

多曲线季节指数模型分析广东省登革热流行时间特征及应用探讨

曾四清¹, 钟豪杰², 方艳², 代吉亚², 肖建鹏¹, 刘涛¹

1. 广东省疾病预防控制中心 广东省公共卫生研究院, 广东 广州 511430; 2. 广东省疾病预防控制中心公共卫生应急部

摘要: **目的** 建立广东省登革热疫情月份分布特征分析模型, 并利用该模型进行疫情预测和防控效果评估。 **方法** 收集广东省 1990–2016 年各月登革热监测报告发病数, 拟合多曲线季节指数模型。 **结果** 各月 (x) 发生登革热概率的拟合模型方程为: $S = 36.585 - 17.436x + 5.425x^2 - 0.318x^3$ ($S \geq 0\%$, $R^2 = 0.980$, $P = 0.000$); 多曲线季节指数模型为: $S = -0.001 + 0.008x - 0.003x^2 + 0.000x^3$ ($S \geq 0$, $x = 1, 2, 3, 4, 5, 6$, $R^2 = 0.992$, $P = 0.012$); $S = e^{-14.021 + 1.733x}$ ($S \geq 0$, $x = 7$, $R^2 = 0.981$, $P = 0.001$); $S = -189.883 + 41.304x - 2.181x^2$ ($S \geq 0$, $x = 8, 9, 10, 11$, $R^2 = 0.998$, $P = 0.048$); $S = 1.274E12 \times e^{-2.612x}$ ($S \geq 0$, $x = 12$, $R^2 = 0.995$, $P = 0.043$)。结果发现, 7–11 月为广东省登革热高发风险、高流行强度月份; 1995 年和 2014 年两年是广东省登革热异常暴发年; 从 2013 年开始, 广东省登革热疫情月度发生态势和流行特征发生悄然变化, 发生异常疫情的月份时间提前、频次增加、强度增大, 并可能引起下半年的疫情高发态势。采用该模型评估表明, 广东省 2015、2016 年登革热早防早控措施效果明显。 **结论** 多曲线模型拟合登革热流行季节指数效果较好, 该模型既可以用于登革热疫情的月份分布特征分析, 开展早期预测预警, 以便做到早发现早防控, 还可用于评估防控措施的效果。

关键词: 登革热; 季节指数; 多曲线模型; 预测; 评估

中图分类号: R512.8 文献标识码: B 文章编号: 1006-3110(2018)09-1137-05 DOI: 10.3969/j.issn.1006-3110.2018.09.034

Analysis of temporal distribution characteristics of dengue fever in Guangdong Province by multi-curve seasonal index model and its application probe

ZENG Si-qing*, ZHONG Hao-jie, FANG Yan, DAI Ji-ya, XIAO Jian-peng, LIU Tao

* Guangdong Provincial Institute of Public Health, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, Guangdong 511430, China

Abstract: **Objective** To establish a seasonal index model to describe the monthly distribution characteristics of dengue fever in Guangdong Province so as to predict the epidemic situation and evaluate the effect of prevention and control measures. **Methods** The reported monthly incidence data of dengue fever in Guangdong Province during 1990–2016 were collected, and multi-curve models were selected to simulate the seasonal index of dengue fever epidemic. **Results** The equation of simulation model for the monthly occurring rate of dengue fever was $S = 36.585 - 17.436x + 5.425x^2 - 0.318x^3$ ($S \geq 0\%$, $R^2 = 0.980$, $P = 0.000$). The equations of simulation model for seasonal index of dengue fever epidemic ran as follows: $S = -0.001 + 0.008x - 0.003x^2 + 0.000x^3$ ($S \geq 0$, $x = 1, 2, 3, 4, 5, 6$, $R^2 = 0.992$, $P = 0.012$); $S = e^{-14.021 + 1.733x}$ ($S \geq 0$, $x = 7$, $R^2 = 0.981$, $P = 0.001$); $S = -189.883 + 41.304x - 2.181x^2$ ($S \geq 0$, $x = 8, 9, 10, 11$, $R^2 = 0.998$, $P = 0.048$); $S = 1.274E12 \times e^{-2.612x}$ ($S \geq 0$, $x = 12$, $R^2 = 0.995$, $P = 0.043$). The results revealed that there were much higher occurring risk and higher incidence level of dengue fever in July–November in Guangdong Province as well as great abnormal outbreaks of dengue fever in 1995 and 2014. The monthly incidence tendency and temporal distribution characteristics of dengue fever epidemic in Guangdong Province had gradually changed since 2013. The month with abnormal outbreaks of dengue fever moved forward, the outbreak frequency increased, the intensity was enlarged, and these changes tended to lead to a high epidemic in the second half of a year. The results evaluated by the model revealed that the actions of early prevention and control had significant effects on preventing the epidemics of dengue fever in Guangdong Province in 2015–2016.

