Tracking and predicting hand, foot, and mouth disease (HFMD) epidemics in China by Baidu queries

Q. Y. XIAO[1,2]†, H. J. LIU[3]*† AND M. W. FELDMAN[3,4]

[1] *School of Humanities, Economics and Law, Northwestern Polytechnical University, Xi'an, Shaanxi Province 710049, China*
[2] *Department of Management Engineering, Engineering University of CAPF, Xi'an, Shaanxi Province 710049, China*
[3] *Institute for Population and Development Studies, School of Public Policy and Administration, Xi'an Jiaotong University, Xi'an, Shaanxi Province 710049, China*
[4] *Morrison Institute for Population and Resource Studies, Stanford University, Stanford, CA 94305, USA*

SUMMARY

Hand, foot, and mouth disease (HFMD) is highly prevalent in China, and more efficient methods of epidemic detection and early warning need to be developed to augment traditional surveillance systems. In this paper, a method that uses Baidu search queries to track and predict HFMD epidemics is presented, and the outbreaks of HFMD in China during the 60-month period from January 2011 to December 2015 are predicted. The Pearson correlation coefficient (R) of the predictive model and the mean absolute percentage errors between observed HFMD case counts and the predicted number show that our predictive model gives excellent fit to the data. This implies that Baidu search queries can be used in China to track and reliably predict HFMD epidemics, and can serve as a supplement to official systems for HFMD epidemic surveillance.

Key words: Epidemics, HFMD, prediction, search query.

BACKGROUND

Hand, foot, and mouth disease (HFMD) is an infectious disease triggered by an enterovirus. Its victims are mainly children under five years old. HFMD, which is mainly prevalent in East and Southeast Asia, infects many people in such countries as China, Malaysia, Japan, and Taiwan each year, where it is a serious public health problem [1]. According to a report from the Department of Disease Control and Prevention of the Chinese National Health and Family Planning Commission (CNHFPC), from January 2011 to December 2015, there were 1 05 27 500 HFMD infections in China, and 1967 deaths. HFMD has caused great suffering to afflicted children and their families. If rapid and low-cost disease surveillance and prediction were possible, the department of public health could institute early prevention tactics in places where children congregate (e.g. kindergartens) [2].

Prediction of HFMD epidemics in China has aroused wide interest, and various predictive models have been constructed. These include a dynamic model [3], an autologistic regression model [4], a gray system GM (1,1) model [5, 6], a neural network model [7] and an auto-regressive integrated moving

* Author for correspondence: Dr. H. Liu, Institute for Population and Development Studies, School of Public Policy and Administration, Xi'an Jiaotong University, Xian Ning West Road No. 28, Xi'an, Shaanxi Province 710049, China.
(Email: liuhuij@mail.xjtu.edu.cn)
† These authors contributed equally to this manuscript.

average (ARIMA) model [1, 6, 8–11]. Pan et al. found the ARIMA model preferable to the GM (1,1) gray system model in predicting HFMD by comparing their outputs [6]. More research on predicting China's HFMD epidemic has been based on the ARIMA model than on any other model. However, internet data have rarely been used for surveillance and prediction of HFMD outbreaks in China. Tracking and predicting infectious diseases using query data from internet searches has the advantages over other methods of high speed and low cost [12, 13].

Previous research has shown that search engine query data can predict such epidemics as seasonal flu [12, 14], human immunodeficiency virus (HIV) [15], rotavirus vaccination (RV) [16], West Nile virus (WNV) [16], respiratory syncytial virus (RSV) [17] and methicillin-resistant staphylococcus aureus (MRSA) [18]. However, research in Australia carried out by Page et al. found that network queries did not help to predict suicide [19]. Thus not all diseases can be monitored or tracked using internet search query data. Furthermore, the studies mentioned above are based on query data from Google or Yahoo, which are relevant in the context of English culture. However, it remains to be tested whether similar conclusions apply to other cultures, contexts, or search engines.

Baidu is the largest Chinese search engine and commands a marked lead among search engines with a 55% market share in China [20]. It should be noted that the epidemic outbreaks of influenza and erythromelalgia (EM) in China were successfully predicted by Baidu query data [13, 21]. This suggests that internet query data can be used to predict some infectious diseases in the Chinese context. However, in previous studies some predictive models took Baidu queries as the only independent variable [21], while the others added historical disease cases as another independent variable [13, 22]. Whether it is necessary to add historical disease cases needs to be tested.

