

· 临床研究 ·

差分自回归移动平均模型在南通市手足口病疫情预测中的应用

练 维¹, 魏 叶², 韩颖颖³, 帅小博^{4*}

¹南通市疾病预防控制中心科研质量管理科, ²急性传染病防制科, ³慢性非传染病防制科, 江苏 南通 226007; ⁴南通市崇川区疾病预防控制中心检验科, 江苏 南通 226000

[摘要] 目的:应用差分自回归移动平均模型(ARIMA)预测南通市手足口病疫情趋势。方法:以2010年1月—2019年6月南通市手足口病分月报告病例数据为基础,构建符合季节性时间序列的ARIMA(p,d,q)×(P,D,Q)_s模型,用2019年7—12月全市手足口病月发病率为验证数据进行验证,检验模型的预测效果。结果:2010—2019年南通市共报告手足口病90 766例,年平均发病率为124.36/10万,疫情有明显季节性,呈双峰特征,为夏季(5、6、7月)高峰和冬季(11、12月)次高峰;近年来南通市手足口病的病原谱以其他肠道病毒为主;利用ARIMA(1,0,1)(1,1,1)₁₂模型,预测2019年7—12月手足口病发病率分别为7.08/10万、1.81/10万、3.74/10万、7.21/10万、10.71/10万和11.29/10万,与实际发病率相比,两者差异无统计学意义(Z=0.48, P=0.63)。结论:差分自回归移动平均模型能较好地预测手足口病的发病趋势,可用于短期的预警监测。

[关键词] 差分自回归移动平均模型;手足口病;预测

[中图分类号] R181.8

[文献标志码] A

[文章编号] 1007-4368(2021)01-059-06

doi:10.7655/NYDXBNS20210111

Application of autoregressive integrated moving average model in prediction of hand-foot-mouth disease epidemic in Nantong

LIAN Wei¹, WEI Ye², HAN Yingying³, SHUAI Xiaobo^{4*}

¹Department of Research and Quality Management, ²Department of Acute Communicable Diseases Control and Prevention, ³Department of Chronic Non-communicable Diseases Control and Prevention, Nantong Center for Disease Control and Prevention, Nantong 226007; ⁴Department of Clinical Laboratory, Chongchuan District Center for Disease Control and Prevention of Nantong City, Nantong 226000, China

[Abstract] **Objective:** To predict the trend of hand-foot-mouth disease in Nantong by using the autoregressive integrated moving average (ARIMA) model. **Methods:** Based on the monthly reported case data of hand-foot-mouth disease in Nantong from January 2010 to June 2019, a ARIMA (p, d, q) × (P, D, Q)_s model was constructed in accordance with the seasonal time series. The monthly incidence of hand-foot-mouth disease in the city from July to December 2019 was used as verification data to test the prediction effect of the model. **Results:** A total of 90 766 cases of hand-foot-mouth disease were reported in Nantong from 2010 to 2019. The average annual incidence rate was 124.36/100 000. The epidemic showed seasonality of incidence with two peaks. One was the summer peak and the other was the winter sub-peak. In recent years, the pathogenic spectrum of hand-foot-mouth disease in Nantong was dominated by other intestinal viruses. Using the ARIMA (1, 0, 1) (1, 1, 1)₁₂ model, the incidence of hand-foot-mouth disease from July to December 2019 was predicted to be 7.08/100 000, 1.81/100 000, 3.74/100 000, 7.21/100 000, 10.71/100 000, and 11.29/100 000. Compared with the actual incidence, there was no significant difference. **Conclusion:** The autoregressive integrated moving average model can better predict the incidence of hand-foot-mouth disease, and can be used for short-term early warning and monitoring.