Conclusions Multi-curve models are fit for the simulation of seasonal index of dengue fever. The models can not only be used to describe the monthly temporal distribution characteristics of dengue fever, early detect and predict the situation of the epidemic to achieve early discovery and early prevention and control, but also be used to evaluate the effect of prevention and control measures.

Key words: dengue fever; seasonal index; multi-curve model; prediction; evaluation

基金项目: 广东省科技计划项目 (2014A040401041)

作者简介: 曾四清 (1968–), 男, 湖北黄冈人, 硕士, 主任医师, 研究方向: 传染病预防控制。

登革热是目前面临的一个全球性公共卫生挑战,严重威胁了热带-亚热带地区公共卫生安全及社会经济发展。近三十年来,登革热在超过 100 个国家和地区流行^[1];近五十年来,全球登革热的发病率上升了 30 倍^[2]。广东省是登革热高发地区。目前,关于广东省登革热的研究主要集中在流行病学特征及影响因素、登革病毒、蚊媒等方面^[3-10]。本研究主要利用 1990-2016 年广东省登革热监测病例每月数据,拟合多曲线季节指数模型,分析登革热疫情的月份分布特征,并利用该模型进行登革热疫情预测和防控效果评估。

1 资料与方法

1.1 资料来源 登革热监测病例资料来源于《广东省疫情资料汇编》、中国疾病预防控制中心信息系统,收集广东省 1990-2016 年各年各月登革热监测报告发病数。

1.2 方法

1.2.1 月发生概率和季节指数计算方法 通过月发生概率和季节指数分析登革热发病水平以年为周期的按月发生和波动规律。月发生概率为当月监测报告病例的年度数占观察年度数的百分比。季节指数计算方法:用某月的平均数除以总的月平均数即得到相应月份的季节指数(计算公式(1))^[11]。式中,

$$S_K = (\sum_{i=1}^n x_{ik} / n) / (\sum_{i=1}^n \sum_{k=1}^m x_{ik} / nm)$$

公式(1)

$k=1,2,\cdots,m;m=12$ (月份数); $i=1,2,\cdots,n;n$ 为年度数。

1.2.2 月发生概率和季节指数模型拟合方法 对月发生概率和季节指数分别进行月趋势多曲线拟合。用于拟合筛选的模型包括:生长模型 $S=e^{(b_0+b_1x)}$ 、指数模型 $S=b_0e^{b_1x}$ 、二次模型 $S=b_0+b_1x+b_2x^2$ 、三次模型 $S=b_0+b_1x+b_2x^2+b_3x^3$ 、对数模型 $S=b_0+b_1\ln(x)$ 等。式中, x 为月份数, $x=1,2,\cdots,12$ 。计算每种模型的决定系数 R^2 ,用 F 检验作模型拟合优度分析,筛选最佳模型^[12]。

1.2.3 应用季节指数模型进行预测和评估 根据筛选出的多曲线拟合模型,计算各月季节指数的预测值

及其 95%可信区间,进而计算各月发病数的理论预测值及其 95%可信区间,将实际每年各月的监测报告病例数与预测病例数的 95%可信区间上限进行对比评估。理论上超过的概率小于 2.5%,超过此范围者视为小概率疫情事件。据此,如果某月报告病例数超出其预测值 95%可信区间上限,则该月疫情视为异常;如果某年报告病例数超出其预测值 95%可信区间上限,则该年疫情视为异常。