For prediction of infectious disease epidemics in a given geographic region using internet query data, the search engine that owns the largest market share in the area should be chosen, so as to guarantee the representativeness of the data. Baidu is the most popular search engine in China and is preferred by 86·7% of internet users [21]. As estimated by Tech in Asia, compared with 3·3 billion daily searches of Google, Baidu's daily search volume reaches around 5 billion [23].

A recent study by Huang et al. [22] has estimated HFMD prevalence in Guangdong province, China, with Baidu search data. Taking into consideration spatial inconsistency between disease cases and internet queries in different areas, which might create a bias for epidemic prediction using search engine data, Huang et al. evaluated Baidu search data with the biased sentinel hospital-based area disease estimation (B-SHADE) model [24] by sampling those subspaces with a high correlation coefficient between HFMD cases and Baidu queries [22]. However, whether a revision of the Baidu index is necessary for HFMD prediction with big internet datasets is still not clear. In addition, no study has yet focused on the HFMD epidemic at the national level in China. Here we use Baidu queries to analyze whether online behaviors can predict HFMD outbreaks in China.

DATA AND METHODS

Query keywords

The query data chosen to monitor infectious diseases depend on the key words used to filter search records. A typical method directly generates key word combinations from the symptoms of diseases. Zeng and Wagner propose that patients' psychological status can be divided into four stages: the perception of symptoms, the explanation of symptoms, the expression of perception, and the search for solutions [25]. Patients (or their family members) at the second or third phase who seek online medical aid usually set key words through the description of symptoms [25], so that epidemics can be predicted by referring to the frequency of key words relevant to symptoms [12, 14, 17, 21]. For instance, Ginsberg et al. selected 45 key words related to symptoms of influenza-like illness (ILI), which they used to search Google records and detect the epidemic in the USA [12].

Disease symptoms are not the only choice for key words to filter search queries in epidemic detection. Yuan et al. [13] successfully predicted flu epidemics by Baidu queries with eight key words, including 'prevent influenza', 'influenza symptoms', 'type A influenza vaccine', 'flu symptom', 'flu epidemic', 'influenza virus', 'influenza pandemic or type A influenza' (in Chinese). Interestingly, these words use generic nouns but not ILI symptoms. Huang et al. [22] used the same method to choose key words for the prediction of HFMD epidemics from Baidu queries, but extended the number of key words to 11. Hulth et al. [14] detected an influenza outbreak in Sweden by counting 20 types of web queries that contained 'influenza' or symptoms of ILI (in Swedish).

Here not only ILI symptoms but also the name of the infectious disease were used as key words to filter search queries.

Prediction of epidemics can be achieved in various ways. A number of researchers choose the names of diseases as key words. For instance, Polgreen et al. [26] used the key words 'flu' and 'influenza' to predict flu epidemics in USA; Dukic et al. [18] predicted MRSA hospitalization rates through quarterly variation in Google in the USA by searching for 'MRSA' and 'staph'; Jena et al. also reported that in the USA the annual incidence of HIV diagnosis is highly correlated with the frequency of Google searches for 'HIV' [15].

Actually, which psychological action stage that patients belong to is not as clear as Zeng and Wagner [25] claimed. Seeking online medical aid might occur at any time during the whole duration of an infectious disease. When afflicted by some common illnesses, people often search for relevant information with the names of diseases as key words. Moreover, in real life, although many patients (or their family members) know about a disease, after a doctors' diagnosis they still seek help through the internet. Thus using the name of the infectious disease as a key word might be a simple but effective way to filter patients' (or their family members') search queries. As HFMD is a common infectious disease, the present study uses '手足口病' (HFMD in Chinese) as the only key word to detect and collect relevant queries of Baidu in China.

