

DOI:10.12138/j.issn.1671-9638.20193911

· 论 著 ·

## 基于乘积季节性 ARIMA 模型对神经内科医院感染发病率的预测研究

王青青<sup>1</sup>, 范馨月<sup>2</sup>, 查筑红<sup>1</sup>, 黄冰<sup>1</sup>, 程永素<sup>1</sup>, 罗光英<sup>1</sup>, 曾妮<sup>1</sup>, 姚尧<sup>1</sup>

(1. 贵州医科大学附属医院医院感染管理科, 贵州 贵阳 550004; 2. 贵州大学数学与统计学院计算机科学系, 贵州 贵阳 550025)

**[摘要]** **目的** 建立神经内科病房医院感染预警模型, 预测神经内科患者发生医院感染的风险, 为早期防控提供依据。**方法** 收集贵州省某三级甲等医院神经内科病房医院感染发病率数据, 构建乘积季节性 ARIMA(p, d, q) × (P, D, Q)<sub>s</sub> 模型, 对建立的模型进行参数估计、模型诊断, 选择最优预测模型。利用构建的最佳模型对神经内科病房医院感染发病率进行预测, 并对预测效果进行评价。**结果** 以该院 2014—2017 年神经内科月度医院感染发病率数据作为训练样本, 获得最优预测模型 ARIMA(2, 1, 2) × (1, 1, 1)<sub>4</sub>。以 2018 年 1—5 月数据作为模型预测验证样本, 结果显示, 模型预测值的动态趋势与实际情况基本一致, 实际发病率均在预测值的 95% 置信区间内。用此模型对 2018 年 6—12 月神经内科医院感染发病率作预测, 预测结果显示预测值均位于 95% 的置信区间内。**结论** ARIMA(2, 1, 2) × (1, 1, 1)<sub>4</sub> 模型能较好地模拟神经内科病房医院感染发病率变化趋势, 具有良好的预测效果。

**[关键词]** 神经内科; 医院感染; ARIMA 模型; 预测

**[中图分类号]** R181.3<sup>+</sup>2

## Prediction of incidence of healthcare-associated infection in department of neurology based on ARIMA model

WANG Qing-qing<sup>1</sup>, FAN Xin-yue<sup>2</sup>, ZHA Zhu-hong<sup>1</sup>, HUANG Bing<sup>1</sup>, CHENG Yong-su<sup>1</sup>, LUO Guang-ying<sup>1</sup>, ZENG Ni<sup>1</sup>, YAO Yao<sup>1</sup> (1. Department of Healthcare-associated Infection Management, The Affiliated Hospital of Guizhou Medical University, Guiyang 550004, China; 2. Department of Computer Science, School of Mathematics and Statistics, Guizhou University, Guiyang 550025, China)

**[Abstract]** **Objective** To establish an early warning model of healthcare-associated infection(HAI) in the department of neurology, predict the risk of HAI in patients in department of neurology, and provide basis for early prevention and control. **Methods** Data on incidence of HAI in neurology ward of a tertiary first-class hospital in Guizhou Province were collected, the ARIMA(p, d, q) × (P, D, Q)<sub>s</sub> model was constructed, parameter estimation and model diagnosis were performed for the established model, and the optimal prediction model was selected. The best constructed model was used to predict the incidence of HAI in the department of neurology, and the prediction efficacy was evaluated. **Results** The data of monthly incidence of HAI in department of neurology in this hospital from 2014 to 2017 was as training specimens, the optimal prediction model ARIMA(2, 1, 2) × (1, 1, 1)<sub>4</sub> was obtained. Data of January-May 2018 was as validation sample for model prediction, the results showed that the dynamic trend of predicted value of model was basically consistent with the actual condition, the actual incidence was within the 95% confidence interval of the predicted value. This model was used to predict the incidence of HAI in department of neurology from June to December 2018, the predicted results showed that the predicted values were within 95% confidence interval. **Conclusion** The ARIMA(2, 1, 2) × (1, 1, 1)<sub>4</sub> model can better simulate the trend of HAI rate in the department of neurology, and it has preferable prediction effect.

