

基于向量自回归模型的清远市手足口病发病预测分析

杜玉忠, 张铭驱, 范秀红, 卢文涛, 曾茜茜, 黄燕, 黄燕琼

广东省清远市疾病预防控制中心, 广东 清远 511500

摘要: **目的** 运用向量自回归模型(vector autoregression model, VAR)分析清远市手足口病发病与气象因素的动态关系, 为手足口病的防控提供科学依据。 **方法** 用 2013 年 1 月 1 日-2017 年 4 月 30 日的气象数据和手足口病发病数据建立 VAR 模型, 评价模型的稳定性及拟合效果, 通过脉冲响应函数和方差分解分析对模型进行定量分析, 进一步对手足口病发病情况进行预测并评价预测效果。 **结果** 模型总拟合优度为 0.96, 调整后拟合优度为 0.95。手足口病发病数对日最低气温、日地表最低气温和日平均相对湿度等气象因素的冲击响应是正向的。模型对未来 1 周、2 周、3 周和 4 周的手足口病发病预测其平均绝对百分误差(mean absolute percentage error, MAPE)分别为 17.75%、13.69%、10.74% 和 15.85%, 希尔不等式系数(Theil inequality coefficient, TIC)分别为 0.08、0.06、0.05 和 0.09。 **结论** 气象因素和手足口病发病的 VAR 模型可以较好地清远市手足口病发病的短期预测。

关键词: 手足口病; 气象因素; 向量自回归模型

中图分类号: R512.5 **文献标识码:** B **文章编号:** 1006-3110(2019)02-0247-04 DOI:10.3969/j.issn.1006-3110.2019.02.038

Prediction of hand, foot and mouth disease incidence in Qingyuan City based on vector autoregression model

DU Yu-zhong, ZHANG Ming-qu, FAN Xiu-hong, LU Wen-tao, ZENG Xi-xi, HUANG Yan, HUANG Yan-qiong

Qingyuan Municipal Center for Disease Control and Prevention, Qingyuan, Guangdong 511500, China

Abstract: **Objective** To analyze the dynamic relationship between hand, foot and mouth disease (HFMD) incidence and meteorological factors in Qingyuan City by using vector autoregression model (VAR) so as to provide a scientific basis for prevention and control of HFMD. **Methods** The VAR model was established with the meteorological data and HFMD incidence data from January 1, 2013 to April 30, 2017. The stability and fitting effect of the model were evaluated. The model was analyzed quantitatively by using impulse response function and variance decomposition analysis, and the results were used to further predict the incidence of HFMD and evaluate the prediction effect. **Results** The total goodness-of-fit of the model was 0.96, and the adjusted goodness-of-fit was 0.95. The impact of HFMD incidence on the daily minimum temperature, daily minimum surface temperature and daily average relative humidity was positive. The mean absolute percentage errors in the prediction of HFMD in the next week and next 2, 3 and 4 weeks were 17.75%, 13.69%, 10.74% and 15.85% respectively, and Theil inequality coefficients were 0.08, 0.06, 0.05 and 0.09 respectively. **Conclusions** The VAR model of meteorological factors and HFMD incidence can be used for short-term prediction of HFMD incidence in Qingyuan City.

Key words: hand, foot and mouth disease; meteorological factor; vector autoregression model

针对手足口病的发病预测, 已有众多学者用不同统计学方法进行了探讨^[1-3], 但运用向量自回归模型

作者简介: 杜玉忠(1977-), 男, 山东曹县人, 硕士, 副主任医师, 主要从事传染病预防控制工作。

(vector autoregression model, VAR) 进行手足口病发病预测的文献较少。有研究^[4]认为虽然 VAR 模型参数估计标准误差较大, 模型稳健性不及一元自回归模型, 但可以提供更为精确的预测结果。鉴于 VAR 具有不用

- [7] Greenhalgh D, Doyle M, Lewis FA. Mathematical treatment of AIDS and condom use[J]. IMA J Math Appl Med Biol, 2001, 18(3): 225-262.
- [8] 马知恩, 靳祯. 总人口在变化的流行病动力学模型[J]. 华北工学院学报, 2001, 22(4): 262-271.
- [9] 孙立哲, 马文丽, 孙汉顺, 等. HIV 感染动力学模型概述[J]. 生物信息学, 2010, 8(4): 302-306.
- [10] 彭志行, 陈峰. HIV/AIDS 传播动力学模型研究进展[J]. 中国卫生统计, 2011, 28(6): 730-734.
- [11] 叶海平, 丁永生. 一类总人口在变化的 HIV/AIDS 传播的动力学模