Data

Two types of data are needed: one is China's HFMD case numbers, and the other is Baidu queries for the entire country. We collected the two kinds of data for the 60-month period from January 2011 to December 2015. The count of HFMD cases is publicly available through 'the report on nationally notifiable infectious diseases' delivered monthly by the CNHFPC, whose website is http://www.moh.gov.cn/jkj/index.shtml. These data are monthly aggregated cases. The query data concerning HFMD were obtained through the 'Baidu index', a sharing platform of big data, whose website is http://index.baidu.com. The name of the disease was the only key word we used to analyze the queries in the Baidu search. As Baidu query data are available on a daily basis, the average value over a given month was treated as the monthly count for that month.

Methods

Since there may be spatial heterogeneity in Baidu searching and HFMD cases due to China's vast diversity, we used Moran's I [27], Gi [28] and Q-statistic [29] to test for spatial autocorrelation, spatial local heterogeneity, and spatial stratified heterogeneity. Results indicate that both Baidu queries and HFMD cases are spatially heterogeneous (data not presented here but available on request). However, the spatial distributions of HFMD cases and Baidu queries are highly correlated and exhibit similar trends over years, which implies that the potential impact of spatial heterogeneity on the relationship between Baidu searching behaviors and HFMD cases is small.

To further test the potential impact of spatial inconsistency between HFMD cases and Baidu queries in subareas, two kinds of models have been constructed: one deals with potential bias between HFMD cases and Baidu queries by re-evaluating the Baidu index with the B-SHADE model as was done by Huang et al. [22]; the other model ignores such diversities. We carried out the test under two conditions: first, HFMD epidemics in China in 2014 and 2015 (dependent variable) were predicted by the only independent variable, namely the Baidu index or the revised Baidu index separately. Second, we added historical HFMD cases into the predictive models as the other independent variable. Mean absolute percentage errors (MAPE) was used to evaluate predictive accuracy of different models.

As Table 1 shows, under the first condition the model that did not re-evaluate the Baidu index with the B-SHADE model had better predictive accuracy. Under the second condition the predictive accuracy was very high. It seems that handling spatial inconsistency between HFMD cases and Baidu queries in subareas did not significantly improve the predictive accuracy. Re-evaluating the Baidu index is not helpful for HFMD prediction at national level in China. On the other hand, after adding historical HFMD cases to the predictive model, MAPE in 2014 was significantly reduced, but MAPE in 2015 increased a little. Thus adding historical disease cases might not generally increase accuracy of prediction, but the variation in MAPE is decreased, which could produce more stable predictions.

We used log-linear regression to construct a predictive model of HFMD epidemics, with the number of HFMD cases as the dependent variable, and the number of Baidu queries and the number of historical

Table 1. *MAPE comparisons of HFMD prediction by two kinds of models*

Variable of historical HFMD cases	Spatial inconsistence in subareas	MAPE in 2014	MAPE in 2015
Not added	Handled	0·484	0·281
	Not handled	0·465	0·245
Added	Handled	0·271	0·300
	Not handled	0·278	0·296

Note: Added, adding historical HFMD cases as an independent variable in the predictive model; Not added, remains unchanged; Handled, spatial inconsistency between HFMD cases and Baidu queries in subareas are handled and re-evaluated by the B-SHADE model; Not handled, remains unchanged.

HFMD cases as the independent variables. The model we use is presented below:

$$\log y_t = \alpha + \beta_1 \log y_{t-1} + \beta_2 \log x_t + \varepsilon \quad (1)$$

where y_t represents number of cases at t time, y_{t-1} the number of cases at t–1 time, χ_t query number at t time, α, β_1, and β_2 are coefficients to be estimated, and ε is the residual error.

Since epidemics are dynamic, whether a prediction at a given time is correct or not may depend on the time. As time goes by, if the predictive curves at different stages fit the actual case number, the conclusion can be more persuasive than fit only at a single predictive stage. The predictive model of big data proposed by Ginsberg *et al*. adopts the strategy of analyzing and predicting in stages [12]. Following that model, we used our model in equation (1) at three periods to predict the outbreak of China's HFMD epidemics in stages.

During a period of 60 months, the queries of Baidu at different stages were used to track and predict HFMD epidemics in China. In the first phase, data for the first 24 months (from January 2011 to December 2012) were used to build a model to predict the HFMD epidemic from the 25th to the 36th month. In the second phase, data for the first 36 months (from January 2011 to December 2013) were used to predict the HFMD epidemic from the 37th to the 48th month. For the third phase, data for the first 48 months (from January 2011 to December 2014) were used to predict the HFMD epidemic from the 49th to the 60th month.

