

· 论 著 ·

传染病综合指数预测研究

尹锡玲, 代文灿, 王松, 周勇, 朱克京, 梁小冬, 李德云, 谭爱军

珠海市疾病预防控制中心, 广东 珠海 519000

摘要: **目的** 应用自回归移动平均模型 (ARIMA) 预测传染病风险, 计算传染病综合指数 (IDI), 为公众提供传染病风险预报。**方法** 收集珠海市 2014 年第 1 周至 2018 年第 14 周流感样病例、手足口病、其他感染性腹泻资料和 2016 年第 1 周至 2018 年第 14 周布雷图指数资料, 分别建立 ARIMA 模型预测传染病风险指数, 并采用熵权法分析 4 类传染病权重, 计算 IDI; 以珠海市 2018 年第 15—19 周 4 种传染病指标进行验证。**结果** 预测期间处于夏季, 则 $IDI = ROUND(0.33 \times \text{流感样病例比例风险指数} + 0.47 \times \text{手足口病发病率风险指数} + 0.10 \times \text{其他感染性腹泻发病率风险指数} + 0.10 \times \text{布雷图指数风险指数})$ 。预测 2018 年第 15—19 周珠海市全市和香洲区 IDI 均为 2 (较安全), 斗门区和金湾区 IDI 均为 1 (安全)。第 15—19 周全部风险值一致率分别为 97.50%、95.00%、97.50%、85.00% 和 77.50%。**结论** IDI 模型可用于传染病流行风险的短期预报。

关键词: 传染病; 预测; 自回归移动平均模型; 熵权法

中图分类号: R181.3 **文献标识码:** A **文章编号:** 2096-5087 (2019) 09-0897-04

Application of infectious disease index to prediction of infectious diseases

YIN Xi-ling, DAI Wen-can, WANG Song, ZHOU Yong, ZHU Ke-jing, LIANG Xiao-dong, LI De-yun, TAN Ai-jun
Zhuhai Center for Disease Control and Prevention, Zhuhai, Guangdong 519000, China

Abstract: Objective To establish a prediction model for infectious disease index (IDI) by autoregressive integrated moving average (ARIMA), and to provide forecast of infectious diseases to the public. **Methods** The data of the percentage of influenza-like illness (ILI), the incidence rates of hand-foot-mouth disease (HFMD) and other infectious diarrhea (OID) from the 1st week of 2014 to the 14th week of 2018, and Breteau index (BI) from the 1st week of 2016 to the 14th week of 2018 were collected. ARIMA models were built to predict the risk indicators of ILI, HFMD, OID and BI. The weights of the four indicators were evaluated seasonally by the entropy weight method. Then the IDI was calculated and the data of ILI, HFMD, OID and BI from 15th to 19th week in 2018 was used for verification. **Results** The forecast was in summer, so $IDI = ROUND(0.33 \times \text{risk index of ILI percentage} + 0.47 \times \text{risk index of HFMD incidence} + 0.10 \times \text{risk index of OID incidence} + 0.10 \times \text{risk index of BI})$. The predicted IDI would be 2 (less safe) in the whole city and Xiangzhou District, and 1 (safe) in Doumen District and Jinwan District. The consistency rates of IDI prediction was 97.50%, 95.00%, 97.50%, 85.00% and 77.50% from 15th to 19th week in 2018, respectively. **Conclusion** It was feasible to use IDI for short-term risk prediction of infectious diseases.

Key words: Infectious disease; Prediction; Autoregressive integrated moving average model; Entropy weight method

预防和应对传染病的有效措施不仅取决于对疫情现状的及时掌握, 还取决于对疫情发生、发展的预测^[1]。基于传染病监测数据早期预测传染病的暴发^[2], 并及时向公众预警, 对于提高公众健康意识, 及时采取防制措施, 减少发病和死亡具有积极意义。目前的

传染病预测研究多集中于单一病种, 采用指标值来反映传染病传播风险^[3-5], 而对传染病综合评价指数的研究较少。综合指数可以综合多个指标定量反映传染病疫情变化, 是一种可以无量纲比较的相对数^[6]。本研究收集珠海市发病率较高、聚集性疫情较多的流行性感、手足口病、其他感染性腹泻和登革热疫情资料, 以流感样病例比例、手足口病发病率、其他感染性腹泻发病率和布雷图指数作为风险评价指标, 采用自回归移动平均模型 (autoregressive integrated