1.3 统计分析 以上统计分析采用 SPSS 19.0 完成。

2 结果

2.1 各年监测病例数月份分布 1990-2016 年广东省每年各月登革热监测报告病例数见表 1。27 年月平均病例数最高是 10 月份,其次是 9 月份、8 月份;月平均病例数最低是 1 月份,其次是 3 月份、4 月份。

2.2 月发生概率模型 每月发生概率原始计算结果见表 1。月发生概率最高是 9 月份和 10 月份,最低是 1 月份和 2 月份。经过多曲线模型拟合,筛选得到各月发生登革热概率的拟合模型为三次方模型,方程为: $S=36.585-17.436x+5.425x^2-0.318x^3$ ($S\geq 0\%$),决定系数 $R^2=0.980$,显著性检验 $F=127.602$, $P=0.000$ 。模拟曲线见图 1。根据模型计算各月登革热发生概率预测值及其可信区间,结果见表 2。

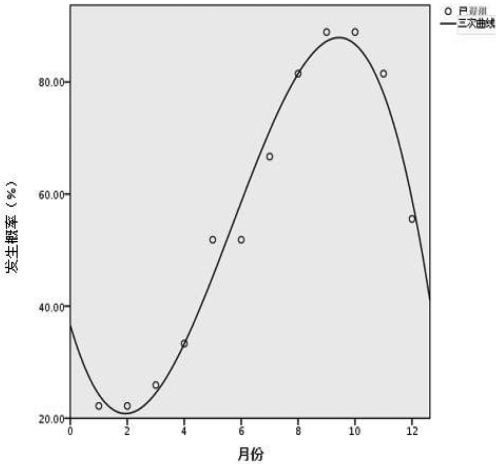


图 1 广东省登革热各月发生概率拟合曲线图

表 1 广东省 1990-2016 年登革热监测病例数月份分布

年度	总病例数	月份											
		1	2	3	4	5	6	7	8	9	10	11	12
1990	374	0	0	0	0	0	0	0	26	172	147	25	4
1991	371	0	0	0	1	2	1	4	12	215	132	4	0
1992	2	0	0	1	0	1	0	0	0	0	0	0	0
1993	359	1	0	0	0	0	0	0	0	17	218	112	11

续表 1

年度	总病例数	月份											
		1	2	3	4	5	6	7	8	9	10	11	12
1994	4	0	0	0	0	0	0	0	0	0	4	0	0
1995	6 812	0	1	0	0	1	2	179	2 219	3 679	609	96	26
1996	0	0	0	0	0	0	0	0	0	0	0	0	0
1997	632	0	0	0	0	0	0	0	9	379	209	34	1
1998	480	0	0	0	0	0	0	1	2	59	368	45	4
1999	290	0	0	0	0	0	0	2	150	32	91	15	0
2000	384	0	0	0	0	1	1	0	75	182	87	38	0
2001	344	0	0	0	0	0	0	0	73	119	124	27	1
2002	1 348	0	0	0	0	2	8	92	342	629	232	43	0
2003	42	0	0	0	0	0	4	13	5	12	8	0	0
2004	47	0	0	0	0	0	0	1	1	37	7	1	0
2005	6	0	0	0	0	0	0	2	0	3	0	1	0
2006	1 010	0	0	0	0	0	3	10	275	425	223	73	1
2007	397	0	1	1	1	1	3	1	38	111	237	2	1
2008	87	0	0	0	3	0	2	3	7	3	44	22	3
2009	19	3	0	0	0	1	0	0	5	3	5	0	2
2010	139	1	0	0	2	1	0	4	11	44	66	9	1
2011	49	0	0	0	0	1	2	2	3	7	26	8	0
2012	474	0	1	3	1	4	1	10	20	123	227	84	0
2013	2 894	0	5	2	2	2	8	151	196	727	1 449	349	3
2014	45 189	2	0	1	3	5	24	252	1 841	18 582	22 755	1 646	78
2015	1 685	8	6	6	5	8	8	15	58	1 141	344	77	9
2016	439	3	14	6	7	9	10	25	37	203	78	43	4
发生年数		6	6	7	9	14	14	18	22	24	24	22	15
月发生率(%)		22.22	22.22	25.93	33.33	51.85	51.85	66.67	81.48	88.89	88.89	81.48	55.56
月平均病例数		0.67	1.04	0.74	0.93	1.44	2.85	28.41	200.19	996.44	1 025.56	102.00	5.52
季节指数		0.003 4	0.005 3	0.003 8	0.004 7	0.007 3	0.014 5	0.144 1	1.015 4	5.054 2	5.201 9	0.517 4	0.028 0