A comparative method was used to analyze prediction of the HFMD epidemic based on the Baidu queries. First, international predictive models that predict all kinds of epidemics by query data from search engines were compared. The fits of different models, especially using the correlation coefficient R, were estimated. The greater the value of R, the better the prediction; an F-test was used to test for significance level. We compared the different methods according to the prediction of HFMD epidemics in China. For different models, comparison of their mean relative errors between HFMD's predictive value and the actual case number (or incidence of the disease) is pointless. Therefore, the predictive-effect indicators of different models were further transformed into MAPE, and the indicators were compared. The smaller the absolute mean relative error, the better the prediction. All the statistical analysis was carried out using the software SPSS 19.

PREDICTIVE MODELS AND CONCLUSIONS

Our predictive model of China's HFMD epidemic applies Formula (1) in three stages.

Stage 1

By fitting HFMD cases to Baidu query data from the first 24 months, we obtain the predictive model shown in equation (2):

$$\log y_t = 0 \cdot 425 + 0 \cdot 263 \log y_{t-1} + 0 \cdot 944 \log x_t \quad (2)$$

$R = 0.954$, $adj.R^2 = 0.901$, $F = 105.460$ $(P < 0.001)$.

Stage 2

By fitting HFMD cases to Baidu query data from the first 36 months, we obtain the predictive model shown in equation (3):

$$\log y_t = 0 \cdot 656 + 0 \cdot 225 \log y_{t-1} + 0 \cdot 924 \log x_t \quad (3)$$

$R = 0.950$, $adj.R^2 = 0.896$, $F = 152.107$ $(P < 0.001)$.

Stage 3

By fitting HFMD cases to Baidu query data from the first 48 months, we obtain the predictive model shown in equation (4):

$$\log y_t = 0 \cdot 718 + 0 \cdot 318 \log y_{t-1} + 0 \cdot 770 \log x_t \quad (4)$$

$R = 0.956$, $adj.R^2 = 0.910$, $F = 238.056$ $(P < 0.001)$.

Analysis of the model in three stages shows that there is a strong correlation between the number of

Baidu queries and the number of China's HFMD cases. R, the correlation with the predictive model, varies between 0·950 and 0·956; adjusted R^2 is between 0·896 and 0·910; the F-test is highly significant ($P < 0.001$). Thus our models fit very well. These three models were applied to predict HFMD epidemics in the 12 months subsequent to the dates of the data used to estimate the parameters.

Comparisons between HFMDs predicted and actual numbers are shown in Figures 1b, d, and f. The Pearson correlation coefficient between the predicted values and actual cases in the three stages ($R = 0.950$; $P < 0.001$) indicates that the predictive ability of our models is very strong. Figure 1 shows that the prevalence curves of Baidu queries and HFMD cases move in the same direction, and every 12 months (each year) there is a clear peak that represents that year's flashpoint for HFMD. Thus, months such as the 6th (June 2011), the 17th (May 2012), the 30th (June 2013), the 41st (May 2014), and the 54th (June 2015) show that HFMD has a periodic peak, typical of seasonal infectious disease. This finding is consistent with the properties of HFMD's epidemiological features proposed by Xing et al. [30] and Hu et al. [31]. Clearly, Baidu search query data can be used to track and predict the epidemic of HFMD quite well in China. These findings can be used to provide warnings before outbreaks and thus prevent or reduce the risks of infection.

Table 2 compares different models for epidemic prediction by internet queries. In our study, the average correlation coefficient R in the three models is 0·93, which is significant at 0·1% level. This is better than those in models using Google, Yahoo, or Baidu queries to predict flu [12, 25], HIV [15], RV [16], HFMD [22], and is similar to those using Baidu or Google queries to predict EM [21] and MRSA [18]. Thus, our method of predicting HFMD from Baidu search queries has high accuracy. Additionally, both common and rare infectious diseases could be successfully predicted, which suggests that internet queries are broadly applicable for tracking and predicting epidemics. The comparison in Table 2 also reveals that the predictive effect of Baidu queries in Chinese is better than that of Google and Yahoo in English. Unlike Chinese, English is rich in compound words and easy to misspell. Also Baidu has a higher share of the Chinese search engine market than Google and Yahoo have in the USA.

