

本文引文格式:李阳,陈坚,廖建英,等.基于ARIMA时间序列的广西百色市艾滋病流行趋势分析[J].右江民族医学院学报,2023,45(1):134-138.

【论著与临床报道】

基于ARIMA时间序列的广西百色市艾滋病流行趋势分析

李阳^{1,2},陈坚³,廖建英²,谢艺红¹

1. 广西医科大学,广西 南宁 530000;
2. 右江民族医学院,广西 百色 533000;
3. 广西百色市疾病预防控制中心,广西 百色 533000)

摘要:目的 分析广西百色市2011年至2021年艾滋病报告病例的分布特征,建立ARIMA时间序列模型,对艾滋病报告病例数进行短期预测,为百色市艾滋病防治策略和措施的制定提供科学依据。**方法** 对2011年至2021年中国疾病预防控制中心信息系统艾滋病监测子系统现住址为百色的报告病例进行描述性分析,并采用Eviews 10.0统计学软件建立ARIMA时间序列模型进行预测。**结果** 百色市2011年至2021年艾滋病报告病例数总体呈上升趋势,病例有明显的地区聚集性,东南部高发,重点分布在右江区、田阳区、平果市、田东县、靖西市等地;报告病例无明显季节性。经过数据序列平稳化处理 and 参数估计,确定ARIMA(2,1,2)为最优拟合模型。模型的平均拟合准确率为85.29%,平均相对误差为14.71%,说明该模型的拟合效果良好。**结论** 百色市艾滋病报告病例数呈逐年上升趋势,且有明显的地区聚集性,ARIMA时间序列模型能较好地预测百色市艾滋病报告病例数,应加强高发地区的防控和危险因素监测,并制定针对性强的综合防控策略和措施。

关键词:获得性免疫缺陷综合征;ARIMA模型;时间序列;预测

中图分类号:R512.91

文献标识码:A

文章编号:1001-5817(2023)01-0134-05

doi:10.3969/j.issn.1001-5817.2023.01.023

Analysis of AIDS epidemic trend in Baise city of Guangxi based on ARIMA time series

Li Yang^{1,2}, Chen Jian³, Liao Jianying², Xie Yihong¹

1. Guangxi Medical University, Nanning 530000, Guangxi, China;
2. Youjiang Medical University for Nationalities, Baise 533000, Guangxi, China;
3. Baise Centre for Disease Control and Prevention, Baise 533000, Guangxi, China)

Abstract: **Objective** To analyze the distribution characteristics of reported AIDS cases from 2011 to 2021 in Baise city, Guangxi Zhuang Autonomous Region, and establish the model of ARIMA time series to predict the number of reported AIDS cases in the short term, so as to provide scientific basis for the formulation of strategies and measures for AIDS prevention and treatment in Baise city. **Methods** Descriptive analysis was conducted on the reported cases whose current address was in Baise according to the AIDS surveillance subsystem of China Information System for Disease Control and Prevention from 2011 to 2021. The model of ARIMA time series was established with the statistical software Eviews 10.0 for prediction. **Results** From 2011 to 2021, the reported AIDS cases in Baise showed an overall increasing trend, with obvious regional clustering and high incidence in the southeast, mainly distributed in Youjiang district, Tianyang district, Pingguo city, Tiandong county, Jingxi city and other places. There was no significant difference in seasonality in the reported ca-

基金项目:百色市科学研究与技术开发计划项目(百科 20184201,百科 20222037)

第一作者简介:李阳(1978-),男,医学学士,讲师,研究方向:艾滋病的预防控制,E-mail:527293391@qq.com

通讯作者简介:谢艺红(1978-),女,医学博士,教授,研究方向:流行病与卫生统计方法应用,E-mail:gxxieyihong@163.com

ses. After stabilization of data series and estimation of parameter, ARIMA (2, 1, 2) was identified as the optimal fitting model. The average fitting accuracy of the model was 85.29%, and the average relative error was 14.71%, indicating that the model had a good fitting effect. **Conclusion** The number of reported AIDS cases in Baise city turns out to be increasing year by year and there is obvious regional aggregation. The model of ARIMA time series can predict the number of reported AIDS cases in Baise city. It is necessary to strengthen the prevention and control and monitoring of risk factors in high-incidence areas, and to formulate well-targeted comprehensive strategies and measures for prevention and control of AIDS.

