

# 全国人间狂犬病疫情的时间序列分析

任江萍,陈直平,孙继民,陈恩富,施旭光,张蓉,刘营,凌峰

**摘要:**目的 分析全国狂犬病疫情的时间分布特征,探讨用自回归移动平均模型(Autoregressive Integrated Moving Average Model, ARIMA)预测疫情发展趋势的可行性。**方法** 建立2004—2015年全国狂犬病月度发病率时间序列,用R3.3.2软件建立最优ARIMA模型,对2016年1~11月发病率进行预测,并评价预测效果。**结果** 2007年以来我国狂犬病年发病率呈下降趋势,8~10月为高峰季节;建立的最优模型为ARIMA(2,1,1)(2,0,0)<sub>12</sub>,其平均绝对标准化误差(The mean absolute error, MASE)为0.755;2016年1~11月发病率预测结果显示,平均相对误差为15.61%。**结论** 我国狂犬病疫情存在季节性发病高峰,ARIMA模型能用于我国狂犬病疫情的短期预测。

**关键词:**时间序列分析;ARIMA模型;狂犬病

中图分类号:R383.2 文献标识码:A 文章编号:1002-2694(2018)03-0239-04

## The time-series analysis of human rabies in China

REN Jiang-ping, CHEN Zhi-ping, SUN Ji-min, CHEN En-fu, SHI Xu-guang,  
ZHANG Rong, LIU Ying, LING Feng

(Zhejiang Center for Disease Control and Prevention, Zhejiang 310051, China)

**Abstract:** In order to analyze the temporal distribution characteristics of human rabies in China and explore the incidence forecasting, R3.3.2 software were used to establish the optimal ARIMA model of monthly incidences from 2004 to 2015 in China. Then the monthly incidences from January to November 2016 were forecasted with the model and the prediction accuracy was evaluated. The annual incidences of human rabies in China were decreasing tendency since 2007, and the monthly incidence reached highly during August to October. The optimal model was ARIMA(2,1,1)(2,0,0)<sub>12</sub> and the mean absolute error (MASE) was 0.755. The mean relative error was 15.61% when predicting the incidences from January to November 2016. In conclusion, there exists seasonal variation for human rabies in China and ARIMA model can be applied for the short-term forecasting.

**Keywords:** time-series analysis; ARIMA model; rabies

Corresponding author: Ling Feng, Email: fengl@cdc.zj.cn

狂犬病是狂犬病病毒感染引起的人兽共患病。除南极洲外全球各大洲均有病例报告,每年约有6万例狂犬病死亡,其中超过95%的死亡在亚洲和非洲<sup>[1]</sup>。WHO估计约99%的人狂犬病感染来源为犬。此病目前无特异疗法,病死率几乎100%。为了解我国狂犬病疫情时间分布特征及流行趋势,现用全国疫情资料进行分析研究。

ARIMA模型即Box-Jenkins模型。一般形式:ARIMA(p,d,q)×(P,D,Q)s,其中d和D分别为

差分和季节性差分次数;p和P分别是自回归(AR)模型和季节性AR模型的参数个数;q和Q分别是移动平均(MA)模型和季节性MA模型的参数个数;s是季节性期数。对有季节性变化的时间序列应建立ARIMA(p,d,q)×(P,D,Q)s模型。

### 1 资料与方法

1.1 资料来源 2004年1月至2016年11月全国狂犬病月发病数来自国家卫计委(<http://www.nhfpc.gov.cn/zhus/index.shtml>),2004—2015年全国人口数据来自国家统计局(<http://www.stats.gov.cn/>)。

**1.2 分析方法** ① 序列建立及特征分析:据月发病数及人口数计算月发病率,并建立2004年1月至2016年12月发病率时间序列。用LOESS光滑对序列做季节性分解,同时计算各月的季节指数,据季节指数确定发病高峰。② 序列的平稳化:用自相关图、偏自相关图及单位根检验对序列进行平稳性检验;对非平稳序列用差分法平稳化。平稳序列用Ljung-Box检验是否为白噪声;对白噪声序列选择合适的分析方法;对非白噪声序列建立模型。③ 模型的识别及建立:据平稳时间序列的自相关图(ACF)和偏自相关图(PACF),初步估计模型阶数。按AIC值最小原则选择最优模型。④ 模型拟合及评价:根据建立的模型拟合模型参数、绘制残差的QQ图,检验是否满足正态分布。用Ljung-Box检验残差是否满足独立性要求。⑤ 预测:用平均绝对标准化误差(mean absolute error, MASE)衡量模型的拟合优度,MASE<1时模型拟合较好。用建立的模型预测2016年1—11月全国狂犬病月发病率,计算预测值和实际值间的相对误差以判断模型的预测精度。