**[Key words]** department of neurology; healthcare-associated infection; ARIMA model; prediction

**[收稿日期]** 2018-05-17

**[基金项目]** 贵州省大数据重点实验室开放课题(2017BDKfJJ012); 贵州大学“本科教学工程建设”项目(JG201723)

**[作者简介]** 王青青(1985-), 女(汉族), 贵州省镇远县人, 副主任医师, 主要从事医院感染流行病学研究。

**[通信作者]** 范馨月 E-mail: fan.xinyue@163.com

医院感染 (healthcare-associated infection, HAI) 不仅增加住院患者死亡风险, 同时延长患者住院时间和增加患者经济负担, 已成为医学界亟待解决的重大问题<sup>[1-2]</sup>。神经内科患者具有年龄大、发病急、基础疾病多、病情严重等特点, 并常伴有不同程度的意识、运动障碍, 免疫功能低下, 卧床时间长等特点, 是医院感染的高危人群<sup>[3-5]</sup>。因此, 了解神经内科患者医院感染发生、发展规律, 并在此基础上采取可能的预防和控制措施具有重要的现实意义<sup>[6]</sup>。医院感染发病率的预测是将被动预防向主动预防转化的重要环节。准确预测医院感染发病率未来变化情况, 对于医院感染的预防与控制有着极为重要的作用<sup>[7]</sup>。自回归滑动平均混合模型 (autoregressive integrated moving average model, ARIMA) 是 Box-Jenkins 方法中的重要时间序列预测模型, 目前用于短期预测效果较为理想, 被广泛应用于医疗卫生、环境、金融等领域的预测<sup>[8]</sup>。本研究基于贵州省某三级甲等医院 2014—2018 年神经内科医院感染发病率数据, 采用乘积季节性模型对该院神经内科医院感染的发病情况进行预测, 以期发现医院感染的潜在规律, 发出早期预警, 从而为神经内科患者医院感染预防关口前移提供科学依据。

## 1 资料与方法

1.1 资料来源 收集贵州省某三级甲等医院 2014 年 1 月—2018 年 5 月入住神经内科 > 48 h 患者月度医院感染发病率资料。

1.2 医院感染诊断标准 参照中华人民共和国卫生部 2001 年发布施行的《医院感染诊断标准(试行)》<sup>[9]</sup>。

1.3 模型构建方法 由于医院感染具有季节性感染的特殊性, 因此, 采用乘积季节性模型拟合时间序列具有较好的预测效果。模型中分别代表非季节性和季节性自回归项、差分和移动平均项系数, 为季节周期。以 2014—2017 年神经内科月度医院感染发病率数据作为训练样本, 2018 年 1—5 月数据作为模型预测验证样本。建模主要过程分为四步, (1) 时间序列平稳化检验和处理: ARIMA 模型要求数据符合时间序列平稳性, 因此需对已有时间序列数据进行数据平稳化处理。采用差分方法使该序列满足零均值且方差不随时间变化, 根据差分次数确定差分阶数。(2) 模型识别: 通过绘制平稳后时间序列的自相关函数 (autocorrelation function, ACF) 和偏自相关函数 (partial autocorrelation function,

PACF) 图对目标序列进行定阶。(3) 参数估计与诊断检验: 从估计的多个模型中, 选择其中最优模型进行预测。模型的筛选依据采用拟合优度统计量比较模型的优劣, 采用赤池信息准则 (Akaike information criterion, AIC) 和贝叶斯信息准则 (Bayesian information criterion, BIC) 优先选择值最小参数进行拟合优化模型, 对模型的参数进行显著性检验和残差进行白噪声检验。(4) 模型预测: 确定出最优模型, 对模型进行预测, 得到原序列将来感染发病率趋势。

1.4 统计分析 应用 SPSS 18.0 软件对每年及各月医院感染发病率进行卡方检验, 应用 Matlab2017a 软件构建 ARIMA(p, d, q) × (P, D, Q)<sub>s</sub> 模型。

## 2 结果

2.1 医院感染情况 2014—2018 年该院神经内科每年医院感染发病率分别为 0.67%、2.01%、1.35%、0.59% 和 0.34%, 差异有统计学意义 ( $\chi^2 = 82.26, P < 0.001$ ), 趋势性  $\chi^2$  检验发现医院感染发病率存在下降趋势 ( $\chi^2 = 22.43, P < 0.001$ )。见图 1、表 1。

