

应用 ARIMA 模型预测厦门市 2017 年乙肝月发病率

杨姝怡¹, 马姗姗², 苏成豪²

1. 厦门大学公共卫生学院, 福建 厦门 361000 2. 复旦大学附属中山医院厦门医院, 福建 厦门 361000

摘要:目的 分析厦门市乙型肝炎发病的流行病学特征, 建立 ARIMA 模型并预测 2017 年厦门市乙肝月发病率, 为厦门市乙肝防治提供理论依据。方法 对 2005 - 2016 年厦门市乙肝报告发病数据进行统计描述, 通过 ARIMA 模型预测 2017 年的乙肝月发病情况。结果 厦门市 2005 - 2016 年乙肝报告年均发病率为 45.70/10 万, 乙肝年发病率呈逐渐下降趋势; 男性乙肝发病高于女性; 各年龄组乙肝发病水平存在差异; 15 岁以下儿童乙肝发病率显著低于其他高年龄组的发病水平。建立模型 ARIMA (1,1,1)×(0,1,0)₁₂, 模型的拟合效果较好, 2017 年厦门市乙肝月发病率预测值区间为 [0.92/10 万, 4.76/10 万]。结论 厦门市 2005 - 2016 年乙肝发病总体呈现下降趋势, 经预测 2017 年厦门市乙肝发病率与前几年相比波动不大, 有轻微下降趋势, 可用 ARIMA 模型对乙肝的短期发病情况进行预测。

关键词: 乙肝; ARIMA; 预测

中图分类号: R181.2 文献标志码: A 文章编号: 1003 - 8507(2018)10 - 1729 - 05

The application of ARIMA model in predicting the monthly incidence of HBV in Xiamen city 2017

YANG Shu-yi*, MA Shan-shan, SU Cheng-hao

*School of Public Health, Xiamen University, Xiamen, Fujian 361000, China

Abstract: Objective Analyzing the epidemiologic characters of HBV in Xiamen city, building ARIMA model and predicting the monthly incidence of HBV in Xiamen city in 2017, and providing theory evidence for preventive treatment of HBV in Xiamen city. **Methods** Statistical describing the HBV case in Xiamen city during 2005-2016, building ARIMA model and applying it to predict the monthly incidence of HBV in 2017. **Results** The average HBV reported incidence is 45.70/100000 in Xiamen city since 2005-2016. There is a general downtrend of incidence of HBV, and the incidence rate is higher in men than women. The incidence rate shows difference in age groups; children under 15 show significant lower incidence level than groups which age older. The model ARIMA (1,1,1)×(0,1,0)₁₂, has been established and fits well, and the prediction interval of monthly incidence of HBV in Xiamen city is [0.92/100000, 4.76/100000]. **Conclusion** The HBV incidence totality trends down from 2005-2016 in Xiamen city, and it remains barely unchanged in 2017 comparing with lately previous years, and has a slightly descending trend. ARIMA model could be used in predicting HBV incidence level in short period.

Keywords: Hepatitis B; ARIMA; Forecasting

乙肝是危害国人健康最为严重的传染病之一, 我国是 HBV 感染率较高的国家, 而厦门市是全国乙肝及肝癌的高发区^[1-3]。时间序列分析是分析动态数据、揭示数据规律性的统计学分支学科, 主要依赖变量自身历史资料建立出能反映序列动态规律性的模型^[4], 最早应用于计量经济学, 现也应用于医学领域^[5-7]。本研究以 2005 - 2016 厦门市乙肝月发病率数据为基础, 建立 ARIMA 模型并对 2017 年乙肝月发病率进行预测, 为乙肝控制政策的制定提供依据。

1 材料与方法

1.1 数据来源 2005 - 2016 年厦门乙肝报告发病

作者简介: 杨姝怡(1993 -), 女, 硕士研究生在读, 研究方向: 乙型肝炎流行病学

通讯作者: 苏成豪 E-mail: 1272208372@qq.com

数据及人口数据均来源于中国疾病监测信息报告系统。按照现住址及发病日期导出厦门市 2005 - 2016 乙肝报告发病病例 Excel 表格, 删除发病日期非当年的病例, 删除疾病类型为病原携带者、阳性检测和疑似病例。