## 2 结 果

**2.1 疫情概况** 2004年1月至2016年11月的月发病率时间序列(图1A)。用LOESS光滑对序列进行季节性分解结果显示,疫情存在明显的季节性,其中8~10月为高峰(32.4%),季节指数分别为1.277、1.285、1.224(图1B);自2007年后发病率呈逐年下降趋势(图1C)。

**2.2 模型识别及建立** 用2004—2015年月发病率数据建立模型。其自相关图(ACF)显示为非平稳序列,需进行差分(图2)。对序列进行1阶差分后,显示为平稳序列(Dickey-Fuller = -9.352,  $P < 0.01$ );进行Box-Ljung检验显示为非白噪声序列( $C^2 = 4.906$ ,  $P = 0.0268$ )。由于有季节性高峰,故用ARIMA(p,d,q)(P,D,Q)<sub>12</sub>模型,因此d=1,D=0。据1阶差分序列的ACF图和PACF图的图形类别,逐步选择合适的p、q、P及D值,以AIC值最小者为最优模型,结果显示最优模型为ARIMA(2,1,1)(2,0,0)<sub>12</sub>, AIC=-57.07。

**2.3 模型拟合及评价** 对建立的ARIMA(2,1,1)(2,0,0)<sub>12</sub>进行拟合,所有系数均有统计学意义(表1)。绘制模型残差的QQ图显示满足正态分布(图3)。模型残差ACF图及PACF图显示无明显的相关性(图4);Box-Ljung检验结果也显示模型残差无

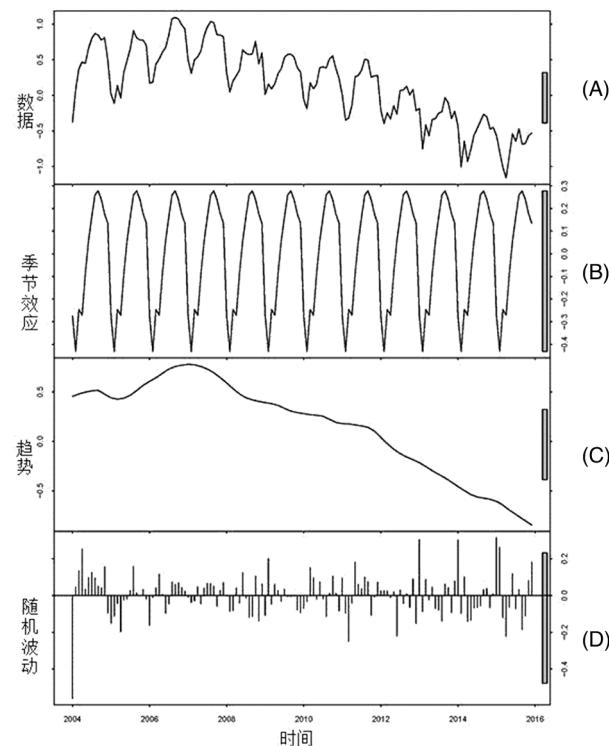


图1-A:月发病率时间分布;图1-B:季节效应;图1-C:年度趋势分布;图1-D:残差随机波动

Fig.1-A: temporal distribution graph; Fig.1-B: seasonality effect graph; Fig.1-C: annual trend; Fig.1-D: residual distribution

图1 全国狂犬病月发病率序列季节性分解

Fig.1 The seasonality decomposition diagram of monthly rabies incidence time series of China

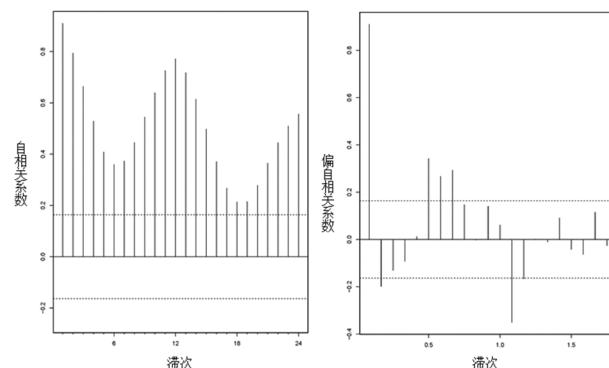


图2 全国狂犬病2004—2015年月发病率时间序列自相关图ACF(左)和偏自相关图PACF(右)

Fig.2 The autocorrelation function (ACF, left) and the partial autocorrelation function (PACF, right) of the monthly rabies incidences in China (2004—2015)

明显的相关性( $C^2 = 0.339$ ,  $P = 0.560$ )、是白噪声序列。

表 1 ARIMA(2,1,1)(2,0,0)<sub>12</sub>模型参数拟合值Tab.1 The fitting values of parameters in ARIMA(2,1,1)(2,0,0)<sub>12</sub>

