



## The application of meteorological data and search index data in improving the prediction of HFMD: A study of two cities in Guangdong Province, China

Shaoxing Chen <sup>a,b</sup>, Xiaojian Liu <sup>c</sup>, Yongsheng Wu <sup>c</sup>, Guangxing Xu <sup>d</sup>, Xubin Zhang <sup>d</sup>, Shujiang Mei <sup>c</sup>, Zhen Zhang <sup>c</sup>, Michael O'Meara <sup>e</sup>, Mary Clare O'Gara <sup>f</sup>, Xuerui Tan <sup>g</sup>, Liping Li <sup>a,\*</sup>

<sup>a</sup> Injury Prevention Research Center, Shantou University Medical College, Shantou, Guangdong 515041, China

<sup>b</sup> Department of Community monitoring, The First Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

<sup>c</sup> Shenzhen Center for Disease Control and Prevention, Shenzhen, China

<sup>d</sup> Shantou Center for Disease Control and Prevention, Shantou, China

<sup>e</sup> Department of Information Technology, Shantou University Medical College, Shantou, Guangdong 515041, China

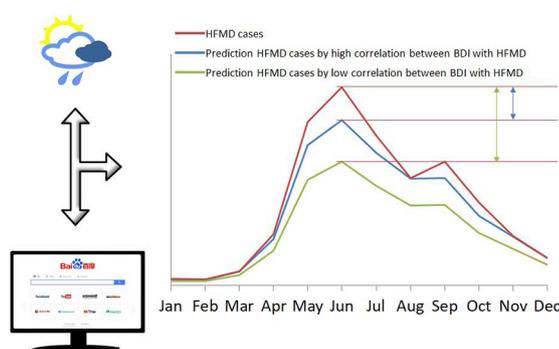
<sup>f</sup> Department of Nursing, Shantou University Medical College, Shantou, Guangdong 515041, China

<sup>g</sup> Department of Cardiology, The First Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

### HIGHLIGHTS

- A multi-centered study showed search index data can improve the performance of the HFMD prediction models.
- Comparing with the lag effect of meteorological factors on HFMD, search index data is more real-time.
- Meteorological indicators showed a linear relationship with HFMD, while BDI was non-linear in both cities.
- The prediction accuracy of the model with high correlation coefficient between BDI and HFMD can be enhanced remarkably.

### GRAPHICAL ABSTRACT



### ARTICLE INFO

#### Article history:

Received 15 July 2018

Received in revised form 22 October 2018

Accepted 22 October 2018

Available online 23 October 2018

Editor: Scott Sheridan

#### Keywords:

HFMD

Prediction

GAM

Meteorological data

Search index data

### ABSTRACT

Hand, foot and mouth disease (HFMD) is a public health issue in China, and its incidence in Guangdong Province is higher than the national average. Previous studies have found climatic factors have an influential role in the transmission of HFMD. Internet search technology has been shown to predict some infectious disease epidemics and is a potential resource in tracking epidemics in countries where the use of Internet search index data is prevalent. This study aims to improve the prediction of HFMD in two Chinese cities, Shantou and Shenzhen in Guangdong Province, applying both meteorological data and Baidu search indices to create a HFMD forecasting model. To this end, the relationship between meteorological factors and HFMD was found to be linear in both cities, while the relationship between search engine data and HFMD was not consistent. The results of our study suggest that using both Internet search and meteorological data can improve the prediction of HFMD incidence. Using comparative analysis of both cities, we posit that improved quality search indices enhance prediction of HFMD.

© 2018 Elsevier B.V. All rights reserved.

\* Corresponding author.

E-mail addresses: [xjliu@szcdc.net](mailto:xjliu@szcdc.net) (X. Liu), [cdc@szcdc.net](mailto:cdc@szcdc.net) (Y. Wu), [sjmei@szcdc.net](mailto:sjmei@szcdc.net) (S. Mei), [zhangz@szcdc.net](mailto:zhangz@szcdc.net) (Z. Zhang), [michaelomeara@stu.edu.cn](mailto:michaelomeara@stu.edu.cn) (M. O'Meara), [maryog@stu.edu.cn](mailto:maryog@stu.edu.cn) (M.C. O'Gara), [lppli@stu.edu.cn](mailto:lppli@stu.edu.cn) (L. Li).

## 1. Introduction

Hand, foot and mouth disease (HFMD) is a common infectious disease among children caused by various human enteroviruses (Cardosa et al., 2003; Wong et al., 2010). The susceptible populations of this disease are mainly preschool children, particularly children under five years old (Xing et al., 2014). Enterovirus 71 (EV71) and Coxsackievirus A16 (CA16) have been confirmed to be the most common causative agents, and some strains of non-polio enteroviruses have also been implicated as causative pathogens (Cardosa et al., 2003; Li et al., 2005; Mandary and Poh, 2018; Wang et al., 2011). Most patients spontaneously recover within 7–10 days due to the typically mild and self-limiting nature of HFMD (Kumar et al., 2015). However, in some instances, death can result from systemic complications, especially when associated with EV71 infection (Wang et al., 2011; Xing et al., 2014).

Since its original identification in New Zealand in 1957, HFMD has been frequently reported worldwide (Kumar et al., 2015; Sarma, 2013; Wang et al., 2011; Zhu et al., 2010). Indeed, it has become a public health priority in China since the first large-scale epidemic in Fuyang City, Anhui Province in 2008 (Zhang et al., 2010; Zhu et al., 2010). Since the epidemic developed over a relatively short time span, EV71-associated HFMD received considerable attention from clinicians and public health officials, such that HFMD was classified as a category C notifiable infectious disease by the Ministry of Health of China on May 2, 2008 (Zhu et al., 2010). According to the national network's surveillance data, from 2008 to 2015, the average annual incidence of HFMD was 133.85/per 100,000 (range: 37.01–205.06), while that of Guangdong Province was 266.14/per 100,000 (range: 51.73–403.50). This is a particular concern in Guangdong Province where the incidence is almost twice the national average (Deng et al., 2013; Zhang et al., 2016a, b, c).

Despite the lengthy history of HFMD globally, there remains a paucity of effective prevention and treatment measures. Should a vaccine prove safe and effective for all children, it may provide an economical and reliable approach to the prevention of HFMD (Aswathyraj et al., 2016 and Liu et al., 2014; Mao et al., 2016). To this end, China has developed an EV-A71 inactivated vaccine which successfully entered the market in 2015 (Mao et al., 2016). Despite this, the outbreak of HFMD has not reduced significantly in China. In 2015 and 2016, the total national HFMD cases were officially reported as 2.00 million and 2.44 million respectively. The reasons put forward to explain this phenomenon include poor uptake of the vaccine to EV71 as the vaccine is non-subsidized and non-mandatory for all children (Wang et al., 2017a, b). While the vaccine is >90% protective for HFMD by EV71 (Zhu et al., 2014), it has no protective effect for HFMD by CA16 or other enteroviruses (Aswathyraj et al., 2016; Li et al., 2015). Additionally, the monovalent EV71 vaccine elicits humoral immunity but lacks long-term immunogenicity (Pinn and Chit, 2017). Therefore, developing a model to predict HFMD remains necessary.