[Key words] ARIMA model; hand-foot-mouth disease; prediction

[J Nanjing Med Univ, 2021, 41(01):059-064]

[基金项目] 南通市卫计委青年基金项目(WQ2015073)

*通信作者(Corresponding author), E-mail:68175632@qq.com

手足口病(hand-foot-mouth disease, HFMD)是一种病毒性的急性传染病,1980年在我国广州首次报道^[1],2008年被列为法定丙类传染病,它由包括柯萨奇病毒、新型肠道病毒、埃可病毒等在内的20余种肠道病毒引起,其中又以肠道病毒71型(EV71)和柯萨奇病毒A组16型(CoxA16)最为常见,一直为优势流行毒株,这两者可交替或者共同传播^[2]。但近年来,由柯萨奇病毒A组6型(CoxA6)、柯萨奇病毒A组9型(CoxA9)、柯萨奇病毒A组10型(CoxA10)、埃可病毒11型(ECV11)等引发的疫情有上升趋势^[3-5],值得密切关注,5岁以下尤其是3岁以下的儿童是防控的重点人群^[6]。本文对2010—2019年南通市手足口病发病情况进行分析,以2010年1月—2019年6月南通市手足口病分月报告病例数据为基础,通过建立差分自回归移动平均(autoregressive integrated moving average, ARIMA)模型^[7],预测2019年7—12月发病率,并与当时的实际发病率比较,验证模型效果,为南通市手足口病的防控提供策略和技术支撑。

1 资料和方法

1.1 资料

资料来源于中国疾病预防控制中心信息系统2010年1月1日—2019年12月31日南通市常住人口的手足口病监测数据和人口数据。

1.2 方法

以2010年1月1日—2019年6月30日南通市手足口病疫情月报告发病率为基础建立时间序列。手足口病月发病率时间序列为季节性时间序列,故采用乘积季节模型,即ARIMA(p, d, q) × (P, D, Q)_s。其中d为平稳化过程中差分的阶数,p、q分别为自回归阶数和移动平均阶数。P、Q分别为季节自回归和移动平均阶数,D为季节差分阶数,s为季节性周期循环长度。模型建立的步骤分为:①确保时序的平稳性。平稳性是ARIMA模型中的一个重要假设,一般可通过时间序列图直观判断。对于不平稳时序,则需要通过数据变换和差分使序列满足平稳性假定,并使用ADF统计检验来验证平稳性假定。②模型识别和定阶。根据平稳化后序列的季节差分自相关(ACF)函数图和偏自相关(PACF)函数图,进行模型的识别和定阶。③模型的参数估计和检验。使用非线性最小二乘法估计模型的参数,Ljung-BoxQ统计量检验,P>0.05,提示残差是白噪声序列,反之为非白噪声序列。④评价模型预测效果。利用非参数检验法(两个

独立样本检验法),对2019年下半年实际月发病率与预测月发病率进行比较,评价模型预测效果。

1.3 统计学方法

运用Excel2007建立和管理数据库,运用SPSS 25.0软件进行统计分析和模型构建,分类资料采用例数和构成比进行统计描述;采用非参数检验法对率进行比较:利用 χ^2 检验对不同型别手足口病阳性检出率进行比较,利用两个独立样本检验法对手足口病实际发病率与预测发病率两组数据进行比较,以P<0.05为差异有统计学意义。

2 结果

2.1 手足口病发病情况

2010—2019年南通市共报告手足口病90 766例,年平均发病率为124.36/10万,其中2018年发病率最高,为235.29/10万(图1)。

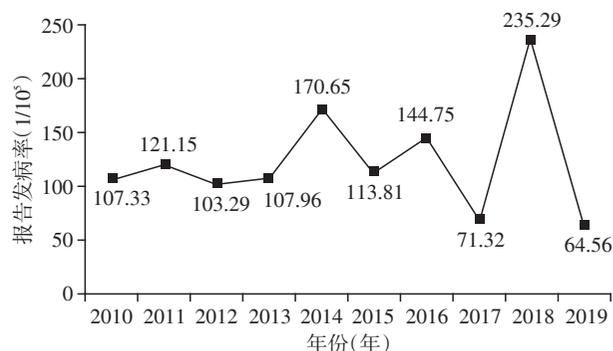


图1 2010—2019年南通市手足口病发病率

Figure 1 The morbidity of HFMD in Nantong from 2010 to 2019

2.2 病原学检测结果

2010—2019年90 766例中实验室诊断病例数为2 950例,其中EV71阳性757例,占25.75%;CoxA16阳性925例,占31.4%;其他肠道病毒阳性1 268例,占42.9%,不同的阳性型别检出率比较,差异有统计学意义($\chi^2=206.95, P<0.01$)。不同年份间EV71、CoxA16和其他肠道病毒阳性检出率差异均有统计学意义(χ^2 值分别为365.20、119.06、334.47, P均<0.01,表1)。

