Science of the Total Environment 553 (2016) 366–371

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Boosted regression tree model-based assessment of the impacts of meteorological drivers of hand, foot and mouth disease in Guangdong, China

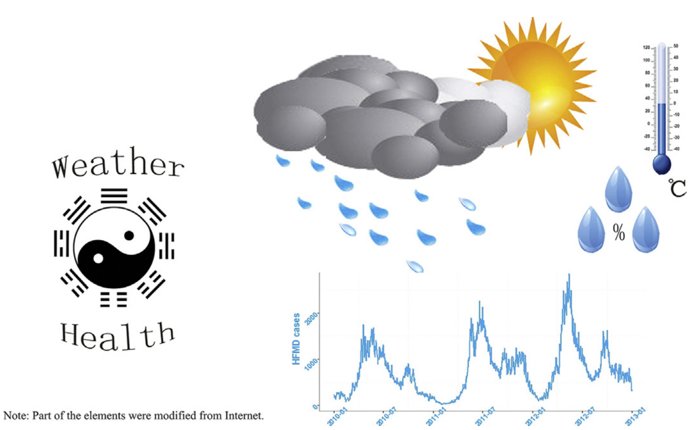
Wangjian Zhang [a], Zhicheng Du [a], Dingmei Zhang [a], Shicheng Yu [b], Yuantao Hao [a,*]

[a] Department of Medical Statistics and Epidemiology, Health Information Research Center, Guangdong Key Laboratory of Medicine, School of Public Health, Sun Yat-sen University, Guangzhou, 510080 Guangdong Province, China
[b] Chinese Center for Disease Control and Prevention, Beijing 102206, China

HIGHLIGHTS

- Variance of HFMD cases was explained most by meteorological factors about 1 week ago.
- The optimal lag at which the variance of HFMD cases was most explained was determined.
- The facilitating effects of meteorological factors were verified and quantified.
- Threshold points for each meteorological factor were identified.
- The contribution of each meteorological factor to the epidemic of HFMD was assessed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 6 November 2015
Received in revised form 2 February 2016
Accepted 3 February 2016
Available online xxxx

Editor: D. Barcelo

Keywords:
HFMD
Meteorological factors
Nonlinear
Facilitating effect

ABSTRACT

Background: Hand, foot and mouth disease (HFMD) is a common childhood infection and has become a major public health issue in China. Considerable research has focused on the role of meteorological factors in HFMD development. Nonlinear relationship, delayed effects and collinearity problems are key issues for achieving robust and accurate estimations in this kind of weather-health relationship explorations. The current study was designed to address these issues and assess the impact of meteorological factors on HFMD in Guangdong, China.
Methods: Case-based HFMD surveillance data and daily meteorological data collected between 2010 and 2012 was obtained from China CDC and the National Meteorological Information Center, respectively. After a preliminary variable selection, for each dataset boosted regression tree (BRT) models were applied to determine the optimal lag for meteorological factors at which the variance of HFMD cases was most explained, and to assess the impacts of these meteorological factors at the optimal lag.
Results: Variance of HFMD cases was explained most by meteorological factors about 1 week ago. Younger children and those from the Pearl-River Delta Region were more sensitive to weather changes. Temperature had the largest contribution to HFMD epidemics (28.99–71.93%), followed by precipitation (6.52–16.11%), humidity (3.92–17.66%), wind speed (3.84–11.37%) and sunshine (6.21–10.36%). Temperature between 10 °C and 25 °C,

* Corresponding author.
E-mail addresses: zhangwj35@mail2.sysu.edu.cn (W. Zhang), duzhch3@mail2.sysu.edu.cn (Z. Du), zhdingm@mail.sysu.edu.cn (D. Zhang), shicheng_yu@hotmail.com (S. Yu), haoyt@mail.sysu.edu.cn (Y. Hao).

http://dx.doi.org/10.1016/j.scitotenv.2016.02.023

as well as humidity between 70% and 90%, had a facilitating effect on the epidemic of HFMD. Sunshine duration above 9 h and wind speed below 2.5 m/s also contributed to an elevated risk of HFMD. The positive relationship between HFMD and precipitation reversed when the daily amount of rainfall exceeded 25 mm.

Conclusions: This study indicated significantly facilitating effects of five meteorological factors within some range on the epidemic of HFMD. Results from the current study were particularly important for developing early warning and response system on HFMD in the context of global climate change.

1. Introduction

Hand, foot and mouth disease (HFMD) is a common childhood infection and has become a major public health issue in China, affecting over two million children annually (China NHAF, 2014; Zhu et al., 2011). It is a particular concern in Guangdong Province, with an incidence higher than four times the national average and exceeding 30 cases per 10,000 in 2012 (Deng et al., 2013; Zhang et al., 2014). Because of the absence of HFMD-targeted vaccination or specific treatments, quantification of the driving effects of environmental agents is essential for the early warning and response system on HFMD (Xu et al., 2015). Recently, there has been an increasing interest to assess the impacts of meteorological factors on HFMD (Chen et al., 2014; Feng et al., 2014; Li et al., 2014; Liu et al., 2015; Urashima et al., 2003; Wei et al., 2015; Zhuang et al., 2014).

