

Supplementary Material

Additional file 1: Methods

1 Synthesize Baidu Index

This study preliminarily screened keywords through the keyword popularity on Baidu website. Through Spearman correlation analysis, we conducted secondary filtering of keywords based on their correlation with influenza case numbers, ultimately identifying key terms including "influenza", "influenza A", "influenza A virus", "viral influenza", "influenza prevention", "influenza treatment" and "seasonal influenza". To avoid multicollinearity, we synthesized these keywords into a composite Baidu Index as the optimal surveillance factor, with higher weights assigned to keywords showing stronger correlations. The synthesis formula was as follows:

$$M^W(t) = \sum_{i=1}^7 \frac{C_i}{\sum_{j=1}^7 C_j} M_i \quad (1)$$

In Formula (1), $M^W(t)$ represents the synthesized Baidu Index, M_i denotes the Baidu Index of the i -th keyword, C_i is the maximum correlation between the Baidu Index of the i -th keyword and influenza cases.

2 Model Establishment and Parameter Selection

The MEM involves determining the start, end, and duration of disease epidemics using the Maximum Cumulative Sum Percentage (MAP) function based on historical data. The epidemic start threshold was calculated using surveillance data from the pre-epidemic phase of historical data. Specifically, the n largest surveillance metrics from the pre-epidemic phase ($n = 30/N$, N represents the number of epidemic seasons) were selected, and the upper limit of the one-sided 95% confidence interval (CI) of their arithmetic mean was computed as the epidemic start threshold. Similarly, the end threshold is calculated using data from the post-epidemic phase. Additionally, the medium, high, and very high epidemic thresholds are determined by analyzing the n largest monitoring indicators during the epidemic phase ($n = 30/N$, N represents the number of epidemic seasons) and calculating the upper limits of the one-sided 40% CI , one-sided 90% CI , and one-sided 97.5% CI of their geometric mean, respectively. The formula for the MAP function is:

$$t_j^r = \max_{k=1, \dots, s-r+1} \left\{ \sum_{i=k}^{k+r-1} t_{i,j} \right\}$$

$$t_j^s = \sum_{i=1}^s t_j^s \quad (2)$$

$$p_j^r = \frac{t_j^r}{t_j^s}$$

In Formula (2), t_j^r represents the maximum cumulative sum of surveillance data over r consecutive weeks in the j -th epidemic season, t_j^s denotes the total cumulative sum of surveillance data for the j -th epidemic season, p_j^r is the maximum percentage of surveillance data within r consecutive weeks of the j -th epidemic season. Here, s represents the total number of weeks in each epidemic season, with the epidemic phase starting at week k and lasting for r weeks, ending at week $(k + r - 1)$.

30 The parameter δ in the MEM model serves as a critical criterion for distinguishing between epidemic and non-epidemic
31 periods, with a recommended range of 2% to 4%. In this study, a trial-and-error approach was employed, iterating
32 through values of δ from 2% to 4% in increments of 0.1%, and the value of δ corresponding to the maximum Youden's
33 index was selected as the optimal parameter for inclusion in the MEM model.

34 3 Epidemic Intensity Assessment

35 After the model was determined, the epidemic intensity of the target monitoring years was evaluated using the start
36 and end thresholds of epidemics, as well as the thresholds for moderate, high, and extremely high epidemic intensities.
37 The epidemic intensity was divided into five levels: Baseline Level: Weekly reported cases \leq epidemic start/end
38 threshold; Low Epidemic Intensity: Epidemic start threshold $<$ weekly reported cases \leq medium epidemic intensity
39 threshold; Medium Epidemic Intensity: Medium epidemic intensity threshold $<$ weekly reported cases \leq high epidemic
40 intensity threshold; High Epidemic Intensity: High epidemic intensity threshold $<$ weekly reported cases \leq very high
41 epidemic intensity threshold; Very High Epidemic Intensity: Weekly reported cases $>$ very high epidemic intensity
42 threshold;