DOI: 10.19485/j.cnki.issn2096-5087.2019.09.008

作者简介: 尹锡玲, 硕士, 主管医师, 主要从事疾病预防控制中心工作

通信作者: 谭爱军, E-mail: zhtanaijun@gmail.com

moving average model, ARIMA) 和熵权法计算传染病综合指数 (infectious disease index, IDI), 现将结果报道如下。

1 资料与方法

1.1 资料来源 2014年第1周至2018年第19周流感样病例资料来源于流感监测信息系统; 2014年第1周至2018年第19周手足口病、其他感染性腹泻的发病率(1/10万)来源于传染病报告管理信息系统; 2016年第1周至2018年第19周布雷图指数监测资料来源于蚊媒密度现场快速评估; 珠海市常住人口数来源于2016年免疫规划年报表。

1.2 方法 构建ARIMA模型, 对流感样病例比例、手足口病发病率、其他感染性腹泻发病率和布雷图指数进行单病种/单因素预测, 将预测结果对照指标分级标准判定风险等级, 根据不同季节各指标权重下的IDI计算公式得到IDI, 比较IDI的预测值与实际值以评估预测的一致性。

1.2.1 建立ARIMA预测模型 利用2014年第1周至2018年第14周流感样病例比例、手足口病发病率和其他感染性腹泻发病率, 以及2016年第1周至2018年第14周布雷图指数, 构建ARIMA(p, d, q)预测模型, 对2018年第15—19周的上述4项指标进行预测。

1.2.2 划分传染病风险等级 2014—2017年珠海市流感样病例比例、手足口病发病率和其他感染性腹泻发病率资料均呈偏态分布, 采用百分位数 P_{40} 、 P_{60} 、 P_{80} 划分风险等级: $\leq P_{40}$ 为安全, 风险指数=1; $> P_{40}$ 且 $\leq P_{60}$ 为较安全, 风险指数=2; $> P_{60}$ 且 $\leq P_{80}$ 为低风险, 风险指数=3; $> P_{80}$ 为高风险, 风险指数=4。布雷图指数风险等级判定标准: ≤ 5 为安全, 风险指数=1; > 5 且 ≤ 10 为较安全, 风险指数=2; > 10 且 ≤ 20 为低风险, 风险指数=3; > 20 为高风险, 风险指数=4。

1.2.3 评估指标权重 采用熵权法^[7], 利用珠海市2016—2017年按季节及周次流感样病例比例、手足口病发病率、其他感染性腹泻发病率和布雷图指数评估指标权重。2—4月为春季, 5—7月为夏季, 8—10月为秋季, 11至次年1月为冬季。指标信息熵计算公式如下:

$$E_j = -K \sum_{i=1}^m P_{ij} \ln(P_{ij})$$

式中, E_j 表示信息熵, 常数 $K=1/\ln(m)$, P_{ij} 表

示第 j 个属性下第 i 个方案的贡献度, m 为方案数, n 为属性数。若某个指标的 E_j 越小, 表示变异程度越大, 提供的信息量越多, 在综合评价中起到的作用越大, 权重也就越大。

通过信息熵计算各指标的权重(W_j)公式如下:

$$W_j = \frac{1 - E_j}{\sum_{j=1}^n (1 - E_j)}$$

1.2.4 计算IDI 根据各个指标的权重(W_j)和风险指数计算IDI, $IDI = \text{ROUND}(W_1 \times \text{流感样病例比例风险指数} + W_2 \times \text{手足口病发病率风险指数} + W_3 \times \text{其他感染性腹泻发病率风险指数} + W_4 \times \text{布雷图指数风险指数})$ 。若 $IDI=1$, 风险等级为安全; $IDI=2$, 风险等级为较安全; $IDI=3$, 风险等级为低风险; $IDI=4$, 风险等级为高风险。

1.2.5 评价预测效果 预测值与实际值进行对比, 计算预测一致率。

2 结果

2.1 4种传染病风险指数和不同季节权重 不同季节4种传染病风险预测指标的权重见表1, 风险指数分级标准见表2。本次预测2018年第15—19周传染病风险, 处于夏季, 则 $IDI = \text{ROUND}(0.33 \times \text{流感样病例比例风险指数} + 0.47 \times \text{手足口病发病率风险指数} + 0.10 \times \text{其他感染性腹泻发病率风险指数} + 0.10 \times \text{布雷图指数风险指数})$ 。

表1 不同季节4种传染病风险预测指标权重

季节	流感样病例比例	手足口病发病率	其他感染性腹泻发病率	布雷图指数
春	0.06	0.53	0.13	0.28
夏	0.33	0.47	0.10	0.10
秋	0.10	0.73	0.05	0.12
冬	0.19	0.51	0.30	