To date, studies have presented the benefits of Internet search query data on diseases for improving real-time tracking and surveillance systems (Du et al., 2017; Li et al., 2017; Madoff et al., 2011; Yuan et al., 2013). In 2017, Li and colleagues found that the dengue warning system combining internet-based query data with traditional surveillance improved prediction capability in Guangzhou (Li et al., 2017). Woo and colleagues demonstrated the feasibility of using Internet search queries to enhance influenza surveillance in South Korea in 2016 (Woo et al., 2016). Based upon this previous application of search index data, it is worth considering whether similar efficacy can result for HFMD. Few studies have focused on HFMD epidemics and Internet search query. For examples, the study by Du and colleagues used a seasonal autoregressive integrated moving average (ARIMA) model to examine whether including search query data can improve the prediction of HFMD based on a provincial-scale data. Another study by Huang and Wang showed that the model which integrated search data, meteorological factors and historical cases exhibit the most stable estimation

result based on one city data. Available studies are based on single-center investigation, and research methods employed are not consistent. It is accepted that meteorological factors play a role in HFMD transmission (Wang et al., 2011; Zhang et al., 2016a, b, c). Previous studies have demonstrated temperature and humidity are of significant influence with regards HFMD (Cheng et al., 2018; Onozuka and Hashizume, 2011; Yang et al., 2017). As other studies have previously used climate variables as predictors to build forecasting models (Feng et al., 2014; Xiao et al., 2017), this research aims at improving the prediction of HFMD in two cities of Guangdong Province, China, Shantou and Shenzhen. Both meteorological data and Baidu search indices were applied to create an HFMD forecasting model.

## 2. Materials and methods

### 2.1. Study sites

Lying proximal to each other in China's southern coast, Shenzhen and Shantou have both experienced serious epidemics of HFMD in recent years in Guangdong Province. The cities share a typical humid subtropical climate, according to the Köppen classification (Peel et al., 2007), with an annual average temperature of 22 °C and an annual average relative humidity of 75%. Shenzhen is one of four major cities in China and comprises eight districts with a total area of 1996 km<sup>2</sup>. At the end of 2017 the population in Shenzhen was 12.5 million, among which local residents account for approximately 30%. Comprising seven districts, Shantou, in comparison, has an area of 2064 km<sup>2</sup>. By 2017, its population was 5.6 million, over 90% of which being local residents.

An investigation of Internet penetration in 2015 showed that Guangdong Province has the largest scale of netizens (internet users) and ranks among the top three provinces of Internet penetration in China (China Report Web, 2017a, b). In 2015, the Internet penetration in Shenzhen reached 82.3%, while that of Shantou was in line with the national average (50.3%) (China Economic Net, 2016; China Report Web, 2017a, b).

### 2.2. Data source

Weekly HFMD cases, Baidu Index (BDI) and meteorological factors (mean temperature and relative humidity) for the 262-week period from January 2011 to December 2015 in Shenzhen and Shantou were gathered. All weekly data were collated from daily data.

#### 2.2.1. HFMD case data

HFMD has been a legally notifiable infectious disease in China since 2008 (China CDC, 2014; Zhu et al., 2010). The HFMD case count was obtained from the Chinese Center for Disease Control and Prevention (China CDC), which are governmental reports summarizing counts of patients diagnosed at health care facilities with a variety of diseases. All individual-level data were anonymized. The clinical criteria for the diagnosis of HFMD were in accordance with Chinese Ministry of Health 2010 guidelines (China Ministry of Health, 2010).

#### 2.2.2. Meteorological data

The meteorological dataset was obtained from China Meteorological Data Sharing Service System, a component of the National Science and Technology Infrastructures Platform of the China Meteorological Administration (<http://data.cma.cn>). It included daily average temperature (TEM) (°C) and relative humidity (RHU) (%).

#### 2.2.3. Search index data

Although Google is a widely and frequently-used search engine, it is not available in China. Of regional search engines in the Chinese language, Baidu is the most widely used in China (China Internet Network Information Center, 2014; Yuan et al., 2013). Therefore BDI was used as

the Internet search index in this study. The shared platform of Baidu index provides search behavior data for numerous search terms (<http://index.baidu.com>). The data are available at the national, provincial and municipal levels daily. In this study, searches on Baidu using the Chinese keywords for 'hand, foot, and mouth disease' (手足口病), which is most widely searched (Xiao et al., 2017), and the search frequency recorded from the Baidu's database was counted for Shenzhen and Shantou respectively.

### 2.3. Statistical methods

Descriptive analysis was applied to present the mean (M) and standard deviation (SD) of the weekly HFMD cases and variables, as well as their time series. Cross-correlation analysis was then performed to examine the correlation between predictors and the HFMD cases with 1 to 12 weeks' lag. In order to compare the results of the two cities, the test of spatial stratified heterogeneity should be compulsory at the early stage of an exploratory spatial data analysis (Wang et al., 2016). Therefore, Cochran-Mantel-Haenszel test (CMH) was performed to analyze the heterogeneity of odds ratios for the indicators in two cities. Considering the population of the two cities, a  $q$ -statistic method was used to measure the degree of spatial stratified heterogeneity and to test its significant (Wang et al., 2016). Finally, a generalized additive model (GAM) (Anderson-Cook, 2007) with a negative binomial family was used to identify the associations between weekly HFMD cases and average temperature, relative humidity and BDI. As noted by previous scholars, GAM is useful in identifying exposure-response relationships for many types of data, particularly in exploring nonparametric relationships (Hastie and Tibshirani, 1995).

All statistical analyses were two-sided and values of  $P < 0.05$  were considered statistically significant. All statistical analyses were conducted in R software Version 3.3.2 (R Development Core Team, 2017), using packages of 'base', 'mgcv', 'ggplot2', 'stringr', and 'forecast'.

Three models were built accordingly. Model 1, Model 2 and Model 3 were expressed as follows:

$$Y_t = \beta_0 + s(\text{week}, df) + s(\text{TEM}_{t-d}, df) + s(\text{RHU}_{t-m}, df) + s(\text{BDI}_{t-n}, df) \quad (\text{Model 1})$$

$$Y_t = \beta_0 + s(\text{week}, df) + s(\text{TEM}_{t-d}, df) + s(\text{RHU}_{t-m}, df) \quad (\text{Model 2})$$

$$Y_t = \beta_0 + s(\text{week}, df) + s(\text{BDI}_{t-n}, df) \quad (\text{Model 3})$$

In these equations,  $Y_t$  represents the predicted mean of HFMD cases during the week  $t$ ;  $s(\text{week}_{t-d}, df)$  denotes the cubic spline of a week with corresponding degrees of freedom ( $df$ ) to control seasonality;  $s(\text{TEM}_{t-d}, df)$  and  $s(\text{RHU}_{t-m}, df)$  indicate the cubic spline of average temperature in the previous  $d$  weeks and relative humidity in the previous  $m$  weeks respectively, within the corresponding  $df$ ;  $d$  and  $m$  indicate the number of lag weeks when the meteorological factors show the maximum cross correlation with HFMD cases;  $s(\text{BDI}_{t-n}, df)$  illustrates the cubic spline of BDI in the previous  $n$  weeks with the corresponding  $df$ .

The  $df$  for each variable was determined by minimizing the sum of the absolute values of partial autocorrelation function (PACF) of residuals (Hastie and Tibshirani, 1995). In this study,  $df$  for *week*, *TEM*, *RHU* and *BDI* was 2, 2, 1 and 2 in Shenzhen, while  $df$  for *week*, *TEM*, *RHU* and *BDI* was 5, 2, 1 and 4 in Shantou.

The disease dataset was divided into two subsets: the first part, from the 1st week of 2011 to the 41th week of 2015, was used for model training and construction, and the subsequent part, from the 42nd to the 52nd week of 2015, for external validity assessment. First, deviance explained and root mean squared error (RMSE) was applied to compare the fit of the three models. Then, leave-one-out cross validation was performed to validate the performance of each model. Each time, a single week's data were temporarily removed, and its value was then predicted using the other data in the training datasets. This process was

repeated until the predicted values and errors of all weeks in the training datasets were obtained. In the final step, the RMSE and adjusted R-square were used to test the relationship between predicted cases and observed cases.