2.3 时间分布情况

2010—2019年,每个月份均有病例发生,经统计,2014年和2018年发病水平较高,总体上2月份发病数最少,6月份发病数最多,有明显季节性,呈双峰特征,为夏季(5、6、7月)高峰和冬季(11、12月)次高峰,从2013年起,有隔年高发、偶数年份增强的流行特征(表2)。

表1 2010—2019年南通市手足口病病原学检测结果

Table 1 The results of etiological detection of HFMO in Nantong from 2010 to 2019

年份	实验室诊断病例	EV71[n(%)]	Coxsackievirus A16[n(%)]	其他肠道病毒[n(%)]
2010年	364	170(46.7)	122(33.5)	72(19.8)
2011年	370	141(38.1)	98(26.5)	131(35.4)
2012年	428	126(29.4)	206(48.1)	96(22.4)
2013年	304	71(23.4)	67(22.0)	166(54.6)
2014年	278	72(25.9)	85(30.6)	121(43.5)
2015年	277	50(18.1)	47(17.0)	180(64.9)
2016年	309	49(15.9)	111(35.9)	149(48.2)
2017年	268	76(28.4)	74(27.6)	118(44.0)
2018年	352	2(0.6)	115(32.7)	235(66.7)
2019年	326	8(2.5)	138(42.3)	180(55.2)
合计	2 950	757(25.7)	925(31.4)	1 268(42.9)

2.4 预测模型构建

以2010年1月—2019年6月南通市手足口病月报告病例数据为基础,构建符合季节性时间序列的ARIMA(p,d,q)×(P,D,Q)_s模型,获得2019年7—12月发病预测值,然后用2019年7—12月全市手足口病月发病率为验证数据进行验证,并绘制实际值和预测值序列图。估计预测值与实际值相对误差来判断模型的预测效果。

2.4.1 序列平稳化和预测模型识别

2010年1月—2019年6月南通市手足口病月发病率时间序列图存在以12个月为1个周期的季节性波动趋势(图2),不能满足平稳化的要求,需对数据进行平稳化处理;通过对原序列进行一阶季节差分处理后,时间序列ACF和PACF函数无明显截尾和拖尾的现象(图3、4),亦无线性衰减,差分后的时

间序列图接近平稳,提示差分后序列适合时间序列模型,确定模型为ARIMA(p,d,q)×(P,D,Q)₁₂。

2.4.2 模型的参数估计和检验

根据序列平稳化处理过程,确定模型的d和D值分别为0和1;根据手足口病的季节性调整,确定s为12;根据ACF函数图和PACF函数图,确定模型的p和q值均为1。模型中的P、Q值分别取0~2由低阶到高阶逐个进行摸索试验。采用Ljung-Box方法检验残差白噪声,若为非白噪声模型则排除。通过筛选,模型ARIMA(1,0,1)(1,1,1)₁₂标准化的贝叶斯信息量(BIC)值为3.59,在模型中最小,R²=0.83,较大,残差序列的自相关系数及偏自相关系数均落在95%CI中,Ljung-BoxQ=19.77,P=0.14>0.05,提示残差是白噪声序列,确认ARIMA(1,0,1)(1,1,1)₁₂模型最优,可以使用。对该模型的参数进行检验,差

表2 2010—2019年南通市手足口病各月份发病情况

Table 2 Monthly incidence of HFMO in Nantong from 2010 to 2019

(例)

