

山西省百日咳发病的 SARIMA 模型预测*

张岳琴¹ 白丽霞^{1,2,△} 光明³ 朱镭² 周浩¹ 李秀辉¹ 弓培慧¹ 康娅楠¹

【摘要】目的 建立季节性自回归移动平均模型(seasonal auto-regressive integrated moving average, SARIMA),并探讨其对山西省百日咳发病的预测效果。**方法** 用 R 4.1.3 软件对 2010—2019 年山西省百日咳月报告发病率进行 SARIMA 模型构建与拟合,用所拟合的最优模型对 2020—2022 年山西省百日咳月发病率进行预测并与实际值比较,分析该模型的预测效果。**结果** 2010—2021 年山西省共报告百日咳病例 1956 例,年平均发病率为 46.50/1000 万。发病率总体上呈现上升趋势,2021 年时达到峰值 140.67/1000 万,2010—2019 年百日咳报告发病主要集中在夏秋两季,2020 年百日咳报告发病集中于上半年,2021 年集中于下半年。2020 年发病率与 2019 年相比减少了 72.98%。SARIMA(5,2,2)(0,1,2)₁₂ 模型对于短期百日咳疫情有较好的预测性能。**结论** SARIMA(5,2,2)(0,1,2)₁₂ 模型对于山西省短期百日咳发病有较好的预测性能,对于指导百日咳疫情防控有一定积极意义。

【关键词】 SARIMA 百日咳 发病率

【中图分类号】 R516.6

【文献标识码】 A

DOI 10.11783/j.issn.1002-3674.2024.04.015

百日咳是由百日咳鲍特菌引起的传染性极强的呼吸道传染病^[1-3],人类是百日咳鲍特菌的唯一宿主^[4],自然人群普遍易感^[5]。百日咳为疫苗可预防的疾病^[6],但感染和疫苗接种均不能提供终身免疫,其仍是影响儿童健康和导致其死亡的主要疾病之一^[7],是全世界的公共卫生问题^[8-11]。本研究使用 2010—2021 年山西省百日咳月报告发病率进行 SARIMA 模型构建与验证,并对山西省 2022 年百日咳的发病情况进行预测,为山西省百日咳的防控提出科学依据。

资料与方法

1. 资料来源

2010—2021 年山西省百日咳监测数据来源于中国疾病预防控制中心与山西省疾病预防控制中心传染病报告信息管理系统。用 2010 年 1 月—2019 年 12 月数据进行模型的拟合,预测 2020—2022 年山西省百日咳月发病率,并与实际月发病率进行比较。

2. 统计学处理

用 Excel 2021 建立山西省百日咳月报告发病率数据库,用 R 4.1.3 进行模型的构建,检验水准 $\alpha=0.05$ 。

3. 模型参数及建模步骤

SARIMA 季节性自回归移动平均模型是时间序列预测分析模型中常用的一种模型,表示为 SARIMA(p, d, q)(P, D, Q)_s, 共有 7 个结构参数。其中, p 表示非季节自回归的阶数,从偏自相关系数

(partial autocorrelation coefficient) 推断; d 表示时间序列数据需要进行差分的阶数; q 表示非季节移动平均阶数; P 表示季节自回归的阶数,从 PACF 推断; D 表示季节差分的阶数,一般为 0 或 1,做了季节差分就是 1; Q 表示季节移动平均阶数; s 表示用来建模的数据的季节性。

建模步骤:①数据处理:绘制原始数据的时间序列图,观察其序列是否平稳,若为非平稳序列则对其进行差分变换使其平稳;②随机性检验:对该时间序列进行随机性检验,若该序列为随机序列,则不能进行建模,反之则进行建模;③模型参数识别:通过自相关系数(autocorrelation coefficient, ACF)和 PACF 识别模型类型并初步确定 p, d, q 和 P, D, Q 的阶值;④模型识别和模型诊断:通过 Akaike 信息准则(akaike information criterion, AIC)来选择最适合的模型, AIC 越小模型越优。使用 Box-Ljung 检验对模型残差进行白噪声检验,若为白噪声序列,则模型建立有效;⑤模型验证:利用最优的模型预测 2020 年 1 月—2021 年 12 月百日咳发病率,并与实际发病率进行比较。⑥模型预测:用最优模型预测 2022 年山西省百日咳月发病率。