注: 冬季不监测布雷图指数。

2.2 建立传染病ARIMA预测模型

2.2.1 参数估计和模型诊断 针对珠海市全市及各区分4种传染病指标分别拟合最佳模型, 平均绝对百分比误差(MAPE)为13.28~73.98, 最大绝对百分比误差(Max APE)为73.45~1302.81, 标准化贝叶斯信息准则为-2.22~3.16, R^2 为0.40~0.92, 残差为随机误差, 所建模型拟合优度均较好。

2.2.2 模型预测 预测2018年第15—19周流感样病

表2 风险指数分级标准

指标	风险指数			
	1 (安全)	2 (较安全)	3 (低风险)	4 (高风险)
流感样病例比例 (%)	≤1.86	> 1.86 且 ≤2.42	> 2.42 且 ≤3.53	> 3.53
手足口病发病率 (1/10 万)	≤5.78	> 5.78 且 ≤10.54	> 10.54 且 ≤25.84	> 25.84
其他感染性腹泻发病率 (1/10 万)	≤4.50	> 4.50 且 ≤5.17	> 5.17 且 ≤7.42	> 7.42
布雷图指数	≤5	> 5 且 ≤10	> 10 且 ≤20	> 20

例比例、手足口病发病率、其他感染性腹泻发病率和布雷图指数，实际值均在预测值的 95% 置信区间

(95%CI) 内，随着预测时间延长，预测值的 95%CI 范围扩大。见表 3。

表3 珠海市 2018 年第 15—19 周 4 种传染病指标的 ARIMA 模型预测值与实际值

时间	流感样病例比例 (%)		手足口病发病率 (1/10 万)		其他感染性腹泻发病率 (1/10 万)		布雷图指数	
	预测值 (95%CI)	实际值	预测值 (95%CI)	实际值	预测值 (95%CI)	实际值	预测值 (95%CI)	实际值
第 15 周								
全市	3.22 (2.27 ~ 4.43)	2.68	1.63 (0.08 ~ 4.40)	2.00	6.86 (3.44 ~ 12.32)	6.14	2.42 (0.16 ~ 4.69)	1.57
香洲区	4.32 (2.96 ~ 6.09)	3.53	2.11 (0.06 ~ 5.97)	2.63	9.42 (4.72 ~ 16.92)	8.33	3.08 (0 ~ 6.35)	1.14
斗门区	1.20 (0.75 ~ 1.81)	1.06	0.87 (0 ~ 8.76)	0.96	2.49 (0 ~ 6.98)	2.87	3.12 (0 ~ 7.23)	2.27
金湾区	1.15 (0.60 ~ 1.82)	0.93	0.91 (0 ~ 9.81)	1.53	4.63 (1.25 ~ 9.43)	3.98	1.30 (0 ~ 3.76)	0.42
第 16 周								
全市	3.26 (1.98 ~ 5.09)	2.53	2.03 (0.11 ~ 7.41)	2.66	7.07 (2.90 ~ 14.55)	6.40	2.42 (0.16 ~ 4.69)	2.38
香洲区	4.39 (2.55 ~ 7.08)	3.34	2.64 (0.19 ~ 9.82)	3.26	9.70 (4.03 ~ 19.81)	8.96	3.08 (0 ~ 6.35)	3.03
斗门区	1.22 (0.68 ~ 2.01)	0.99	0.87 (0 ~ 13.56)	2.29	2.61 (0 ~ 7.99)	3.25	3.12 (0 ~ 7.23)	2.94
金湾区	1.17 (0.43 ~ 2.15)	0.84	0.90 (0 ~ 13.49)	1.23	4.70 (0.98 ~ 10.25)	2.76	1.30 (0 ~ 3.76)	1.06
第 17 周								
全市	3.31 (1.78 ~ 5.66)	3.07	2.50 (0.69 ~ 10.42)	3.73	7.29 (2.54 ~ 16.70)	5.83	2.47 (0 ~ 5.67)	2.38
香洲区	4.47 (2.27 ~ 7.94)	4.14	3.21 (0.94 ~ 13.50)	4.16	9.98 (3.54 ~ 22.59)	8.24	3.14 (0 ~ 7.75)	3.03
斗门区	1.23 (0.62 ~ 2.20)	1.15	0.87 (0 ~ 16.99)	4.01	2.70 (0 ~ 8.41)	2.10	3.26 (0 ~ 8.37)	2.94
金湾区	1.19 (0.32 ~ 2.43)	0.76	0.89 (0 ~ 16.31)	1.84	4.77 (0.76 ~ 11.02)	3.68	1.48 (0 ~ 4.48)	1.06
第 18 周								
全市	3.36 (1.62 ~ 6.19)	3.21	2.99 (1.57 ~ 13.34)	9.47	7.52 (2.24 ~ 18.83)	5.93	2.47 (0 ~ 5.67)	3.13
香洲区	4.55 (2.06 ~ 8.75)	4.28	3.80 (2.00 ~ 16.98)	10.86	10.28 (3.15 ~ 25.34)	8.15	3.14 (0 ~ 7.75)	4.00
斗门区	1.25 (0.58 ~ 2.37)	1.43	0.87 (0 ~ 19.81)	9.93	2.75 (0.00 ~ 8.60)	2.29	3.26 (0 ~ 8.37)	3.20
金湾区	1.21 (0.24 ~ 2.67)	0.83	0.89 (0 ~ 18.69)	3.98	4.84 (0.58 ~ 11.74)	4.29	1.48 (0 ~ 4.48)	1.70
第 19 周								
全市	3.41 (1.50 ~ 6.70)	3.21	3.49 (2.63 ~ 16.17)	11.26	7.76 (2.02 ~ 20.97)	4.86	2.51 (0 ~ 6.43)	3.13
香洲区	4.62 (1.90 ~ 9.54)	4.26	4.39 (3.25 ~ 20.32)	13.22	10.58 (2.84 ~ 28.10)	6.97	3.19 (0 ~ 8.84)	4.00
斗门区	1.26 (0.54 ~ 2.55)	1.41	0.87 (0 ~ 22.27)	9.93	2.79 (0 ~ 8.70)	1.34	3.36 (0 ~ 8.95)	3.20
金湾区	1.23 (0.17 ~ 2.89)	0.77	0.88 (0 ~ 20.78)	6.74	4.91 (0.44 ~ 12.43)	3.37	1.60 (0 ~ 4.85)	1.70