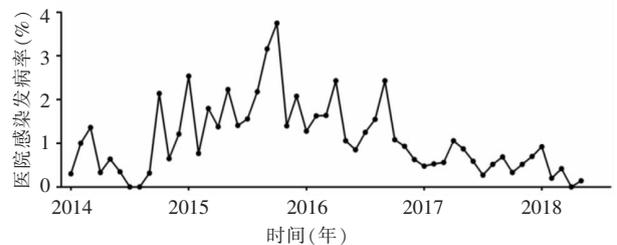


图 1 2014—2018 年某院神经内科医院感染发病率变化趋势  
Figure 1 Change trend in HAI incidence in department of neurology in a hospital from 2014 to 2018

2.2 时间序列平稳化 通过原始时序图 (图 1) 发现, 该三甲医院 2014—2018 年神经内科各月医院感染发病率呈现出一定季节性和周期性, 为非平稳时间序列, 需要对其进行平稳化处理。对原数据作一阶差分后消除了序列的长期趋势, 发病率时序图基本趋于平稳。见图 2。

2.3 模型识别与定阶 对原序列进行一阶差分后的 ACF 和 PACF 图, 见图 3。差分后通过 ADF 检验, 确定 ARIMA(p, d, q) × (P, D, Q)<sub>s</sub> 中的 d 和 D 均为 1。此时残差序列自相关函数和偏相关函数在可信区间内, 模型的计算值和实际值拟合度较高。对其进行残差相关性检验, 检验值基本落在 95% 的置信区间内。

表 1 2014—2018 年神经内科各月医院感染发病率[% (n/N)]

Table 1 The monthly HAI incidence in department of neurology in a hospital from 2014 to 2018(%[n/N])

月份	2014 年	2015 年	2016 年	2017 年	2018 年
1	0.30(1/331)	2.54(8/315)	1.28(6/467)	0.48(3/629)	0.92(6/649)
2	1.00(2/200)	0.77(2/260)	1.63(6/368)	0.53(3/564)	0.20(1/506)
3	1.36(4/294)	1.80(6/333)	1.64(8/489)	0.56(4/718)	0.42(3/721)
4	0.33(1/303)	1.38(5/363)	2.43(11/453)	1.06(7/663)	0.00(0/678)
5	0.64(2/313)	2.23(8/359)	1.06(5/471)	0.87(6/689)	0.14(1/693)
6	0.35(1/287)	1.41(5/355)	0.85(4/472)	0.59(4/681)	-
7	0.00(0/325)	1.56(6/384)	1.25(6/480)	0.27(2/750)	-
8	0.00(0/310)	2.18(8/367)	1.55(8/517)	0.52(4/771)	-
9	0.32(1/309)	3.16(9/285)	2.43(11/452)	0.69(5/727)	-
10	2.14(6/280)	3.75(12/320)	1.08(6/555)	0.33(2/611)	-
11	0.65(2/307)	1.40(5/357)	0.93(6/647)	0.52(3/578)	-
12	1.21(4/331)	2.08(8/384)	0.63(4/632)	0.70(4/568)	-
合计	0.67(24/3 590)	2.01(82/4 082)	1.35(81/6 003)	0.59(47/7 949)	0.34(11/3 247)