结 果

1. 2010—2021 年山西省百日咳流行概况

2010—2021 年山西省共报告百日咳病例 1956 例,年平均报告发病率为 46.50/1000 万。报告发病率总体上呈现上升趋势,2021 年达到峰值 140.67/1000 万,2020 年较 2019 年大幅下降。2010—2019 年百日咳报告发病主要集中在夏秋两季,2020 年百日咳报告发病集中于上半年,2021 年百日咳报告发病集中于下半年。

* 基金项目:山西省医学重点科技项目计划(2021XM25)

1. 山西医科大学公共卫生学院流行病学教研室(030001)

2. 山西省儿童医院(山西省妇幼保健院)

3. 山西省疾病预防控制中心

△通信作者:白丽霞, E-mail: bailixia111@163.com

2. SARIMA 模型构建

(1) 数据处理

首先将 2010 年 1 月—2019 年 12 月山西省百日咳发病率绘制成时间序列图,将观测值序列分解为随机波动效应、季节效应和长期趋势效应 3 个分量,可以看出数据具有季节性特征,每年出现一个流行大高峰和一个小高峰,总体呈波动上升趋势(图 1)。由图 2 及使用“`adf.test()`”函数进行平稳性检验, $P = 0.85 (>0.05)$,可知该序列不平稳,需对其进行处理使其平稳。用“`diff()`”函数对原始数据进行二阶差分和一次周期为 12 的季节差分,消除时间趋势和季节效应。使用“`adf.test()`”函数对差分后的时间序列进行平稳性检验, $P = 0.01 (<0.05)$,说明差分处理后的时间序列满足平稳性要求。对时间序列进行 Box-Ljung 检验, $P < 0.05$,表示该时间序列为非白噪声序列,可以进行模型构建。

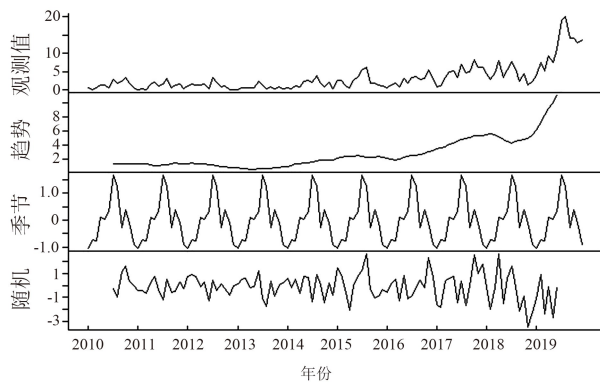


图 1 2010—2019 年山西省百日咳发病的时间序列图及其分解

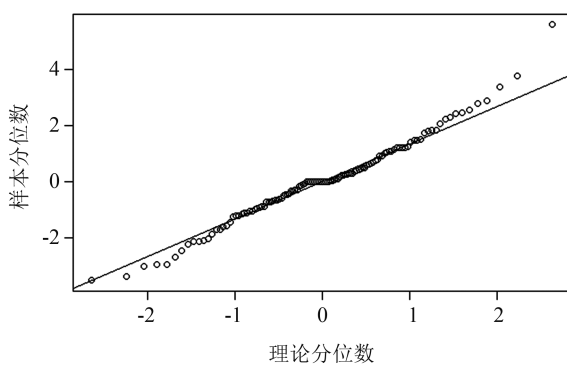


图 2 模型残差

(2) 模型识别

通过对数据的处理可以判断模型是以 12 个月为周期的复合模型 $ARIMA(p, 2, q)(P, 1, Q)_{12}$ 。自相关图与偏自相关图均为拖尾,自相关图在 2 阶后开始落入置信区间, $q = 2$ (图 3);偏自相关函数 p 难以确定,取值通常为 1~5;参数 P 和 Q 较难判断,取值通常为 0~2,一般从低阶到高阶逐步尝试。根据 AIC 来选择最适合的模型,AIC 越小模型越优。