注:总月平均病例数为 197.15。

表 2 广东省登革热各月发生概率(%)预测值及其可信区间

月份	实际值	预测值	残差	预测值 95%可信区间
1	22.22	24.26	-2.04	11.03~37.48
2	22.22	20.87	1.35	9.59~32.15
3	25.93	24.51	1.42	13.33~35.68
4	33.33	33.28	0.05	22.03~44.53
5	51.85	45.27	6.58	34.17~56.37
6	51.85	58.57	-6.72	47.66~69.48
7	66.67	71.27	-4.60	60.36~82.18
8	81.48	81.46	0.02	70.36~92.56
9	88.89	87.24	1.65	75.99~98.49
10	88.89	86.70	2.19	75.53~97.87
11	81.48	77.93	3.55	66.65~89.21
12	55.56	59.02	-3.46	45.79~72.24

2.3 季节指数模型分析 登革热监测发病数季节指

数原始计算结果见表 1,其分布见图 2。

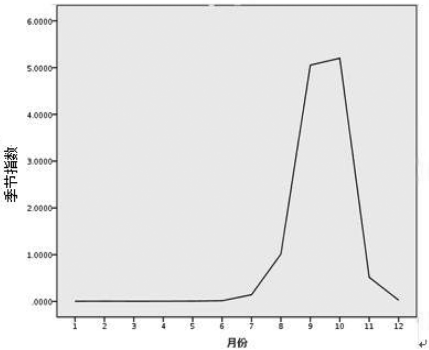


图 2 广东省登革热监测病例数季节指数

经过观察原始季节指数线图(图 2)特征,反复尝试分段(为保证每个模型中有一定数量的数据点值,并反映趋势的延续性,分段有重叠)拟合多种曲线,根据决定系数大小,分别筛选出 1-6 月、5-9 月、8-11 月、10-12 月的拟合模型为:三次方模型、生长模型、二次

方模型和指数模型。当有多个模型重叠模拟同一月份的季节指数时,根据最小残差原则确定该月最终采纳的预测模型和预测值。

2.3.1 1-6 月季节指数拟合三次方模型 见图 3。模型方程为 $S = -0.001 + 0.008x - 0.003x^2 + 0.000x^3$ ($S \geq 0$), $R^2 = 0.992$, $F = 80.847$, $P = 0.012$ 。用该模型计算 1-6 月季节指数预测值及其可信区间,结果见表 3。

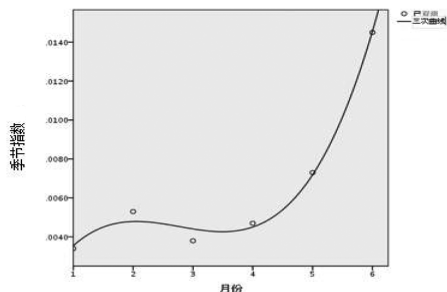


图 3 广东省登革热病例数 1-6 月季节指数拟合三次方曲线

2.3.2 5-9 月季节指数拟合生长模型 见图 4。模型方程为 $S = e^{-14.021 + 1.733x}$ ($S \geq 0$), $R^2 = 0.981$, $F = 150.928$, $P = 0.001$ 。选用该模型计算 7 月份季节指数预测值及其可信区间,结果见表 3。