月份	2010年	2011年	2012年	2013年	2014年	2015年	2016年	2017年	2018年	2019年	合计
1月	206	338	696	516	296	723	667	824	304	370	4 940
2月	68	61	156	169	113	273	97	192	67	64	1 260
3月	238	267	372	676	345	282	306	236	191	169	3 082
4月	732	470	760	1 205	1 020	751	699	265	714	197	6 813
5月	1 378	998	1 425	1 169	1 774	1 267	1 084	517	3 471	471	13 554
6月	1 351	1 245	1 068	1 195	1 940	1 524	1 547	845	5 470	799	16 984
7月	887	909	679	791	1 530	914	1 117	732	3 017	799	11 375
8月	249	375	163	261	737	367	649	232	882	312	4 227
9月	590	464	338	275	700	214	480	247	1 056	377	4 741
10月	572	801	379	380	1 284	317	1 012	291	799	458	6 293
11月	942	1 467	654	587	1 242	632	1 480	466	595	421	8 486
12月	922	1 428	839	638	1 447	1 046	1 429	361	622	279	9 011
合计	8 135	8 823	7 529	7 862	12 428	8 310	10 567	5 208	17 188	4 716	90 766

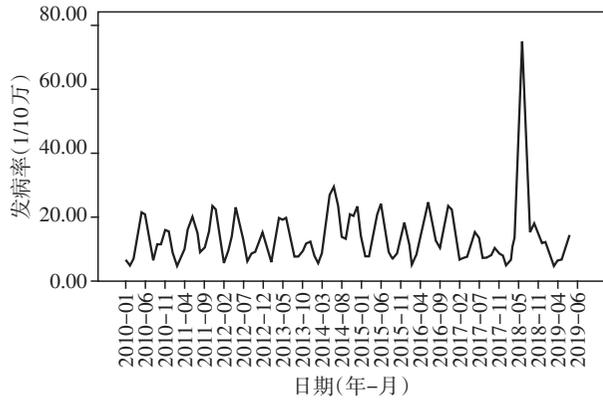


图2 2010年1月—2019年6月南通市手足口病月发病率时间序列图

Figure 2 Time series plot of the monthly morbidity of HFMO in Nantong from January 2010 to June 2019

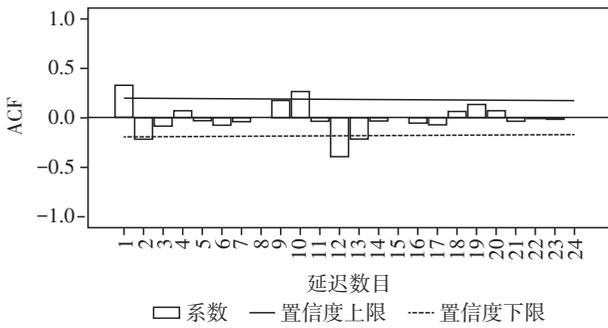


图3 2010年1月—2019年6月南通市手足口病月发病率季节差分自相关(ACF)函数图

Figure 3 Seasonal difference autocorrelation function graph of the monthly morbidity of HFMO in Nantong from January 2010 to June 2019

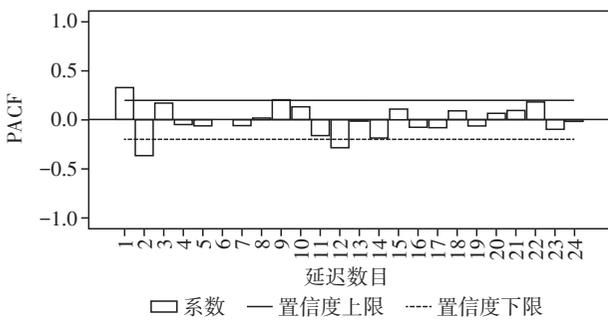


图4 2010年1月—2019年6月南通市手足口病月发病率季节差分偏自相关(PACF)函数图

Figure 4 Seasonal difference partial autocorrelation function graph of the monthly morbidity of HFMO in Nantong from January 2010 to June 2019