(3) 模型诊断

根据“`auto.arima()`”函数得出的最优模型为 SA-

RIMA(0,1,3)(1,0,0)₁₂。分别对 45 个备选模型和 1 个系统自动拟合的模型进行参数估计和假设检验。AIC 统计量最小的模型拟合效果最好,因此综合以上信息最优模型为 $ARIMA(5, 2, 2)(0, 1, 2)_{12}$, AIC = 434.53, 其均方根误差 (root mean squared error, RMSE) 为 1.50, 绝对百分比误差 (mean absolute percentage error, MAE) 为 1.10。对该模型的残差进行 Box-Ljung 检验, $P = 0.69 (>0.05)$,表示残差为白噪声。同时其残差的 QQ 图显示该模型的残差为正态分布(图 2)。表明该模型能够较好地提取出残差中的信息,模型拟合效果较好。

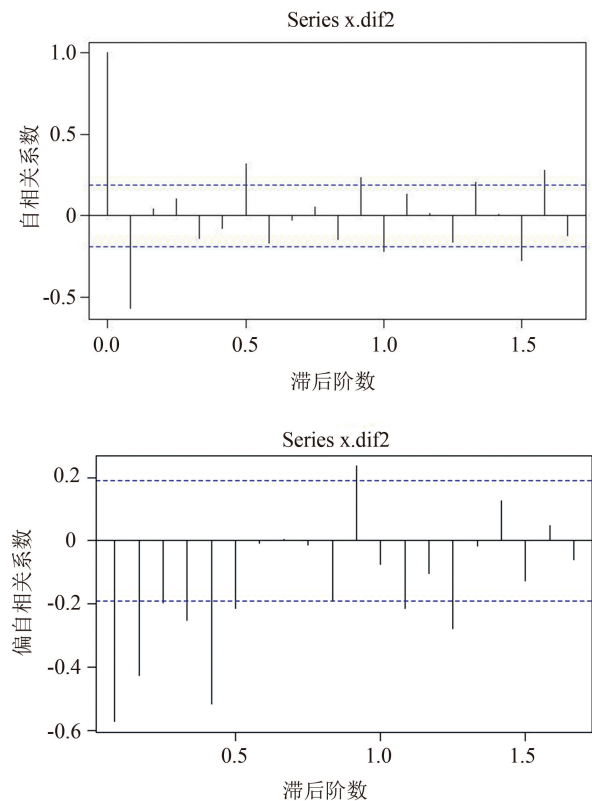


图 3 2010—2019 年山西省百日咳发病率的自相关图与偏自相关图

(4) 模型验证

使用“`forecast()`”函数对 2020 年 1 月—2021 年 12 月百日咳发病率进行预测。2020 年预测值与实际值的平均绝对误差为 -24.15, 平均相对误差为 -3647.25, 2021 年预测值与实际值的平均绝对误差为 -33.91, 平均相对误差为 -3773.36。2020 年百日咳发病受新冠肺炎疫情影响出现异常,均未落在预测值的 95% 置信区间内;而 2021 年 10、11、12 月份实际值落在预测值的 95% 置信区间内,说明此模型有较好的预测性能(表 1)。