1.2 统计分析方法 使用 SPSS13.0 统计分析软件对厦门乙肝的报告发病情况进行描述性分析。使用时间序列方法对月发病率进行建模, 并对乙肝未来发病趋势进行预测。

1.3 病例分类及定义

1.3.1 急性乙肝 (1) 近期出现无其他原因可解释的乏力和消化道症状, 可有尿黄、眼黄和皮肤黄疸。(2) 肝脏生化检查异常, 主要是血清 ALT 和 AST 升高, 可有血清胆红素升高。(3) HBsAg 阳性。(4) 有证

据表明 6 个月内曾检测血清 HBsAg 阴性。(5) 抗 - HBc IgM 阳性 1 : 1 000 以上。(6)胆组织学符合急性病毒性肝炎改变。(7)恢复期血清 HBsAg 阴转, 抗 - HBs 阳转。同时符合(1)和(3)或者同时符合(2)和(3)可诊断为疑似急性乙肝病例。疑似病例且同时符合(4)、(5)、(6)、(7) 中的任何 1 项可确诊急性乙肝病例。

1.3.2 慢性乙肝 (1) 急性 HBV 超过 6 个月仍 HBsAg 阳性或发现 HBsAg 阳性超过 6 个月。(2) HBsAg 阳性持续时间不详, 抗 HBcIgM 阴性。(3) 慢性肝病患者的体征如肝病面容, 肝掌、蜘蛛痣和肝、脾肿大等。(4)血清 ALT 反复或持续升高, 血浆白蛋白降低(或)球蛋白升高, 或胆红素升高等。(5)肝脏病理学有慢性病毒性肝炎的特点。(6)血清 HBsAg 阳性或可检出 HBV DNA, 并排除其他导致 ALT 升高的原因。同时符合(1)(3)或(2)(3)或(2)(4)者可诊断为疑似慢性乙肝病例。疑似病例同时符合(1)(4)(6)或(1)(5)(6)或(2)(4)(6)或(2)(5)(6)者可确诊慢性乙肝病例。

1.3.3 未分类乙肝 根据我国传染病报告信息管理系统的要求, 乙肝按照急性乙肝、慢性乙肝进行分类报告, 不能明确分类的报告为未分类乙肝^[8]。

1.4 时间序列模型简介 ARIMA (Autoregressive Integrated Moving Average) 称为自回归滑动平均混合模型, 试图解决以下 2 个问题 (1)分析时间序列的随机性、平稳性和季节性(2)在对时间序列分析的基础上, 选择合适的模型进行预测。包括自回归、滑动平均及差分求和 3 个过程。

该方法的基本思想是: 将预测对象随时间推移而形成的数据序列视为一个随机序列, 即除去个别的因偶然原因引起的观测值外, 时间序列是一组依赖于时间的随机变量。这组随机变量所具有的依存关系或自相关性表征了预测对象发展的延续性, 而这种自相关性一旦被相应的数学模型描述出来, 就可以从时间序列的过去值及现在值预测未来值^[9]。

ARIMA 方法包括模型识别、模型估计、模型诊断、建立控制图表 4 个步骤, 见图 1。

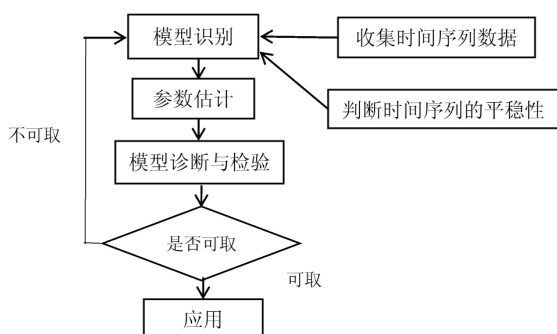


图 1 时间序列分析流程图

2 结果

2.1 流行情况 厦门市 2005 - 2016 年合计报告乙肝病例 16 756 例。病例分类中, 急性病例占 19.41% (3 253/16 756), 慢性病例占 63.51% (10 642/16 756), 未分类病例占 17.07% (2 861/16 756)。2005 - 2016 厦门市乙肝年均发病率为 45.70/10 万, 乙肝年发病率总体呈逐年下降趋势; 男性乙肝发病高于女性, 各年龄组乙肝发病存在差异, 15 岁以下儿童乙肝发病水平显著低于高年龄组, 结果见表 1。