- 型[J]. 生物数学学报, 2010, 25(1): 51-60.
- [12] 张少林. 人口流动性对感染性疾病扩散与传播的影响[J]. 生物数学学报, 2006, 27(2): 253-260.
- [13] Centers for Disease Control and Prevention(2012). Monitoring selected national HIV prevention and care objectives by using HIV surveillance data United States and 6 U. S. dependent areas-2010[EB/OL]. (2013-07-18) [2018-06-10]. <http://www.cdc.gov/hiv/topics/surveillance/resources/reports>.
- [14] Shermer M. How the survivor bias distorts reality[J]. Sci Am, 2014, 311(2): 94-99.

收稿日期: 2018-10-11

区分内生变量和外生变量、对模型系统不须加任何约束等众多优点,本文拟采用 VAR 建立清远市手足口病发病与气象因素的数量关系,并利用建立的模型进行发病预测,以期对手足口病的发病预测和控制提供一种新的思路。

1 资料与方法

1.1 数据来源 手足口病发病数据来自《中国疾病预防控制中心信息系统》,收集清远市范围内 2013 年 1 月 1 日-2017 年 4 月 30 日每日手足口病(hand, foot and mouth disease, HFMD)发病资料。从中国气象局获取同期气象数据,包括日最高气温(Tmax)、日最低气温(Tmin)、日地表最低气温(GST_min)、日累计降水量(PRE)、日平均气压(PRS)、日平均相对湿度(RH_avg)、日平均风速(Win_avg)和平均日照时数(SSD)等资料,对上述手足口病发病和同期气象资料进行处理,计算出气温的温差(DTEM)并对全部数据换算为每周的平均水平。

1.2 统计方法 变量的平稳性检验采用 ADF 检验,VAR 模型参考相关文献^[5]。采用 Excel 2007 建立数据库,采用 Eviews 10.0 进行模型拟合及预测。

2 结果

2.1 基本情况 2013 年 1 月 1 日-2017 年 4 月 30 日共收集 226 周数据,手足口病发病 57 132 例,平均每周发病 23 例(中位数)。其中 2013 年发病 9 735 例,2014 年发病 16 594 例,2015 年发病 13 913 例,2016 年发病 15 397 例,2017 年 1-4 月发病 1 493 例,各变量基本情况见表 1。

表 1 清远市 2013-2017 年手足口病发病及同期气象因素基本情况

变量	最小值	最大值	均值	标准差
每日 HFMD(例)	2.00	172.00	36.14	35.77
Tmin(℃)	2.93	26.36	17.10	6.62
DTEM(℃)	2.39	17.34	8.17	2.62
GST_min(℃)	0.71	26.78	17.02	6.99
PRE(mm)	0.00	1 635.00	241.06	275.95
PRS(hPa)	986.97	1 013.65	1 000.75	6.82
RH_avg(%)	45.43	93.14	78.85	7.98
SSD(h)	0.00	10.27	4.06	2.58
Win_avg(m/s)	1.15	3.24	1.88	0.41

2.2 变量的平稳性检验 VAR 模型拟合前提,各序列须满足平稳性条件,对各变量进行单位根检验,结果

各变量均为平稳序列,见表 2。

表 2 清远市 2013-2017 年手足口病发病及同期气象因素单位根检验结果

变量	差分次数	(C,T,K)	ADF 值	1% 临界值	5% 临界值	10% 临界值	结论
HFMD	0	(c,0,4)	-4.57	-3.46	-2.87	-2.57	平稳
Tmin	0	(c,0,7)	-4.23	-3.46	-2.87	-2.57	平稳
DTEM	0	(c,0,0)	-12.79	-3.46	-2.87	-2.57	平稳
GST_min	0	(c,0,9)	-5.60	-3.46	-2.87	-2.57	平稳
PRE	0	(c,0,0)	-15.08	-4.00	-3.43	-3.14	平稳
PRS	0	(c,0,11)	-6.28	-3.46	-2.87	-2.57	平稳
RH_avg	0	(c,0,2)	-4.83	-3.46	-2.87	-2.57	平稳
SSD	0	(c,0,2)	-5.00	-3.46	-2.87	-2.57	平稳
Win_avg	0	(c,0,2)	-6.13	-3.46	-2.87	-2.57	平稳

