

2011–2018 年我国登革热疫情时间序列分析及空间自相关分析

杨慧欣 赵晨皓 雒静静 胡芳芳 张思文 王太君 甄清

130021 长春, 吉林大学公共卫生学院

通信作者: 甄清, E-mail: zq415@sina.com

DOI: 10.16462/j.cnki.zhjbkz.2019.10.018

【摘要】 目的 了解 2011–2018 年我国登革热疫情时空分布特征, 对 2019 年我国登革热的发病情况进行预测。方法 基于中国疾病预防控制中心信息系统中 2011–2018 年我国登革热的病例数据, 借助 R 3.6.0 软件, 使用自回归积分滑动平均模型 (autoregressive integrated moving average model, ARIMA) 对登革热的发病趋势进行描述和预测。基于国家人口与健康科学数据共享服务平台提供的 2011–2016 年全国以及各省市登革热发病率、发病人数数据, 采用 GeoDa 1.12 软件进行全局、局部空间自相关分析, 确定登革热时空热点区域。结果 预测 2019 年全年登革热发病数为 14 302 人。2012 年 (Moran's $I = -0.088$, $P = 0.037$)、2013 年 (Moran's $I = -0.121$, $P = 0.040$) 和 2014 年 (Moran's $I = -0.076$, $P = 0.045$) 年全国登革热发病呈现全局空间负相关关系, 2016 年 (Moran's $I = 0.078$, $P = 0.048$) 登革热发病呈现全局空间正相关关系。局部自相关分析结果显示, 登革热发病高聚集区域主要在我国东南沿海地区。结论 2019 年我国登革热的流行无明显波动趋势, 且疫情呈空间聚集性分布。

【关键词】 登革热; 时间序列; 全局自相关分析; 局部自相关分析; 发病预测

【中图分类号】 R181

【文献标识码】 A

【文章编号】 1674-3679(2019)10-1250-05

Time series analysis and spatial autocorrelation analysis of dengue data in China from 2011 to 2018 YANG Hui-xin, ZHAO Chen-hao, LUO Jing-jing, HU Fang-fang, ZHANG Si-wen, WANG Tai-jun, ZHEN Qing

Jilin University School of Public Health, Changchun 130021 China

Corresponding author: ZHEN Qing, E-mail: zq415@sina.com

【Abstract】 Objective To understand the spatial and temporal distribution characteristics of dengue fever in China from 2011 to 2018, and predict the incidence of dengue fever in China in 2019. **Methods** Based on the case data of dengue fever in China from 2011 to 2018 in the Chinese Disease Prevention and Control Information System, the trend of dengue fever was described and predicted by using the autoregressive integrated moving average model (ARIMA) with R 3.6.0 software. Based on the data of the incidence of dengue fever in the country, provinces and cities from 2011 to 2016 provided by the national scientific data sharing platform for population and health, global and local spatial autocorrelation analysis was performed using GeoDa 1.12 software to determine the dengue fever hotspots. **Results** The incidence of dengue fever was 14 302 in 2019, showing no disease outbreaks. The incidence of dengue fever in 2012 (Moran's $I = -0.088$, $P = 0.037$), 2013 (Moran's $I = -0.121$, $P = 0.040$) and 2014 (Moran's $I = -0.076$, $P = 0.045$) showed a global spatial negatively correlation. In 2016 (Moran's $I = 0.078$, $P = 0.048$), the incidence of dengue fever was positively correlated with global space. The results of local autocorrelation analysis showed that the high incidence of dengue fever was mainly in the southeast coastal areas of China. **Conclusions** In 2019, the epidemic of dengue fever in China showed no obvious fluctuation trend, and the epidemic situation showed spatial clustering distribution.

