

Original Articles



Integrating Internet Search Data and Surveillance Data to Construct Influenza Epidemic Thresholds in Hubei Province: A Moving Epidemic Method Approach

Caixia Dang^{1,2,&}, Feng Liu^{1,2,&}, Hengliang Lyu¹, Ziqian Zhao¹, Sijin Zhu¹, Yang Wang¹,
Yuanyong Xu^{1,#}, Yeqing Tong^{3,#}, and Hui Chen^{1,#}

1. Chinese People's Liberation Army Center for Disease Control and Prevention, Beijing 100071, China;
2. Department of Epidemiology, School of Public Health, China Medical University, Shenyang 110000, Liaoning China; 3. Department of Infectious Diseases, Hubei Center for Disease Control and Prevention, Wuhan 430000, Hubei, China

Abstract

Objective Enhancing influenza surveillance through multi-source data integration is critical for timely epidemic response. This study aimed to establish data-driven influenza monitoring thresholds in Hubei Province by combining Baidu search indices with traditional surveillance data.

Methods Influenza case reports (2020-2023) and Baidu search indices were analyzed using the Moving Epidemic Method (MEM). Cross-validation with maximum Youden's index determined optimal model parameters. Epidemic thresholds (start/end) and intensity levels (moderate/high/extremely high) were calculated to assess the 2023/2024 influenza season.

Results The epidemic thresholds for Hubei Province were 4,921 cases (start) and 4,387 cases (end), with intensity thresholds of 19,630 (moderate), 56,674 (high), and 73,048 (extremely high) weekly cases. The 2023/2024 epidemic lasted 9 weeks (Week 45/2023–Week 2/2024), with prefecture-level cities showing 8-11 week durations concentrated between Week 43/2023–Week 2/2024. While Wuhan reached extremely high intensity, Xianning and Xiangyang peaked at high intensity, and Shiyan/Huanggang/Enshi maintained moderate levels.

Conclusion This study demonstrates the feasibility of MEM in synthesizing digital and surveillance data for influenza threshold determination. The established tiered threshold system provides a scientific framework for dynamic risk assessment and targeted public health interventions in Hubei Province, exemplifying the operational value of big data integration in epidemic preparedness.

Key words: Moving Epidemic Method; Influenza surveillance; Epidemic thresholds; Multi-source data

Biomed Environ Sci, 2025; 38(9): 1-6

doi: [10.3967/bes2025.109](https://doi.org/10.3967/bes2025.109)

ISSN: 0895-3988

www.besjournal.com (full text)

CN: 11-2816/Q

Copyright ©2025 by China CDC

[&]These authors contributed equally to this work.

[#]Correspondence should be addressed to Hui Chen, Associate Investigator, Tel: 18149338893, E-mail: chfmmu@outlook.com; Yuanyong Xu, Investigator, E-mail: xyy_827@sina.com; Yeqing Tong, Investigator, E-mail: t_yeqing@163.com

Biographical notes of the first authors: Caixia Dang, MPH, female, majoring in spatial analysis and public health, E-mail: dangcx418@163.com; Feng Liu, MPH, majoring in spatial analysis and public health, E-mail: 13551712380@163.com

INTRODUCTION

Influenza, an acute respiratory infectious disease caused by the influenza virus, exhibits distinct seasonal patterns in China, with peak activity occurring in winter and spring in northern regions, and in winter and summer in southern areas^[1]. The World Health Organization (WHO) emphasizes that early warning and epidemic intensity assessments are critical public health strategies for influenza prevention and control. Internet-based flu surveillance, with real-time data and low costs, effectively complements traditional methods. The Baidu Search Index, which reflects flu-related queries, strongly correlates with influenza trends, aiding in regional activity assessment and outbreak tracking^[2].

In 2013, Vega et al.^[3] applied The Moving Epidemic Method (MEM) to establish influenza epidemic thresholds and assess epidemic intensity^[3,4]. The MEM relies solely on historical incidence data, demonstrating strong applicability. By updating thresholds in real time, the onset of an epidemic can be identified more accurately. The MEM has been widely and effectively used in temperate regions; however, its application in subtropical countries has rarely been reported. Hubei Province, located in Central China, has experienced a transient surge in influenza activity following the COVID-19 pandemic, necessitating the urgent development of early warning systems to monitor future influenza trends.