However, the assessment of the relationship between HFMD and meteorological factors is complicated by two major issues. First, meteorological factors tend to have nonlinear impacts on the HFMD burden. There may exist some threshold points on two sides of which the weather-HFMD relationship can be substantially different (Chen et al., 2014; Huang et al., 2013; Lin et al., 2013; Wu et al., 2014). Second, interaction is a common issue among meteorological predictors (Huang et al., 2014; Lin et al., 2009; Zhang et al., 2016). Results may be biased if the collinearity problem is not reasonably addressed (Wang et al., 2010). Predictor selection procedures including the traditional stepwise method can reduce the influence of collinearity but simultaneously sacrifice information on those removed predictors.

The current study was designed to address these issues with a novel modeling technique and assess the impact of meteorological factors on HFMD in Guangdong, China.

Boosted regression tree (BRT) model is a recently developed technique, combining the advances of the traditional regression models and the machine-learning methods (Tonkin et al., 2015). It accommodates complex linear and nonlinear responses to multiple categorical and continuous predictors while is relatively insensitive to collinearity problems (Elith et al., 2008; Main et al., 2015). The BRT model was particular suitable for this case, therefore, was the major statistical method of the current study.

Results were also stratified by areas and age groups to study the difference and consistence of the meteorological factor-HFMD relationship across subpopulations.

2. Materials and methods

2.1. Study settings

Guangdong is one of the biggest provinces in Southern China, with an area of 179,800 km^2 and a population of 104 million (from 2010 census data). It can be generally divided into two parts: the Pearl River Delta Region and the Non-Pearl River Delta Region. According to statistical yearbooks of Guangdong, the Pearl River Delta Region has a much higher level of social-economic development, accounting for 80% GDP of the whole province with <50% population (Province SBOG, 2012).

2.2. Data sources

HFMD surveillance data collected between 2010 and 2012 was obtained from China Center for Disease Control and Prevention (China CDC). Information including birth date, onset time and ZIP code for the current address was reported to the real-time surveillance system once a patient was diagnosed with HFMD according to the National Clinic Guide (Version 2010) (China THMO, 2009). According to our data, over 99.5% HFMD cases were children under 15 years old who therefore was the study population of the current research. Children were further classified into 4 groups (<1, 1–3, 3–6, and >6 years) since children of different age groups tended to differ in daily activities, as well as in susceptibility and forms of being cared (Huang et al., 2013).

Daily meteorological data was obtained from the National Meteorological Information Center (http://cdc.nmic.cn/). Originally, eleven meteorological factors including mean vapor, precipitation, maximum wind speed, mean air pressure, mean wind speed, mean temperature, mean humidity, minimum temperature, sunshine duration, maximum temperature and minimum humidity were considered for the analysis.

2.3. Statistical analyses

Correlation between the eleven meteorological factors was examined to evaluate the collinearity problems. While boosted regression tree models were robust towards collinear predictors, to aid interpretability, we only retained those previously reported in literatures when predictors were highly correlated ($r > 0.75$) (Tonkin et al., 2015).

Boosted regression tree (BRT) model with Poisson distribution was used to explore the impacts of climate parameters on HFMD. BRT model combines algorithms of regression trees that use recursive binary splits to eliminate interactions among the predictors and boosting that built a large ensemble of small regression trees to display the nonlinear relationship between the response and its predictors as well as improve predictive performance (Ayanu et al., 2015; Tonkin et al., 2015). For this boosting method, 10-fold cross validation was used to select the best model - determining the optimal number of trees that should be included in the model to achieve the best predictive performance. The gbm.step procedure in *dismo* package in R (3.1.2) takes a stepwise approach to model selection (Elith and Leathwick, 2015).

BRT models require the specification of three parameters: the learning rate, tree complex and bag fraction. The learning rate shrinks the contribution of each tree as it is added to the growing model. The tree complex determines the maximum order of interaction in each tree, and bag fraction specifying the proportion of the training set that is used for model building in each step. Three parameters in the present study were specified at 0.005, 5, and 0.5, respectively, as recommended (Elith et al., 2008).

Since the delayed effects of climate parameters on HFMD are well documented (Wang et al., 2013; Zhang et al., 2015), lag selection is essential for the assessment of weather-HFMD relationships. The optimal lag of climate predictors was defined as the lag at which the variance of HFMD case numbers was most explained in the cross validation. The maximum lag was set at 14 days which was previously reported to be the upper bound of the incubation of HFMD (Ministry of Health, 2014). Then the relationship between the climate parameters at the