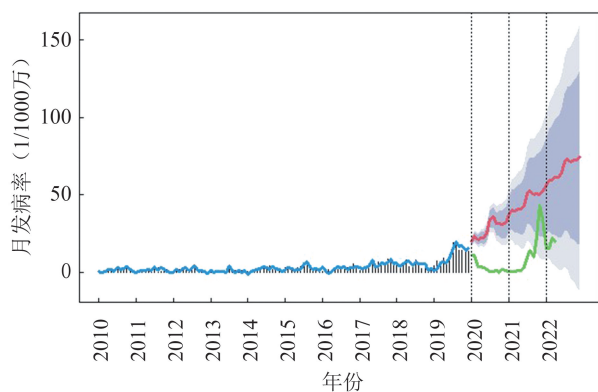
(5) 模型预测

用“`forecast()`”函数进行模型预测,2022 年山西省百日咳发病仍呈现上升趋势(图 4),提示山西省应提前做好相应的防控准备。

表 1 2020—2021 年山西省百日咳发病率预测结果

年份	月份	实际值 (1/1000 万)	预测值 (1/1000 万)	预测值的 95% 置信区间 (1/1000 万)	绝对误差	相对误差
2020	1	11.17	19.35	16.22~22.47	-8.18	-73.31
2020	2	10.88	23.16	19.18~27.15	-12.28	-112.89
2020	3	3.72	20.60	15.81~25.38	-16.88	-453.89
2020	4	3.44	22.42	16.90~27.95	-18.98	-552.61
2020	5	2.86	21.97	15.84~28.11	-19.11	-667.41
2020	6	1.43	25.07	18.13~32.01	-23.64	-1651.39
2020	7	0.29	33.65	25.11~42.18	-33.36	11653.95
2020	8	0.29	35.47	25.57~45.37	-35.18	-12289.67
2020	9	0.86	31.21	20.00~42.43	-30.35	-3533.88
2020	10	0.00	31.39	18.94~43.84	-31.39	-
2020	11	1.43	30.66	17.07~44.25	-29.23	-2041.91
2020	12	0.86	32.02	17.12~46.92	-31.16	-3628.20
2021	1	0.29	37.23	19.76~54.69	-36.94	-1289.13
2021	2	0.57	40.37	20.70~60.03	-39.80	-6945.57
2021	3	0.29	39.50	17.67~61.32	-39.21	-13687.48
2021	4	0.86	40.93	17.04~64.82	-40.07	-4662.21
2021	5	0.86	40.58	14.74~66.41	-39.72	-4621.48
2021	6	3.44	43.47	15.46~71.48	-40.03	-1164.43
2021	7	9.74	50.68	19.95~81.40	-40.94	-420.29
2021	8	14.04	52.44	19.03~85.85	-38.40	-273.55
2021	9	9.45	49.79	13.70~85.87	-40.34	-426.64
2021	10	23.21*	50.62	11.93~89.31	-27.41	-118.13
2021	11	43.26*	50.20	8.97~91.43	-6.94	-16.04
2021	12	34.67*	51.77	7.82~95.71	-17.10	-49.34

*: 表示该月实际发病率落在预测值的 95% 置信区间内, - : 表示无法计算该值。



注: 悬垂线为 2010—2019 年实际值, 为 2010—2019 年模型拟合值, 为 2020—2022 年模型预测值, 为 2020—2021 年实际值, 深色阴影部分为预测值 85% 置信区间, 浅色阴影部分为预测值 95% 置信区间。

图 4 2020—2022 年山西省百日咳月发病率预测图

讨 论

传染病发病率的准确模拟和预测对资源利用和未来流行病的规划具有重要影响^[12]。各种数学模型被广泛应用于传染病预测预警, 如指数平滑 [ETS (exponential smoothing) 模型], 自回归和移动平均模型 (ARMA 模型), 神经网络模型等。其中 ARIMA (或 SARIMA) 模型和同时考虑总体趋势和季节变化的

(exponential smoothing, ETS) 模型, 被认为是具有趋势、周期性和季节性的时间序列建模的有用工具, 被广泛应用于各种传染病短期发病趋势的预测^[13-15]。

20 世纪 50—60 年代, 实施大规模的免疫接种后, 百日咳的发病率和病死率显著下降^[16-17]。在 20 世纪 90 年代以来一些高疫苗覆盖率的国家出现了百日咳发病的逐年升高, 被称为“百日咳重现”^[18-24]。在本研究中我们同样看到 2010—2021 年山西省百日咳报告发病在总体上呈现波动上升趋势。

本研究建模数据为 2010—2019 年的山西省百日咳月发病率资料, 用此模型预测 2022 年山西省百日咳发病呈现上升趋势, 提示我们在 2022 年山西省新冠肺炎疫情相对平稳的状态下应警惕百日咳疫情的发生, 提前做好相应的防控准备。