2.3 预测效果评价 预测 2018 年第 15—19 周珠海市全市及香洲区 IDI 为 2 (较安全)，斗门区、金湾区 IDI 为 1 (安全)；第 15—17 周预测 IDI 与实际 IDI 一致；第 18 周香洲区预测 IDI 为 2 (较安全)，实际 IDI 为 3 (低风险)；第 19 周全市及香洲区预测 IDI 为 2 (较安全)，实际 IDI 为 3 (低风险)；第 15—19 周全部风险值一致率分别为 97.50%、95.00%、

97.50%、85.00%和 77.50%。

3 讨论

本研究在百分位数法划分风险等级的基础上，通过熵权法和 ARIMA 模型计算 IDI，同时预测了珠海市全市及 3 个行政区 5 个周的 IDI，预测 IDI 与实际 IDI 基本一致，研究方法可行。许多传染病存在季节

性流行特征^[8], 在建模时应根据季节性特点, 并结合本地的具体情况。熵权法在指标量化评估方面有较多应用^[9], 可根据指标变异性的来确定客观权重, 排除主观性影响, 且可按照季节进行权重评定。评估指标权重结果基本反映本地相关传染病的季节性特点: 手足口病各个季节的权重均较大, 实际上手足口病在珠海市常年均有发生^[10], 夏秋季常发生聚集性疫情; 流感样病例在夏季、冬季权重较大; 其他感染性腹泻在冬季、春季权重较大。

ARIMA 模型对数据的要求不高, 具有较好的拟合效果, 适用于近期或短期预测^[11-15]。若外推时间延长, 预测值的 95%CI 增大, 预测精度降低, 因此实际预报中应逐 2~3 周计算相应指数。本研究所建 ARIMA 模型拟合优度较好, 根据风险指数分级标准转换为 IDI 后, 预测前 3 周全部风险值一致率均 $\geq 95\%$, 预测 IDI 与实际 IDI 一致, 预测效果良好; 珠海市全市及香洲区预测 IDI 与实际 IDI 均高于斗门区和金湾区, 原因在于香洲区是中心城区, 人口相对稠密, 更容易发生传染病疫情。预测期间 IDI 较高的原因主要在于香洲区流感样病例比例及其他感染性腹泻发病率较高, 处于低风险至高风险范围, 与实际疫情流行情况基本一致。