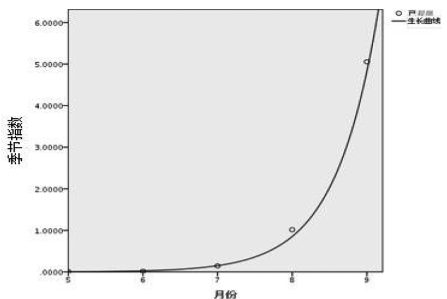


图 4 广东省登革热病例数 5-9 月季节指数拟合生长曲线

2.3.3 8-11 月季节指数拟合二次方模型 见图 5。二次模型方程为 $S = -189.867 + 41.301x - 2.181x^2$ ($S \geq 0$), $R^2 = 0.998$, $F = 215.821$, $P = 0.048$ 。用该模型计算 8-11 月季节指数预测值及其可信区间,结果见表 3。

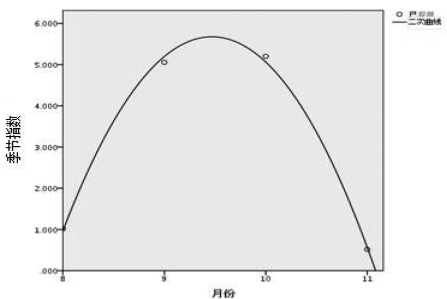


图 5 广东省登革热病例数 8-11 月季节指数拟合二次曲线

2.3.4 10-12 月季节指数拟合指数模型 见图 6。指数模型方程为 $S = 1.274E12 \times e^{-2.612x}$ ($S \geq 0$), $R^2 = 0.995$, $F = 221.050$, $P = 0.043$ 。选用该模型计算 12 月份季节指数预测值及其可信区间,结果见表 3。

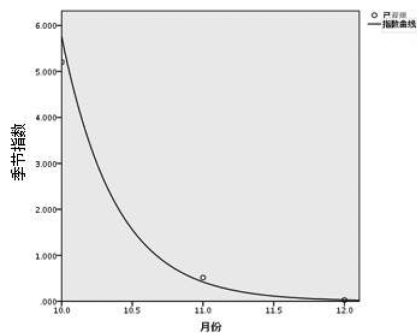


图 6 广东省登革热病例数 10-12 月季节指数模拟指数曲线

综合分析得到广东省登革热发病季节指数多曲线模型如下:

$$S = \begin{cases} -0.001 + 0.008x - 0.003x^2 + 0.000x^3 & (S \geq 0, x = 1, 2, 3, 4, 5, 6) \\ e^{-14.021 + 1.733x} & (S \geq 0, x = 7) \\ -189.867 + 41.304x - 2.181x^2 & (S \geq 0, x = 8, 9, 10, 11) \\ 1.274E12 \times e^{-2.612x} & (S \geq 0, x = 12) \end{cases}$$

表 3 广东省登革热病例季节指数拟合模型预测值及其可信区间

月份	季节指数 实际值	季节指数预测值 (95%可信区间)	残差	预测病例数 95%可信区间	当月实际病例数 超过 95%上限的年度
1	0.0034	0.00354 ^a (-0.00003/0 * ~0.00712)	-0.00014	0~1	2009, 2014, 2015 2016(4 年)
2	0.0053	0.00479 ^a (0.00158~0.00800)	0.00051	0~2	2013, 2015, 2016 (3 年)
3	0.0038	0.00441 ^a (0.00132~0.00750)	-0.00061	0~1	2012, 2013, 2015 2016(4 年)
4	0.0047	0.00451 ^a (0.00142~0.00759)	0.00019	0~1	2008, 2010, 2013 2014, 2015, 2016(6 年)
5	0.0073	0.00719 ^a (0.00398~0.01040)	0.00011	1~2	2012, 2014, 2015 2016(4 年)
6	0.0145	0.01456 ^a (0.01099~0.01814)	-0.00006	2~4	2002, 2013, 2014 2015, 2016(5 年)
7	0.1441	0.15091 ^b (0.03187~0.71462)	-0.00681	6~141	1995, 2013, 2014 (3 年)
8	1.0154	0.96834 ^c (-2.76548/0 * ~4.70217)	0.04706	0~927	1995, 2014 (2 年)
9	5.0542	5.19536 ^c (1.86645~8.52428)	-0.14116	368~1681	1995, 2014 (2 年)
10	5.2019	5.06073 ^c (1.73182~8.38965)	0.14117	341~1654	2014 (1 年)
11	0.5174	0.56445 ^c (-3.16937/0 * ~4.29828)	-0.04705	0~847	2014 (1 年)
12	0.0280	0.03099 ^d (0.00043~2.22734)	-0.00299	0~439	无