表 1 2005 - 2016 年厦门市乙型肝炎流行情况

类别	发病率(1/10 万)	病例数(%)
年份		
2005	100.70	2 308(13.77)
2006	120.02	2 828(16.88)
2007	78.20	1 822(10.87)
2008	68.74	1 671(9.97)
2009	48.88	1 217(7.26)
2010	55.16	1 390(8.30)
2011	22.00	777(4.64)
2012	29.56	1 067(6.37)
2013	23.98	883(5.27)
2014	24.99	937(5.59)
2015	22.99	876(5.23)
2016	25.39	980(5.85)
性别		
男	62.09	11 745(70.09)
女	28.24	5 011(29.91)
病例分类		
急性	8.87	3 253(19.41)
慢性	29.03	10 642(63.51)
未分类	7.80	2 861(17.07)
年龄组(岁)		
0~	3.24	167(1.00)
15~	58.97	6 351(37.90)
30~	71.09	5 965(35.60)
45~	47.90	2 733(16.31)
60~	55.98	1 540(9.19)

2.2 乙肝发病趋势预测 本研究选用 2005 - 2016 年的月发病数据进行建模。2005 - 2016 年厦门市乙肝月发病率在 0.47/10 万~20.20/10 万之间, 该序列存在明显下降趋势, 为非平稳序列, 见图 2。

2.2.1 数据预处理和模型识别 从序列图观察, 月发病率具有明显下降趋势和周期性, 对序列进行对数转换、一阶差分、一阶季节差分处理, 形成新序列(见图 3)。对差分后序列进行单位根检验得到 Dickey - Fuller = - 18.701, P = 0.00, 差异具有统计学意义, 表明差分后序列为平稳序列。对差分后的序列做 ACF、PACF 图(见图 4), 可见 ACF、PACF 均为拖尾, 确定