异均有统计学意义($P < 0.05$, 表3)。

2.4.3 预测模型的效果

通过ARIMA(1,0,1)(1,1,1)₁₂模型对2019年7—12月手足口病发病率进行预测,预测结果分别

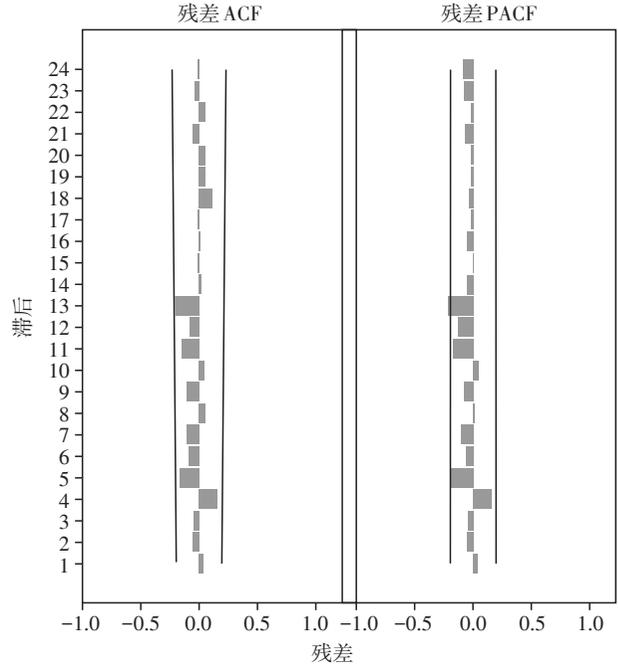


图5 模型ARIMA(1,0,1)(1,1,1)₁₂残差序列的自相关与偏自相关图

Figure 5 Autocorrelation and partial autocorrelation function graphs of ARIMA(1,0,1)(1,1,1)₁₂ model residual sequence

表3 ARIMA(1,0,1)(1,1,1)₁₂模型参数估计

Table 3 Parameter estimation of ARIMA(1,0,1)(1,1,1)₁₂ model

参数	估计	标准误	t值	P值	
AR	滞后1	0.52	0.10	4.98	<0.001
MA	滞后1	-0.58	0.10	-5.90	<0.001
季节性差分		1.00			
MA,季节性	滞后1	0.55	0.22	2.54	0.01

为7.08/10万、1.81/10万、3.74/10万、7.21/10万、10.71/10万和11.29/10万,绘制实际值和预测值序列图(图7)。

采用两个独立样本检验法,比较2019年7—12月手足口病实际发病率(10.96/10万、4.30/10万、5.15/10万、6.25/10万、5.83/10万和3.82/10万)与预测值的差异,经分析两者间的差异无统计学意义($Z=0.48, P=0.63$),预测数据与真实值比较接近,因此该模型有较好的预测效果。

3 讨论

本研究结果显示,2010—2019年南通市手足口病年平均发病率为124.36/10万,低于同时期苏州^[8]、无锡^[9]和全国^[6]平均发病水平,整体呈现隔年高发、