A comparison of MAPE between the number of HFMD cases and the predictive value of different methods was made to analyze the prediction of HFMD epidemics using internet query in China. Previous predictive studies, which adopted methods such as dynamic modeling [3], autologistic regression [4], gray system GM (1,1) [5, 6], and neural networks [7], did not report errors in the predicted value; therefore the predictive value of these different models cannot be compared using MAPE. Huang et al. [22] used MAPE to describe the difference of fitted results and the real HFMD data. But what we care about is the divergence between the predictive value and the actual case number. Fortunately, most predictive studies that use the ARIMA model provide MAPE of HFMD epidemic prediction. As Table 3 shows, the error in the predicted value varies between 0·15 and 1·3, and it is higher than 0·3 in most studies [1, 8, 10,11]. In the current study, which predicts the outbreaks of HFMD epidemics within 60 months, the absolute mean relative error is 0·28, indicating that our method has good predictive accuracy over a longer tracking and predictive period.

DISCUSSION

Infectious diseases, whose pathogens can spread in a variety of ways, may reach high prevalence over a vast region. Such diseases need to be monitored and predicted without delay. Internet search engines provide disease information or medical consultation online, and caregivers of patients may search using relevant themes when disease occurs. By analyzing the address of IP or hardware, query behaviors can be counted and located geographically. Therefore, the prevalence status of epidemics can be detected by analysis of online behavior data [12]. This is very important in China, which is densely populated and has many seasonal and regional infectious diseases.

The current study, using Baidu queries in the Chinese context, manages to track and predict outbreaks of HFMD epidemics. Although there is spatially stratified heterogeneity in Baidu searching behavior and HFMD cases due to regional differences across China, our study shows that the spatially stratified heterogeneity has little negative effect on the relationship between HFMD cases and Baidu queries. Also, spatial inconsistency between HFMD cases and Baidu queries at the provincial level has little effect on prediction of HFMD epidemics with internet searching data at the national level. When detecting the outbreak of an epidemic with internet engine data, historical disease cases should be embedded into the
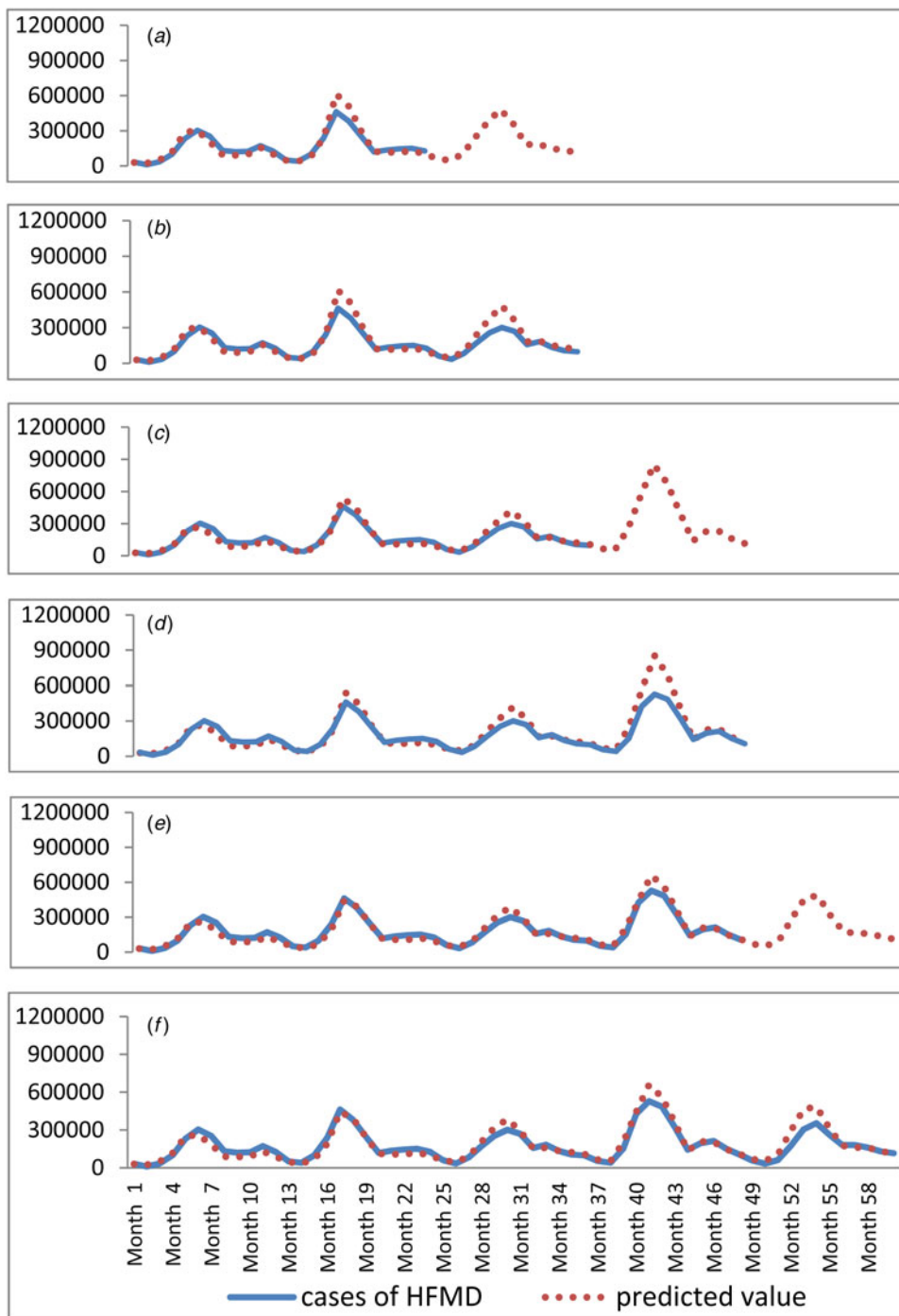
Fig. 1. Tracking and predicting outbreaks of HFMD epidemics in China during 60 months. (*a–f*) Describe the tracking and prediction of HFMD epidemics in Mainland China by Baidu queries during 60 months (from January 2011 to December 2015). (*a*) Predicts HFMD cases from the 25th to the 36th month, (*c*) predicts HFMD cases from the 37th to the 48th month, and (*e*) predicts HFMD cases from the 49th to the 60th month. (*b, d, f*) Compare the predicted and actual number within the corresponding period.