注:(C,T,K)表示常数项、趋势项及滞后阶数,滞后阶数的确定采用施瓦兹准则 SC(Schwarz Criterion)判定。

2.3 VAR 模型的建立 以 HFMD 为因变量,采用模型滞后结构确定准则进行模型滞后阶数的筛选,经对 0-10 阶滞后 VAR 的各准则结果比较,最终确定滞后阶数为 2,建立的 VAR 模型见表 3。模型总的拟合优度为 0.96,调整后的拟合优度为 0.95,且所有单位根位于单位圆内,表明模型结构稳定,拟合效果好,见图 1。

表 3 清远市 2013-2017 年手足口病发病及同期气象因素 VAR 模型结果

变量	系数	标准差	T 统计量
HFMD(-1)	1.09	0.07	15.58
HFMD(-2)	-0.17	0.07	-2.53
DTEM(-1)	1.33	0.54	2.47
DTEM(-2)	-0.65	0.53	-1.24
GST_min(-1)	-3.51	1.60	-2.18
GST_min(-2)	-2.20	1.40	-1.57
PRE(-1)	-0.01	0.01	-3.21
PRE(-2)	-0.01	0.01	-0.42
PRS(-1)	-0.07	0.26	-0.27
PRS(-2)	0.16	0.26	0.62
RH_avg(-1)	8.58	17.08	0.5
RH_avg(-2)	-17.01	15.82	-1.07
SSD(-1)	-2.79	0.57	-4.92
SSD(-2)	0.13	0.58	0.22
Tmin(-1)	4.50	1.61	2.79
Tmin(-2)	2.14	1.44	1.48
Win_avg(-1)	-0.11	2.13	-0.04
Win_avg(-2)	-1.33	2.08	-0.64
C	-89.43	272.77	-0.32

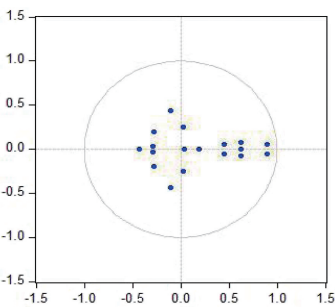


图 1 滞后阶数为 2 的 AR 特征根多项式逆根图

2.4 Granger 因果检验 Granger 因果检验结果表明,从总体上看,日温差、地表最低气温、累计降水量、平均

气压、平均相对湿度、日照时数、日最低气温和平均风速为手足口病发病的 Granger 原因($P<0.001$),表明建立的 VAR 模型是合理的。

2.5 脉冲响应分析 HFMD 受自身冲击后,至第 2 期达到最高点,随后开始呈下降趋势,总体表现为正向、长期影响;HFMD 受 DTEM 的冲击后,表现为负向影响,至第 6 期影响达到最高点,随后呈长期、缓慢减小影响;HFMD 受其他各因素冲击后,表现为正向影响的分别为 GST_min、RH_avg 和 Tmin,表现为负向影响的分别为 PRE、PRS、SSD 和 Win_avg,见图 2。