注: * 负数为理论值,实际最小值为 0; a 三次方模型、b 生长模型、c 二次方模型、d 指数模型。

2.4 模型应用 将各月筛选出的模拟季节指数预测值的 95%上限求和,合计为 28.915 09,按照总月平均数为 197.15 计算,全年病例数 95%预测值的上限合计为 5 700 例。对照表 1 可见,1995 年和 2014 年两年的病例数超过此数,可以视为异常暴发年份。其余年份为常规流行。

以各月预测病例数 95%可信区间上限为标准(表 3),将每年各月的实际病例数(表 1)与之对照,当月实

实际病例数超过 95% 上限的年度见表 3。27 年中, 1995 年有 3 个月(7、8、9 月)、2002 年有 1 个月(6 月)、2008 年有 1 个月(4 月)、2009 年有 1 个月(1 月)、2010 年有 1 个月(4 月)、2012 年有 2 个月(3、5 月)、2013 年有 5 个月(2、3、4、6、7 月)、2014 年有 9 个月(1、4、5、6、7、8、9、10、11 月)、2015 年有 6 个月(1、2、3、4、5、6 月) 和 2016 年有 6 个月(1、2、3、4、5、6 月) 的实际发生病例数超过预测值上限。理论上, 超过预测值上限发生的概率小于 2.5%, 属于小概率异常疫情事件。

3 讨论

登革热是登革热病毒引起、伊蚊传播的一种急性传染病。郭汝宁等^[3] 2013 年研究发现, 广东省登革热发病呈明显的周期性, 间隔 4~7 年出现 1 个流行高峰; 在历经 2007-2011 年为期 5 年的相对低水平发病之后, 自 2012 年开始广东省登革热发病呈上升趋势, 提示之后两年广东省可能面临较高的登革热发病风险^[3]。本研究模型拟合结果与该研究者的分析预测提示及实际发生情况高度吻合。研究发现, 广东省登革热全年 12 个月均有可能发生, 但各月发生风险和强度明显不同。1、2、3 月份发生风险较低、强度弱, 4、5、6、12 月为中等风险、中等强度, 7、8、9、10、11 月为高风险、高强度。同时, 上半年疫情异常波动的概率更大, 可以根据 1-6 月份的发生态势, 对下半年的疫情形势进行预警, 以便及早采取应对措施。

研究发现, 1995 年和 2014 年两年是广东省登革热疫情异常暴发年; 而疫情异常出现的月份在 1995 年之后有几年的相对静息期, 从 2013 年开始, 广东省登革热疫情月度发生态势和流行特征发生悄然变化, 异常发生月份时间提前、频次增加、强度增大, 提示可能引起下半年的疫情高发态势。2013 年, 广州市发生了近十几年规模较大的登革热疫情^[13]。对此, 本应高度警惕, 早防早控。预防策略方面应提前做好常规防控和异常暴发防控两手准备。但由于各种原因, 2014 年上述早防早控及暴发处置措施未能落实到位, 从而出现了 2014 年下半年的全省范围的更大规模的登革热流行^[14]。经历了 2014 年的经验和教训之后, 尽管 2015 年、2016 年上半年的来势比 2014 年有过之而无不及, 但由于防控措施及时有力, 并没有引起下半年的大规模流行, 防控成效显著。因此, 采用多曲线模型分段拟合广东省登革热流行的季节指数, 取得了较好的模拟效果。该模型既可以用于登革热疫情的月份分布特征分析, 开展早期预测预警, 为早发现、早防控提供科学依据, 还可以用于评估防控措施的效果。研究提示, 早