predictive model, which might make predictive outputs more stable.

How to choose the key words to filter search records is a central problem in predicting epidemics with internet queries. Even for the same infectious disease (e.g. flu), different researchers have adopted different key words [12–14, 25]. The unfamiliarity of the public with a disease might pose a challenge to

Table 2. *Comparison of different models of epidemic prediction by internet queries*

Study name	R (P value)	Infectious disease	Search engine	Language	Country
Ginsberg *et al.* [12]	0·85 (P < 0·05)	Flu	Google	English	USA
Polgreen *et al.* [26]	0·61 (P < 0·001)	Flu	Yahoo	English	USA
Hulth *et al.* [14]	0·95	Flu	Vardguiden.se	Swedish	Sweden
Yuan *et al.* [13]	0·96 (P < 0·001)	Flu	Baidu	Chinese	China
Jena *et al.* [15]	0·83 (P < 0·001)	HIV	Google	English	USA
Desai *et al.* [16]	0·88 (P < 0·001)	RV	Google	English	USA
Dukic *et al.* [18]	0·93 (P < 0·001)	MRSA	Google	English	USA
Gu *et al.* [21]	0·93 (P < 0·001)	EM	Baidu	Chinese	China
Huang *et al.* [22]	0·86 (P < 0·001)	HFMD	Baidu	Chinese	China
This study	0·95 (P < 0·001)	HFMD	Baidu	Chinese	China

Note: R is the correlation coefficient between predicted values of the model and disease case counts; P value, significance level. If R (or R^2) of predictive models is obtained more than once, we only give the average value.