模型为 ARMA(1,1)。由于经过一阶差分和一阶季节性

差分 模型为 ARIMA(1,1,1) × (? , 1, ?)₁₂。

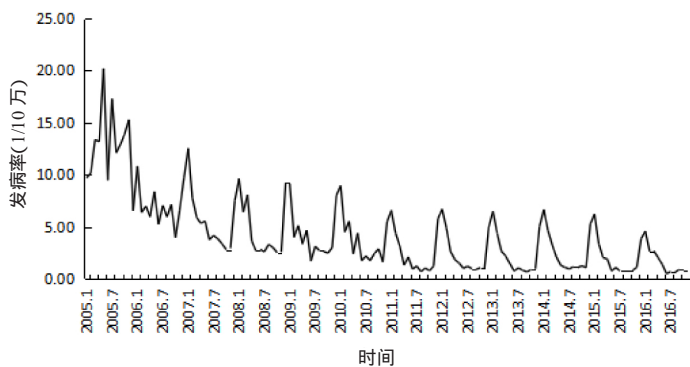


图2 厦门市 2005 - 2016 年乙肝报告发病率时间序列分布

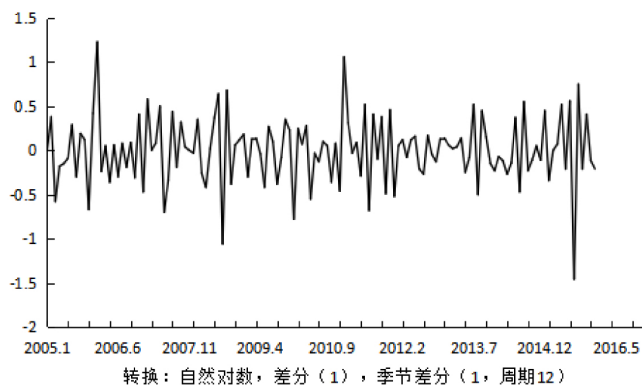


图3 厦门市 2005 - 2016 年乙肝月发病率差分后序列图

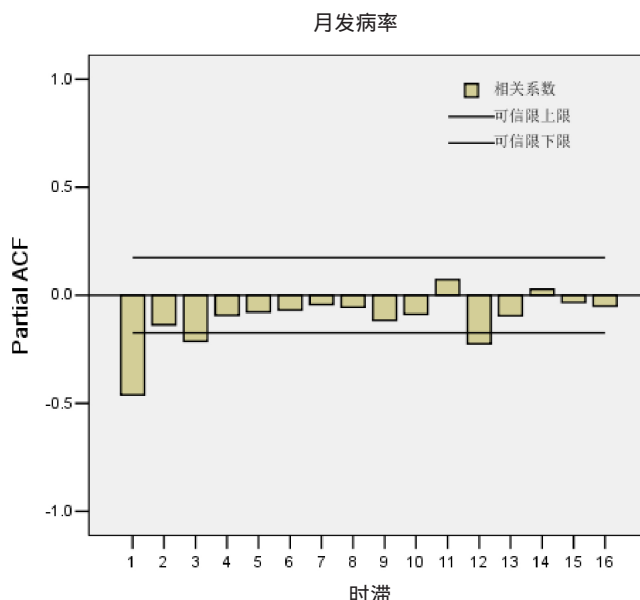
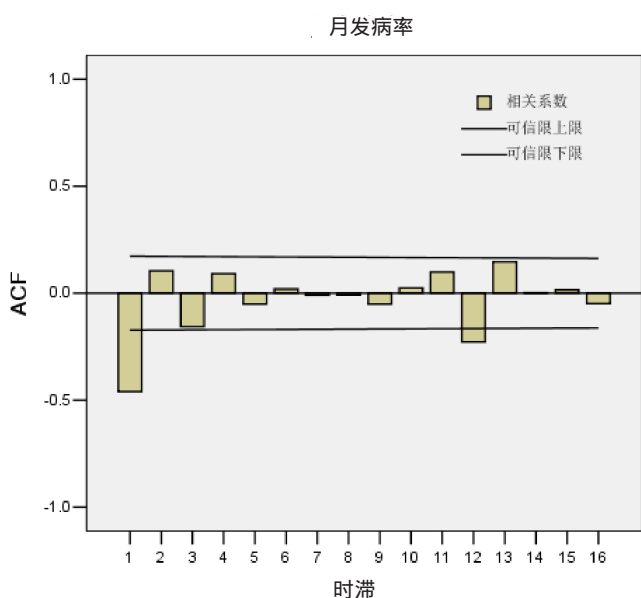


图4 厦门 2005 - 2016 乙肝月发病率经一阶差分、一阶季节性差分后自相关图和偏自相关图

2.2.2 参数估计及模型选择 在上一步基础上尝试建模,可能的组合、所估计参数及模型参数 *t* 检验情况见表 2,可以看到仅模型 ARIMA(1,1,1) × (0,1,0)₁₂ 模型所估计的参数通过了 *t* 检验。对各模型进行拟合优度检验情况,根据最小 AIC、BIC 准则选出最优模

型亦为 ARIMA(1,1,1) × (0,1,0)₁₂。对该模型的残差序列进行白噪声检验,残差自相关检验结果表明,在任何时滞上的 *P* 值均远大于 0.05,可认为残差为白噪声序列,模型通过诊断,见表 3。

