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\textsuperscript{a,b}, Xiaojian Liu\textsuperscript{c}, Yongsheng Wu\textsuperscript{c}, Guangxing Xu\textsuperscript{d}, Xubin Zhang\textsuperscript{d}, Shujiang Mei\textsuperscript{c}, Zhen Zhang\textsuperscript{c}, Michael O'Meara\textsuperscript{e}, Mary Clare O'Gara\textsuperscript{f}, Xuerui Tan\textsuperscript{g}, Liping Li\textsuperscript{a,*}

\textsuperscript{a} Injury Prevention Research Center, Shantou University Medical College, Shantou, Guangdong 515041, China
\textsuperscript{b} Department of Community monitoring, The First Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China
\textsuperscript{c} Shenzhen Center for Disease Control and Prevention, Shenzhen, China
\textsuperscript{d} Shantou Center for Disease Control and Prevention, Shantou, China
\textsuperscript{e} Department of Information Technology, Shantou University Medical College, Shantou, Guangdong 515041, China
\textsuperscript{f} Department of Nursing, Shantou University Medical College, Shantou, Guangdong 515041, China
\textsuperscript{g} 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

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.

Keywords: HFMD Prediction GAM Meteorological data Search index data
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 notified 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 (Aswathraj 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 (Aswathraj 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 km². 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 km². 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 (°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:

\[
Y_t = \beta_0 + s(\text{week}, \text{df}) + s(\text{TEM}_{t-1}, \text{df}) + s(\text{RHU}_{t-2}, \text{df}) + s(\text{BDI}_{t-n}, \text{df}) \tag{Model 1}
\]

\[
Y_t = \beta_0 + s(\text{week}, \text{df}) + s(\text{TEM}_{t-1}, \text{df}) + s(\text{RHU}_{t-1}, \text{df}) \tag{Model 2}
\]

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

In these equations, \( Y_t \) represents the predicted mean of HFMD cases during the week \( t \); \( s(\text{week}, \text{df}) \) denotes the cubic spline of a week with corresponding degrees of freedom (df) to control seasonality; \( s(\text{TEM}_{t-1}, \text{df}) \) and \( s(\text{RHU}_{t-1}, \text{df}) \) 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; \( s(\text{BDI}_{t-n}, \text{df}) \) 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 were 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, deviation 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 (\( \chi^2 = 70.762 \) and \( \chi^2 = 69.833, P < 0.000 \)). 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 times series of weekly HFMD cases, TEM, RHU and BDI in Shenzhen and Shantou respectively. The HFMD cases, BDI and meteorological variables all presented observed seasonal patterns. It can be seen that the curve of HFMD cases was double-peak every year from 2011 to 2015 inclusive. The BDI curve of Shenzhen double-peak 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.
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: 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 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 with the lag effect of meteorological factors on HFMD, BDI data is more real-time.

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 −345332, 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

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

<table>
<thead>
<tr>
<th>Sites</th>
<th>Variables</th>
<th>Mean</th>
<th>SD</th>
<th>Max</th>
<th>Median</th>
<th>Min</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TEM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RHU</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BDI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shenzhen</td>
<td>Weekly HFMD count</td>
<td>752.89</td>
<td>726.53</td>
<td>3300.00</td>
<td>552.00</td>
<td>10.00</td>
<td>0.68*</td>
</tr>
<tr>
<td></td>
<td>TEM (°C)</td>
<td>23.18</td>
<td>5.42</td>
<td>30.33</td>
<td>24.29</td>
<td>8.22</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>RHU (%)</td>
<td>73.72</td>
<td>10.55</td>
<td>99.43</td>
<td>75.57</td>
<td>37.00</td>
<td>0.44*</td>
</tr>
<tr>
<td></td>
<td>BDI</td>
<td>276.84</td>
<td>225.76</td>
<td>1161</td>
<td>198.50</td>
<td>42.00</td>
<td>0.51**</td>
</tr>
<tr>
<td>Shantou</td>
<td>Weekly HFMD count</td>
<td>181.89</td>
<td>148.85</td>
<td>871.00</td>
<td>137.00</td>
<td>2.00</td>
<td>0.66**</td>
</tr>
<tr>
<td></td>
<td>TEM (°C)</td>
<td>22.65</td>
<td>5.75</td>
<td>31.40</td>
<td>23.63</td>
<td>8.60</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>RHU (%)</td>
<td>75.72</td>
<td>8.72</td>
<td>92.70</td>
<td>76.57</td>
<td>48.50</td>
<td>0.43**</td>
</tr>
<tr>
<td></td>
<td>BDI</td>
<td>98.27</td>
<td>51.96</td>
<td>285.00</td>
<td>78.00</td>
<td>0.00</td>
<td>0.44*</td>
</tr>
</tbody>
</table>

