

网络出版时间: 2019-9-3 17:27 网络出版地址: http://kns.cnki.net/kcms/detail/34.1065.r.20190830.1434.024.html

◇ 预防医学研究 ◇

## ARIMA 模型在临床红细胞需求预测中的应用

刘芸男<sup>1</sup>, 彭荣荣<sup>1</sup>, 杨冬燕<sup>2</sup>, 杨小丽<sup>1</sup>

**摘要** 目的 利用临床红细胞用量的数据建立 ARIMA 模型, 预测未来临床红细胞需求量, 为采供血机构制定采集及招募计划提供科学依据。方法 收集 2006~2016 年重庆市血液中心每月临床红细胞用量数据, 利用 SPSS 软件对 2006 年 1 月~2016 年 6 月的临床红细胞用量数据建立 ARIMA 模型, 经模型识别、参数估计与检验, 确定最优模型; 运用该模型预测 2016 年 7~12 月红细胞临床需求量, 验证预测效果。结果 最优模型为 ARIMA(3, 1, 0)(0, 1, 1)<sub>12</sub>, 残差的自相关函数和偏自相关函数基本都在 95% CI 内, 同时 Ljung-Box Q 统计量为 19.274,  $P > 0.05$ , 无统计学意义, 提示残差为白噪声。2016 年 7~12 月临床红细胞预测值均在 95% CI 内, 模型拟合值与实际值曲线趋势大致相同, 平均相对误差为 2.55%, 预测精度较高, 结果可靠。结论 ARIMA 模型能较好地拟合重庆市血液中心临床红细胞需求量变化趋势, 适用于临床红细胞需求量的短期预测, 有助于了解未来临床红细胞需求, 为血液采集、招募及库存管理提供依据, 突破采供血区域政策的限制, 实现全市范围内血液资源的有效整合与合理调配。

**关键词** ARIMA 模型; 临床红细胞; 需求预测

**中图分类号** R 457.1; R 193.3

**文献标志码** A **文章编号** 1000-1492(2019)10-1611-05  
doi: 10.19405/j.cnki.issn1000-1492.2019.10.023

根据世界卫生组织的研究报告, 当无偿献血率达 1%~3% 时才能满足该国基本的供血需求, 2017 年我国无偿献血率为 1.05%, 可以看出我国血液保障能力尚处于“紧平衡”状态<sup>[1]</sup>。近年来重庆市社会经济快速发展, 百姓医疗服务需求不断增长, 引发临床用血量的迅速增加, 尤其是互助献血取消后, 临

床血液需求缺口加大。然而, 目前重庆市对临床血液需求的预测缺乏科学的方法, 主要依据相关人员既往经验粗略估算, 此法存在较多局限。故该研究采用自回归移动平均模型 (autoregressive integrated moving average model, ARIMA) 建立需求预测模型, 在全面了解临床血液需求的基础上, 突破采供血的区域政策限制, 在各区域间实现血液资源的整合与合理调配, 避免区域性血液资源短缺或过剩现象的发生。

## 1 材料与方法

**1.1 数据来源及处理** 获取重庆市血液中心 2006~2016 年红细胞类制品出库记录, 红细胞类制品包括辐照悬浮红细胞、辐照悬浮少白细胞红细胞、去白细胞悬浮红细胞等。临床红细胞用量以单位 (U) 计算, 1 U 红细胞类制品由 200 ml 全血分离制备。

**1.2 ARIMA 模型** 时间序列分析是一种考虑对象本身的历史数据随时间发展变化的规律, 并用该变量以往的资料建立统计模型推算未来值的预测方法<sup>[2]</sup>, 即以时间代替各种影响因素, 从中探寻和分析事件的变化特征和发展趋势, 并以此来建立时间序列模型的方法。ARIMA(p, d, q) 模型又被称为 Box-Jenkins 模型, 是时间序列分析中最常用的方法。若时间序列包含了季节性特征, 则需建立季节性 ARIMA 模型, 即 ARIMA(p, d, q)(P, D, Q)<sub>s</sub>, 其中 p、q 为自回归和移动平均阶数, P、Q 为季节性自回归和移动平均阶数, d、D 为非季节性和季节性差分次数, s 为季节周期。一般表达式为:  $\phi_p(B)\Phi_P(B^s)$   
 $\sum_{i=0}^d X_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t$ , 其中  $\{\varepsilon_t\}$  为白噪声序列,  $\phi_p(B)$  与  $\Theta_Q(B^s)$  分别为季节 P 阶自回归算子和 Q 阶移动平均算子。