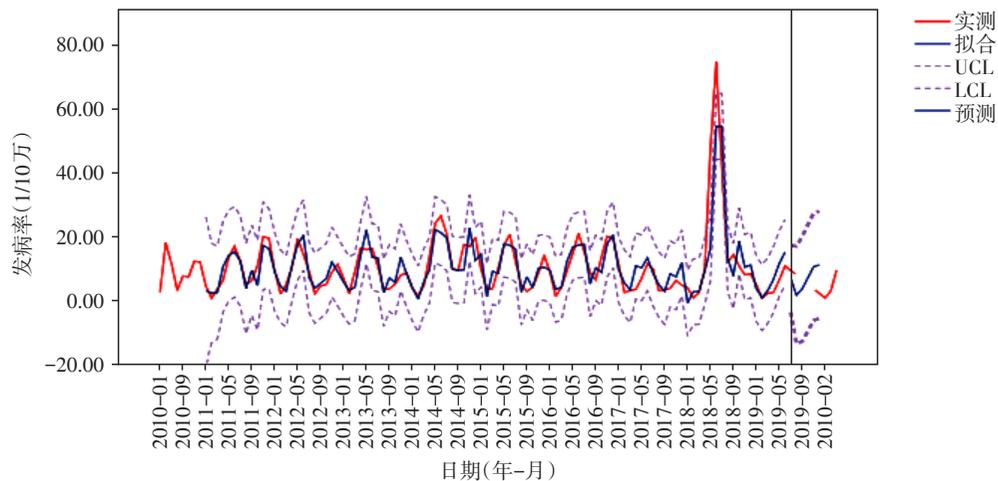


图6 ARIMA(1,0,1)(1,1,1)₁₂模型拟合图

Figure 6 Fitted figure of ARIMA(1,0,1)(1,1,1)₁₂ model

偶数年份增强的流行特点。值得关注的是相比往年,2017年发病率有较明显下降,但随后2018年出现最高峰,2019年又出现类似情况,按照本市手足口病的流行规律,2020年仍有很大概率出现发病高峰,这可能与易感人群的积累有关^[10]。因此做到关口前移,提前做好宣传防控并加大疫苗接种的覆盖率很有必要。

从病原学检测结果来看,2010—2012年引起南通市手足口病的主要病原体为EV71或CoxA16,从2013年起转变为其他肠道病毒,这种变化趋势与国内其他地方报道相一致^[11-12]。从2010—2019年各月份发病数来看,疫情有明显季节性,呈双峰特征,为夏季(5、6、7月)高峰和冬季(11、12月)次高峰,亦符合我国大部分省份手足口病的发病模式^[13],究其原因,主要是5、6、7月气温适宜,湿度较大,有利于肠道病毒的生存和繁殖,而11、12月儿童的自身免疫力下降。

由于肠道病毒各型之间没有交叉免疫力,且病原体类型较多,传播途径复杂,造成了手足口病反复感染,发病率一直居高不下。因此提前建立预警模型,进行预警分析,进而采取相应措施就显得很有必要。ARIMA模型主要用于拟合具有平稳性(或者可以被转换成平稳序列)的时间序列,最早由Box和Jenkins提出,属于时间序列模型,在手足口病、肺结核、猩红热等传染病的预测中得到了广泛的应用,具有较高的实际价值^[14]。本文以2010年1月—2019年6月南通市手足口病月发病数为基础,通过序列平稳化、预测模型的识别、预测模型的参数估计和检验,构建了ARIMA(1,0,1)(1,1,1)₁₂模型,最终模型的预测结果与实际拟合度较高,表明该模型对手足口病的流行趋势预测有很好的适用性。尽

管其具有拟合度较好、适用性强的特点,尤其对周期性、季节性的传染性疾病有明显优势,但也存在一定局限性,比如只能应用于短期预测,一般不超过时间序列的10%^[15],受外部因素影响^[16]等。有资料显示,手足口病的发病与温度、湿度等气象因素以及人口密度呈正相关^[17],并由此衍生了空间累加非线性时空、空间贝叶斯风险评估等一系列模型,提示在今后的实际应用中,可将更多的影响因素纳入手足口病的预警系统,进一步优化模型,从而提高模型的整体预警效果。

【参考文献】

- [1] 王凌燕. 长春市2011—2015年手足口病疫情特征分析及趋势预测[D]. 吉林: 吉林大学, 2017
- [2] 苍旻, 谢国锦, 李松涛, 等. 血清维生素D与重症手足口病患儿临床表现的关系[J]. 南京医科大学学报(自然科学版), 2019, 39(9): 1341-1344
- [3] 陈慧, 袁芳, 马江涛, 等. 2010—2015年宁夏手足口病非EV71/CVA16肠道病毒病原谱分析[J]. 现代预防医学, 2017, 44(4): 729-733
- [4] 许晓庆, 陈国清, 王瑶, 等. 2016—2017年盐城市手足口病CVA9型肠道病毒VP1基因分子进化特征[J]. 江苏预防医学, 2019, 30(3): 269-271
- [5] 宋岩, 仲一, 付凌姣, 等. 2016年北京市朝阳区手足口病相关非EV71非CVA16型肠道病毒病原学特征[J]. 首都公共卫生, 2019, 13(3): 154-157
- [6] 张静. 2008—2017年中国手足口病流行趋势和病原变化动态数列分析[J]. 中华流行病学杂志, 2019, 40(2): 147-154
- [7] KABACOFF R I. R语言实战[M]. 2版. 北京: 人民邮电出版社, 2016: 333-341
- [8] 杭惠, 夏瑜, 陈立凌. 2011—2015年苏州市手足口