Note: SD: standard deviation, TEM: average temperature, RHU: relative humidity, BDI: Baidu index.  
* P < 0.05.  
** P < 0.01.

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

<table>
<thead>
<tr>
<th>Factor</th>
<th>Test of homogeneity of the odds ratio</th>
<th>Sites</th>
<th>OR</th>
<th>95% confidence interval</th>
<th>P value</th>
<th>df</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chi-squared</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEM</td>
<td>1.609</td>
<td>All</td>
<td>2.05</td>
<td>16.634</td>
<td>9.937</td>
<td>27.844</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>RHU</td>
<td>0.004</td>
<td>All</td>
<td>0.55</td>
<td>2.852</td>
<td>1.961</td>
<td>4.146</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>BDI</td>
<td>70.762</td>
<td>All</td>
<td>0.00</td>
<td>29.278</td>
<td>14.656</td>
<td>58.487</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ST</td>
<td>0.57</td>
<td>0.871</td>
<td>1.438</td>
<td>0.591</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: TEM: average temperature, RHU: relative humidity, BDI: Baidu index, OR: odds ratio, SZ: Shenzhen, ST: Shantou.
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. 1. The time series of weekly HFMD cases, BDI, TEM, and RHU from 2011 to 2015 in Shenzhen (A) and Shantou (B).](image-url)
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. 2. Lag-specific predictors-RR curves at various lags. (A) Shenzhen. (B) Shantou.](image)

![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.](image)
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

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

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Correlation coefficient</th>
<th>q-Statistic</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEM</td>
<td>0.68</td>
<td>0.66</td>
<td>0.03</td>
</tr>
<tr>
<td>RHU</td>
<td>0.45</td>
<td>0.42</td>
<td>-0.10</td>
</tr>
</tbody>
</table>

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

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.

<table>
<thead>
<tr>
<th>Sites</th>
<th>Model</th>
<th>Model description</th>
<th>Deviance explained (%)</th>
<th>RMSE</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shenzhen</td>
<td>Model 1</td>
<td>TEM + RHU + BDI</td>
<td>79.1</td>
<td>370.598</td>
<td>3530.228</td>
</tr>
<tr>
<td></td>
<td>Model 2</td>
<td>TEM + RHU</td>
<td>62.1</td>
<td>515.808</td>
<td>3676.422</td>
</tr>
<tr>
<td></td>
<td>Model 3</td>
<td>BDI</td>
<td>47.8</td>
<td>563.906</td>
<td>3760.721</td>
</tr>
<tr>
<td>Shantou</td>
<td>Model 1</td>
<td>TEM + RHU + BDI</td>
<td>71.6</td>
<td>97.477</td>
<td>2851.441</td>
</tr>
<tr>
<td></td>
<td>Model 2</td>
<td>TEM + RHU</td>
<td>63.1</td>
<td>106.392</td>
<td>2914.410</td>
</tr>
<tr>
<td></td>
<td>Model 3</td>
<td>BDI</td>
<td>25.2</td>
<td>181.183</td>
<td>3099.783</td>
</tr>
</tbody>
</table>
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. 20160640).

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

References