ARIMA 模型建立的基本步骤: ① 数据平稳化处理。通过时序图初步判断序列是否平稳, 若为不平稳序列, 则针对序列不平稳的趋势性或周期性进行差分或季节性差分处理, 实现序列的平稳化。② 模型识别。对平稳序列做自相关图, 根据自相关函数和偏自相关函数拖尾、截尾情况估计 p、d、q 值, 建

2019-03-28 接收

基金项目: 重庆市决策咨询与管理创新计划项目 (编号: cstc2016jccxBX0064)

作者单位: <sup>1</sup> 重庆医科大学公共卫生与管理学院、医学与社会发展研究中心、健康领域社会风险预测治理协同创新中心, 重庆 400016

<sup>2</sup> 重庆市血液中心业务部, 重庆 400015

作者简介: 刘芸男, 女, 硕士研究生;

杨小丽, 女, 教授, 硕士生导师, 责任作者, E-mail: 872463319@qq.com

立备选模型; 并根据贝叶斯准则 ( bayesian information criterion , BIC) 选择最优模型<sup>[3]</sup>。③ 模型检验。根据残差 ACF 图与 PACF 图以及 Ljung-Box Q 统计量 判断残差序列是否为白噪声。若是 则模型可用于预测; 若否 则需重新对序列进行识别、估计和检验 以获得最优拟合模型。④ 预测并验证。运用最终选定的 ARIMA 模型预测需求值 , 计算 95% CI 以及相对误差 并与同期的实际值比较 , 以验证模型的拟合效果。

1.3 统计学处理 运用 Excel 软件建立数据库 , 按月对临床红细胞用量进行统计与整理; 并采用 SPSS 19.0 软件进行数据分析及模型构建 , 以  $P < 0.05$  为差异有统计学意义。

2 结果

2.1 红细胞类用量时间序列图及平稳处理 绘制 2006 年 1 月 ~ 2016 年 6 月红细胞类临床用量时序图( 图 1) 。 2006 ~ 2016 年红细胞类制品用量波动幅度较大 , 总体呈上升趋势; 同时序列还存在明显的季节周期性 , 在 12 个月为一个周期的序列中 , 每年的 3、4、11、12 月份用量相对较高 , 2、6、7、8 月份则较低。鉴于序列具有趋势性与周期性的不平稳特征 , 故对数据采取差分及季节性差分处理。经过一次差分和一次季节性差分后序列如图 2 所示 , 观测值均围绕 0 值上下随机波动 , 故可认为处理后的序列为平稳序列。

2.2 模型的识别 鉴于 2006 年 1 月 ~ 2016 年 6 月红细胞临床用量序列存在明显的趋势性和季节性 , 故选用季节性 ARIMA(  $p, d, q$  ) (  $P, D, Q$  )  $s$  模型。本研究序列的季节周期为 12 个月 , 故  $s$  取 12; 因对原

序列进行了一阶差分和一阶季节差分 , 故  $d = 1, D = 1$ 。绘制一阶差分和一阶季节差分后 ACF 图和 PACF 图( 图 3) , ACF 延迟数目在 1、4、5、11、12 阶时均有突出 , 截尾及拖尾特征不明显 , 判定  $q = 0$  , 而 PACF 在 3 阶之后快速下降为 0 , 呈三阶截尾 , 判断  $p = 3$ ; ACF 延迟数目在 12 阶时样本自相关函数显著不为 0 , PACF 则是在 12 阶附近显著不为 0 , 因此  $P = 1$  或 0 ,  $Q = 1$ 。因此 , 识别的备选模型为 ARIMA