注:医院感染发病率=新发生医院感染的患者例数/同期住院患者例数×100%;表中 n 表示神经内科新发生医院感染的患者例数;N 表示同期神经内科住院患者例数

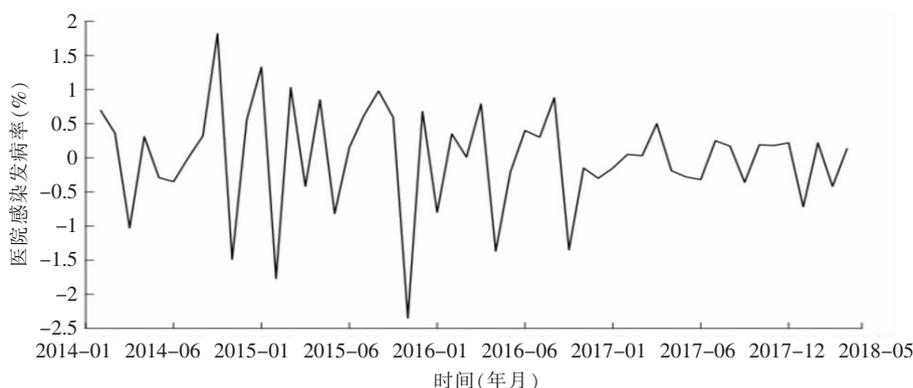


图 2 2014—2018 年某院神经内科医院感染发病率数据一阶差分后时序图

Figure 2 Temporal graph after first order difference of data of HAI incidence in neurology department of a hospital from 2014 to 2018

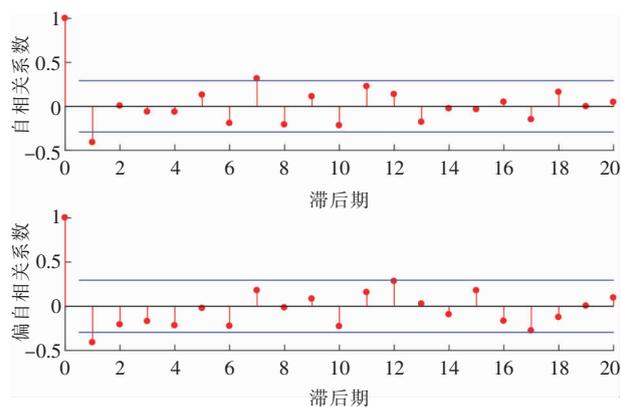


图 3 一阶差分后 ACF 和 PACF 图

Figure 3 The ACF and PACF after first order difference

AIC 数值越小,模型精度越好。通过实验数据分析,自回归部分阶数和移动平均阶数可以选取  $p=2, q=2$ 。综上得出的最优预测模型为  $ARIMA(2, 1, 2) \times (1, 1, 1)_4$ 。通过计算  $AIC = 125.68$ ;  $BIC = 133.16$ 。 $ARIMA(2, 1, 2) \times (1, 1, 1)_4$  预测感染发病率的残差、QQ 图检验、ACF 和 PACF 见图 4。模型估计结果的残差序列满足随机性检验。

2.4 模型诊断 对 2018 年 1—5 月神经内科各月医院感染发病率进行预测,结果显示,模型预测值的动态趋势与实际情况基本一致,实际发病率均是在预测值的 95% 置信区间内,说明了该模型拟合效果较好,可用以对未来进行较好的跟踪和预测。见表 2,图 5。

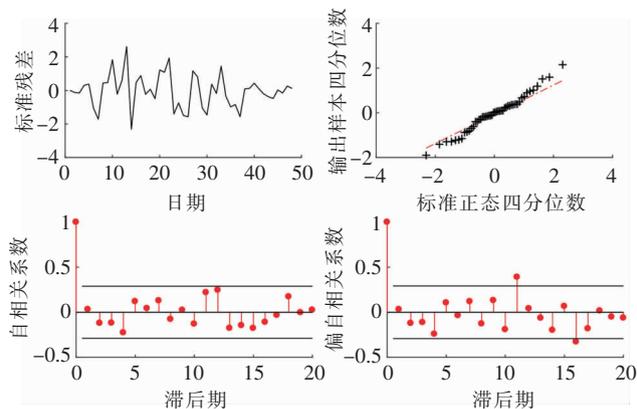


图 4 预测模型残差、QQ 图检验、ACF 和 PACF 图

Figure 4 Prediction model residuals, QQ plot, ACF and PACF

表 2 2018 年 1—5 月神经内科医院感染发生率实际值与预测值比较

Table 2 Comparison of actual and predicted incidence of HAI in department of neurology from January to May 2018

月份	实际值(%)	预测值(%)	差率
1	0.92	0.85	0.08
2	0.20	0.22	0.10
3	0.42	0.40	0.05
4	0.00	0.03	0.00
5	0.14	0.16	0.14