实际预报时可根据 IDI 值对应的风险大小进行提示, 当预测某地存在疫情风险时, 应提高警惕, 根据具体风险病种, 提供有针对性的建议。如流感高发, 则告诫老人和小孩尽量不要去人多的地方, 做好个人防护等; 其他感染性腹泻高发, 则建议加强卫生宣传, 注意饮食卫生, 饭前便后要洗手, 严格水源管理等; 若布雷图指数过高, 建议采取翻盆倒罐、防蚊灭蚊措施等。当预测风险为高风险时, 除以上针对群体性的建议外, 应在高风险地区优先考虑扩大公共卫生应急能力。

本研究构建的 IDI 可用于传染病流行风险的短期预测, 但人类活动是传染性病原体地理传播的重要驱动因素^[16], 由于不同的人接触感染的机会不同, 表现在年龄、性别、职业等的差别所导致的传染病的发病情况也不同, 可考虑从人群特征的角度进行预测。进一步优化模型, 从多角度、多层次^[17-18]评估近期群体传染病疫情趋势, 是下一步的研究重点。

参考文献

- [1] GENEROUS N, FAIRCHILD G, DESHPANDE A, et al. Global disease monitoring and forecasting with Wikipedia [J]. *PLoS Comput Biol*, 2014, 10 (11): 1-16.
- [2] 余艳妮, 聂绍发, 廖青, 等. 传染病预测及模型选择研究进展 [J]. *公共卫生与预防医学*, 2018, 29 (5): 89-92.
- [3] 董选军, 楼洁云, 余运贤, 等. 流感流行综合指数的探索建立与应用 [J]. *中国卫生统计*, 2016, 33 (4): 665-666.
- [4] 逯建华, 何建凡, 房师松, 等. 深圳市流感指数的制定及应用 [J]. *实用预防医学*, 2016, 23 (5): 628-629.
- [5] 林良强, 张韶华, 梁焯南, 等. 深圳市登革热风险指数建立和应用评估 [J]. *中华卫生杀虫药械*, 2018, 24 (1): 29-32.
- [6] 孙振球, 王乐三. 综合评价方法及其医学应用 [M]. 北京: 人民卫生出版社, 2014.
- [7] 冉茂亮, 黄浩, 钟颖. 基于熵权法的评价模型构建 [J]. *科技风*, 2018 (14): 207-208.
- [8] 刘守钦, 张军, 周林. 2004—2012 年济南市常见传染病季节性特征研究 [J]. *现代预防医学*, 2015, 42 (5): 921-929.
- [9] 孙丽红, 杨鹏, 董瑶, 等. 基于熵权法的老年人肌力模糊综合评价 [J]. *中国康复理论与实践*, 2017, 23 (12): 1464-1469.
- [10] 张雪宝, 叶中文, 练雁飞, 等. 2009—2012 年珠海市手足口病流行特征分析 [J]. *实用心脑血管病杂志*, 2018, 26 (1): 289-291.
- [11] MEHMANDAR M, SOORI H, MEHRABI Y. Predicting and analyzing the trend of traffic accidents deaths in Iran in 2014 and 2015 [J]. *Int J Crit Illn Inj Sci*, 2016, 6 (2): 74-78.
- [12] 胡碧波, 傅克本, 许亮亮, 等. 应用 ARIMA 模型预测结核病发病率研究 [J]. *预防医学*, 2018, 30 (10): 1011-1015.
- [13] 王怡, 张震, 范俊杰, 等. ARIMA 模型在传染病预测中的应用 [J]. *中国预防医学杂志*, 2015, 16 (6): 424-428.
- [14] 李广智, 刘峰, 王维华, 等. 自回归求和移动平均模型在流感发病预测中的应用 [J]. *中国热带医学*, 2016, 16 (12): 1190-1193.
- [15] 王晨, 郭倩, 周罗晶. ARIMA 模型与指数平滑法预测门诊量效果比较 [J]. *预防医学*, 2018, 30 (11): 1152-1155.
- [16] KRAEMER M U G, GOLDING N, BISANZIO D, et al. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings [J]. *Sci Rep*, 2019, 9 (1): 1-11.
- [17] MILINOVICH G J, AVRIL S M, CLEMENTS A C, et al. Using internet search queries for infectious disease surveillance: screening diseases for suitability [J]. *BMC Infect Dis*, 2014, 14 (690): 1-9.
- [18] BAI Y L, HUANG D S, LIU J, et al. Effect of meteorological factors on influenza-like illness from 2012 to 2015 in Huludao, a northeastern city in China [J]. *Peer J*, 2019, 7: 1-14.

收稿日期: 2019-03-15 修回日期: 2019-06-11 本文编辑: 徐文璐