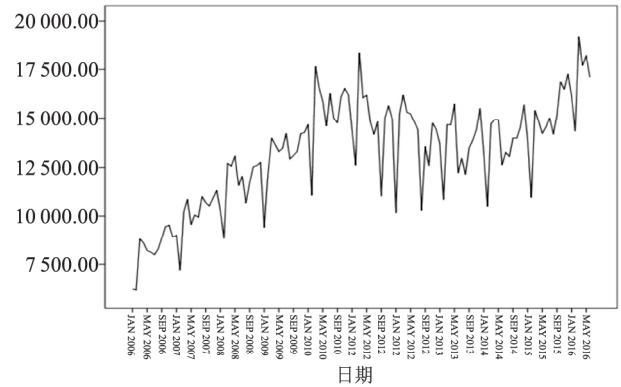


图 1 原始数据序列图

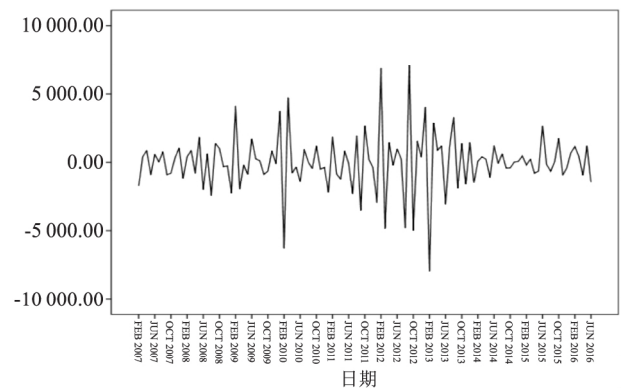


图 2 差分 and 周期调整后序列图

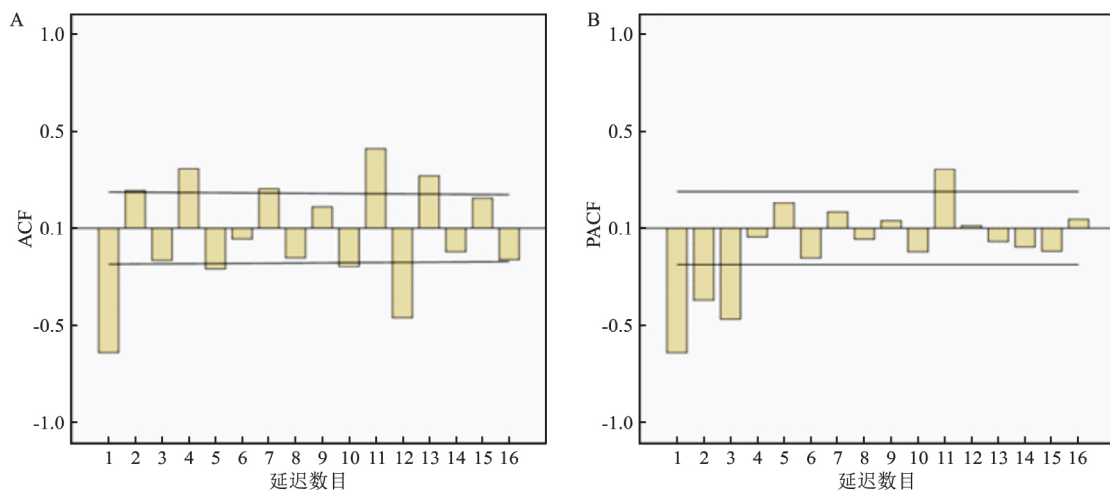


图 3 原始序列一阶差分和一阶季节差分后 ACF 图和 PACF 图

A: ACF; B: PACF

(3, 1, 0) (0, 1, 1)<sub>12</sub> 和 ARIMA(3, 1, 0) (1, 1, 1)<sub>12</sub>。对备选模型的参数估计和假设检验结果见表 1, 依据标准化 BIC 值越小, 模型拟合效果越好的准则, 确定最优模型为 ARIMA(3, 1, 0) (0, 1, 1)<sub>12</sub>, 且该模型与 SPSS 专家建模器推荐的模型一致。