【Key words】 Dengue fever; Time series; Global spatial autocorrelation analysis; Spatial clustering analysis

登革热主要是由登革热病毒引起的,依靠埃及伊蚊和白纹伊蚊传播的一种急性传染性疾病,每年造成约 3.9 亿人发病^[1-2]。自 2011 年以来,中国登革热发病率明显增加,特别是 2014 年,全国报告了 44 896 例登革热病例,数量超过以往 30 年的总和,流行规模更是达到 1986 年以来的新高^[3-4]。由于我国传染病发病数据的更新和获取存在滞后性,故国内大多数研究采用的是 2016 年及以前的数据。本研究分析 2011–2018 年全国登革热发病情况,旨在了解中国大陆最新登革热疫情时空分布特征,预测 2019 年中国登革热的发病人数,以便有的放矢,明确今后的防控工作重点。

1 资料与方法

1.1 研究资料 本研究采用中国疾病预防控制中心信息系统中 2011–2018 年我国登革热的病例数据及“国家人口与健康科学数据共享服务平台”中国公共卫生科学数据中心提供的 2011–2016 年中国 31 个省、直辖市和自治区(未包括台湾省、香港地区和澳门地区)的登革热疫情数据。

1.2 研究方法

1.2.1 描述性分析 利用 Excel 2019 软件清理登革热发病信息,利用 R 3.6.0 软件绘制 2011–2018 年我国登革热的发病趋势图。以 2011–2016 年全国各月份发病人数资料,分别采用 Excel 2019 软件绘制全国登革热发病年龄变化图,R 3.6.0 软件绘制全国登革热月份分布热图。

1.2.2 时间序列分析 以 2011–2018 年全国各月份登革热的发病人数及发病率数据,利用 R 3.6.0 软件中时间序列分析的自回归积分滑动平均模型 (autoregressive integrated moving average model, ARIMA) 拟合与预测 2019 年我国登革热的发病人数。

1.2.3 空间自相关分析 基于 2011–2016 年全国各省市登革热发病率、发病人数数据,使用 GeoDa 1.12 软件进行全局、局部空间自相关分析, Moran 散点图以及 LISA 聚集图描述发病的热点区域。

1.3 统计学方法 采用 R 3.6.0 统计软件中时间序列分析的 ARIMA 模型对经 Box-Cox 变换(公式 1 $\lambda = -0.095\ 322\ 82$) 后的数据进行拟合、预测。

$$W_t = \begin{cases} \frac{(y_t^\lambda - 1)}{\lambda} & \lambda \neq 0 \\ \log(y_t) & \lambda = 0 \end{cases} \quad (1)$$

本研究采用 GeoDa 1.12 软件计算 Moran's I 指数,取值范围为 $-1 \sim 1$ 。当 $|Z| > 1.96$, $P < 0.05$,认为差异有统计学意义,存在空间自相关^[5]。全局 Mo-

ran's I 的计算公式(2)如下^[6]:

$$\text{Moran's I} = \frac{n \times \sum_i \sum_j w_{ij} \times (y_i - \bar{y})(y_j - \bar{y})}{(\sum_i \sum_j w_{ij} \times \sum_j (y_i - \bar{y})^2)} \quad (2)$$

通过绘制 LISA 聚集图来反映一个区域与邻近区域的相关程度,即全国范围内登革热的冷、热点的分布情况^[7]。检验水准 $\alpha = 0.05$ 。

2 结果

2.1 登革热的流行病学分析

2.1.1 2011–2018 年全国登革热的发病趋势 根据 2011–2018 年全国登革热发病率(1/10 万)情况绘制发病率时间序列图。从图中可以看出从 2011–2014 年全国登革热发病率呈平稳趋势。2014 年发生全国登革热大爆发。此后每两年发生一次登革热的发病率升高,2014–2018 年登革热的发病呈平稳波动趋势。见图 1。

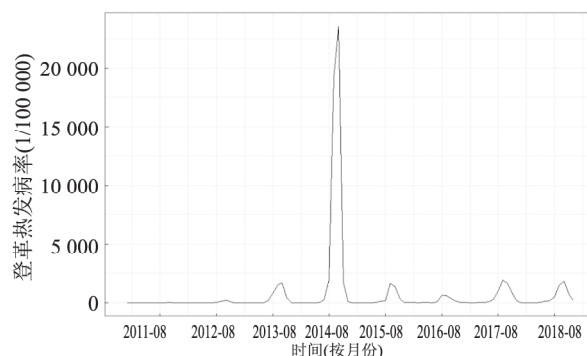


图 1 2011–2018 年全国登革热发病率时间序列

Figure 1 Time series of national dengue fever incidence rates from 2011 to 2018

2.1.2 2011–2016 年全国登革热患者发病年龄分布变化 2011–2016 年全国登革热患者的年龄范围为 0.5~91 岁,发病的主要年龄段为 20~34 岁。病例发病年龄分布变化显示:50 岁以上的人群发生登革热的比重逐年增加,而 20~34 岁的青壮年发病人数呈下降趋势。见图 2。