病流行病学特征分析[J]. 现代预防医学, 2016, 43(21):3850-3854

[9] 魏崇崇,施超,葛桂芝. 无锡市手足口病流行特征和病原学分析[J]. 预防医学, 2019, 31(11):1156-1158

[10] 于栓宝,罗垵炜,周永红,等. 湖南省安化县2013—2016年手足口病住院负担估计[J]. 中华流行病学杂志, 2019, 40(1):79-83

[11] 李平,庞燕,邝仕壮. 2014—2016年海口市手足口病流行特征及病原体分析[J]. 检验医学与临床, 2019, 16(5):610-617

[12] 刘静文,耿进妹,陈纯,等. 2011—2015年广州市手足口病流行病学特征及病原学监测分析[J]. 现代预防医学, 2017, 44(5):772-777

[13] 樊欢,嵇红,霍翔,等. 2009—2018年江苏省EV71分子流行病学及其基因特征分析[J]. 中华微生物学和免疫学杂志, 2019, 39(9):686-692

[14] 刘小娟,李燕,沙小兰,等. 2008—2018年银川市肺结核流行特征调查分析及趋势预测[J]. 中国全科医学, 2019, 22(39):1-6

[15] 丁小萍,胡德英,万青,等. 综合医院非精神科住院病人自杀行为预测ARIMA模型的构建及其应用[J]. 护理研究, 2019, 33(19):3302-3307

[16] 刘芸男,彭荣荣,杨冬燕,等. ARIMA模型在临床红细胞需求预测中的应用[J]. 安徽医科大学学报, 2019, 54(10):1611-1615

[17] 姚一泓,霍达,贾蕾,等. 手足口病预警模型及其应用现状[J]. 国际病毒学杂志, 2018, 25(2):136-139

[收稿日期] 2020-02-04

(上接第58页)

2005[S]. 北京:中国标准出版社, 2005

[7] Evaluation of the linearity of quantitative measurement procedures: a statistical approach. CLSI document EP06[S]. Wayne, PA: Clinical and Laboratory Standards Institute, 2003

[8] Interference testing in clinical chemistry. CLSI document EP07[S]. Wayne, PA: Clinical and Laboratory Standards Institute, 2018

[9] User demonstration of performance for precision and accuracy. CLSI document EP15[S]. Wayne, PA: Clinical and Laboratory Standards Institute, 2014

[10] 临床实验室对商品化定量试剂盒分析性能的验证. 中华人民共和国卫生行业标准 WS/T 420-2013[S]. 北京:中国标准出版社, 2013

[11] 降钙素原测定试剂盒. 中华人民共和国医药行业标准 YY/T 1588-2018[S]. 北京:中国标准出版社, 2018

[12] LEVINE R J, MAYNARD S E, QIAN C, et al. Circulating angiogenic factors and the risk of preeclampsia [J]. N Engl J Med, 2004, 350(7):672-683

[13] 张展,李鹤,张琳琳. 胎盘因素、免疫因素、血管因素与子痫前期发病机制相关性探讨[J]. 临床检验杂志, 2012, 30(10):813-816

[14] SIBIUDE J, GUIBOURDENCHE J, DIONNE M D, et al. Placental growth factor for the prediction of adverse outcomes in patients with suspected preeclampsia or intra-uterine growth restriction [J]. PLoS One, 2012, 7(11): e50208

[15] 谢雪钦. 免疫层析试纸条标记探针研究进展[J]. 食品安全质量检测学报, 2014, 5(7):2138-2145

[16] 权秋宁,武雅俐,张娟玲,等. 时间分辨荧光免疫层析法应用于中孕产前筛查的价值[J]. 中国优生与遗传杂志, 2015, 23(1):52-53

[收稿日期] 2019-07-09