表 1 备选模型的参数估计和假设检验

参数	ARIMA(3, 1, 0) (0, 1, 1) <sub>12</sub>			ARIMA(3, 1, 0) (1, 1, 1) <sub>12</sub>		
	β	T 值	P 值	β	T 值	P 值
AR1	-0.890	-9.798	0.000	-0.951	-10.256	0.000
AR2	-0.618	-5.686	0.000	-0.667	-5.992	0.000
AR3	-0.370	-4.115	0.000	-0.399	-4.470	0.000
SAR1	—	—	—	0.994	0.228	0.820
SMA1	0.786	7.302	0.000	0.244	1.466	0.146
BIC	14.338			14.376		

**2.3 模型检验** 对最优模型 ARIMA(3, 1, 0) (0, 1, 1)<sub>12</sub> 的残差序列作自相关函数和偏自相关函数图 (图 4) 残差的自相关函数和偏自相关函数基本都在 95% CI 内, 同时 Ljung-Box Q 统计量为 19.274, P > 0.05, 无统计学意义, 说明残差序列不存在自相关, 通过白噪声检验。综上判定该模型适用于重庆市血液中心临床红细胞需求量的预测。

**2.4 预测及应用** 运用 ARIMA(3, 1, 0) (0, 1, 1)<sub>12</sub> 模型对 2016 年 7 ~ 12 月红细胞临床用量进行预测, 绘制预测值与实际值时序拟合图 (图 5)。预测值与实际值曲线趋势基本相同, 且预测值均在 95% CI 内, 平均相对误差较小, 为 2.55%, 见表 2。

3 讨论

ARIMA 模型综合考虑了序列的趋势性和周期性变化等因素, 借助模型参数进行量化表达, 通过反复识别和模型诊断、比较, 以获得最佳模型<sup>[4]</sup>。ARIMA 模型具有数据收集简单、适用性强、预测精确度高等特点<sup>[5]</sup>, 近年来在医学卫生领域已得到广泛应用, 如艾滋病、疟疾、乙肝等疾病发病趋势及卫生支出的预测等<sup>[2, 6-8]</sup>。

临床红细胞用量是一组根据某年某月的时间间

隔顺序记录下的有序数据, 分析显示该时间序列数据呈一定的趋势性与周期性特征, 说明时间序列分析的 ARIMA 模型可用于血液需求预测, 且已得到相关研究<sup>[9]</sup>的证实。有文献<sup>[10]</sup>报道, ARIMA 模型在预测临床血液需求方面优于其他时间序列模型。

红细胞成分血具有纯度和浓度高、临床疗效好、病毒感染风险低等特点, 但保存期限较短一般为 21 ~ 35 d。近年来, 由于医疗服务能力的提高和医疗服务需求的快速增长加剧了血液的供需矛盾, 结构性、季节性、区域性缺血现象时有发生。尤其是在省级血液中心覆盖的、高端医疗资源集中的地区情况更为严重<sup>[11]</sup>。