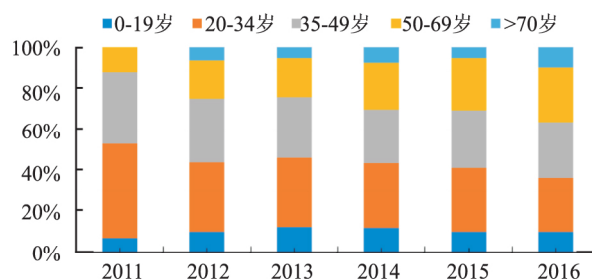


图 2 2011–2016 年全国登革热患者发病年龄分布变化

Figure 2 Age distribution change of dengue patients in China from 2011 to 2016

2.1.3 2011–2016 年全国登革热发病月份分布 通过分析 2011–2016 年全国各月份登革热的发病率发现登革热疫情有明显的季节性,1~5 月全国基本无登革热的流行,6 月份发病数开始增多,9~10 月为发病高峰。见图 3。

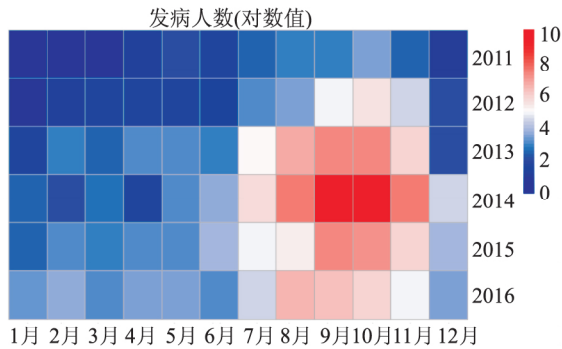


图 3 2011–2016 年我国登革热的发病月份分布

Figure 3 Distribution of the incidence of dengue fever in China from 2011 to 2016

2.2 全国登革热的时间序列

2.2.1 模型的选择 按照公式 (1) 对 2011–2018 年全国登革热发病人数数据进行 Box-Cox 变换,并对变换后的时间序列进行一阶差分,得到了稳定序列。选用季节性 ARIMA 模型对数据进行拟合,为了估计模型的参数,分别对数据进行差分和季节性差分。共同估计 $ARIMA(P, D, Q)(p, d, q)_{12}$ 的各项系数,最终选择 $ARIMA(1, 1, 1)(0, 1, 1)_{12}$ 作为最佳的模型。见图 4。

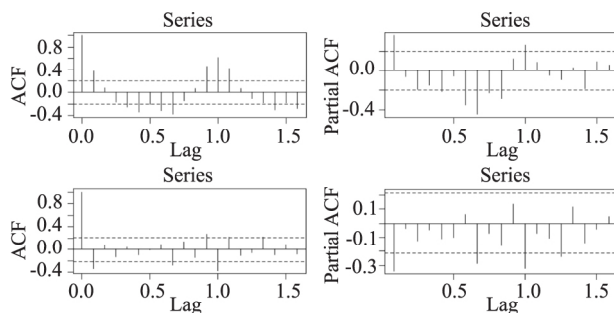


图 4 差分与季节性差分后的自相关和偏自相关图

Figure 4 Autocorrelation and partial autocorrelation plots after difference and season difference

