International Journal of Infectious Diseases 34 (2015) 66–70

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

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

Detecting the association between meteorological factors and hand, foot, and mouth disease using spatial panel data models

Hao Wang [a], Zhaohui Du [a], Xianjun Wang [b], Yunxia Liu [a], Zhongshang Yuan [a], Yanxun Liu [a,*], Fuzhong Xue [a,*]

[a] Department of Biostatistics, School of Public Health, Shandong University, PO Box 100, 44 West Wenhua Road, Jinan 250012, Shandong, China
[b] Shandong Center for Disease Control and Prevention, Jinan, Shandong, China

ARTICLE INFO

Article history:
Received 19 August 2014
Received in revised form 14 January 2015
Accepted 6 March 2015

Corresponding Editor: Eskild Petersen, Aarhus, Denmark

Keywords:
Spatial panel data model
Hand, foot, and mouth disease
Meteorological factors

SUMMARY

Objectives: The aim of this study was to quantify the relationship between meteorological factors and the occurrence of hand, foot, and mouth disease (HFMD) among children in Shandong Province, China, at a county level, using spatial panel data models.

Methods: Descriptive analysis was applied to describe the epidemic characteristics of HFMD from January 2008 to December 2012, and then a global autocorrelation statistic (Moran's *I*) was used to detect the spatial autocorrelation of HFMD in each year. Finally, spatial panel data models were performed to explore the association between the incidence of HFMD and meteorological factors.

Results: Moran's *I* at the county level were high, from 0.30 to 0.45 ($p < 0.001$), indicating the existence of a high spatial autocorrelation on HFMD. Spatial panel data models are more appropriate to describe the data. Results showed that the incidences of HFMD in Shandong Province, China were significantly associated with average temperature, relative humidity, vapor pressure, and wind speed.

Conclusions: Spatial panel data models are useful when longitudinal data with multiple units are available and spatial autocorrelation exists. The association found between HFMD and meteorological factors makes a contribution towards advancing knowledge with respect to the causality of HFMD and has policy implications for HFMD prevention and control.

1. Introduction

Hand, foot, and mouth disease (HFMD) is a common infectious disease, particularly in those under the age of 5 years.[1] It is mainly caused by the enteroviruses, especially coxsackievirus A16 (CA16) and enterovirus 71 (EV71).[2] Direct contact with respiratory droplets, feces, and blister fluid of an infective patient, or contact with a contaminated environment, e.g., water, food, or surfaces, can result in the spread of the disease.[3] In recent years, the HFMD outbreak trend has increased among children in China,[4–6] and this is regarded as an important public health problem by health officers.[1]

Many studies have been performed to analyze the association between meteorological factors and climate-sensitive infectious diseases due to their potential as early warning tools.[7–9] The relationship between meteorological factors and HFMD has been investigated in a number of studies.[10–14] For instance, Hu et al.[10] found that climate factors were potential determinants of the incidence of HFMD in most areas in China, using geographically weighted regression. A study by Huang et al.[12] provided evidence that the incidence of HFMD in children was associated with the high average temperature and high relative humidity in Guangzhou. However, to our knowledge, most previous studies have been conducted focusing only on a spatial dimension or on a time dimension approach. Spatial dimension analysis alone is performed on the basis of data at a certain temporal point, and time series analysis alone usually focuses on a certain region, which might result in a loss of information by ignoring the heterogeneity in both time and space. The space–time scan statistic provides a method for detecting possible spatial–temporal clusters of disease during a given study period, but it cannot be used to examine the risk factors of disease.[15]

Compared to traditional methods based on cross-sectional or time-series data alone, spatial panel data models are more informative,[16] and can enable researchers to control for both spatial dependency and unknown heterogeneity.[17] In this study,

* Corresponding authors. Tel.: +86 18264170182/+86 13906405997.
E-mail addresses: liu-yx@sdu.edu.cn (Y. Liu), xuefzh@sdu.edu.cn (F. Xue).

http://dx.doi.org/10.1016/j.ijid.2015.03.007

spatial panel data models were performed to analyze the relationship between meteorological variables and the incidence of HFMD among children in 140 counties in Shandong Province, China.

2. Materials and methods

2.1. Study site

Shandong Province, located between latitude 34°25′ and 38°23′ North, and longitude 114°36′ and 112°43′ East, is a coastal province in East China with a population of approximately 98 million (Figure 1). Shandong Province features a monsoon climate of medium latitudes.

2.2. HFMD surveillance data

County-level HFMD data for Shandong Province for the period January 2008 to January 2012 were obtained from the China Information System for Disease Control and Prevention (CISDCP, http://www.cdpc.chinacdc.cn). The clinical criteria for diagnosis of an HFMD case were provided in a guidebook published by the Chinese Ministry of Health in 2009.[18]

Meteorological county-level monthly climate factor data for Shandong Province for the period January 2008 to December 2012 were obtained from the China Meteorological Data Sharing Service System. Monthly meteorological variables assessed in this study included average atmospheric pressure (AAP), average temperature (AT), average vapor pressure (AVP), average relative humidity (ARH), monthly precipitation (MP), average wind speed (AWS), and monthly total sunshine hours (MSH).