注:差率 = (|实际值 - 预测值|)/实际值

2.5 模型预测 用此模型对 2018 年 6—12 月神经内科医院感染发病率进行预测,预测结果见图 5,结果显示预测值均位于 95% 的置信区间内。

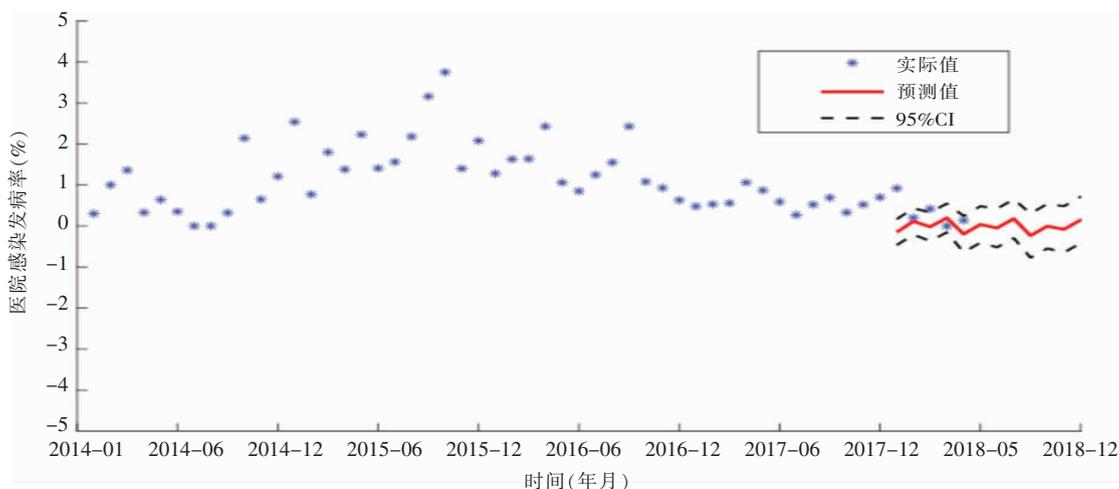


图 5 ARIMA 模型对神经内科医院感染发病率预测图

Figure 5 ARIMA model for predicting incidence of HAI in department of neurology

### 3 讨论

医院感染给患者和家庭,以及医院均带来不同程度的疾病负担。预防和控制医院感染,降低医院感染发病率,是医院感染管理的核心工作。目前,我国对医院感染控制的监测模式多采用回顾性调查,前瞻性研究尚不多<sup>[10-11]</sup>。利用统计预测分析技术对医院感染进行预测性数据挖掘,建立医院感染预警模式是医院感染管理工作今后发展的重要方向。

ARIMA 模型是时间序列分析中重要而基本的模型之一,是针对有季节性变动的的时间序列提出的建模方法,综合考虑了序列的趋势变化、周期变化及

随机干扰,并借助模型参数进行了量化表达,具有实用性强、预测精准度高等特点<sup>[12-13]</sup>,已被广泛应用于各个学科,尤其在卫生领域前瞻性预测方面具有广阔的应用前景<sup>[14-15]</sup>。本研究采用时间序列 ARIMA (2,1,2) × (1,1,1)<sub>4</sub> 模型对某三级甲等医院神经内科 2014—2018 年各月医院感染发病率数据进行分析,结果显示,该模型的预测精度较高,能很好地拟合原始序列的趋势性和周期性,可以用于神经内科医院感染发病趋势的分析和预测。使用该预测模型对 2018 年 1—5 月数据进行验证,结果显示,实际值与预测值重合较好,波动形势基本一致,实际值均在预测值 95% 可信区间内,预测拟合效果较可靠。在此基础上,对 2018 年 6—12 月神经内科医院感染发

病率进行预测,结果显示 2018 年 7 月份和 12 月份是两个高发月份,因此在此期间应该加强落实医院感染控制措施,有效减少医院感染的发生。

此外,本研究所构建的 ARIMA 模型与李红等<sup>[16]</sup>构建的模型不同,说明了不同时间段、不同地域构建医院感染发病率预测模型的必要性。另外,单次分析建立的 ARIMA 模型,不能作为永久不变的预测工具,只能用于短期预测。在实际工作中,应收集足够多的时间序列数据,用新的实际值对已建立的模型进行验证,并不断加入新的实际值,以修正或重新拟合医院感染发病率预测模型,更能反映实际情况。