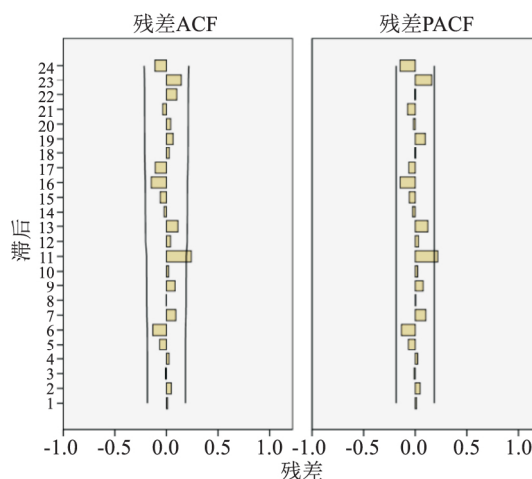


图 4 模型残差自相关函数和偏相关函数图

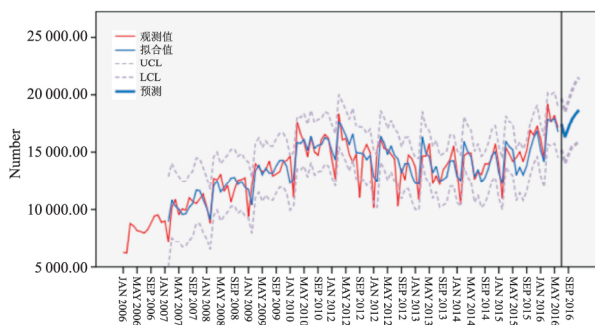


图 5 实际值与预测值比较图

表 2 2016 年 7 月 ~ 2016 年 12 月预测结果 (U, %)

项目	7 月	8 月	9 月	10 月	11 月	12 月
实际值	15 970.00	16 173.00	17 069.00	17 587.00	18 486.00	17 988.50
预测值	17 438.94	16 296.00	17 122.87	17 910.54	18 343.41	18 704.48
95% CI 下限	15 200.43	14 044.07	14 782.83	15 470.40	15 650.73	15 953.86
95% CI 上限	19 677.44	18 547.94	19 462.92	20 350.68	21 036.09	21 455.10
相对误差	9.20	0.76	0.32	1.84	-0.77	3.99
平均相对误差	2.55					

本研究将血液供需矛盾最为突出的血液中心纳入研究视野,以重庆市血液中心为研究对象,对该血液中心2006年1月~2016年6月每月向医院提供的红细胞用量进行分析建模,得出最优模型为  $ARI-MA(3,1,\rho)(0,1,1)_{12}$ ;运用该模型对重庆市血液中心2016年7~12月每月红细胞临床用量进行预测,结果显示预测值和实际值的动态趋势基本一致,平均相对误差为2.55%。已有研究<sup>[6,10,12]</sup>表明,ARIMA模型预测结果平均相对误差若小于5%则说明模型预测精准性较高,有较好的预测效果。因此该模型可以用于重庆市血液中心未来红细胞临床用量的预测。

建立红细胞临床需求预测模型,有助于科学地制定招募采血计划,提高血液供给及需求之间的契合度,使红细胞成分制品既能满足临床需求,又能避免过期浪费。然而,ARIMA模型也具有一定的局限性,它是依靠历史的统计数据建立的数学模型,并未考虑国家重大政策改变和调整、突发公共卫生事件等外部因素的影响<sup>[6]</sup>。因此,ARIMA模型仅适用于短期结果预测,在红细胞成分制品未来临床需求的预测中,预测未来1年的趋势是可行的<sup>[13]</sup>。预测时间过长会增大预测误差,影响预测精度。由于健康中国战略的实施、二孩政策的推行、人口老龄化的快速发展,尤其是互助献血的取消等因素都会对预测效果产生影响。因此应加强临床红细胞需求数据长期的收集,保证数据的及时更新,以便对模型进行修正或重新拟合,确保预测的精度。

### 参考文献

[1] 经济日报-中国经济网. 2017年全国无偿献血1459万人次,

血液保障处于“紧平衡”状态[EB/OL]. 2018-06-12. [http://bgimg.ce.cn/cyssc/newmain/yc/jxsw/201806/12/t20180612\\_29415345.shtml](http://bgimg.ce.cn/cyssc/newmain/yc/jxsw/201806/12/t20180612_29415345.shtml).

[2] 向颖,郁红月,张维,等.重庆市HIV感染和艾滋病新发报告总数预测模型的拟合和比较分析[J/OL]. 第三军医大学学报. <https://doi.org/10.16016/j.1000-5404.201810141>.

[3] 朱佳佳,胡登利,洪秀琴,等.基于时空大数据的甲型肝炎发病率分布特征分析及预测模型[J]. 中华疾病控制杂志, 2018, 22(11): 1144-7.

