researchers demonstrated the effectiveness of three doses of CoronaVac in preventing death among individuals over 50 years and in protecting children aged 3–12 years from COVID-19. The COVEP project in Indonesia highlighted the benefits and safety of inactivated COVID-19 vaccines during pregnancy. A study conducted in Chile indicated diminishing protection 180 days post-vaccination, emphasizing the need for a booster dose. Additionally, a Zimbabwean study confirmed the effectiveness of the inactivated vaccine among people living with HIV.

Results from research funded by COVEP underscored scientifically reliable evidence bolstering the effectiveness of China-produced vaccines in addressing COVID-19 domestically and internationally. Furthermore, these findings affirm that China's COVID-19 vaccination strategy was progressing appropriately, although there was a need to enhance efforts to achieve maximal vaccination coverage among the elderly population.

LESSONS LEARNED

COVEP established an effective support system for facilitating independent international research studies during the COVID-19 pandemic. The successful completion of ten projects illustrates that COVEP could serve as a viable model for future international research initiatives. Key insights gained from the COVEP implementation are presented in Table 2, and are organized into categories of structure, relationships, and cooperative agreements.

The steering committee, in collaboration with international partners, guided the COVEP initiative and monitored progress to ensure objectives were achieved. The diverse viewpoints offered by committee members were invaluable for the multifaceted tasks involved in COVEP. They also promoted the funding

TABLE 2. Key lessons learned during the implementation of COVEP.

Program aspect	Key points
Program structure	• Steering committee with international partners playing the role of supporting, promoting, and monitoring • Partnership with academic institutions for providing technical support
Relation to the institutions	• Investigator independence for the most highly capable investigators and for maintaining the credibility of their findings • Regular communication including periodic written reports • Support to projects from both project and financial officers
Cooperative agreements	• Bilingual, clear and transparent, freely-negotiated documentation • End-to-end legal support of funded projects with a force majeure clause • Flexibility to negotiate changes in research design and scope

Abbreviation: COVEP=COVID-19 Vaccines Evaluation Program.

opportunities worldwide and assisted in selecting highly qualified international investigators. Additionally, partnership with Fudan University contributed technical support to all COVEP-supported projects and enhanced the breadth of assistance available to international projects.

Independence of the investigators and projects was crucial in attracting highly qualified PIs to COVEP. Such autonomy enhanced the credibility of COVEP project findings by reassuring journals and their readers that the funding source did not influence manuscript preparation or publication decisions.

COVEP cooperative agreements outlined the mutual responsibilities of the implementing institutions and China CDC. These agreements were bilingual, transparent, and resulted from free negotiations. They offered comprehensive legal support for funded projects. Given the unpredictability of the pandemic, these projects required the flexibility to request methodological changes if the research environment shifted unexpectedly. To safeguard investigators from unforeseen circumstances beyond their control, the agreements included a force majeure clause.

IMPLICATIONS

COVEP is consistent with the primary functions and strategic objectives of the China CDC in advancing the national disease control and prevention system (*12*). Major areas of alignment comprise advancing collaborative efforts with premier public health institutions in higher education, boosting the China CDC emergency response team's capability for remote and international support in managing acute infectious diseases, and enhancing capabilities in global public health governance and international collaboration through the development of talent equipped for global public health emergency responses, active participation in international public health assistance, and bolstering international public health cooperation and exchanges.

CONCLUSIONS

COVEP-sponsored research provided critical, policy-relevant findings regarding the key performance metrics of vaccines produced in China, which were challenging or impossible to obtain domestically in a timely manner due to the low number of cases in China at the time. COVEP established a viable model for international collaboration, facilitating research projects funded by China CDC and fostering partnerships with world-renowned experts in vaccines, epidemiology, and public health. Through effective collaboration with the financial, management, and legal departments of China CDC, COVEP developed protocols that support the execution of these international research partnerships. This initiative aligns with China CDC's strategic plans and contributes to the advancement of the national disease control and prevention framework.

Conflicts of interest: No conflicts of interest.

Acknowledgements: All staff members at COVEP and China CDC for their relentless dedication to the administration and management of this vital program.

Funding: This work was supported by the Chinese Center for Disease Control and Prevention through the COVID-19 Vaccines Evaluation Program (COVEP).

doi: 10.46234/ccdcw2024.120

[#] Corresponding author: Zijian Feng, fengzj@chinacdc.cn.

[1] National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases (NITFID). National Immunization Program, Chinese Center for Disease Control and Prevention, Beijing, China; [2] Institutes of Biomedical Sciences, Fudan University, Shanghai, China; [3] Office of International Cooperation, Chinese Center for Disease Control and Prevention, Beijing, China; [4] Office of Science and Technology, Chinese Center for Disease Control and Prevention, Beijing, China; [5] Office of Finance, Chinese Center for Disease Control and Prevention, Beijing, China; [6] Chinese Preventive Medicine Association, Beijing, China.