This study aims to apply the MEM to assess the intensity of influenza epidemics in Hubei Province during the 2023-2024 season, establish corresponding epidemic thresholds, and evaluate the performance of the MEM in this region. The findings will provide a reference for defining influenza epidemic thresholds and assessing epidemic intensity in Hubei Province, thereby offering a scientific basis for relevant health authorities to formulate effective prevention and control strategies, develop related policies, and allocate medical resources rationally.

METHODS

Study Design and Participants

The MEM involves historical incidence data and is primarily applicable to diseases with unimodal epidemics, preferably with ≥ 3 historical epidemic seasons^[3]. It evaluates the influenza epidemic status

and intensity of the target surveillance year by establishing graded thresholds. The influenza incidence data from 2020 to 2023 for various prefecture-level cities in Hubei Province were obtained from the Hubei Provincial Center for Disease Control and Prevention. According to the actual epidemic situation of reported influenza cases in various regions of Hubei from 2020 to 2023, only one epidemic peak was observed each year. Following Yang's framework for MEM applicability^[5], we selected regions with ≥ 3 influenza epidemic seasons between 2020-2024, including Hubei Province at the provincial level, and Wuhan, Xiangyang, Huanggang, Shiyan, Xianning, and Enshi Tujia and Miao Autonomous Prefecture at the prefecture level.

This study obtained the Baidu search index of influenza-related keywords through the Baidu website (<https://index.baidu.com/v2/index.html#/>). To avoid multicollinearity, we synthesized these keywords into a composite Baidu Index as the optimal surveillance factor, assigning higher weights to keywords that showed stronger correlations^[6] (Additional file 1: Methods 1).

MEM Modeling Process

The MEM determines the start, end, and duration of disease epidemics using the Maximum Cumulative Sum Percentage (MAP) function based on historical data (Additional File 1: Methods 2). The model employs cross-validation by comparing epidemic and non-epidemic periods of the target season with pre- and post-epidemic thresholds estimated from historical epidemic seasons. The performance of the model was assessed using sensitivity, specificity, Youden's index, and the Matthews correlation coefficient^[4].

After model determination, the epidemic intensity of the target monitoring years was evaluated using the start and end thresholds, as well as thresholds for moderate, high, and extremely high epidemic intensity levels (Additional File 1: Methods 3). Statistical analysis and modeling were performed using the MEM package in R version 4.4.1, and the surveillance database was managed in Excel 2016.

RESULTS

Correlation Between Baidu index and Influenza Incidence

Spearman correlation analysis results between the Baidu Index and influenza incidence

(Supplementary Figure 1) revealed that "influenza virus" had the highest correlation, while "seasonal influenza" showed the lowest. The synthesized Baidu Index showed a correlation coefficient of 0.45, indicating potential utility as an early warning indicator for influenza.

Model Fitting and Performance

Based on observed epidemic patterns from 2020 to 2023, a complete influenza season was defined as running from week 40 of each year to week 39 of the following year, and this structure was incorporated into the MEM model. At the provincial level, fitting the 2020–2023 data yielded an optimal $\delta = 3.4$, with cross-validation results of sensitivity = 0.93, specificity = 0.97, and Youden's index = 0.9. Similarly, the optimal parameters were selected for the synthesized Baidu Index and at the prefecture- and city-level (Supplementary Table 1).

The model showed excellent fit, with Youden's index exceeding 0.8 in most cases, and sensitivity and specificity above 0.9 for nearly all cities, confirming the MEM's suitability for influenza surveillance and early warning in Hubei Province.

Epidemic Thresholds and Patterns

The MEM model indicated that the onset threshold for influenza epidemics in Hubei Province during 2023–2024 was 4,921 cases, and the end threshold was 4,387 cases. Thresholds for medium, high, and very high epidemic intensities were 19,630, 56,674, and 73,048 cases, respectively. The epidemic period lasted for nine weeks (week 45 of 2023 to week 2 of 2024), with a peak in week 52 (late December). Among the prefecture-level cities, the epidemic duration ranged from 8 weeks in Huanggang to 11 weeks in Enshi Tujia and Miao Autonomous Prefecture, concentrated mainly between week 43 of 2023 and week 2 of 2024.