指标 index	AR1 <sup>a</sup>	AR2 <sup>b</sup>	MA1 <sup>c</sup>	SAR1 <sup>d</sup>	SAR2 <sup>e</sup>
系数 coefficient	0.665	0.202	-0.985	0.452	0.372
标准误 standard error	0.096	0.098	0.031	0.080	0.086
t 值 t value	6.927 **	2.061 *	3.177 **	5.65 **	4.326 **

注释:a 和 b 分别为 AR 模型的第 1 和第 2 个参数,c 为 MA 模型的第 1 个参数,d 和 e 分别是季节性 AR 模型的第 1 和第 2 个参数; \* P<0.05; \*\* P<0.01。

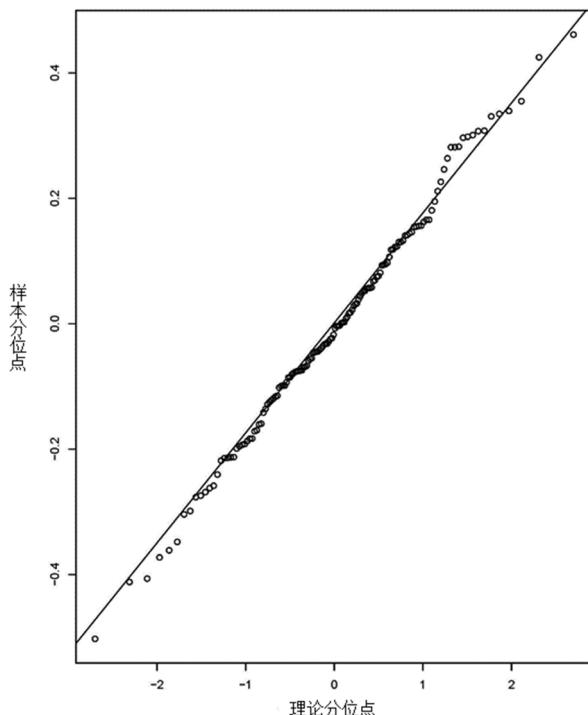
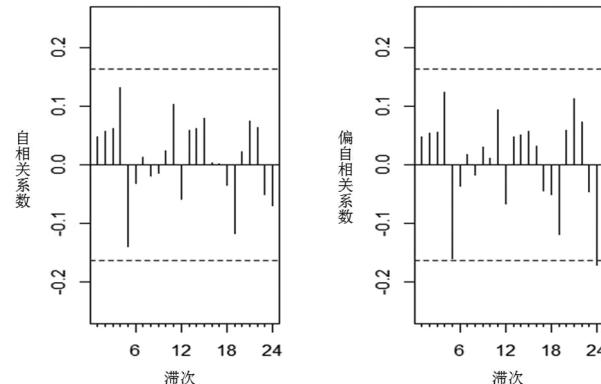
图 3 ARIMA(2,1,1)(2,0,0)<sub>12</sub> 模型的残差 QQ 图Fig.3 Residual QQ plot of ARIMA(2,1,1)(2,0,0)<sub>12</sub>

图 4 模型残差的自相关图 ACF (左) 和偏自相关图 PACF (右)

Fig.4 The autocorrelation function (ACF, left) and the partial autocorrelation function (PACF, right) of the model residual

2.4 预测 R 软件分析结果显示模型 ARIMA(2,1,1)(2,0,0)<sub>12</sub> 的 MASE=0.755, 此值<1 表明拟合效果较好。用此模型预测 2016 年 1—11 月发病率, 结果相对误差 -17.25%~62.48%, 平均相对误差为 15.61% (表 2)。

表 2 全国狂犬病 2016 年 1—11 月份月发病率预测结果

Tab.2 The predicted incidences of rabies in China, from Jan. to Nov., 2016

月份 Month	实际值 actual value	预测值(95% CI) predicted value(95% CI)	绝对误差 absolute error	相对误差 relative error
1	0.400	0.511(0.146~0.875)	0.111	0.276
2	0.335	0.355(0~0.795)	0.0199	0.059
3	0.233	0.378 (0~0.882)	0.145	0.625
4	0.211	0.309 (0~0.858)	0.098	0.465
5	0.473	0.391 (0~0.975)	-0.082	-0.173
6	0.393	0.501(0~1.112)	0.108	0.274
7	0.429	0.498(0~1.132)	0.069	0.160
8	0.458	0.570 (0~1.222)	0.112	0.244
9	0.553	0.543 (0~1.211)	-0.010	-0.018
10	0.560	0.537(0~1.217)	-0.023	-0.041
11	0.378	0.521(0~1.212)	0.143	0.377