### 3. Results

During 2011 to 2015, a total of 200,210 and 47,656 HFMD cases were reported in Shenzhen city and Shantou city respectively. A summary of HFMD cases, meteorological variables and Baidu Index are presented in Table 1. There was an average of 752.89 and 181.89 HFMD cases every week in the two cities over the study period. The mean value for Shenzhen city weekly data was TEM (23.18), RHU (73.72%) and BDI (276.84). The respective mean value for Shantou city weekly data was TEM (22.65), RHU (75.72%) and BDI (98.27). Both cities showed significant correlations between HFMD cases and TEM or RHU or BDI. The correlation coefficient between TEM or RHU and HFMD count is similar in both cities, while the BDI and HFMD count in Shenzhen (0.78) is much higher than that of Shantou (0.54).

In Table 2, both Breslow-Day and Tarone's statistics show no significant evidence for heterogeneity of odds ratios for HFMD cases and TEM or RHU in two cities ( $\chi^2 = 1.609, P = 0.205$  and  $\chi^2 = 0.004, P = 0.952$ ). Therefore, after controlling the effect of city, TEM and RHU are risk factors for HFMD cases (OR = 16.634, CI: 9.937–27.844,  $P < 0.000$  and OR = 2.852, CI: 1.961–4.146,  $P < 0.000$ ). There is significance for heterogeneity of odds ratios for HFMD cases and BDI in two cities ( $\chi^2 = 70.762$  or  $\chi^2 = 69.853, P < 0.000$ ). Therefore, after controlling the effect of city, BDI is a risk factor for HFMD cases in Shenzhen (OR = 29.278, CI: 14.656–58.487,  $P < 0.000$ ). But in Shantou, BDI has no significant effect on HFMD cases (OR = 0.871, CI: 0.527–1.439,  $P = 0.591$ ). Fig. 1 shows the time series of weekly HFMD cases, BDI, TEM and RHU in Shenzhen and Shantou respectively. The HFMD cases, BDI and meteorological variables all presented obvious seasonal patterns. It can be seen that the curve of HFMD cases was double-peaked every year from 2011 to 2015 inclusive. The BDI curve of Shenzhen double-peaked during the study period, and the BDI curve of Shantou only presented similarly from 2014 to 2015.

For the lag-specific predictor-response relationship (Fig. 2A), TEM in the previous week, RHU in the previous four weeks and BDI in the current week have the highest correlation with HFMD cases in Shenzhen. The TEM in the previous week, RHU in the previous five weeks and BDI in the current week have the highest correlation with HFMD cases in Shantou (Fig. 2B). The RHU curves are shown to be relatively smooth in two cities. According to many previous studies, HFMD activities are better explained when meteorological variables are used with one or two week's lag, which is due to the incubation period for enteroviruses and the potential delay in parental awareness of and response to clinical symptoms of children (Wang et al., 2017a, b; Hii et al., 2011). In this study, TEM with one week's lag and RHU with two week's lag were included into the model.

Fig. 3 shows the dose-response relationship between HFMD cases and meteorological variables or BDI. In both cities, TEM and RHU were approximately linearly associated with HFMD cases, while BDI was non-linear associated with HFMD cases. The HFMD cases appeared to increase with TEM or RHU, but began to decrease slowly once TEM or RHU reached a high threshold. For Shenzhen, the relative risk of HFMD cases initially increased with BDI increment, and then began to stabilize when the BDI was  $>800$ . For Shantou, the relationship between BDI and HFMD cases showed an S curve. The risk of HFMD cases initially increased with BDI, peaked at approximately 50, and then began to decrease slowly until it troughed at a BDI of 140. The test of spatial stratified heterogeneity (Table 3) shows that the stratification is no significant between HFMD and TEM ( $q = 0.03, P = 0.83$ ) or RHU ( $q = 0.10, P = 0.86$ ) after considering the populations of two cities.

Fig. 4 shows the three fitting models during the training process.

**Table 1**  
Descriptive statistics and correlation matrix among weekly HFMD count, TEM, RHU and BDI in Shantou and Shenzhen from 2011 to 2015.

Sites	Variables	Mean	SD	Max	Median	Min	Correlation		
							TEM	RHU	BDI
Shenzhen	Weekly HFMD count	752.89	726.53	3308.00	552.00	10.00	0.68**	0.45**	0.78**
	TEM (°C)	23.18	5.42	30.33	24.29	8.22	1.00	–	–
	RHU (%)	73.72	10.55	99.43	75.57	37.00	0.44**	1.00	–
	BDI	276.84	225.76	1161	198.50	42.00	0.51**	0.24*	1.00
Shantou	Weekly HFMD count	181.89	148.85	871.00	137.00	2.00	0.66**	0.42**	0.54**
	TEM (°C)	22.65	5.75	31.40	23.63	8.60	1.00	–	–
	RHU (%)	75.72	8.72	92.70	76.57	48.50	0.45**	1.00	–
	BDI	98.27	51.96	285.00	78.00	0.00	0.44**	0.21*	1.00

Note: SD: standard deviation, TEM: average temperature, RHU: relative humidity, BDI: Baidu index.

\*  $P < 0.05$ .

\*\*  $P < 0.01$ .

In Shenzhen, the order of the performance of the three models in [Table 4](#) was: [Model 1](#) (Deviance explained: 79.1% and RMSE: 370.598); [Model 2](#) (Deviance explained: 62.1% and RMSE: 515.808); [Model 3](#) (Deviance explained: 47.8% and RMSE: 563.906). In Shantou, the order of the performance of the three models in [Table 2](#) was: [Model 1](#) (Deviance explained: 71.6% and RMSE: 97.477); [Model 2](#) (Deviance explained: 63.1% and RMSE: 106.392); [Model 3](#) (Deviance explained: 25.2% and RMSE: 181.183). Having considered all of the values of the three models, it could be concluded that [Model 1](#) demonstrated the best estimation result for both cities.

The predictions of HFMD outbreaks that occurred from the 42nd week to the 52nd week of 2015 for three models are shown in [Fig. 5](#). As seen in [Table 4](#), [Model 1](#) is the better predictor of HFMD cases (RMSE: 199.265 and adjusted R-square: 0.853 for SZ; RMSE: 59.686 and adjusted R-square: 0.771 for ST) outperforming [Model 2](#) (RMSE: 311.564 and adjusted R-square: 0.612 for SZ; RMSE: 79.486 and adjusted R-square: 0.659 for ST) and [Model 3](#) (RMSE: 246.271 and adjusted R-square: 0.813 for SZ; RMSE: 88.120 and adjusted R-square: 0.228 for ST) in both cities. Interestingly, for Shenzhen city, [Model 3](#) appeared a better predictor than [Model 2](#).