The province as a whole reached a high epidemic level, and Wuhan City reached a very high epidemic level due to urbanization and population density. Xianning and Xiangyang had a high epidemic level, whereas Shiyan, Huanggang, and Enshi had a moderate level. The peak period of the synthesized Baidu Index coincided with the peak of influenza epidemics, and public attention to influenza increased simultaneously from late November to the beginning of the following year, suggesting that the Baidu Index can serve as an auxiliary early warning indicator for influenza (Figure 1 and 2, Table 1).

The historical evaluation results from 2020 to

2023 showed that the epidemic levels in Hubei Province and its prefecture-level cities were at baseline during 2020/2021, low levels in 2021/2022 (Huanggang, Xianning, and Xiangyang reached moderate levels), and extremely high levels across all areas in 2022/2023 (Supplementary Table 2).

DISCUSSION

This study is the first to apply the MEM model to influenza epidemic intensity in Hubei after COVID-19, introducing the Baidu Index as an auxiliary early-warning indicator to establish a warning system. The findings indicate a high prevalence level in 2023–2024, with notable variation among prefecture-level cities. While peaks occurred within the seasonal range, high monitoring values require attention.

The Moving Epidemic Method (MEM) model indicated that the influenza epidemic period in Hubei Province during the 2023–2024 season lasted for nine weeks, peaking in week 52. This finding is consistent with previous studies demonstrating that influenza in Hubei typically exhibits a single-peak epidemic pattern, with peaks generally occurring between December and March of the following year^[7]. The epidemic duration in prefecture-level cities was concentrated mainly between week 43 of 2023 and week 2 of 2024. Factors such as the winter-spring transition, high population density, air pollution due to heating in northern regions, and population mobility during the Spring Festival collectively facilitated influenza transmission, which also confirms the scientific rationale for completing vaccination before October^[8]. Historical evaluation results from 2020 to 2023 showed that the epidemic levels in Hubei Province and its prefecture-level cities were at baseline during 2020/2021, while all areas reached extremely high levels in 2022/2023. Widespread non-pharmaceutical interventions (NPIs) during and after the COVID-19 pandemic helped suppress influenza transmission^[9]. Post-pandemic analysis showed extremely high epidemic levels in 2022/2023, with overall influenza activity significantly higher than historical levels despite remaining within the seasonal range^[10].

Determining the onset of influenza epidemics is important for medical resource allocation and vaccination guidance. By integrating real-time Baidu Index data with epidemic thresholds, surveillance efficiency can be improved, trends in influenza activity predicted in advance, and prevention

strategies adjusted in a timely manner.

LIMITATIONS

This study only covered single-peak epidemic regions. The generalizability of the findings is constrained by the single-region data, primarily

applicable to central and northern areas exhibiting single-peak epidemic patterns. Future research should expand to provinces with dual-peak epidemic patterns. The Baidu Index may overestimate actual cases due to the influence of public opinion. Therefore, comprehensive analysis is required for practical applications.

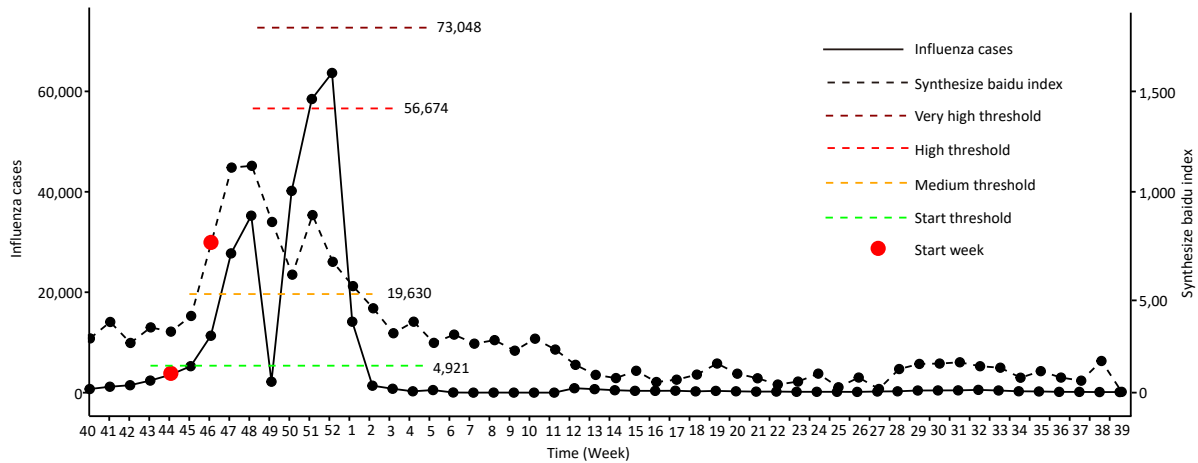


Figure 1. Influenza cases and Synthesized Baidu Index epidemiological intensity in Hubei Province, 2023 to 2024.