### 3 讨 论

全球除印度外中国年报告狂犬病死亡最多<sup>[1]</sup>。50年代以来,我国人间狂犬病出现3次高峰<sup>[2-3]</sup>。第1个高峰50年代中期年死亡最高达1 900多例。第2次高峰为80年代初期,1981年死亡7 037例是最高年份,80年代年均死亡5 537例。第3次高峰在21世纪初期,疫情重现快速增长趋势。主要宿主动物是犬,调查显示93.74%病例的感染来源是犬<sup>[4]</sup>。广西、贵州、广东、湖南和四川是近年来高发省份。此病潜伏期数天至≥1年,但多为20~90 d<sup>[2]</sup>。可在每年高峰季节到来前的3~4个月,对<15岁儿童、农民及流动人口等高危人群及高危地区,加强健康教育,提高暴露后的规范处置率,以减少发病。

时间序列分析据该观察值历史数据来预测其发展趋势,探索其时间分布特征及影响因素。ARIMA模型应用于经济、工程、气象、水利等众多领域。医学卫生领域也用于预测疾病发病率、人群寿命、医疗卫生费用和食物中毒等,还用于研究气候及环境因素对人群健康的影响<sup>[5-11]</sup>。

本研究用2004年1月至2016年11月全国狂犬病月发病率数据进行时间序列分析。显示2007年以来发病率呈下降趋势,疫情有明显季节性(8~10月高峰),与既往报道的时间分布特征类似<sup>[2,4]</sup>。

建立ARIMA模型至少要有50个时间点或7~8个周期的数据。本研究用了144个数据。通过数据差分、模型识别,建立最优模型ARIMA(2,1,1)(2,0,0)<sub>12</sub>。已知该模型的MASE=0.755,拟合效果较好。对2016年1~11月发病率预测显示效果较好,平均相对误差15.61%。由于ARIMA模型预测时是假定事件影响因素无明显变化,因此ARIMA模型常用于时间序列数据的短期预测。

### 参 考 文 献:

- [1] Jackson AC. Human rabies: a 2016 update[J]. Curr Infect Dis Rep, 2016, 18(11):38. DOI: 10.1007/s11908-016-0540-y
- [2] Zhou H, Vong S, Liu K, et al. Human rabies in China, 1960—2014: a descriptive epidemiological study[J]. PLoS Negl Trop Dis, 2016, 10(8):e4874. DOI: 10.1371/journal.pntd.0004874
- [3] Mao WC, Lan RW, Meng NX, et al. Rabies epidemical and prevention and control research progress[J]. J Med Pest Control, 2015(3):283-286. (in Chinese)
- 毛伟成,蓝荣伟,蒙南新,等.狂犬病流行与防治研究进展[J].医学动物防制, 2015(3):283-286.
- [4] Song M, Tang Q, Rayner S, et al. Human rabies surveillance and control in China, 2005—2012[J]. BMC Infect Dis, 2014, 14: 212. DOI: 10.1186/1471-2334-14-212
- [5] Wang KW, Deng C, Li JP, et al. Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network[J]. Epidemiol Infect, 2017, 1-12. DOI: 10.1017/S0950268816003216
- [6] Liu L, Luan RS, Yin F, et al. Predicting the incidence of hand, foot and mouth disease in Sichuan province, China, using the ARIMA model[J]. Epidemiol Infect, 2016, 144(1): 144-151. DOI: 10.1017/S0950268815001144
- [7] Anwar MY, Lewnard JA, Parikh S, et al. Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence[J]. Malar J, 2016, 15(1):566. DOI: 10.1186/s12936-016-1602-1
- [8] Zhang X, Pang Y, Cui M, et al. Forecasting mortality of road traffic injuries in China using seasonal autoregressive integrated moving average model[J]. Ann Epidemiol, 2015, 25(2):101-106. DOI: 10.1016/j.annepidem.2014.10.015
- [9] Chadsuthi S, Iamsirithaworn S, Triampo W, et al. Modeling seasonal influenza transmission and its association with climate factors in Thailand using time-series and ARIMAX analyses[J]. Comput Math Methods Med, 2015, 2015: 436495. DOI: 10.1155/2015/436495
- [10] Zhang Y, Wang SN, Liu Y, et al. Application of ARIMA model on predicting monthly hospital admissions and hospitalization expenses for respiratory diseases[J]. Chin J Health Statistic, 2015, 32(2):197-200. (in Chinese)
- 张越,王胜难,刘媛,等.应用ARIMA模型对呼吸系统疾病月住院量及住院费用的预测[J].中国卫生统计,2015,32(2):197-200.
- [11] Wang L, Yao WQ. Forecasting the epidemic trend of hand-foot-and-mouth disease with time sequence model in Liaoning, [J]. Chin J Health Statistic, 2016, 33(5):847-849. (in Chinese)
- 王伶,姚文清.利用时间序列模型分析预测辽宁手足口病疫情趋势[J].中国卫生统计,2016,33(5):847-849.