#### 4. Discussion and conclusions

HFMD is considered as a major public health concern in China, seriously threatening the health of children under the age of 5 years. Effective prevention and treatment measures for HFMD are required, therefore establishing a timely and accurate predictive model warranted. In this study, a predictive model of HFMD in two cities both using meteorological factors and internet search engine query data was developed aiming to improve the traditional monitoring methods. Our major findings are as follows: (1) A multi-centered study to use both meteorological data and search index data to predict HFMD suggested that search data can improve the performance of the prediction models. (2) The prediction accuracy of the model with high correlation coefficient between BDI and HFMD can be improved remarkably. (3) The relationships between temperature or relative humidity and HFMD were approximately linear in both cities, while the relationship between search engine data and HFMD was non-linear. (4) Comparing

**Table 2**  
Test of homogeneity of the odds ratio.

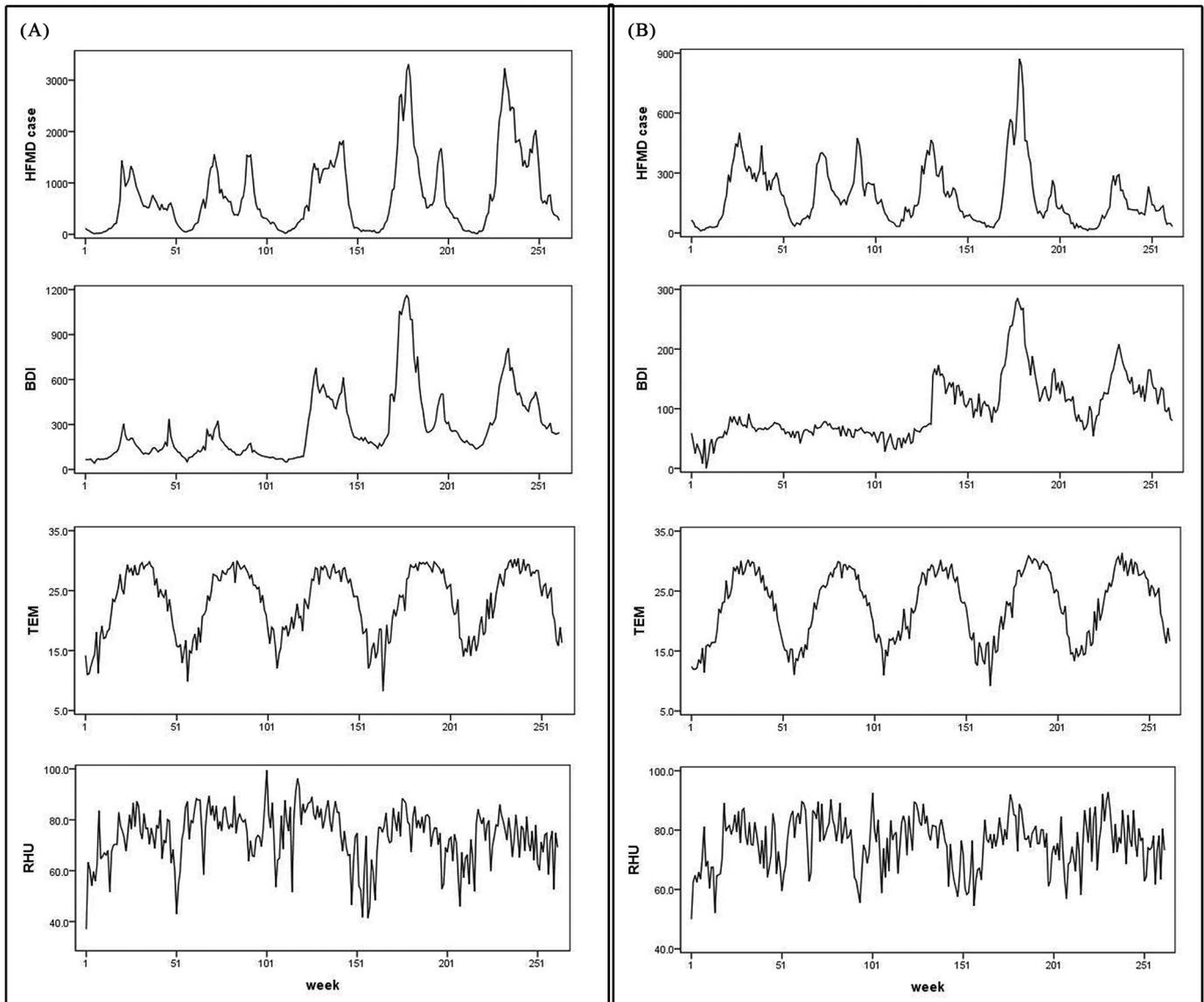
Factor	Test of homogeneity of the odds ratio			Sites	OR	95% confidence interval		P value
	Chi-squared	df	P value			Lower	Upper	
TEM	1.609	1	0.205	All	16.634	9.937	27.844	0.000
RHU	0.004	1	0.952	All	2.852	1.961	4.146	0.000
BDI	70.762	1	0.000	SZ	29.278	14.656	58.487	0.000
				ST	0.871	0.527	1.438	0.591

Note: TEM: average temperature, RHU: relative humidity, BDI: Baidu index, OR: odds ratio, SZ: Shenzhen, ST: Shantou.

with the lag effect of meteorological factors on HFMD, BDI data is more real-timely.

With the popularity of the Internet and the advancement of information technology in China, Internet-based surveillance systems have been widely used to monitor the prevalence of infectious diseases in recent years. For example, Internet search query data are used to monitor trends in diseases such as influenza, dengue fever, and HIV/AIDS ([Chiu et al., 2017](#); [Liu et al., 2016](#); [Li et al., 2017](#); [Yuan et al., 2013](#)). In our study, the results of the two cities of Shenzhen and Shantou, suggest that BDI is an advantageous supplementary measure to monitor HFMD cases which is consistent with recent studies. Du and colleagues, found that a model including BDI achieved the best goodness-of-fit, with an AIC value of  $-345.332$ , and delivered the most accurate prediction in terms of the mean absolute percentage error (MAPE), of 101.745%, for Guangdong Province ([Du et al., 2017](#)). Another recent study confirmed that a model which integrated search data, meteorological factors and historical cases demonstrated the most stable estimation result in Guangzhou city ([Huang and Wang, 2018](#)). By comparing the performance of the prediction models in our study, it was found that [Model 1](#) (adding both BDI and meteorological factors as variables) was the best predictor of HFMD. This conclusion not only supported the results from the above studies ([Du et al., 2017](#); [Huang and Wang, 2018](#)), but also improved the results of the multi-center study. To the best of our knowledge, this current study is the first multi-centered study to use both BDI and meteorological factors to establish a model for predicting HFMD. This is one of the important differences in our analysis compared to previous studies.

Our study may also be the first to compare the performance of HFMD models with or without BDI in two proximal cities, namely, Shenzhen and Shantou. The value of adjusted R-square increased from 0.612 in [Model 2](#) (adding solely meteorological factors as variables) to 0.853 in [Model 1](#) in Shenzhen, and from 0.659 to 0.771 in Shantou. It may also be interpreted that using BDI does not significantly improve the accuracy of model prediction for Shantou. Meanwhile, by comparing the performance of [Model 2](#) and [Model 3](#) (adding only BDI as variable), the latter confirms itself as a stronger predictor than the former for Shenzhen, while it was the opposite case in Shantou. This phenomenon may reflect Internet penetration and may indirectly influence the degree of effective use of search engines. Previous studies based on Internet search engines as predictors have focused on areas with higher Internet penetration ([Chiu et al., 2017](#); [Li et al., 2017](#)). Therefore we posit that the application of BDI in areas with a low Internet penetration, and possible resultant superficial or low quality Internet search or low information technology literacy, will not necessarily significantly increase forecast accuracy. At the same time, after controlling the effect of city, we found that the effect of meteorological factors (TEM and RHU) on HFMD cases was homogeneous, while the effect of BDI on HFMD cases was heterogeneous in two cities. In Shenzhen, the effect of BDI on HFMD cases was significant, but it was not significant in Shantou. So, these findings also supported our one of the conclusions which there



**Fig. 1.** The time series of weekly HFMD cases, BDI, TEM, and RHU from 2011 to 2015 in Shenzhen (A) and Shantou (B).

was different effect on improving the prediction accuracy of model with different correlation coefficient between BDI and HFMD cases.