在本研究中, 我们发现 2020 年山西省百日咳的报告发病的实际值与预测值相差很大, 且 2020 年百日咳的报告发病与 2019 年相比减少了 72.98%, 其原因与 2020 年 1 月起新冠肺炎疫情密切相关。百日咳与新型冠状病毒肺炎均可经呼吸道飞沫传播^[25], 所以防控新冠肺炎的措施, 对于控制百日咳的发病也同样有效, 如戴口罩、居家隔离、网络教学、疫期消毒、随时健康监

测、随时进行手消毒、少去公共场所等人员聚集的地方等,在控制新冠肺炎疫情的同时也控制住了百日咳的传播;此外,由于受新冠肺炎疫情的影响,人们出行受限以及就医程序变得复杂,一部分症状较轻或不明显的患者选择居家自行吃药缓解,造成漏诊导致百日咳报告发病率的下降。

SARIMA 模型只需要历史观测来开发预测模型,考虑了总体趋势和季节变化^[26],是时间序列分析中最常用的一种模型,以自动化的方式在一个时间序列上执行,最大限度地提高了预测精度。但是因为 SARIMA 模型不能将影响百日咳发病的很多因素,如人口因素(年龄构成、人口密度等)、社会因素(经济条件、医疗水平、受教育程度等)、环境因素等(温度、湿度、污染情况等)完全纳入,可能会影响模型的预测性能。在以后的研究中我们应尽可能将这些因素考虑进来,以提高 SARIMA 预测模型的预测性能。此外,SARIMA 模型建立的预测模型只能进行短期预测,若要进行长期预测,需对数据及时更新并建立新的预测模型。

此外,应用组合模型预测传染病的发病率也逐渐成为我们研究的新突破点,很多研究也表明组合模型的预测性能优于单一模型,在今后的研究中我们需探索更多模型,以期建立更优的预测模型,为百日咳的防控提供更有针对性的指导。