综上所述,医院感染的发生受社会因素、环境因素的影响,既有趋势变动,又有季节变动。选用相对最优的 ARIMA 模型对下一年的不同时间段内的医院感染人数进行初步预测,有利于我们提前做好防范工作,从而使医院感染的防线前移,从根本上减少医院感染的发生,降低总体医院感染发病率。

#### [参 考 文 献]

- [1] Allegranzi B, Nejad SB, Combescore C, et al. Burden of endemic health-care-associated infection in developing countries: systematic review and meta-analysis[J]. *Lancet*, 377(9761): 228 - 241.
- [2] Arefian H, Vogel M, Kwetkat A, et al. Economic evaluation of interventions for prevention of hospital acquired infections: a systematic review[J]. *PLoS One*, 2016, 11(1): e0146381.
- [3] 余秋华,王作艳,张小红,等. 1 106 例神经内科住院患者医院感染调查分析[J]. *中国感染控制杂志*, 2011, 10(5):380 - 382.
- [4] 刘娟,汪志云,张坚磊,等. 神经内科住院患者医院感染危险因素分析[J]. *中华医院感染学杂志*, 2016, 26(1):209 - 210.
- [5] 徐宗俊,龙军,钱树星. 神经内科住院病人医院感染分析[J]. *中华神经医学杂志*, 2006, 5(12):1266 - 1268.
- [6] 吕一欣,任南,吴安华,等. 神经内科住院患者医院感染危险因素研究[J]. *中国感染控制杂志*, 2008, 7(1):29 - 31.

- [7] 李娜,黄艳芳,唐喻莹,等. 统计预测模型在医院感染中的应用研究[J]. *四川医学*, 2017, 38(1): 97 - 101.
- [8] Taneja K, Ahmad S, Ahmad K, et al. Time series analysis of aerosol optical depth over New Delhi using Box-Jenkins ARIMA modeling approach[J]. *Atmo Pollut Res*, 2016, 7(4): 585 - 596.
- [9] 中华人民共和国卫生部. 医院感染诊断标准(试行)[S]. 北京,2001.
- [10] 刘小燕,洗翠尧,王法霞,等. 基层三甲医院 2012—2015 年医院感染现患率及危险因素[J]. *中国感染控制杂志*, 2017, 16(11):1026 - 1029.
- [11] 李金梅,李家斌,王进. 综合医院医院感染横断面调查分析[J]. *中华医院感染学杂志*, 2015, 25(1):102 - 104.
- [12] 李红,潘东峰,郭忠琴,等. 时间序列模型在医院感染发生率拟合预测中的比较研究[J]. *中国卫生统计*, 2013, 30(1):87 - 89.
- [13] 管利华. ARIMA 模型预测医院感染发病状况研究[J]. *实用预防医学*, 2013, 20(10):1247 - 1249.
- [14] 王永斌,李向文,柴峰,等. ARIMA 模型在我国梅毒发病率预测中的应用[J]. *现代预防医学*, 2015, 42(3):385 - 388.
- [15] 吴秋红,张莉,袁军,等. 深圳市南山区 2005—2016 年淋病流行特征及 ARIMA 模型发病趋势预测[J]. *中国感染控制杂志*, 2018, 17(3): 202 - 206.
- [16] 李红,梁沛枫,潘东峰,等. 自回归滑动平均混合模型在医院感染发病率预测中的应用研究[J]. *中华医院感染学杂志*, 2013, 23(11):2693 - 2695.

(本文编辑:左双燕)

**本文引用格式:**王青青,范馨月,查筑红,等. 基于乘积季节性 ARIMA 模型对神经内科医院感染发病率的预测研究[J]. *中国感染控制杂志*, 2019, 18(1):59 - 63. DOI:10. 12138/j. issn. 1671 - 9638. 20193911

**Cite this article as:** WANG Qing-qing, FAN Xin-yue, ZHA Zhu-hong, et al. Prediction of incidence of healthcare-associated infection in department of neurology based on ARIMA model[J]. *Chin J Infect Control*, 2019, 18(1):59 - 63. DOI:10. 12138/j. issn. 1671 - 9638. 20193911