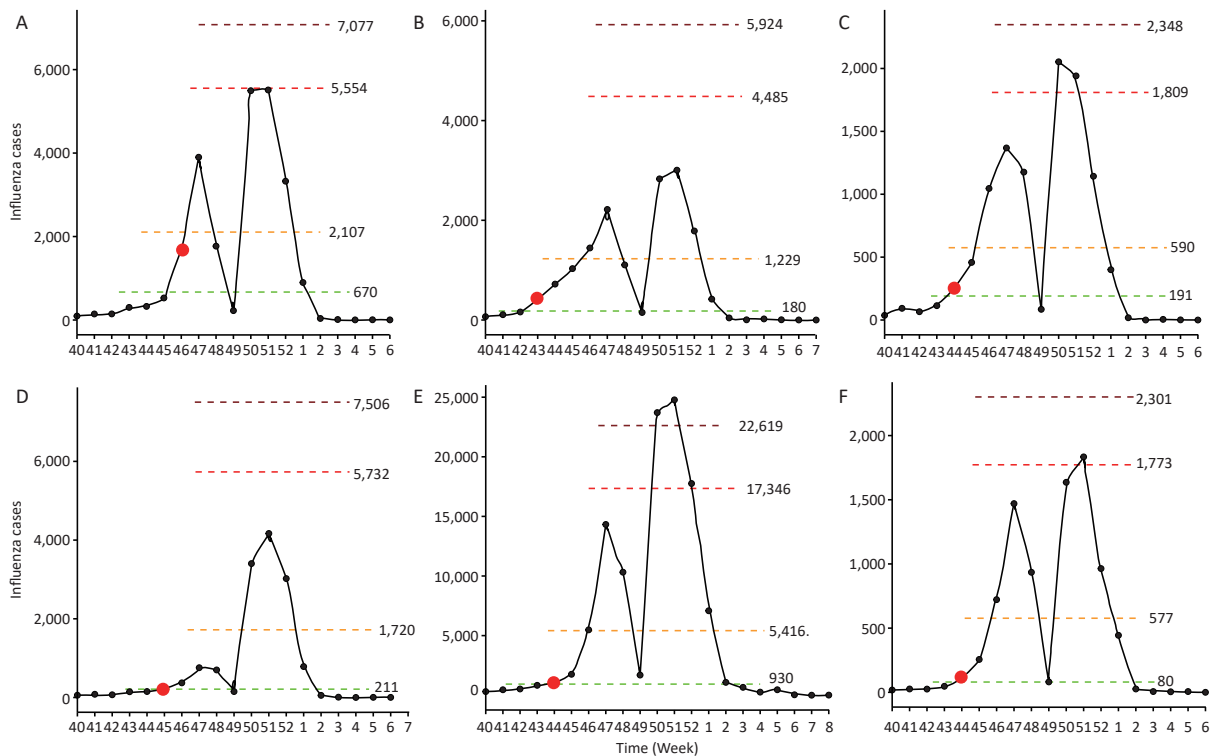


Figure 2. Influenza epidemiological intensity in major cities of Hubei Province, 2023 to 2024. (A) Huanggang; (B) Enshi Tujia and Miao Autonomous Prefecture; (C) Xiangyang; (D) Shiyan; (E) Wuhan; (F) Xianning.