表 2 ARIMA(1,1,1) × (? , 1, ?)₁₂ 不同参数组合拟合后参数估计结果及拟合优度统计量

模型	AR1	MA1	SAR1	SMA1	常数项	AIC	BIC	标准误差	对数似然
ARIMA(1,1,1) × (0,1,0) ₁₂						519.697	528.322	1.731	-256.848
B	0.469	0.941	-	-	0.030				
<i>t</i> 值	5.109	22.683	-	-	1.595				
<i>P</i> 值	0.000	0.000	-	-	0.113				
ARIMA(1,1,1) × (1,1,0) ₁₂						521.343	532.844	1.735	-256.672
B	0.464	0.940	0.072	-	0.032				
<i>t</i> 值	4.918	22.675	0.808	-	1.562				
<i>P</i> 值	0.000	0.000	0.421	-	0.121				
ARIMA(1,1,1) × (1,1,1) ₁₂						523.326	537.702	1.727	-256.663
B	0.481	0.950	0.930	0.982	0.029				
<i>t</i> 值	5.269	23.086	2.181	1.370	1.877				
<i>P</i> 值	0.000	0.000	0.031	0.173	0.063				
ARIMA(1,1,1) × (0,1,1) ₁₂						521.401	532.902	1.735	-256.701
B	0.467	0.942	-0.061	-	0.032				
<i>t</i> 值	4.976	22.971	-0.681	-	1.588				
<i>P</i> 值	0.000	0.000	0.497	-	0.115				

注：“-”代表模型中未应用该项参数及统计量。

表 3 ARIMA(1,1,1)×(0,1,0)₁₂ 模型残差检验结果

时滞	自相关系数	标准误	BOX- ljung 检验		
			检验统计量	自由度	P 值
1	-0.049	0.086	0.325	1	0.569
2	0.104	0.086	1.786	2	0.409
3	0.042	0.086	2.031	3	0.566
4	0.035	0.085	2.199	4	0.699
5	0.014	0.085	2.228	5	0.817
6	0.120	0.085	4.239	6	0.644
7	-0.119	0.084	6.241	7	0.512
8	0.013	0.084	6.266	8	0.617
9	-0.173	0.084	10.515	9	0.310
10	-0.029	0.083	10.638	10	0.386
11	-0.047	0.083	10.957	11	0.447
12	-0.008	0.083	10.966	12	0.532
13	-0.002	0.082	10.967	13	0.614
14	0.139	0.082	13.842	14	0.462
15	0.022	0.082	13.914	15	0.532
16	0.001	0.081	13.914	16	0.605

2.2.3 模型拟合值与实际值的比较 ARIMA (1,1,1)×(0,1,0)₁₂ 模型拟合值和月发病率实际值的拟合情况基本一致,黑线后为 2017 年 1 月 - 12 月的预测值。

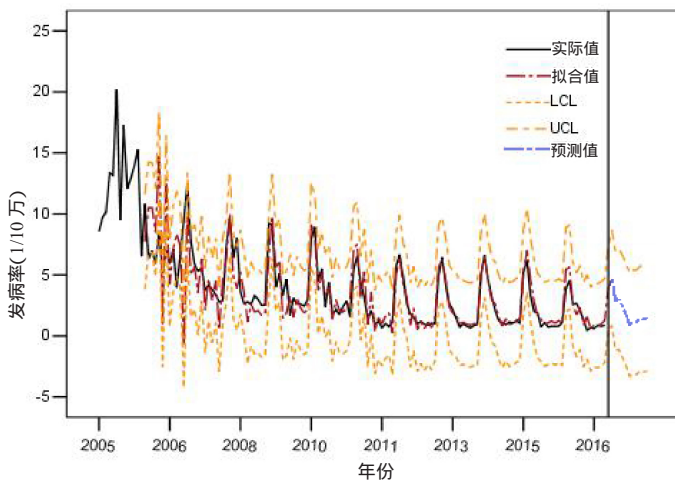


图 5 ARIMA(1,1,1)×(0,1,0)₁₂ 模型拟合预测情况

3 讨论

病毒性肝炎是全球公共卫生面临的重要威胁之一,据《2017 全球肝炎报告》估算,2015 年全球约 2.57 亿人口为乙型肝炎患者,占世界总人口的 3.5%。西太平洋地区(包括中国)的乙肝负担最重,有 1.15 亿患者,占地区人口的 6.2%,这些感染者主要是乙肝疫苗投入使用之前出生的成年人^[10]。