2.2.2 模型的验证与评价 在 p, q, P, Q 四个值为 0 和 1 之间尝试了多个模型,以赤池信息准则 (akaike information criterion, AIC) 为选取依据,选取结果如表 1 所示。在所有的模型中,所选取的 $ARIMA(1, 1, 1)(0, 1, 1)_{12}$ 的 AIC 值最小。该模型均方根误差 (root mean squared error, RMSE) 为 0.423 949 9, 平均绝对百分误差 (mean absolute percentage error, MAPE) 为 11.512 12, 平均绝对比例误

差 (mean absolute scaled error, MASE) 为 0.498 057 9。对模型拟合数据的残差进行 Ljung-Box 检验,检验结果没有统计学意义 (Ljung-Box $Q = 24.483, P = 0.08452$),可以认为残差序列为白噪声。

表 1 模型选择表

Table 1 Model selection table

模型参数	AIC	RMSE	Ljung-Box P 值
(010)(010)	151.250 7	0.552 8	<0.001
(110)(010)	143.086 3	0.519 6	0.055
(011)(010)	142.591 5	0.518 0	0.018
(010)(110)	139.425 3	0.501 8	0.007
(010)(011)	126.968 5	0.441 8	0.035
(111)(010)	140.638 7	0.503 4	0.001
(010)(111)	128.722 3	0.422 1	0.018
(110)(110)	136.696 4	0.489 5	0.110
(011)(011)	125.819 9	0.439 1	0.069
(111)(110)	133.012 6	0.470 6	0.038
(111)(011)	122.575 1	0.423 9	0.085
(110)(111)	127.150 9	0.407 3	0.078
(011)(111)	126.539 9	0.406 1	0.070
(111)(111)	123.633 1	0.391 7	0.051

2.2.3 发病预测 使用 $ARIMA(1, 1, 1)(0, 1, 1)_{12}$ 模型对 2019 年全国各月份登革热发病人数进行预测。已知 2019 年前三个月的登革热发病数分别为 165、78 和 52。用模型预测值与实际发病数计算的 RMSE 为 68.82, 平均绝对误差 (mean absolute error, MAE) 为 52.00。发病数都在 95% CI 范围内。见表 2。

表 2 2019 年全国登革热发病预测表

Table 2 National dengue fever morbidity forecast table in 2019

预测月份	预测值	95% CI 预测值范围
2019.1	50.404 40	14.46~207.82
2019.2	48.918 14	10.71~289.03
2019.3	40.470 22	7.92~279.17
2019.4	55.526 18	9.51~463.63
2019.5	94.503 85	13.93~984.42
2019.6	137.548 70	18.02~1 714.02
2019.7	480.783 14	47.48~9 410.66
2019.8	1 689.394 40	122.16~56 570.20
2019.9	5 003.354 18	267.67~294 224.05
2019.10	5 093.178 19	258.43~332 296.33
2019.11	1 437.776 99	94.09~57 382.31
2019.12	176.286 24	17.39~3 454.57

2.3 2011–2016 年全国登革热发病率的空间自相关分析 2011–2016 年登革热发病分布的 Moran 散点图以及 LISA 聚集图结果表明: 2011–2016 年中国大陆地区登革热疫情呈明显聚集性分布。2012、2013、2014 和 2016 年全局型空间自相关系数 Moran's I 均具有统计学意义 (均有 $P < 0.05$)。局部自相关分析结果表明, 2011 和 2012 年福建省及 2014 年的广西省处于“高-高”区域; 2011 年甘肃省和 2014 年广东省

处于“高-低”区域; 2013、2015 和 2016 年中国大陆所有地区均未呈现“高-高”区域。见图 5。

3 讨论

本研究采用时间序列分析和空间自相关分析方法对 2011–2018 年全国登革热发病情况进行回溯性研究。对登革热发病的流行病学研究结果表明, 各年龄组对登革热都具有易感性, 主要以 20~34 岁的青壮年居多, 这与青壮年社会活动丰富且易出汗有关。但近年该年龄段人群在发病人群中所占比例下降, 而 50 岁以上人群则呈上升趋势。人民生活水平不断提高且交通出行日益便捷, 导致户外活动不断增多, 这可能造成全年龄段人群涉足登革热流行区的机会增多, 而年龄的升高对免疫功能也具有重要的影响^[8]。本文认为这可能导致了 50 岁以上人群的易感性增加。本文表明登革热的发病高峰期主要是 8~10 月间, 这是因为某些气候因素如气温、气压、湿度、降雨量与蚊媒密度变化密切相关^[9-11]。