参 考 文 献

- [1] Kandeil W, van den Ende C, Bunge EM, et al. A systematic review of the burden of pertussis disease in infants and the effectiveness of maternal immunization against pertussis[J]. *Expert review of vaccines*, 2020, 19(7): 621-638.
- [2] Hincapié-Palacio D, Hoyos MC, Ochoa J, et al. Effect of maternal immunization against pertussis in Medellín and the metropolitan area, Colombia, 2016-2017[J]. *Vaccine*, 2018, 36(27): 3984-3991.
- [3] Martínón-Torres F, Halperin SA, Nolan T, et al. Impact of maternal diphtheria-tetanus-acellular pertussis vaccination on pertussis booster immune responses in toddlers: Follow-up of a randomized trial[J]. *Vaccine*, 2021, 39(11): 1598-1608.
- [4] Garlasco J, Bordino V, Marengo N, et al. Pertussis immunisation during pregnancy: Antibody levels and the impact of booster vaccine[J]. *Vaccine*, 2021, 39(35): 4957-4963.
- [5] Jenkins VA, Savic M, Kandeil W. Pertussis in High Risk Groups: An Overview of the Past Quarter Century[J]. *Human vaccines & immunotherapeutics*, 2020, 16(11): 2609-2617.
- [6] Andersen AR, Kolmos SK, Flanagan KL, et al. Systematic review and meta-analysis of the effect of pertussis vaccine in pregnancy on the risk of chorioamnionitis, non-pertussis infectious diseases and other adverse pregnancy outcomes[J]. *Vaccine*, 2022, 40(11): 1572-1582.
- [7] Vaz-de-Lima LRA, Sato APS, Pawloski LC, et al. Effect of maternal Tdap on infant antibody response to a primary vaccination series with whole cell pertussis vaccine in São Paulo, Brazil[J]. *Vaccine*; X, 2021, 7: 100087.
- [8] Godoy P, García-Cenoz M, Rius C, et al. Effectiveness of maternal pertussis vaccination in protecting newborn: A matched case-control study[J]. *Journal of Infection*, 2021, 83(5): 554-558.
- [9] Li WF, Huang SY, Peng HH, et al. Factors affecting pregnant women's decisions regarding prenatal pertussis vaccination: A decision-making study in the nationwide Prenatal Pertussis Immunization Program in Taiwan[J]. *Taiwanese Journal of Obstetrics and Gynecology*, 2020, 59(2): 200-206.
- [10] Yeung KHT, Duclos P, Nelson EAS, et al. An update of the global burden of pertussis in children younger than 5 years: a modelling study[J]. *The Lancet Infectious Diseases*, 2017, 17(9): 974-980.
- [11] Carrasquilla G, Porras A, Martinez S, et al. Incidence and mortality of pertussis disease in infants <12 months of age following introduction of pertussis maternal universal mass vaccination in Bogotá, Colombia[J]. *Vaccine*, 2020, 38(46): 7384-7392.
- [12] 李晨希, 吴群红, 刘欢, 等. SARIMA 模型在广东省急性出血性结膜炎疫情预测中的应用[J]. *职业与健康*, 2021, 37(18): 2522-2525.
- [13] 黄莉, 陈田木, 杨云斌, 等. 应用 SARIMA 模型预测肺结核发病趋势[J]. *现代预防医学*, 2019, 46(5): 794-797.
- [14] 王磊, 侯焯, 周美静, 等. 基于 R 语言的季节性自回归滑动平均(SARIMA)模型对流感样病例发病趋势的预测[J]. *医学动物防制*, 2022, 38(4): 349-353.
- [15] Wang Y, Xu C, Wang Z, et al. Time series modeling of pertussis incidence in China from 2004 to 2018 with a novel wavelet based SARIMA-NAR hybrid model[J]. *PLoS One*, 2018, 13(12): e208404.
- [16] Fernandes EG, Sato APS, Vaz-de-Lima LRA, et al. The effectiveness of maternal pertussis vaccination in protecting newborn infants in Brazil: A case-control study[J]. *Vaccine*, 2019, 37(36): 5481-5484.
- [17] Nguyen HS, Vo NP, Chen SY, et al. The optimal strategy for pertussis vaccination: a systematic review and meta-analysis of randomized control trials and real-world data[J]. *American Journal of Obstetrics and Gynecology*, 2022, 226(1): 52-67.
- [18] Vaz-de-Lima LRA, Sato HK, Fernandes EG, et al. Association between the timing of maternal vaccination and newborns' anti-pertussis toxin antibody levels[J]. *Vaccine*, 2019, 37(36): 5474-5480.
- [19] Santana CP, Luhm KR, Shimakura SE. Impact of Tdap vaccine during pregnancy on the incidence of pertussis in children under one year in Brazil-A time series analysis[J]. *Vaccine*, 2021, 39(6): 976-983.
- [20] Naninck T, Contreras V, Coutte L, et al. Intranasal inoculation with Bordetella pertussis confers protection without inducing classical whooping cough in baboons[J]. *Current Research in Microbial Sciences*, 2021, 2: 100072.
- [21] Mohammed H, Roberts CT, Grzeskowiak LE, et al. Safety of maternal pertussis vaccination on pregnancy and birth outcomes: A prospective cohort study[J]. *Vaccine*, 2021, 39(2): 324-331.
- [22] Fedele G, Carollo M, Palazzo R, et al. Parents as source of pertussis transmission in hospitalized young infants[J]. *Infection*, 2017, 45(2): 171-178.
- [23] Berbers G, van Gageldonk P, Kasstele JV, et al. Circulation of pertussis and poor protection against diphtheria among middle-aged adults in 18 European countries[J]. *Nature Communications*, 2021, 12(1): 2871.
- [24] 黄海涛, 张颖. 中国百日咳的监测现状及其流行特征[J]. *中华实用儿科临床杂志*, 2017, 32(22): 1685-1688.
- [25] Cong J, Ren M, Xie S, et al. Predicting Seasonal Influenza Based on SARIMA Model, in Mainland China from 2005 to 2018[J]. *Int J Environ Res Public Health*, 2019, 16(23): 4760.
- [26] Abd EW, Eassa SM, Metwally M, et al. SARS-CoV-2 Transmission Channels: A Review of the Literature[J]. *MEDICC Rev*, 2020, 22(4): 51-69.

(责任编辑:邓妍)