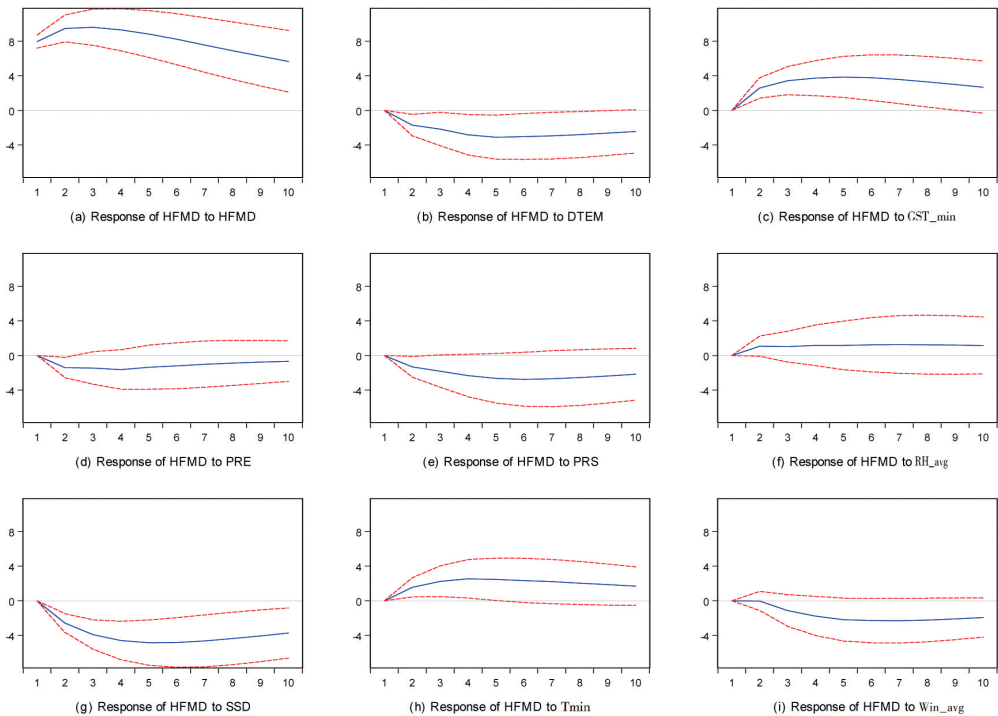


图 2 VAR 模型的脉冲响应函数图

2.6 方差分解分析 对手足口病发病数变化贡献率最大的是自身发病数的变化,贡献率呈逐周递减趋势,自第 7 周开始下降不明显,至第 10 周贡献率仍超过

50%。GSTMIN 和 SSD 对手足口病发病数的贡献率分别达到 9.18% 和 14.10%,其他各因素的贡献率均不超过 6%,见表 4。

表 4 方差分解分析

滞后期	标准差	HFMD	DTEM	GST_min	PRE	PRS	RH_avg	SSD	Tmin	Win_avg
1	7.93	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	13.29	86.70	1.69	3.88	1.15	0.98	0.61	3.66	1.34	0.00
3	17.74	78.31	2.48	6.03	1.33	1.59	0.68	6.87	2.31	0.41
4	21.55	71.91	3.41	7.15	1.49	2.25	0.75	9.18	2.92	0.94
5	24.75	67.38	4.16	7.91	1.45	2.86	0.79	10.77	3.19	1.48
6	27.36	64.21	4.63	8.42	1.38	3.36	0.86	11.91	3.33	1.91
7	29.48	61.90	4.98	8.77	1.31	3.73	0.92	12.71	3.42	2.25
8	31.19	60.21	5.26	8.98	1.25	4.01	0.99	13.31	3.47	2.52
9	32.56	58.97	5.47	9.11	1.21	4.22	1.04	13.76	3.50	2.72
10	33.65	58.04	5.65	9.18	1.17	4.37	1.09	14.10	3.52	2.88

2.7 VAR 模型的预测 利用上述模型采用动态预测的方法分别对未来 1 周、2 周、3 周和 4 周进行预测,并与实际发病数进行比较,若平均绝对百分误差 (mean absolute percentage error, MAPE) 取值在 0~5 之间,则预测精度较高,希尔不等式系数 (Theil inequality coefficient, TIC) 介于 0~1 之间,数值越小预测精度越高^[6]。从模型预测结果来看,模型可以较好地预测未来 1~4 周的发病数据,尤其以预测未来第 3 周 (225 周) 发病数较为精确,其 MAPE 和 TIC 均较小,结果见表 5。

表 5 VAR 模型预测发病数与实际发病数比较情况

预测时间	预测发病	实际发病	MAPE	RMSE	TIC
223 周	10.92	12.86	17.75	1.94	0.08
224 周	13.69	12.71	13.69	1.53	0.06
225 周	15.04	16.14	10.74	1.41	0.05
226 周	17.10	22.43	15.85	2.93	0.09