Table 3. *Comparison of the prediction of HFMD epidemics in China*

Study name	MAPE	Surveillance area	Predictive period
Hu *et al.* [10]	0·37	Nationwide	6 months
Huang *et al.* [11]	0·29	Nationwide	7 months
Liu *et al.* [1]	0·15	Sichuan Province	6 months
Li *et al.* [8]	1·30	Yantian District of Shenzhen	8 months
This study	0·28	Nationwide	60 months

using internet search queries [18]. For example, in the study on RV prediction, Desai *et al.* [16] used not only the correct term 'rotavirus' to filter internet queries, but also common misspellings such as 'rota virus,' 'rotovirus,' 'roto virus,' 'rodo virus,' and 'rhoda virus' [16]. Users may enter different terms for the same disease, depending on their level of education and their cultural and language backgrounds [17]. Thus too many key words may not be a good choice for epidemic prediction by internet queries. In our study, the name of the infectious disease was used as the only key word, and this method appears to work simply but effectively for HFMD prediction.

From 2002 to 2003, there was an outbreak of severe acute respiratory syndrome (SARS) in China, which lasted for 8 months (from 16 November 2002 to 14 July 2003). The disease caused 7435 infections and 646 deaths, covering 32 regions (provinces, municipalities, autonomous regions, and special administrative zones) in China [32]. Important reasons for the panicked response of the Chinese public health department to that public health emergency were inaccurate estimation at the early stages and delayed reporting of the epidemic situation. Therefore, in 2004 the government of China therefore adopted a policy of reporting, notifying, and announcing

infectious disease epidemics. However, there is still much room for improvement of the system. On the one hand, reporting and aggregation of disease cases takes a lot of time, although the CNHFPC notifies the epidemic status of nationally notifiable infectious diseases every month. Taking the year 2015 as an example, the time to notify averaged 11·5 days later than the last day of the month in which the diseases occurred. On the other hand, information about infectious diseases in smaller regions cannot be obtained, since the notification data concerning infectious diseases refer to the whole nation.

Internet queries are a precious resource that can be used to detect epidemics and should receive the attention of public health departments [12]. The official collection and announcement of epidemics have certain organizational procedures that result in low disclosure efficiency. In the USA, the prediction of seasonal flu by internet search is 1–2 weeks faster than the official announcement [12]; in China, prediction of HFMD epidemics by Baidu queries is about 10 days faster than official disclosure, and use of the former could mitigate the low efficiency in the bureaucratic hierarchy, as well as track and predict more flexibly in space and time at lower surveillance cost. It is suggested that the Chinese government should develop a

supplementary system for HFMD prediction using Baidu queries, to augment the traditional surveillance system.

ACKNOWLEDGEMENTS

DECLARATION OF INTEREST

None.