We observed an approximate positively linear relationship between average temperature and HFMD cases. Similar findings have been reported in Guangzhou and Shenzhen in Guangdong Province, China (Huang et al., 2013; Zhang et al., 2016a, b, c). Yet, studies have found that the risk of temperature with HFMD was non-linear, for example in Japan (Onozuka and Hashizume, 2011) and in Zhengzhou, China (Feng, et al., 2014). Regarding relative humidity, the result of our study demonstrated that the relationship with HFMD cases approximated a positively linear association which is consistent with the findings of previous reports in Guangzhou (Huang et al., 2013; Huang and Wang, 2018). However, findings differ from other studies, suggesting interesting variables. For instance, Zhang and colleagues identified two thresholds of relative humidity, the lower threshold at 45% and the higher threshold at 85% (Zhang et al., 2016a, b, c); Yang and colleagues argued the relative risk increased rapidly above a threshold of relative humidity of 84% (Yang et al., 2017). Possible explanations for the inconsistency between meteorological factors and HFMD may be the differing weather and demographic profiles of these different geographical areas. Our results reflect those similar latitude and longitude coordination, and differ greatly from those dissimilar positioning.

Thus, the stratification by population is no significant between HFMD and meteorological factors in our study. In addition, the threshold of the relationship between meteorological factors and HFMD often occurs at extreme weather conditions. Further explanation of this discrepancy could be the differing scales employed in the studies. Monthly or weekly scales may not examine the relationship between meteorological factors and HFMD in the peak period incidence or in the extreme weather period as flexibly as a daily scale (Yang et al., 2017).

In addition, a positive relationship between BDI and HFMD cases was found in Shenzhen in this study, illustrating that the Internet-based search engine may be a useful predictor of HFMD incidence, which corresponds to studies that investigated the relationship between BDI and HFMD (Du et al., 2017) and studies investigating Google Dengue Trends and Dengue cases (Althouse et al., 2011; Chan et al., 2011). However, it seems that BDI has a non-linear relationship with HFMD cases in Shantou. Specifically, BDI between 50 and 150 in Shantou was negatively correlated with HFMD cases yet BDI positively correlated above 150. Table 1 shows that in Shenzhen, significant correlations between HFMD cases and BDI ( $r = 0.78$ ,  $P < 0.001$ ) were much higher than those ( $r = 0.54$ ,  $P < 0.001$ ) of Shantou. The quality of search engine data in the two cities may contribute to this phenomenon (Huang et al., 2016). Our findings support the above conclusion. We found

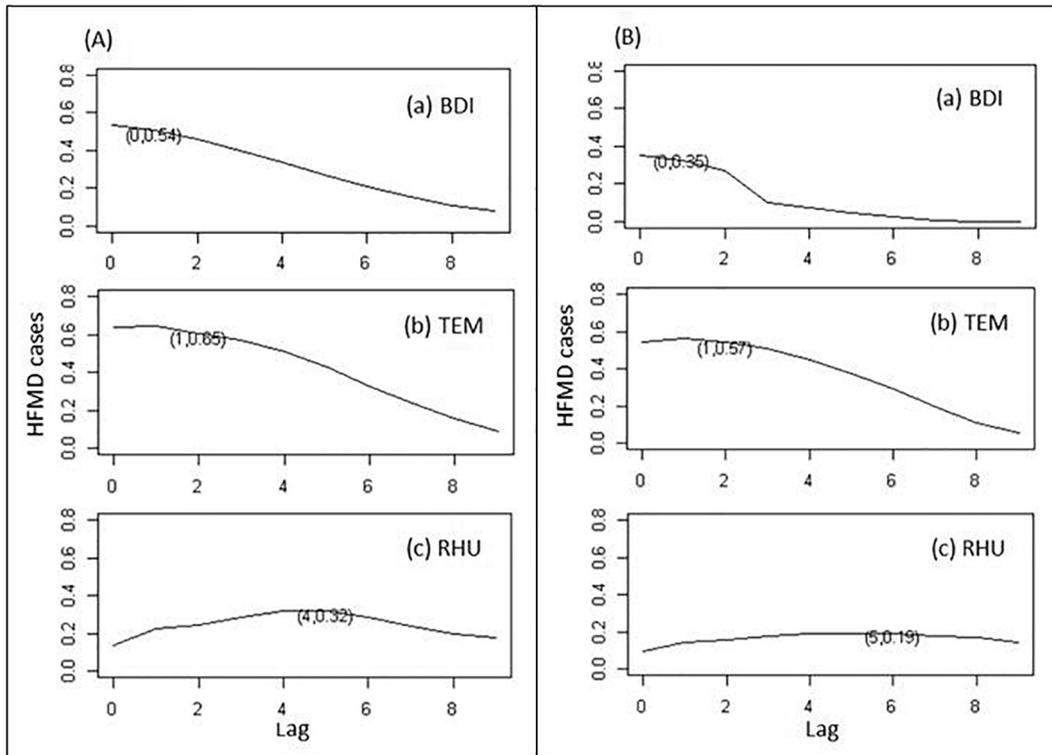


Fig. 2. Lag-specific predictors-RR curves at various lags. (A) Shenzhen. (B) Shantou.

that the model with high correlation coefficient of BDI and HFMD had more accurate prediction ability than those with a weaker correlation coefficient.

HFMD activities are better explained when meteorological variables are used with one week's lag, which is due to the incubation period for enteroviruses and the potential delay in parental awareness of and

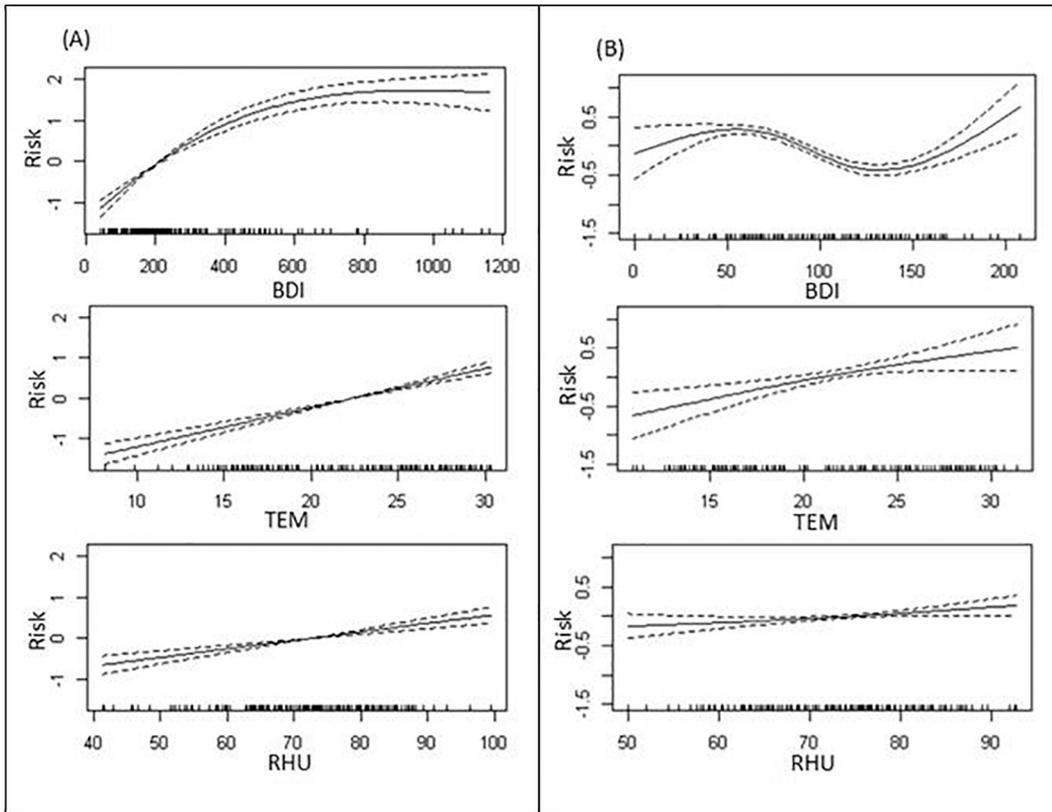


Fig. 3. The relationship between HFMD cases with variables at each lag. (A) Shenzhen, TEM at one week's lag, RHU at two weeks' lag and BDI at the current week. (B) Shantou, TEM at one week's lag, RHU at two weeks' lag and BDI at the current week. Note: Solid lines represent the relative risks of HFMD and dotted lines represent the upper and lower limits of 95% confidence intervals.