ARIMA 模型可以对各种不平稳的时间序列进行差分, 所以可以处理多种不平稳的时间序列数据, 这对于预测传染病时间序列具有重要意义。2019 年 1~3 月实际发病人数均落入预测模型的 95% CI 范围内, 且实际值与预测值的平均相对误差为 30.0%。表明本次构建的预测模型短期预测精度较高, 可为及早采取防控措施提供科学依据。

2011–2016 年中国登革热的空间分析表明, 该病显示出空间聚集性。2012、2013 和 2014 年全国登革热发病呈空间负相关关系, 与我国学者岳玉娟等^[12]的研究所得结论不一致, 这可能是由于数据来源不同造成的差异。利用 Moran 散点图与 LISA 积聚图进一步进行局部空间自相关分析, 均显示福建、云南、广东等地为登革热高发区, 该结论与之前的研究^[12-14]一致, 登革热发病多集中在沿海、经济发达、交通便利、人口流动性大的地区^[15-16]。逐年计算 Moran's I 指数发现, 2012–2014 年以及 2016 年登革热在全国呈非随机分布, 而 2011 年与 2015 年是随机分布的。由于登革热不是永久免疫性疾病, 因此人群免疫力的波动在登革热传播和流行中起着重要的作用^[17]。2014 年, 我国广东省发生了大规模登革热暴发, 全国发病人数达 4.5 万余人, 超过以往 30 年的总和, 此后人群免疫力保持较高水平。因此 2011–2016 年我国登革热分布方式不同的原因可能是人口免疫水平的变化。这提示我们在今后的研究中有必要对传染病空间随机分布期和非随机分布期的变化进行调查, 掌握其变化规律, 为预防疾病暴发提供指导。