REFERENCES

1. **Liu L, et al.** Predicting the incidence of hand, foot and mouth disease in Sichuan province, China using the ARIMA model. *Epidemiology and Infection* 2016; **144**: 144–151.
2. **Wei J, et al.** The effect of meteorological variables on the transmission of hand, foot and mouth disease in four major cities of Shanxi province, China: a time series data analysis (2009–2013). *PLoS Neglected Tropical Diseases* 2015; **9**: 1–19.
3. **Li Y, Zhang J, Zhang X.** Modeling and preventive measures of hand, foot and mouth disease (HFMD) in China. *International Journal of Environmental Research and Public Health* 2014; **11**: 3108–3117.
4. **Bo Y, et al.** Using an autologistic regression model to identify spatial risk factors and spatial risk patterns of hand, foot and mouth disease (HFMD) in Mainland China. *BMC Public Health* 2014; **14**: 358. doi: 10.1186/1471-2458-14-358.
5. **Wang W, et al.** Application of grey system GM(1, 1) model on trend forecast of hand, foot and mouth disease. *Chinese Journal of School Doctor* 2013; **27**: 769–770 (in Chinese).
6. **Pan H, et al.** Comparison of GM(1,1) gray model and ARIMA model in forecasting the incidence of hand-foot-mouth disease in Shanghai. *Chinese Journal of Disease Control & Prevention* 2011; **5**: 445–448 (in Chinese).
7. **Zhang X, et al.** Application of Elman Recurrent Neural Network in the incidence prediction of hand-foot-month disease. *Modern Preventive Medicine* 2012; **39**: 2136–2141 (in Chinese).
8. **Li B, Li X, Gu L.** The applied research of ARIMA model in forecasting and early warning of hand-foot-mouth disease. *China Health Industry* 2014; **26**: 26–27 (in Chinese).
9. **Yu L, et al.** Application of a new hybrid model with seasonal auto-regressive integrated moving average (ARIMA) and nonlinear auto-regressive neural network (NARNN) in forecasting incidence cases of HFMD in Shenzhen, China. *PLoS ONE* 2014; **9**: e98241.
10. **Hu Y, et al.** Application of multiple seasonal autoregressive integrated moving average model in prediction of incidence of hand foot and mouth disease in China. *Surveillance* 2014; **29**: 827–832 (in Chinese).
11. **Huang X, et al.** Prediction of monthly hand foot and mouth disease incidence in China by using autoregressive integrated moving average model. *Disease Surveillance* 2016; **144**: 144–151 (in Chinese).
12. **Ginsberg J, et al.** Detecting influenza epidemics using search engine query data. *Nature* 2009; **457**: 1012–1014.
13. **Yuan Q, et al.** Monitoring influenza epidemics in China with search query from Baidu source. *PLoS ONE* 2013; **8**: e64323. doi: 10.1371/journal.pone.0064323.
14. **Hulth A, Rydevik G, Linde A.** Web queries as a source for syndromic surveillance. *PLoS ONE* 2009; **4**: e4378. doi: 10.1371/journal.pone.0004378.
15. **Jena AB, et al.** Predicting new diagnoses of HIV infection using internet search engine data. *Clinical Infectious Diseases* 2013; **56**: 1352–1353.
16. **Desai R, et al.** Internet search data to monitor impact of rotavirus vaccination in the United States. *Clinical Infectious Diseases* 2012; **54**: 115–118.
17. **Carneiro HA, Mylonakis E.** Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases* 2009; **49**: 1557–1564.
18. **Dukic VM, David MZ, Lauderdale DS.** Internet queries and methicillin-resistant *Staphylococcus aureus* surveillance. *Emerging Infectious Diseases* 2011; **17**: 1068–1070.
19. **Page A, Chang S, Gunnell D.** Surveillance of Australian suicidal behavior using the internet? *Australian and New Zealand Journal of Psychiatry* 2011; **45**: 1020–1022.
20. **Return on Now.** 2015 Search Engine Market Share By Country (http://returnonnow.com/internet-marketing-resources/2015-search-engine-market-share-by-country/). Accessed 15 March 2016.
21. **Gu Y, et al.** Early detection of an epidemic erythromelalgia outbreak using Baidu search data. *Scientific Reports* 2015; **5**: 12649. doi: 10.1038/srep12649.
22. **Huang DC, et al.** Towards identifying and reducing the bias of disease information extracted from search engine data. *PLoS Computational Biology* 2016; **6**: e1004876.
23. **Pconline.** Has Baidu surpassed Google with 5 billion daily queries (http://news.pconline.com.cn/322/3221749.html). Accessed 20 March 2013.
24. **Wang JF, et al.** Area disease estimation based on sentinel hospital records. *PLoS ONE* 2011; **8**: e23428.
25. **Zeng X, Wagner M.** Modeling the effects of epidemics on routinely collected data. *Journal of the American Medical Informatics Association* 2002; **9**: 17–22.
26. **Polgreen PM, et al.** Using internet searches for influenza surveillance. *Clinical Infectious Diseases* 2008; **47**: 1443–1448.
27. **Moran PAP.** Notes on continuous stochastic phenomena. *Biometrika* 1950; **37**: 17–23.
28. **Getis A, Ord JK.** The analysis of spatial association by use of distance statistics. *Geographical Analysis* 1992; **24**: 189–206.
29. **Wang JF, Zhang TL, Fu BJ.** A measure of spatial stratified heterogeneity. *Ecological Indicators* 2016; **67**: 250–256.

30. **Xing W, *et al.*** Hand, foot, and mouth disease in China 2008–2012: an epidemiological study. *Lancet Infectious Diseases* 2014; **14**: 308–318.

31. **Hu Y, *et al.*** The epidemic features of the hand, foot, and mouth disease during 2008–2011 in China. *Chinese Journal of Disease Control & Prevention* 2014; **18**: 693–747 (in Chinese).

32. **Fan X, Ying L.** An exploratory spatial data analysis of SARS epidemic in China. *Advances in Earth Science* 2005; **3**: 282–291 (in Chinese).