**Table 3**  
Test spatial stratified heterogeneity for the relationship between HFMD and meteorological factors by city.

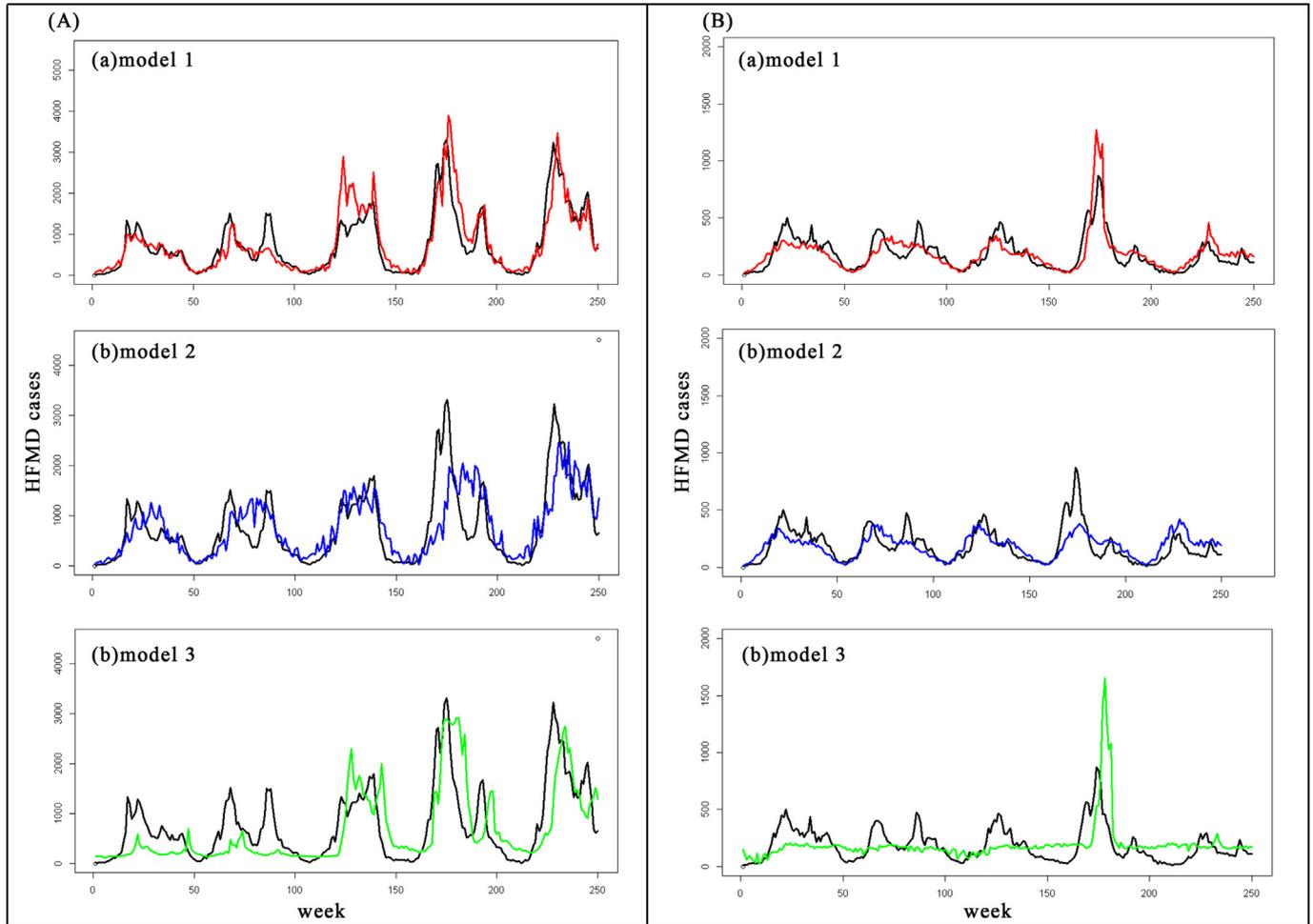
Indicator	Correlation coefficient		q-Statistic	P value
	Shenzhen	Shantou		
TEM	0.68	0.66	0.03	0.83
RHU	0.45	0.42	−0.10	0.86

Note: TEM: average temperature, RHU: relative humidity.

response to clinical symptoms of children (Wang et al., 2017a, b; Hii et al., 2011). According to our understanding, Internet search data is freely available in near real-time. Parents tend to obtain relevant information from the Internet before the HFMD affected children are sent to hospital or during the period of actual hospitalization (Huang and Wang, 2018; Huang et al., 2016). Thus, the timeliness of the prediction effect of BDI on HFMD is an important advantage. However, using BDI does not seek

to replace actual estimates of HFMD cases; rather, it is supplementary to traditional disease surveillance by providing a means of progression towards establishing reliable early warning models, and by allowing for more efficient and rapid control of infectious diseases (Lazer et al., 2014).

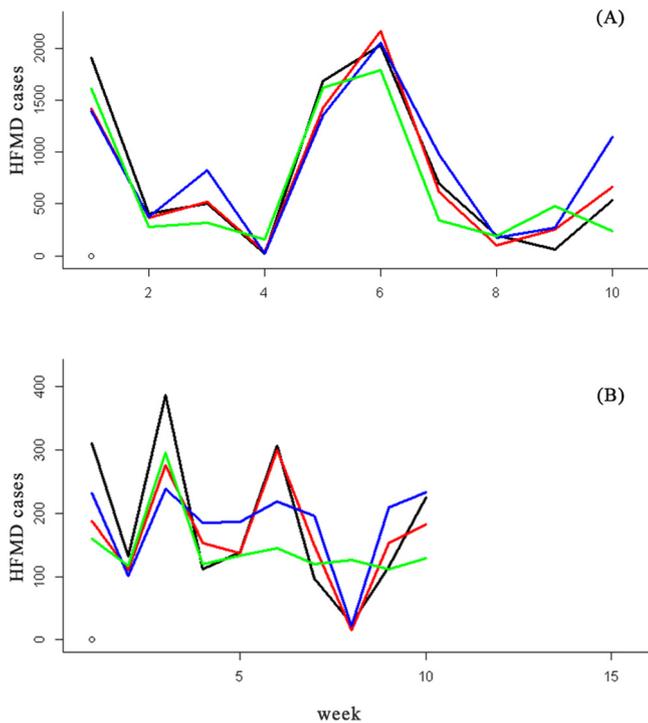
Accurate predicting of the epidemic of HFMD has some benefits. There is the potential for health promotional professionals to use HFMD epidemic trends to design timely awareness campaigns and public health interventions to maximize reach and effectiveness. Involvement of parents, schools and communities in proactively implementing infection prevention and control measures to reduce the spread and impact of viruses may positively impact health spending and clinical outcome. Such measures might include hand and respiratory hygiene, environmental hygiene and avoidance of unnecessary contact with infected persons. Much rigor has contributed to this research; however some limitations are inevitable and require mentioning. Firstly, the current conclusions are based on weekly data, whereas more accurate findings would have emerged if daily data were used



**Fig. 4.** Weekly observed and fitted HFMD cases using three different models from the 1st week of 2011 to the 42nd week of 2015. (A) Shenzhen. (B) Shantou. Note: The black line represents the reported HFMD cases, the red, blue and green lines represent the cases fitted by Model 1, Model 2 and Model 3 respectively.