2.3. Statistical analysis

2.3.1. Global spatial autocorrelation analysis

For the purpose of providing additional information about the presence of spatial dependence in the dependent variable, the HFMD incidences were annualized for 140 counties, and then the global Moran's *I* was calculated for each year. The global Moran's *I*, is a well-known measure to evaluate whether the spatial pattern is clustered, dispersed, or random.[19] Monte Carlo randomization was employed to assess the significance of Moran's *I*. A statistically significant estimate of Moran's *I* (*Z*-score ≥1.96) indicated that neighboring counties had similar incidences of HFMD.

2.3.2. Spatial panel data models

Spatial panel data models were used to quantify the association between the incidence of HFMD and meteorological variables, which could address data with spatial dependence and also enable researchers to consider both spatial and temporal heterogeneity. Spatial panel data typically refers to data containing continuous observations of a number of spatial units. According to Elhorst,[16] spatial panel data models are more informative and contain more variation and less collinearity among the variables compared to cross-sectional or time series models. The use of panel data results in a greater availability of degrees of freedom, and hence increases the efficiency of the estimation.

A simple linear model between a dependent variable *Y* and a set of *K* independent variables *X* is given as follows:

$$y_{it} = X\beta + \epsilon_{it} \tag{1}$$

where *i* is an index for the cross-sectional dimension (spatial units), with $i = 1, ..., N$, and *t* is an index for the time dimension (time periods), with $t = 1, ..., T$. y_{it} is an observation on the dependent variable at *i* and *t*. X_{it} is a 1-by-*K* row vector of observations on the independent variables, and β is a matching *K*-by-1 vector of fixed but unknown parameters. ε_{it} is an independently and identically distributed error term for *i* and *t* with zero mean and variances.[2]

The main drawback of this model is its failure to account for spatial and temporal heterogeneity, because spatial units and time periods tend to have spatial or temporal heterogeneity, and space- and time-specific variables do influence the dependent variable. To incorporate a variable intercept m_i and/or g_t representing the effects of the omitted variables that are peculiar to each spatial unit and/or time period, the basic form of the simple panel data model with spatial and temporal specific effect is:

$$y_{it} = \mu_i + \gamma_t + X\beta + \mu_i + \epsilon_{it} \tag{2}$$

where μ_i denotes a spatial specific effect and γ_t represents temporal specific effects. The spatial specific effects may be treated as fixed effects or as random effects. A random effect is an appropriate specification if a certain number of individuals are randomly drawn from a large population; the fixed effect model is favored[20] when the regression analysis is applied to a precise set of regions. For this reason, since our data contained all counties of the study region, we established fixed effects panel data models that included spatial error autocorrelation or a spatially lagged dependent variable. At the same time, we compared the random effects specification against fixed effects specification using Housman's specification test; it also suggested that fixed effects specification was more appropriate.

If spatial dependence exists, the simple panel data models with specific effects can be extended to include spatially lagged dependent variables or spatial error autocorrelation terms named as the spatial lag and the spatial error model, respectively.

The spatial lag model posits that the dependent variable depends on the dependent variable observed in neighboring units.[16] For example, the spatial and temporal fixed effects panel data model including a spatially lagged dependent variable can be specified as follows:[21]

$$y_{it} = \mu_i + \mu_t + \rho\sum_{j=1}^{N} W_{ij} y_{jt} + X'_{it}\beta + \epsilon_{it} \tag{3}$$

The fixed effects spatial error model can be written as:[21]

$$\begin{aligned} y_{it} &= \mu_i + \mu_t + X'_{it}\beta + \phi_{it} \\ \phi_{it} &= \lambda\sum_{j=1}^{N} W_{ij}\phi_{it} + \epsilon_{it} \end{aligned} \tag{4}$$

where *W* is an $N \times N$ positive non-stochastic spatial weight matrix; ρ is the spatial autoregressive coefficient and λ is usually called the spatial autocorrelation coefficient.

Likelihood ratio (LR) tests can be used to determine the extension of the model with spatial and/or time-period fixed effects. Whether the spatial lag model and/or the spatial error model are more appropriate to describe the data than a simple panel data model can be tested using the Lagrange multiplier (LM) and robust Lagrange multiplier (robust LM) test. If the LM lag is more significant than LM error and the robust LM lag test is significant and robust LM error is not significant, then the spatial lag model is more appropriate, and vice versa. Model evaluation can be conducted based on the goodness of fit; R^2 and log-likelihood are the commonly used effective criteria.[22–24] The spatial panel data models and LR and LM tests used in this study were executed in Matlab R2009a (Mathworks Inc., Natick, MA, USA).

3. Results

3.1. Descriptive analysis

A total of 497 664 HFMD cases were reported in Shandong Province between 2008 and 2012. Table 1 shows that 93.99% of