43

44

45

46

47

48

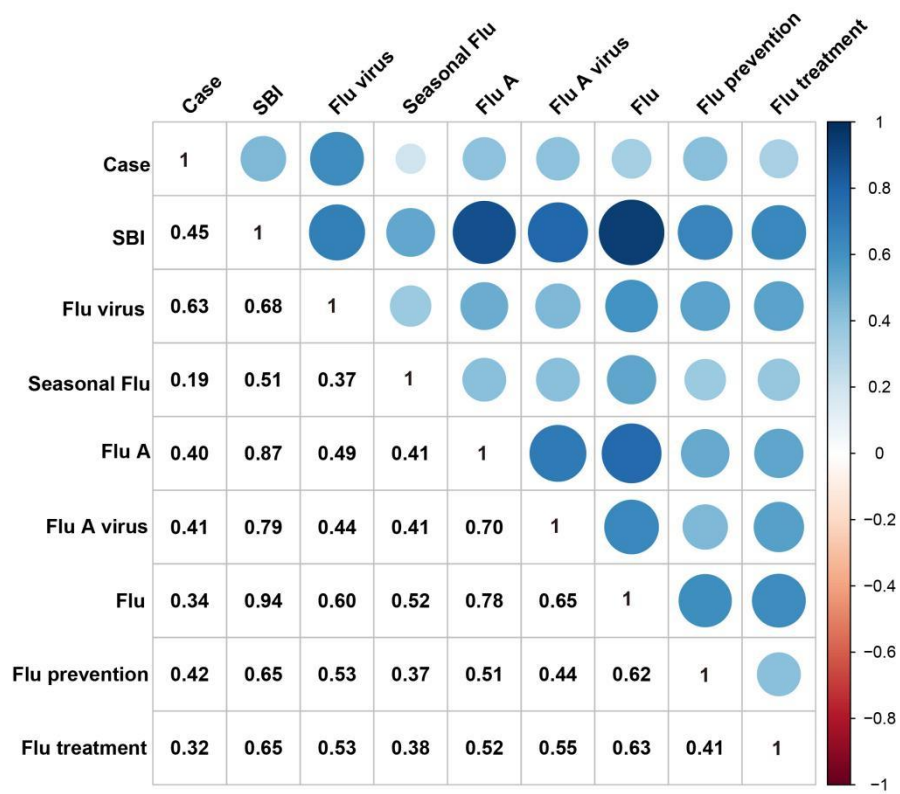
49

50

51

52

53



54

55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83

Supplementary Figure S1.Correlation analysis between Baidu Search Index keywords and influenza cases. SBI, Baidu Search Index,Flu, infiuenza.

84
85**Supplementary Table S1.** Evaluation of cross-validation fit effect for optimal parameters of the MEM model for influenza in Hubei region

Region	Year	Optimal δ	Youden's Index	Matthew's Correlation Coefficient	Sensitivity	Specificity
Hubei Province	2020-2024	3.4	0.9	0.83	0.93	0.97
Wuhan	2020-2024	3.6	0.84	0.65	0.86	0.94
Shiyan	2020-2024	3.5	0.57	0.5	0.65	0.92
Huanggang	2020-2024	3.5	0.87	0.73	0.93	0.94
Xianning	2020-2024	3.6	0.89	0.8	0.92	0.96
Xiangyang	2020-2024	3.5	0.7	0.61	0.76	0.93
Enshi Tujia and Miao Autonomous Prefecture	2020-2024	3.7	0.83	0.64	0.94	0.89
Synthesize Baidu Index	2020-2024	3.3	0.67	0.58	0.73	0.95

86

87

88

89

Supplementary Table S2. Influenza epidemic levels in Hubei province (cities) from 2020 to 2023

Year and Province/City	Peak Value	Peak Week	Epidemic Start Threshold	Medium Epidemic Intensity Threshold	High Epidemic Intensity Threshold	Very High Epidemic Intensity Threshold	Epidemic Level
Hubei Province							
2020/2021	648	51	4,998	15,798	59,679	107,385	baseline level
2021/2022	14,641	41	3,928	15,077	160,025	454,588	low prevalence level
2022/2023	78,888	25	3,115	4,912	25,372	25,372	extremely high prevalence level
Wuhan							
2020/2021	109	34	942	3,600	19,836	42,173	baseline level
2021/2022	2,984	40	708	4,057	58,149	188,634	low prevalence level
2022/2023	23,998	25	594	594	1,482	3,783	extremely high prevalence level
Shiyan							
2020/2021	203	51	219	1,318	8,163	18,280	baseline level
2021/2022	572	43	159	1,036	11,442	33,079	low prevalence level
2022/2023	7,813	25	175	261	645	963	extremely high prevalence level

Huanggang							
2020/2021	63	4	609	1,840	5,801	9,635	baseline level
2021/2022	1,981	42	673	1,465	15,576	44,285	medium prevalence level
2022/2023	7,779	25	427	611	3,875	8,767	high prevalence level
Xianning							
2020/2021	21	4	71	458	1,842	3,405	baseline level
2021/2022	537	41	80	420	4,380	12,339	medium prevalence level
2022/2023	2,848	25	43	154	913	2,005	extremely high prevalence level
Xiangyang							
2020/2021	64	53	195	652	2,084	3,484	baseline level
2021/2022	893	42	118	251	3,299	10,308	medium prevalence level
2022/2023	2,938	25	156	156	951	2,208	extremely high prevalence level
Enshi Tujia and Miao Autonomous Prefecture							
2020/2021	62	38	183	853	4,134	8,303	baseline level

2021/2022	893	42	118	250	3,299	33,079	low prevalence level
2022/2023	2,938	25	156	156	951	1,699	extremely high prevalence level