2005 年厦门市乙肝报告发病率为 100.70/10 万,高于全国同期 75.57/10 万的发病水平^[11],2005 - 2007 全国乙肝报告发病率表现为上升趋势,2007 年为全国乙肝报告发病率最高的年份(89.00/10 万),而厦门

市乙肝发病除 2010 年有增高外表现为下降趋势,至 2007 年发病率已低于全国发病水平(78.20/10 万)。2007 年后,全国乙肝报告发病率存在下降趋势,2016 年乙肝发病率为 68.74/10 万,厦门市 2016 年的乙肝报告发病率为 25.39/10 万,显著低于国家平均水平。提示近 10 年厦门市在乙肝防控工作中已取得了一定成效。

2005 - 2016 年厦门市乙肝报告年均发病率为 45.70/10 万,低于近期研究中 2001 - 2015 年宜昌市 87.26/10 万的年均报告发病率^[12],也低于铜川市 101.29/10 万和阆中市 76.27/10 万的发病水平^[13-14]。另外笔者发现所计算的年均发病率这一结果与张冬娟在“福建省 2004 - 2016 年乙型肝炎特征分析”研究中计算的厦门市 2004 - 2013 年均报告发病率 119.94/10 万的结果出入较大,可能与数据清洗标准的差异有关^[15]。提示在进行乙肝流行病学研究时,应该统一数据清洗标准,使用统一人口资料,才能使研究结果具有可比性。

各种统计预测方法现已广泛应用于人口预测、疾病发展、门诊量、药品需求等医学领域。ARIMA 方法是集自回归、滑动平均、差分求和于一体的混合模型,也是当下传染病监测的重要手段。近年来,越来越多的时间序列数据通过 ARIMA 方法建模,当数据单位为按月计算时,做出模型预测的时长常可为 1 年,如潘欢弘等在手足口病月发病率预测中的应用^[16]、谭恩丽等人对包虫病的月发病率预测中的应用^[17],也有研究中使用 ARIMA 做出预测的时长长达 3 年,如陈莉等对淋病发病趋势预测的研究^[18]。时间序列常具有周期性,季节性为普遍现象也是建模的重要依据,预测数据中亦可拟合出 1 年中各个季节的数据分布,因

(下转第 1744 页)

- [7] 冯冰,甘华芬,王文勇,等.武汉市江江区 2005 - 2014 年法定丙类传染病疫情分析[J].公共卫生与预防医学, 2015, 26(5): 44 - 47.
- [8] 覃大章,钟少华.宾阳县 2008 - 2011 年丙类传染病流行特征分析[J].现代预防医学, 2014, 41(1): 155 - 157, 163.
- [9] 何建英,毛伟成,覃萍,等.来宾市 2010 - 2014 年丙类传染病流行病学分析[J].医学动物防制, 2016, 32(6): 606 - 609.
- [10] 张琦,祝绯飞,陈田,等.杭州市拱墅区 2008 年手足口病流行特征及控制效果分析[J].中国学校卫生, 2009, 30(7): 658 - 659.
- [11] 张行钦,孙智夫,翁永夫.舟山市定海区 2004 - 2008 年丙类传染病流行病学监测分析[J].中国农村卫生事业管理, 2010, 30(10): 852 - 854.
- [12] 刘敬东,刘光定,张延斌.2006 - 2015 年襄阳市法定丙类传染病疫情分析[J].职业与健康, 2017, 33(4): 516 - 518.

收稿日期 2018 - 01 - 31

(上接第 1732 页)

此,使用模型对未来 1 年的趋势进行预测是可行的。另有研究表明,ARIMA 模型预测精度随预测时长增加逐步下降,说明模型不宜进行远期预测^[9]。已有研究尚未对有效预测时长进行规定,还需进一步的规范和探究,这也是本研究的局限性所在。本研究使用 2005 年 1 月 - 2016 年 12 月合计 12 年、144 个月的乙肝发病率数据建立了 ARIMA (1,1,1)×(0,1,0)₁₂ 模型,样本量符合 ARIMA 建模需要,从控制图表来看,模型拟合情况较好,预测结果表明 2017 年厦门市乙肝月发病率预测值区间为[0.92/10 万, 4.76/10 万],其发病水平与 12 年之后厦门市乙肝发病情况相似。若影响乙肝发病的因素基本保持相同,就能得到序列的 95% 部分都保持在上下可信限之间。反过来说,若实际发病数据超出预测值的置信区间,则可认为当地乙肝的流行特征发生了变化。据此,可根据时间序列模型对乙肝进行趋势分析,掌握乙肝发病规律,从而更好地制定防控策略及评估防控手段。