Table 1. Influenza epidemiological intensity at the provincial (city) level in Hubei Province, 2023–2024

Province/City	Epidemic period	Epidemic period (weeks)	Peak value (cases)	Epidemic level during the epidemic period
Hubei Province	Week 45 of 2023 to Week 2 of 2024	9	63,627	High prevalence level
Wuhan	Week 44 of 2023 to Week 2 of 2024	10	24,775	Extremely high prevalence level
Shiyan	Week 45 of 2023 to Week 2 of 2024	9	4,168	Medium prevalence level
Huanggang	Week 46 of 2023 to Week 2 of 2024	8	5,512	Medium prevalence level
Xianning	Week 44 of 2023 to Week 2 of 2024	10	1,834	High prevalence level
Xiangyang	Week 44 of 2023 to Week 2 of 2024	10	2,053	High prevalence level
Enshi Tujia and Miao Autonomous Prefecture	Week 43 of 2023 to Week 2 of 2024	11	3,005	Medium prevalence level
Synthesized Baidu Index	Week 46 of 2023 to Week 2 of 2024	8	1,226.6	Medium prevalence level

CONCLUSION

Preventing influenza resurgence remains a public health priority. Determining the onset of influenza epidemics is of great significance for medical resource allocation and vaccination guidance. By integrating real-time Baidu Index data and establishing influenza epidemic thresholds, it is possible to enhance surveillance efficiency, predict influenza trends in advance, and adjust prevention and control strategies in a timely manner. The MEM model, with its simplicity and effectiveness, meets the needs of public health services and provides a scientific basis for influenza prevention and control.

Funding This study was supported by the National Key Research and Development Program of China (Project No. 2023YFC2307500).

Competing Interests The authors declare no competing interests relevant to this article.

Ethics This retrospective study did not require ethical review.

Author Contributions Conceptualization, supervision, and project administration were conducted by Cheni and Tong. Methodology and validation were carried out by Chen Caixia Dang et al. The data analysis was performed by Caixia Dang, Feng Liu, and Hengliang Lv. Data curation and visualization were performed by Zhao, Zhu, and Wang. The original draft of the manuscript was written by Caixia Dang and Hui Chen. Hui Chen, Yeqing Tong and Yuanyong Xu wrote, reviewed, and edited the manuscript. All the authors have read and approved the final version of the manuscript.

Acknowledgements The authors thank the Hubei

Provincial Center for Disease Control and Prevention for providing access to the databases used in this study.

Data Sharing No additional data available. The supplementary materials will be available in www.besjournal.com.

Received: August 17, 2025;

Accepted: March 7, 2025

REFERENCES

1. Wang DN, Lei H, Wang DY, et al. Association between temperature and influenza activity across different regions of China during 2010-2017. *Viruses*, 2023; 15, 594.
2. Yang LY, Zhang T, Han X, et al. Influenza epidemic trend surveillance and prediction based on search engine data: deep learning model study. *J Med Internet Res*, 2023; 25, e45085.
3. Vega T, Lozano JE, Meerhoff T, et al. Influenza surveillance in Europe: establishing epidemic thresholds by the moving epidemic method. *Influenza Other Resp Viruses*, 2013; 7, 546–58.
4. Vega T, Lozano JE, Meerhoff T, et al. Influenza surveillance in Europe: comparing intensity levels calculated using the moving epidemic method. *Influenza Other Resp Viruses*, 2015; 9, 234–46.
5. Yang WZ, Li ZJ, Lan YJ, et al. A nationwide web-based automated system for outbreak early detection and rapid response in China. *Western Pac Surveill Response J*, 2011; 2, 10–5.
6. Li K, Liu ML, Feng Y, et al. Using Baidu search engine to monitor AIDS epidemics inform for targeted intervention of HIV/AIDS in China. *Sci Rep*, 2019; 9, 320.
7. Al Amad M, Almoayed K. Influenza circulating viruses, positivity rate and risk factors for influenza associated severe acute respiratory infection during 2018/2019 winter season, Yemen. *BMC Infect Dis*, 2022; 22, 111.
8. Buchy P, Badur S. Who and when to vaccinate against influenza. *Int J Infect Dis*, 2020; 93, 375–87.
9. Feng LZ, Zhang T, Wang Q, et al. Impact of COVID-19

outbreaks and interventions on influenza in China and the United States. [Nat Commun](#), 2021; 12, 3249.

10. National Immunization Advisory Committee (NIAC) Technical

Working Group (TWG), Influenza Vaccination TWG. Technical guidelines for seasonal influenza vaccination in China (2023-2024). *Chin J Epidemiol*, 2023; 44, 1507–30. (In Chinese)