防早控十分必要, 效果明显, 应成为应对近几年广东省登革热流行时间特征变化的常态措施。

广东省登革热发病呈明显的周期性、季节性和空间分布非随机性^[3-5, 15]。如果仅以年为时间粒度、以省为地域粒度, 采用单一模型分析, 难以全面、深入、真实了解登革热的流行特征。因此, 本研究以年度-月份为时间粒度, 尝试采用多曲线模型拟合分析方法, 在分析方法上进行探索, 结果证明是可行的, 并具有一定的应用价值。如果时间粒度从月到周, 甚至是日; 地区粒度从省到地市, 甚至是县区, 进行多曲线模型拟合分析, 将能更精准地确定高风险地区/风险点、高风险时段/时间和流行强度, 从而研究编制网格化登革热风险预测预警时空动态地图, 对精准防控登革热将更有实践指导意义。也可以尝试将该方法推广到其它的传染病相关研究。不过, 该方法假设每年各月的病例数服从正态分布或近似正态分布, 以算术平均数代表各月的平均水平; 同时, 对季节指数进行分段拟合曲线模型时, 为了充分反映季节指数时间趋势的延续性, 模型采用的时点数据不宜过少, 因此重叠使用了部分临近的时点数据。此两方面的适用性及其对分析结果的影响有待深入探讨。

参考文献

- [1] 张复春. 登革热: 一个日益严重的全球性公共卫生问题[J]. 实用医学杂志, 2011, 27(19): 3459-3461.
- [2] World Health Organization. Global strategy for dengue prevention and control 2012-2020[R]. Geneva: WHO, 2012: 1-2.
- [3] 郭汝宁, 彭志强, 宋铁, 等. 广东省 1990-2012 年登革热和基孔肯雅热感染状况及流行风险分析[J]. 中华流行病学杂志, 2014, 35(2): 167-169.
- [4] 梁文佳, 何剑峰, 罗会明, 等. 广东省 2001-2006 年登革热流行病学分析[J]. 华南预防医学, 2007, 33(1): 4-7.
- [5] 樊景春, 林华亮, 吴海霞, 等. 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): 389-392.
- [6] 严子锦, 胡志刚, 江毅民, 等. 广州市登革热媒介的监测分析[J]. 中华卫生杀虫药械, 2010, 16(1): 42-44.
- [7] 杨芬, 马绍强, 何剑峰, 等. 广东和香港地区 2004-2006 年输入性登革热病例流行病学分析[J]. 中华流行病学杂志, 2009, 30(1): 42-44.
- [8] 易彬橙, 张治英, 徐德忠, 等. 广东省登革热及媒介种群的空间分布[J]. 第四军医大学学报, 2003, 24(17): 1623-1626.
- [9] 易彬橙, 张治英, 徐德忠, 等. 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.
- [10] 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.
- [11] 杨维中. 传染病预警理论与实践[M]. 北京: 人民卫生出版社, 2012: 78-87.
- [12] 陈平雁. SPSS 13.0 统计软件应用教程[M]. 北京: 人民卫生出版社, 2005: 186-190.
- [13] 罗小华, 黄昱, 甘圳尧, 等. 广州市黄埔区 2013 年登革热流行病学分析及防控效果[J]. 实用预防医学, 2014, 21(7): 807-810.
- [14] Xiao JP, He JF, Deng AP, et al. Characterizing a large outbreak of dengue fever in Guangdong Province, China[J]. Infect Dis Poverty, 2016, 5(1): 44.
- [15] Liu C, Liu Q, Lin H, et al. Spatial analysis of dengue fever in Guangdong province, China, 2001-2006[J]. Asia-Pac J Public Health, 2014, 26(1): 58-66.