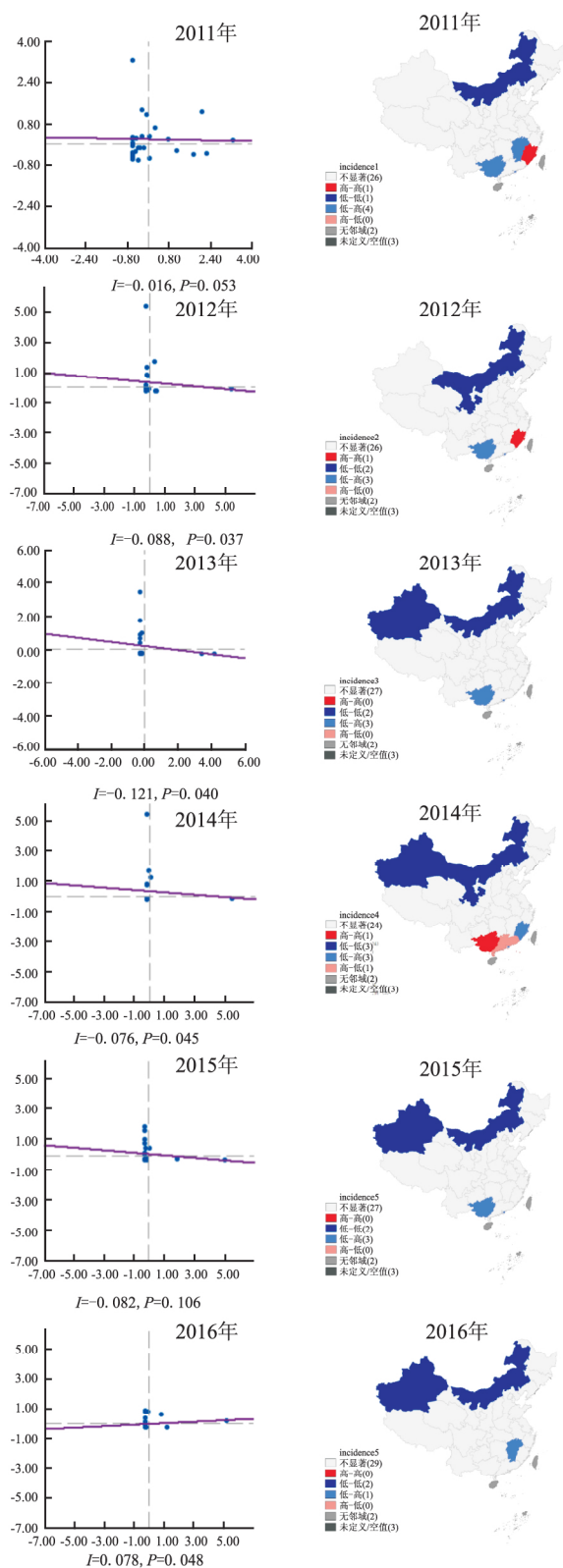


图5 2011–2016 年中国大陆登革热全局型空间自相关分析和局部自相关分析示意图

Figure 5 Schematic diagram of global spatial autocorrelation analysis and local autocorrelation analysis of dengue fever in China from 2011 to 2016