3 讨论

脉冲分析结果显示, GST_min、RH_avg 和 Tmin 对手足口病的发病表现为正向作用,说明地表最低气温、相对湿度和日最低气温越高,手足口病发病的风险越大,分析原因可能与最低气温升高、儿童活动的增多增加了相互接触感染的机会有关^[3];相对湿度的增加较适合病原微生物生长,增加了肠道病毒的感染机会,这也与相关研究结果一致^[7-9]。与相关研究结果^[5]不同,本文分析结果显示 HFMD 受 DTEM、PRE 和 SSD 的冲击后,表现为负向影响,分析原因可能为温差大和日累计降雨量增加,均可减少儿童的活动,降低儿童接触感染机会;日照时间的增加则可能使环境干燥,不适宜肠道病毒繁殖和活动,从而降低了手足口病的发病风险。本文分析结果同时显示 PRS 和 Win_avg 冲击 HFMD 后呈负向影响,这也与安庆玉等^[9]的研究结果一致。手足口病为湿热病,风速等气象因素均可以直接影响病原体的繁殖及其在环境中的生存能力和时间,高风速和高气压的天气条件有利于污染物的扩散,但风速和气压也可能通过影响其他气象因素间接影响手足口病的发生^[10]。

方差分解的结果提示虽然气象因素对手足口病的发病呈现一定的影响作用,但手足口病的防控仍然是以加强手足口病的病例管理、切断传播途径和保护易感人群三个方面的工作为主,如隔离病例、改善环境卫生、增强人群自身免疫力或接种 EV71 疫苗等。

从建立的 VAR 模型来看,利用气象资料建立的 VAR 模型其稳定性及拟合效果均较好,从模型预测结果来看,虽然其精确性仍有改进空间,但模型基本可以较好地预测未来 1~4 周的发病数据,尤其以预测未来第 3 周 (225 周) 发病数较为精确,其 MAPE 和 TIC 均较小,提示利用气象资料及手足口病发病资料建立的清远市手足口病发病预测的 VAR 模型的适用性。

利用 VAR 模型预测手足病的发生,模型的稳定性、拟合效果及预测的精确性受纳入模型变量的影响。本文纳入了多项气象因素,从数量关系上来看,纳入的各气象因素整体上可以作为手足口病发病的影响因素,但各气象资料是否真正对手足口病的发病具有影响作用仍有待进一步研究。

参考文献

- [1] 叶永同, 颜玉炳. 厦门市翔安区手足口病流行特征分析及发病趋势预测方法研究[J]. 实用预防医学, 2016, 23(2):237-239.
- [2] 原凌云, 周以军, 朱妮, 等. 多种数据模型在手足口病发病预测的应用探讨[J]. 实用预防医学, 2018, 25(11):1400-1402.
- [3] 康敏, 马文军, 林锦炎, 等. 广州市气象因素对于手足口病发病的短期效应研究[J]. 中华流行病学杂志, 2012, 33(2):244-245.
- [4] 倪延延, 张晋昕. 向量自回归模型拟合与预测效果评价[J]. 中国卫生统计, 2014, 31(1):53-56.
- [5] 韦懿芸. 向量自回归模型在手足口病发病与气象因素的动态分析中的应用[J]. 中国卫生统计, 2013, 30(6):794-797.
- [6] 霍宗杰, 祝伟. 中国的能源消费、能源消费结构与经济增长—基于 ARDL 模型的实证研究[J]. 当代经济科学, 2010, 32(3):28-35.
- [7] 杜玉忠, 李翠华, 梁剑辉, 等. 清远市手足口病流行趋势与短期气候因素的关系[J]. 职业与健康, 2015, 31(15):2062-2063.
- [8] Ma E, Lam T, Wong C, et al. Is hand, foot and mouth disease associated with meteorological parameters? [J]. Epidemiol Infect, 2010, 138(12):1779-1788.
- [9] 杜志成, 张王剑, 于石成, 等. 广东省手足口病气象危险因素阈值效应的研究[J]. 现代预防医学, 2015, 42(19):3457-3460.
- [10] 安庆玉, 范雪松, 吴隼, 等. 大连市手足口病与气象因素的相关性研究[J]. 现代预防医学, 2014, 41(1):3-6.

收稿日期:2018-01-28