[4] 于林凤,吴静,周锁兰,等. ARIMA季节模型在我国丙肝发病预测中的应用[J]. 郑州大学学报, 2014, 49(3): 345-8.

[5] 钱俊,王永波. 几种软件建立ARIMA模型的应用比较[J]. 统计与决策, 2018, 16(2): 80-4.

[6] 樊雯婧,陆群,邹立巍,等. ARIMA模型在合肥市疟疾发病预测中的应用[J]. 安徽医科大学学报, 2013, 48(3): 252-5.

[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): 646-50.

[8] 于菲,耿顺利,高建民,等. 基于ARIMA和GM(1,1)模型的陕西省个人卫生支出预测[J]. 中国卫生政策研究, 2018, 11(7): 19-23.

[9] 马文文,王梦圆,赵轶轮,等. 血液需求预测模型研究[J]. 中国输血杂志, 2015, 28(6): 740-3.

[10] 黄国军,王乐三,张统宇,等. 自回归滑动平均混合模型在红细胞供应量预测中的应用[J]. 中国输血杂志, 2016, 29(2): 140-3.

[11] 杨冬燕,成晓娇,段恒英,等. 重庆市采供血工作现状分析及对策研究[J]. 中国输血杂志, 2016, 29(9): 962-5.

[12] Peng Y, Yu B, Wang P, et al. Application of seasonal autoregressive integrated moving average model in forecasting the incidence of hand-foot-mouth disease in Wuhan, China[J]. *J Huazhong Univ Sci Technolog Med Sci*, 2017, 37(6): 842-8.

[13] 杨姝怡,马姗姗,苏成豪. 应用ARIMA模型预测厦门市2017年乙肝月发病率[J]. 2018, 45(10): 1729-32.

## Application of autoregressive integrated moving model in forecasting the clinical usage of red blood cells

Liu Yunnan<sup>1</sup>, Peng Rongrong<sup>1</sup>, Yang Dongyan<sup>2</sup>, et al

(<sup>1</sup>School of Public Health and Management, Chongqing Medical University, Research Center for Medical and Social Development, Innovation Center for Social Risk Governance in Health, Chongqing 400016;

<sup>2</sup>Dept of Operational, Chongqing Blood Center, Chongqing 400015)

**Abstract Objective** To provide scientific guides and references for the collection and recruitment plans of blood transfusion services, and to predicate the clinical usage demand of red blood cells (RBC) based on the autoregressive moving average (ARIMA) model. **Methods** Based on the monthly clinical usage data of RBC from 2006 to 2016 obtained from Chongqing blood center, SPSS software was used to built a best ARIMA model, which was determined by model identification, parameters estimation and White-noise test, with data from January 2006 to June

网络出版时间: 2019-9-3 17:27 网络出版地址: <http://kns.cnki.net/kcms/detail/34.1065.r.20190830.1434.025.html>

◇ 临床医学研究 ◇

## 外周血循环肿瘤细胞联合细胞角蛋白 19 片段 21-1 检测 在非小细胞肺癌筛查中的应用价值

官燕飞<sup>1</sup> 陈健锋<sup>1</sup> 袁斌<sup>1</sup> 刘君<sup>2</sup> 陈昂<sup>3</sup> 袁春雷<sup>1</sup> 童辉纯<sup>1</sup> 陈志强<sup>3</sup>

**摘要** 目的 观察外周血循环肿瘤细胞(CTC)、细胞角蛋白 19 片段 21-1(CYFRA21-1)单独或联合检测在肺部健康体检者(健康组)、肺部良性病变患者(良性组)及非小细胞肺癌患者(NSCLC 恶性组)中阳性率及诊断 NSCLC 病变效能的差异,初步评估两项指标单独或联合检测在 NSCLC 筛查中的应用价值。方法 临床回顾并随机选取 172 例肺部健康体检者、156 例肺部良性病变患者、118 例 NSCLC 患者为研究对象。检测三组患者 CTC 及 CYFRA21-1 水平。比较 CTC、CYFRA21-1 单独或联合检测在三组的阳性率差异及诊断 NSCLC 的效能差异;综合评估 CTC、CYFRA21-1 单独或联合诊断 NSCLC 的价值。结果 ① CTC、CYFRA21-1 单独或联合检测在肺部健康组、良性组与恶性组的阳性率差异均有统计学意义( $P < 0.05$ ),阳性率均在恶性组中最高。② CTC 检测阳性率在恶性组的各 TNM 分期中差异有统计学意义( $\chi^2 = 9.18, P = 0.03$ ),CTC 的阳性率随着分期的增加而增加,在 IV 期中最高。③ CTC、CYFRA21-1 单独或联合检测诊