**Table 4**  
Comparison of the performance of the three models.

Sites	Model	Model description	Deviance explained (%)	RMSE	AIC
Shenzhen	Model 1	TEM + RHU + BDI	79.1	370.598	3530.228
	Model 2	TEM + RHU	62.1	515.808	3676.422
	Model 3	BDI	47.8	563.906	3760.721
Shantou	Model 1	TEM + RHU + BDI	71.6	97.477	2851.441
	Model 2	TEM + RHU	63.1	106.392	2914.410
	Model 3	BDI	25.2	181.183	3099.783



**Fig. 5.** Out-of-sample predictions of HFMD cases using the three models from the 42nd week to the 52nd week of 2015. (A) Shenzhen. (B) Shantou. Note: The black line represents the reported HFMD cases, while the red, blue and green lines represent the cases estimated by Model 1, Model 2 and Model 3 respectively.

(Cheng et al., 2018; Yin et al., 2016). Secondly, it should be noted that HFMD is also affected by many other factors, such as socio-economic and demographic factors as well as population density and population mobility (Huang et al., 2015). More predictive factors should be considered for inclusion in future models.

In conclusion, our study suggests that using both Internet search engine and meteorological data can improve the prediction of HFMD incidence, especially with improved internet penetration and information technology literacy.

### Competing interests

None.

### Funding sources

This study was supported by the National Natural Science Foundation of China (No. 81172776), the Guangdong Science and Technology Planning Project (No. 2016A020216026), and the Guangdong Government under the Top-tier University Development Scheme for Research and Control of Infectious Diseases (No. 2016040).

### Acknowledgements

We thank Zilu Zhang and Qjulin Wu (Department of mathematics, Shantou University) for giving suggestions and comments for this manuscript.

### References

Althouse, B.M., Ng, Y.Y., Cummings, D.A., 2011. Prediction of dengue incidence using search query surveillance. *PLoS Negl. Trop. Dis.* 5, e1258. <https://doi.org/10.1371/journal.pntd.0001258>. 21829744.

Anderson-Cook, C.M., 2007. Generalized additive models: an introduction with R, by Simon N. Wood. *Technometrics* 49 (3), 360–361.

Aswathyraj, S., Arunkumar, G., Alidjinou, E.K., et al., 2016. Hand, foot and mouth disease (HFMD): emerging epidemiology and the need for a vaccine strategy. *Med. Microbiol. Immunol.* 205, 397. <https://doi.org/10.1007/s00430-016-0465-y>.

Cardosa, M.J., Perera, D., Brown, B.A., Cheon, D., Chan, H.M., Chan, K.P., et al., 2003. Molecular epidemiology of human enterovirus 71 strains and recent outbreaks in the Asia-Pacific region: comparative analysis of the VP1 and VP4 genes. *Emerg. Infect. Dis.* 9, 462–468.

Chan, E.H., Sahai, V., Conrad, C., Brownstein, J.S., 2011. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl. Trop. Dis.* 5 (5), e1206.

Cheng, Q., Bai, L., Zhang, Y., Zhang, H., Wang, S., Xie, M., et al., 2018. Ambient temperature, humidity and hand, foot, and mouth disease: a systematic review and meta-analysis. *Sci. Total Environ.* 625, 828–836.

China Economic Net, 2016. The Scale of Netizens in Shenzhen Reached 8.97 Million and the Internet Penetration Reached 82.3%. [http://district.ce.cn/zg/201601/21/t20160121\\_8446633.shtml](http://district.ce.cn/zg/201601/21/t20160121_8446633.shtml).

China Internet Network Information Center, 2014. 2013 Chinese Search Engine Market Research Report. <http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/sbg/201401/P020140127366465515288.pdf>.

China Ministry of Health, 2010. Guideline for Diagnosis and Treatment of HFMD (2010 Edition) of Ministry of Health, P.R.C. <http://www.moh.gov.cn/publicfiles/business/htmlfiles/mohyzs/s3586/201004/46884.htm>.

China Report Web, 2017a. Statistics of the Scale and Penetration of Netizens by Provinces in China in 2015. <http://data.chinabaogao.com/it/2017/06232V2032017.html>.

China Report Web, 2017b. Statistics of the Scale and Penetration of Netizens Nationwide from 2011 to 2015. <http://data.chinabaogao.com/it/2017/06222V1632017.html>.

Chinese Center for Disease Control and Prevention (China CDC), 2014. National Incidence and Death Cases of Notifiable Class A or Class B Infectious Disease (2008, 2009, 2010, 2011, 2012). <http://www.chinacdc.cn>.

Chiu, A., Lin, Q., He, D., 2017. News trends and web search query of HIV/AIDS in Hong Kong. *PLoS One* 12 (9), e0185004. <https://doi.org/10.1371/journal.pone.0185004>.

Deng, T., Huang, Y., Yu, S., Gu, J., Huang, C., Xiao, G., Hao, Y., et al., 2013. Spatial-temporal clusters and risk factors of hand, foot, and mouth disease at the district level in Guangdong Province China. *PLoS One* 8 (2), e56943.

Du, Z., Xu, L., Zhang, W., Zhang, D., Yu, S., Hao, Y., 2017. Predicting the hand, foot, and mouth disease incidence using search engine query data and climate variables: an ecological study in Guangdong, China. *BMJ Open* 7, e016263. <https://doi.org/10.1136/bmjopen-2017-016263>.

Feng, H., Duan, G., Zhang, R., Zhang, W., 2014. Time series analysis of Hand-Foot-Mouth disease hospitalization in Zhengzhou: establishment of forecasting models using climate variables as predictors. *PLoS One* 9, e87916. <https://doi.org/10.1371/journal.pone.0087916>.

Hastie, T., Tibshirani, R., 1995. Generalized additive models for medical research. *Stat. Methods Med. Res.* 4 (3), 187–196.

Hii, Y.L., Rocklöv, J., Ng, N., 2011. Short term effects of weather on hand, foot and mouth disease. *PLoS One* 6, e16796. <https://doi.org/10.1371/journal.pone.0016796>.

Huang, D., Wang, J., 2018. Monitoring hand, foot and mouth disease by combining search engine query data and meteorological factors. *Sci. Total Environ.* 612, 1293–1299.

Huang, Y., Deng, T., Yu, S., Gu, J., Huang, C., Xiao, G., Hao, Y., 2013. Effect of meteorological variables on the incidence of hand, foot, and mouth disease in children: a time-series analysis in Guangzhou, China. *BMC Infect. Dis.* 13, 134. <https://doi.org/10.1186/1471-2334-13-134>.

Huang, J., Wang, J., Li, Z., Wang, Y., Lai, S., Yang, W., 2015. Visualized exploratory spatio-temporal analysis of Hand-Foot-Mouth disease in Southern China. *PLoS One* 10 (11), e0143411. <https://doi.org/10.1371/journal.pone.0143411>.

Huang, D., Wang, J., Huang, J., Sui, D.Z., Zhang, H., Hu, M., et al., 2016. Towards identifying and reducing the bias of disease information extracted from search engine data. *PLoS Comput. Biol.* 12 (6), e1004876. <https://doi.org/10.1371/journal.pcbi.1004876>.

Kumar, V.S., Budur, S.V., Odappa, G.H., Bankolli, S.Y., Rao, A.P., 2015. Clinical profile of hand, foot, and mouth disease and its associated complications among children in Shimoga City, southern Karnataka: a hospital-based study. *Indian J. Public Health* 59, 141–144.