参考文献

- [1] 王富珍,崔富强.中国 8 个省 18 个县(区)乙型病毒性肝炎监测病例分析[J].中国计划免疫, 2007, 13(4): 309 - 311.
- [2] 张国民,孙校金,王富珍,等.中国 18~59 岁人群乙型病毒性肝炎流行病学特征分析及乙型肝炎疫苗免疫策略探讨[J].中国疫苗和免疫, 2013, 19(3): 266 - 270.
- [3] 苏成豪,牛建军,张一中,等.厦门市乙型肝炎、肝癌流行趋势及其关系研究[J].旅行医学科学, 2007, 13(3): 30 - 31.
- [4] 史代敏,谢小燕.应用时间序列分析[M].北京:高等教育出版社, 2011.
- [5] 朱琦,于石成,郝元涛.传染病监测数据的统计分析方法[J].中国卫生统计, 2011, 28(2): 217 - 220.
- [6] 吴爱萍,陈银苹,张天哲,等.ARIMA 模型在乙肝发病预测中的应用[J].中国现代医学杂志, 2012, 22(22): 78 - 82.
- [7] 王涛,苑新海,朱宗龙.ARIMA 时间序列在乙肝发病预测中的应用[J].中国卫生统计, 2014, 31(4): 646 - 647, 650.
- [8] 李黎,崔富强,张国民,等.乙型肝炎诊断标准(W/S299 - 2008)解读[J].中华预防医学杂志, 2014, 16(9): 758 - 761.
- [9] 孙振球,徐勇勇.医学统计学[M].北京:人民卫生出版社, 2014.
- [10] 世界卫生组织.2017 全球肝炎报告[R].日内瓦:世界卫生组织, 2017.
- [11] 张国民,缪宁,郑徽,等.中国 2005 - 2016 年乙型病毒性肝炎报告发病的年龄和地区特征[J].中国疫苗和免疫, 2018, 1(2): 121 - 126.
- [12] 徐英,赵露,鲁芳芳,等.2001 - 2015 年宜昌市乙型病毒性肝炎流行病学特征与免疫接种效果[J].职业与健康, 2018, 34(2): 233 - 236.
- [13] 胡改霞,刘新利,蔺淞锋.2011 - 2016 年陕西省铜川市乙型肝炎报告及准确性分析[J].中国疫苗和免疫, 2018, 1(1): 23 - 26.
- [14] 伏利蓉,吴小芳,许华,等.2007 - 2016 年阆中市乙肝流行趋势分析[J].职业卫生与病伤, 2017, 32(6): 325 - 329.
- [15] 张冬娟,洪荣涛,郑金凤,等.福建省 2004 - 2013 年乙型肝炎特征分析[J].中国预防医学杂志, 2015, 16(1): 18 - 22.
- [16] 潘欢弘,朱蒙曼,刘晓青.ARIMA 乘积季节模型在江西省手足口病发病预测中的应用[J].现代预防医学, 2018, 45(1): 1 - 4, 7.
- [17] 谭恩丽,王正峰,周文策,等.我国包虫病报告病例数自回归移动平均模型预测研究[J].中国血吸虫病防治杂志, 2018, 1(1): 47 - 53.
- [18] 陈莉,陈棠晖,黎江,等.海南省 2010 - 2016 年淋病流行特征分析与趋势预测[J].实用预防医学, 2018, 25(1): 27 - 29.
- [19] 方积乾,陆盈.现代医学统计学[M].北京:人民卫生出版社, 2002: 219 - 264.

收稿日期 2017 - 07 - 10