断 NSCLC 的灵敏度、特异度、符合率、阳性预测值、阴性预测值总体差异均有统计学意义( $P < 0.05$ )。灵敏度最高的是联合检测,与 CTC、CYFRA21-1 的单独检测差异均有统计学意义( $P < 0.02$ )。特异度最高的是 CTC 检测,与 CYFRA21-1、联合检测差异均有统计学意义( $P < 0.02$ )。符合率最高的是 CTC 检测法,与 CYFRA21-1 检测差异有统计学意义( $P < 0.02$ )。阳性预测值最高的是 CTC 检测,与 CYFRA21-1、联合检测差异均有统计学意义( $P < 0.02$ )。阴性预测值最高的是联合检测,与 CYFRA21-1 检测差异有统计学意义(均  $P < 0.02$ )。约登指数最大的是联合检测(约登指数等于 0.8645)。CTC 与 CYFRA21-1 联合检测表现出了较佳的综合性能,其灵敏度、特异度、符合率、阳性预测值、阴性预测值分别是 94.07%、92.38%、92.83%、81.62%、97.74%。结论

CTC、CYFRA21-1 单独或联合检测均有助于 NSCLC 的筛查,与单独检测比较,联合检测有效提高了诊断 NSCLC 灵敏度、阴性预测值和约登指数,同时特异度、符合率和阳性预测值也维持在较高水平,表现出了较好的综合性能,建议在临床诊疗中进行联合检测。

**关键词** 循环肿瘤细胞;细胞角蛋白 19 片段;非小细胞肺癌;筛查

中图分类号 R 734.2

文献标志码 A 文章编号 1000-1492(2019)10-1615-05

doi: 10.19405/j.cnki.issn1000-1492.2019.10.024

非小细胞肺癌(non-small cell lung cancer, NSCLC)约占全部肺癌的 85%,5 年生存率仅为

2019-04-02 接收

基金项目:广东省医学科研基金(编号:A2019348);广东省医学科学技术研究基金立项课题(编号:A2017321);中山市社会公益科技研究专项立项(编号:2018B1035、2019B1090)

作者单位:<sup>1</sup>南方医科大学附属中山博爱医院检验科,中山 528400

<sup>2</sup>中山大学肿瘤防治中心病理科,广州 510060

<sup>3</sup>南方医科大学附属中山博爱医院病理科,中山 528400

作者简介:官燕飞,女,副主任医师;

陈昂,男,副主任医师,责任作者,E-mail: 361782947@

qq.com

2016. The model was adopted to predict the clinical usage of RBC from July to December 2016 to verify the forecasting effect. **Results** Autocorrelation function and partial autocorrelation function of residual of model ARIMA(3, 1, 0)(0, 1, 1)<sub>12</sub> were almost in the 95% CI, and Ljung-Box Q test = 19.274,  $P > 0.05$ , indicating that the residual series was white noise. The curve tendencies of forecasting value calculated by the ARIMA model and actual value were roughly the same, and the mean relative error was 2.55%, with high precision and reliable results. **Conclusion** ARIMA model is excellent in fitting the change tendency of clinical RBC usage of Chongqing blood center and appropriated for short-term predication of clinical demand of RBC, which can help to grasp the demand of clinical RBC and provide basis for blood collection, recruitment and inventory management, and it also can break through the administrative regions to achieve the integration and allocation of blood resources throughout the city.

**Key words** ARIMA model; clinical red blood cells; demand prediction