Lazer, D., Kennedy, R., King, G., Vespignani, A., 2014. The parable of Google flu: traps in big data analysis. *Science* 343, 1203–1205. <https://doi.org/10.1126/science.1248506>. 24626916.

Li, L., He, Y., Yang, H., Zhu, J., Xu, X., Dong, J., et al., 2005. Genetic characteristics of human enterovirus 71 and coxsackievirus A16 circulating from 1999 to 2004 in Shenzhen, People's Republic of China. *J. Clin. Microbiol.* 43, 3835–3839.

Li, L., Yin, H., An, Z., Feng, Z., 2015. Considerations for developing an immunization strategy with enterovirus 71 vaccine. *Vaccine* 33, 1107–1112.

Li, Z., Liu, T., Zhu, G., Lin, H., Zhang, Y., He, J., et al., 2017. Dengue Baidu Search Index data can improve the prediction of local dengue epidemic: a case study in Guangzhou, China. *PLoS Negl. Trop. Dis.* 11, e0005354. <https://doi.org/10.1371/journal.pntd.0005354>.

Liu, C., Chow, Y., Chong, P., Klein, M., 2014. Prospect and challenges for the development of multivalent vaccines against hand, foot and mouth diseases. *Vaccine* 32, 6177–6182.

Liu, K., Wang, T., Yang, Z., Huang, X., Milinovich, G., Lu, Y., et al., 2016. Using Baidu Search Index to predict dengue outbreak in China. *Sci. Rep.* 6, 38040. <https://doi.org/10.1038/srep38040>.

Madoff, L., Fisman, D., Kass-Hout, T., 2011. A new approach to monitoring dengue activity. *PLoS Negl. Trop. Dis.* 5, e1215. <https://doi.org/10.1371/journal.pntd.0001215>.

Mandary, M., Poh, C., 2018. Changes in the EV-A71 genome through recombination and spontaneous mutations: impact on virulence. *Viruses* 10, 320.

Mao, Q., Wang, Y.L., Bian, L., Xu, M.L., Liang, Z., 2016. EV-A71 vaccine licensure: a first step for multivalent enterovirus vaccine to control HFMD and other severe diseases. *Emerg. Microbes Infect.* 5, e75. <https://doi.org/10.1038/em.2016.73>.

- Onozuka, D., Hashizume, M., 2011. The influence of temperature and humidity on the incidence of hand, foot, and mouth disease in Japan. *Sci. Total Environ.* 410–411, 119–125.
- Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen–Geiger climate classification. *Hydrol. Earth Syst. Sci.* 11, 1633–1644. <https://doi.org/10.5194/hess-11-1633-2007>.
- Pinn, T., Chit, L., 2017. Impact of genetic changes, pathogenicity and antigenicity on enterovirus-A71 vaccine development. *Virology* 506, 121–129. <https://doi.org/10.1016/j.virol.2017.03.017>.
- R Development Core Team, 2017. The R Project for Statistical Computing. <http://www.rproject.org/>.
- Sarma, N., 2013. Hand, foot, and mouth disease: current scenario and Indian perspective. *Indian J. Dermatol. Venereol. Leprol.* 79, 165–175 Available from: <http://www.ijdv.com/text.asp?2013/79/2/165/107631>.
- Wang, Y., Feng, Z., Yang, Y., Steve, S., Gao, Y., Ira, I.M., et al., 2011. Hand, foot, and mouth disease in China: patterns of spread and transmissibility. *Epidemiology* 22, 781–792.
- Wang, J.F., Zhang, T.L., Fu, B.J., 2016. A measure of spatial stratified heterogeneity. *Ecol. Indic.* 67, 250–256.
- Wang, P., Zhao, H., You, F., Zhou, H., Goggins, W.B., et al., 2017a. Seasonal modeling of hand, foot, and mouth disease as a function of meteorological variations in Chongqing, China. *Int. J. Biometeorol.* 61, 1411. <https://doi.org/10.1007/s00484-017-1318-0>.
- Wang, W., Song, J., Wang, J., Li, Y., Deng, H., Li, M., et al., 2017b. Cost-effectiveness of a national enterovirus 71 vaccination program in China. *PLoS Negl. Trop. Dis.* 11 (9), e0005899. <https://doi.org/10.1371/journal.pntd.0005899>.
- Wong, S., Yip, C., Lau, S., Yuen, K., 2010. Human enterovirus 71 and hand, foot and mouth disease. *Epidemiol. Infect.* 138, 1071–1089.
- Woo, H., Cho, Y., Shim, E., Lee, J.-K., Lee, C.-G., Kim, S.H., 2016. Estimating influenza outbreaks using both search engine query data and social media data in South Korea. *J. Med. Internet Res.* 18, e177. <https://doi.org/10.2196/jmir.4955>.
- Xiao, Q., Liu, H., Feldman, M., 2017. Tracking and predicting hand, foot, and mouth disease (HFMD) epidemics in China by Baidu queries. *Epidemiol. Infect.* 145, 1699–1707. <https://doi.org/10.1017/S0950268817000231>.
- Xing, W., Liao, Q., Viboud, C., Zhang, J., Sun, J., Wu, J., et al., 2014. Hand, foot and mouth disease in China, 2008–12: an epidemiological study. *Lancet Infect. Dis.* 14, 308–318.
- Yang, H., Wu, J., Cheng, J., Wang, X., Wen, L., Li, K., et al., 2017. Is high relative humidity associated with childhood hand, foot, and mouth disease in rural and urban areas? *Public Health* 142, 201–207.
- Yin, F., Zhang, T., Liu, L., Lv, Q., Li, X., 2016. The association between ambient temperature and childhood hand, foot, and mouth disease in Chengdu, China: a distributed lag non-linear analysis. *Sci. Rep.* 6, 27305.
- Yuan, Q., Nsoesie, E.O., Lv, B., Peng, G., Chunara, R., Brownstein, J.S., 2013. Monitoring influenza epidemics in China with search query from Baidu. *PLoS One* 8, e64323.
- Zhang, Y., Zhu, Z., Yang, W., Ren, J., Tan, X.J., Wang, Y., et al., 2010. An emerging recombinant human enterovirus 71 responsible for the 2008 outbreak of Hand Foot and Mouth Disease in Fuyang city of China. *Virology* 7, 94.
- Zhang, W., Du, Z., Zhang, D., Yu, S., Hao, Y., 2016a. Quantifying the adverse effect of excessive heat on children: an elevated risk of hand, foot and mouth disease in hot days. *Sci. Total Environ.* 541, 194–199.
- Zhang, W., Du, Z., Zhang, D., Yu, S., Hao, Y., 2016b. Boosted regression tree model based assessment of the impacts of meteorological drivers of hand, foot and mouth disease in Guangdong, China. *Sci. Total Environ.* 553, 366–371.
- Zhang, Z., Xie, X., Chen, X., Li, Y., Lua, Y., Mei, S., et al., 2016c. Short-term effects of meteorological factors on hand, foot and mouth disease among children in Shenzhen, China: non-linearity, threshold and interaction. *Sci. Total Environ.* 539, 576–582.
- Zhu, Z., Zhu, S., Guo, X., Wang, J., Wang, D., Yan, D., et al., 2010. Retrospective seroepidemiology indicated that human enterovirus 71 and coxsackievirus A16 circulated widely in central and southern China before large-scale outbreaks from 2008. *Virology* 7, 300. <https://doi.org/10.1186/1743-422X-7-300>.
- Zhu, F., Xu, W., Xia, J., Liang, Z., Liu, Y., Zhang, X., et al., 2014. Efficacy, safety, and immunogenicity of an enterovirus 71 vaccine in China. *N. Engl. J. Med.* 370 (9), 818–828. <https://doi.org/10.1056/NEJMoa1304923>.