Submitted: April 23, 2024; Accepted: June 13, 2024

REFERENCES

1. Pollard CA, Morran MP, Nestor-Kalinoski AL. The COVID-19 pandemic: a global health crisis. Physiol Genomics 2020;52(11):549 – 57. https://doi.org/10.1152/physiolgenomics.00089.2020.
2. WHO. COVID-19 vaccine tracker and landscape. 2023. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines. [2024-2-6].
3. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of COVID-19 vaccinations. Nat Hum Behav 2021;5(7):947 – 53. https://doi.org/10.1038/s41562-021-01122-8.
4. China CDC. Tracking the epidemic. 2022. https://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm. [2024-3-14].
5. China CDC. COVID-19 Vaccines Evaluation Program (COVEP) request for letters of interest of real-world studies on the effectiveness of China-made COVID-19 vaccines. 2021. https://en.chinacdc.cn/special/COVID19_Response/discoveries_guidelines/202205/t20220516_259204.html. [2024-2-7].
6. McMenamin ME, Nealon J, Lin Y, Wong JY, Cheung JK, Lau EHY, et al. Vaccine effectiveness of one, two, and three doses of BNT162b2 and CoronaVac against COVID-19 in Hong Kong: a population-based observational study. Lancet Infect Dis 2022;22(10):1435 – 43. https://doi.org/10.1016/S1473-3099(22)00345-0.
7. Yang BY, Wong IOL, Xiao JY, Tsang TK, Liao QY, Cowling BJ. Effectiveness of CoronaVac and BNT162b2 vaccines against severe acute respiratory syndrome coronavirus 2 omicron BA. 2 infections in Hong Kong. J Infect Dis 2022;226(8):1382 – 4. https://doi.org/10.1093/infdis/jiac360.
8. Huang ZY, Xu SF, Liu JC, Wu LL, Qiu J, Wang N, et al. Effectiveness of inactivated and Ad5-nCoV COVID-19 vaccines against SARS-CoV-2 Omicron BA. 2 variant infection, severe illness, and death. BMC Med 2022;20(1):400. https://doi.org/10.1186/s12916-022-02606-8.
9. Huang ZY, Xu SF, Liu JC, Wu LL, Qiu J, Wang N, et al. Effectiveness of inactivated COVID-19 vaccines among older adults in Shanghai: retrospective cohort study. Nat Commun 2023;14(1):2009. https://doi.org/10.1038/s41467-023-37673-9.
10. Musa S, Merdrignac L, Skocibusic S, Nedic R, Cilovic-Lagarija S, Kissling E. BBIBP-CorV vaccine effectiveness against COVID-19 in patients aged 60 years and older during the Delta-dominant period in the Federation of Bosnia and Herzegovina, a test-negative case-control study. Vaccine 2024;42(15):3467 – 73. https://doi.org/10.1016/j.vaccine.2024.04.047.
11. Alzate-Ángel JC, Avilés-Vergara PA, Arango-Londoño D, Concha-Eastman A, Garcés-Hurtado A, López-Carvajal L, et al. How has research on the effectiveness and safety of COVID-19 vaccination been evaluated: a scope review with emphasis on CoronaVac. Front Public Health 2024;12:1321327. https://doi.org/10.3389/fpubh.2024.1321327.
12. Shen HB. Comprehensive strategic planning and enhancement of China CDC contributes to high-quality development of the national disease control and prevention system. China CDC Wkly 2024;6(4):61 – 3. https://doi.org/10.46234/ccdcw2024.013.

Erratum

Vol. 6 No. 1

In the article entitled 'Evolution of HIV/AIDS Prevention and Control Policies in China: A Grounded Theory Approach' [2024, 6 (1): 12–22. doi: 10.46234/ccdcw2024.003], The first author "Huang Li" was mistaken upload with the wrong one and should be the following: "Huang Lin".

Vol. 6 No. 24

In the article entitled 'Community Incidence Estimates of Five Pathogens Based on Foodborne Diseases Active Surveillance — China, 2023' [2024, 6(24): 574–579. doi: 10.46234/ccdcw2024.112], the figures in the summary box on page 574 " Norovirus, 3,188.28 (95% *UI*: 2,518.03, 7,296.96); *Salmonella* spp., 1,295.59 (95% *UI*: 1,002.62, 1,573.11); diarrheagenic *E. coli* (DEC), 782.62 (95% *UI*: 651.19, 932.05); *Vibrio parahaemolyticus*, 404.06 (95% *UI*: 342.19, 468.93); "should be the following: "Norovirus: 3,188.28 (95% *UI*: 2,510.80, 3,872.96); *Salmonella* spp.: 1,295.59 (95% *UI*: 1,020.62, 1,573.11); Diarrheagenic *E.coli*: 782.62 (95% *UI*: 616.34, 950.46); *Vibrio parahaemolyticus*: 404.06 (95% *UI*: 318.41, 491.45)".

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. 26 Jun. 28, 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