参 考 文 献

- [1] Bhatt S, Gething PW, Brady OJ, et al. The global distribution and burden of dengue [J]. *Nature*, 2013, 496(7446): 504-507. DOI: 10.1038/nature12060.
- [2] 国家卫生和计划生育委员会. 登革热诊疗指南(2014 年第 2 版) [J]. *中药新药与临床药理*, 2016, 27(1): 138-142. DOI: 10.3969/j.issn.1003-9783.2016.01.030.
National Health and Family Planning Commission. Guidelines for the diagnosis and treatment of dengue fever (2nd edition, 2014) [J]. *Tradit Chin Drug Res Clin Pharmacol*, 2016, 27(1): 138-142. DOI: 10.3969/j.issn.1003-9783.2016.01.030.
- [3] 熊益权, 陈清. 1978~2014 年我国登革热的流行病学分析 [J]. *南方医科大学学报*, 2014, 34(12): 1822-1825. DOI: 10.3969/j.issn.1673-4254.2014.12.24.
Xiong YQ, Chen Q. Epidemiological analysis of dengue fever in China from 1978 to 2014 [J]. *J South Med Univ*, 2014, 34(12): 1822-1825. DOI: 10.3969/j.issn.1673-4254.2014.12.24.
- [4] Lai S, Zhou H, Xiong W, et al. Changing epidemiology of human brucellosis, China, 1955-2014 [J]. *Emerg Infect Dis*, 2017, 23(2): 184-194. DOI: 10.3201/eid2302.151710.
- [5] Erazo C, Pereira SM, Costa Mda C, et al. Tuberculosis and living conditions in Salvador, Brazil: a spatial analysis [J]. *Rev Panam Salud Publica*, 2014, 36(1): 24-30.
- [6] 王劲峰, 廖一兰, 刘鑫. 空间数据分析教程 [M]. 北京: 科学出版社, 2010.
Wang JF, Liao YL, Liu X. Spatial data analysis course [M]. Beijing: Science Press, 2010.
- [7] 张冰冰, 姜祥坤, 张世英. 空间自相关分析在探究疾病分布热点区域中的应用 [J]. *山东大学学报(医学版)*, 2012, 50(5): 129-132. DOI: 10.6040/j.issn.1671-7554.2012.05.029.
Zhang BB, Jiang XK, Zhang SY, et al. Application of spatial autocorrelation analysis in exploring hotspots of disease distribution [J]. *Journal of Shandong University(Natural Medicine)*, 2012, 50(5): 129-132. DOI: 10.6040/J.issn.1671-7554.2012.05.029.
- [8] Fuentes E, Fuentes M, Alarcón M, et al. Immune system dysfunction in the elderly [J]. *An Acad Bras Cienc*, 2017, 89(1): 285-299. DOI: 10.1590/0001-3765201720160487.
- [9] 王芹, 殷文武, 窦丰满. 2006 年中国登革热疫情监测分析 [J]. *疾病监测*, 2009, 24(1): 22-24. DOI: 10.3784/j.issn.1003-9961.2009.01.008.
Wang Q, Yin WW, Dou FM, et al. Monitoring and analysis of dengue fever in China in 2006 [J]. *Dis Surveillance*, 2009, 24(1): 22-24. DOI: 10.3784/j.issn.1003-9961.2009.01.008.
- [10] 屈志强, 黎祖秋, 汤洪洋. 广西南宁市登革热媒介季节消长及其影响因素研究 [J]. *中国热带医学*, 2018, 18(10): 1007-1011. DOI: 10.13604/j.cnki.46-1064/r.2018.10.08.
Qu ZQ, Li ZQ, Tang HY, et al. Seasonal fluctuation of dengue vector in Nanning, Guangxi, and its influencing factors [J]. *China Tropical Medicine*, 2018, 18(10): 1007-1011. DOI: 10.13604/j.cnki.46-1064/r.2018.10.08.
- [11] 易彬程, 张治英, 徐德忠. 广东省登革热流行与气候因素变化的相关性 [J]. *第四军医大学学报*, 2003, 24(2): 143-146. DOI: 10.3321/j.issn:1000-2790.2003.02.014.
Yi BT, Zhang ZY, Xu DZ, et al. Correlation between dengue fever epidemics and changes in climatic factors in Guangdong Province. [J]. *J Fourth Mil Med Univ*, 2003, 24(2): 143-146. DOI: 10.3321/j.issn:1000-2790.2003.02.014.
- [12] 岳玉娟, 任东升, 刘起勇. 2005-2013 年中国大陆登革热疫情时空分布 [J]. *疾病监测*, 2015, 30(7): 555-560. DOI: 10.3784/j.issn.1003-9961.2015.07.008.
Yue YJ, Ren DS, Liu QY. Temporal and spatial distribution of dengue fever in mainland China from 2005 to 2013 [J]. *Dis Surveillance*, 2015, 30(7): 555-560. DOI: 10.3784/j.issn.1003-9961.2015.07.008.
- [13] 宁文艳, 鲁亮, 任红艳. 2004-2013 年间中国登革热疫情时空变化分析 [J]. *地球信息科学学报*, 2015, 17(5): 614-621. DOI: 10.3724/SP.J.1047.2015.00614.
Ning WY, Lu L, Ren HY, et al. Analysis of temporal and spatial changes of dengue fever in China during 2004-2013 [J]. *Journal of Geo-information Science*, 2015, 17(5): 614-621. DOI: 10.3724/SP.J.1047.2015.00614.
- [14] 张红, 张恒端, 彭志强. 中国 2011-2015 年登革热疫情分析 [J]. *寄生虫与医学昆虫学报*, 2017, 24(2): 118-125. DOI: 10.3969/j.issn.1005-0507.2017.02.008.
Zhang H, Zhang HD, Peng ZQ, et al. Analysis of dengue fever in China from 2011 to 2015 [J]. *Acta Parasitol Med Entomol Sin*, 2017, 24(2): 118-125. DOI: 10.3969/j.issn.1005-0507.2017.02.008.
- [15] Li Z, Yin W, Clements A, et al. Spatiotemporal analysis of indigenous and imported dengue fever cases in Guangdong Province, China [J]. *BMC Infect Dis*, 2012, 12: 132. DOI: 10.1186/1471-2334-12-132.
- [16] 樊景春, 吴海霞, 王君. 广东省 2006-2011 年登革热时空分布特征 [J]. *中国媒介生物学及控制杂志*, 2013, 24(5): 389-391.
Fan JC, Wu HX, Wang J, et al. Temporal and spatial distribution of dengue fever in Guangdong Province from 2006 to 2011 [J]. *Chin J Vector Biol & Control*, 2013, 24(5): 389-391.
- [17] Bateson TF, Schwartz J. Who is sensitive to the effects of particulate air pollution on mortality? a case-crossover analysis of effect modifiers [J]. *Epidemiology*, 2004, 15(2): 143-149.

(收稿日期: 2019-05-30)

(修回日期: 2019-08-05)

(本文编辑: 王军平)