Key words: acquired immune deficiency syndrome; ARIMA model; time series; prediction

艾滋病(Acquired Immune Deficiency Syndrome, AIDS)是由人类免疫缺陷病毒(human immunodeficiency virus, HIV)感染引起的一种危害性极大的慢性传染病。截至2021年10月底,全国报告现存的艾滋病感染者总数为114万例^[1],部分省/自治区疫情严重,其中广西是全国艾滋病重灾区,2010年至2020年每年新报告病例约1万例,报告死亡约5000例,目前累计存活病例数、报告死亡例数均居全国前列^[2]。百色市地处广西西部,辖12个县(市、区),与云南文山、贵州兴义、越南等地相邻,是广西、贵州、云南三省重要的物流中心和交通枢纽,也是大西南进入广西的重要咽喉城市。为了解百色市艾滋病报告病例的变化趋势和分布特征,探讨ARIMA模型在艾滋病报告病例数短期预测的效果,本文对百色市2011年1月至2021年12月的报告病例进行分析,为百色市艾滋病防控策略和措施的制定提供参考依据。

1 资料与方法

1.1 资料来源 数据来源于中国疾病预防控制中心艾滋病监测子系统2011年1月至2021年12月登记的现住址为百色市的报告病例。

1.2 模型构建 采用ARIMA(p, d, q)时间序列模型建模,其中,p, d, q分别代表非季节自回归阶数、普通差分阶数和非季节移动平均阶数。建模过程包括以下几个步骤:①数据序列平稳化处理:通过绘制2011年至2021年艾滋病报告病例数的原始序列图,并进行平稳性单位根检验(ADF检验),以判断该数据的平稳性。如果不平稳,可以进行数据变换(如对数变换、指数变换等),若经数据变换后还是不平稳,则进行数据差分处理,可以消除趋势因子和季节性因子的影响;②模型识别、定阶:将经差分处理后平稳时间序列进一步绘制自相关图(ACF图)和偏自相关图(PACF图),对模型进行初步识别以及模型类型的选择,初步确定p, d, q的取值范围;③对模型的参数进行估计:通过赤池信息准则(AIC)选取最优模型^[3-4];④模型预测:利用所选的最优模型进行艾滋病报告病例数的短期预测^[5-6],对模型的拟合和预测价值进行评价。

1.3 统计学方法 用Excel 2016软件对数据进行整

理,绘制相应的统计图、统计表。用描述性统计分析方法分析2011年至2021年百色市艾滋病报告病例的时间、地区分布特征。采用Eviews 10.0统计学软件进行建模、分析和预测。

2 结果

2.1 基本情况 2011年至2021年百色市共报告艾滋病感染者和病人8015例,各年度报告病例数呈逐年上升趋势(趋势性卡方检验: $\chi^2 = 640.25, P < 0.001$)。其中2011年至2015年相对较为平稳,每年报告的感染者和病人数维持在461~548例之间;2016年以后报告的感染者和病人数逐年上升,2021年报告病例1151例,比2015年的508例上升了226.57%。见图1。



图1 2011年至2021年百色市艾滋病新报告病例变化趋势

2.2 地区分布 报告病例的分布有明显的地区聚集性,2011年至2015年平均报告病例数较高的地区主要集中在平果市、田阳区、右江区、田东县、隆林县;2016年至2021年平均报告病例数较高的地区主要集中在右江区、田阳区、隆林县、田东县、平果市、靖西市;2011年至2021年百色市新增病例年均总体报告率为20.29/10万,年均新增病例报告率较高的地区主要集中在田阳区(31.73/10万)、西林县(27.33/10万)、右江区(26.11/10万)、平果市(24.25/10万)、田东县(21.78/10万)、隆林县(20.03/10万)、德保县(19.34/10万),详见表1和图2。

表 1 2011 年至 2021 年百色市平均报告病例数和新增病例报告率统计表

地区	2011 年至 2015 年年均报告病例数(例)	2016 年至 2021 年年均报告病例数(例)	2011 年至 2021 年年均报告病例数(例)	2011 年至 2021 年新增病例年均报告率/10 万
右江区	70.00	136.67	106.36	26.11
田阳区	78.40	121.33	101.82	31.73
田东县	59.20	97.33	80.00	21.78
平果市	114.40	106.17	109.91	24.25
德保县	35.40	76.67	57.91	19.34
靖西市	36.20	103.00	72.64	14.22
那坡县	17.60	22.17	20.09	12.47
凌云县	11.40	32.50	22.91	11.96
乐业县	9.80	15.33	12.82	8.41
田林县	15.00	50.17	34.18	14.87
西林县	24.40	51.67	39.27	27.33
隆林县	37.00	98.83	70.73	20.03
合计	508.80	911.83	728.64	20.29

2.3 时间分布 报告病例数无明显季节性,3~9 月份报告病例数相对较多,占总报告病例数的 64.74%,

见图 3。

2.4 ARIMA 模型构建

2.4.1 时间序列的平稳化处理 百色市艾滋病报告病例数的原始数据序列图见图 4,该序列呈缓慢的长期上升趋势。用原始数据进行单位根检验,结果显示该序列是具有时间趋势的平稳序列(ADF 检验 P 值 = 0.0004,时间趋势 Trend 系数 P 值 = 0.0005)。对该序列进行非季节性一阶差分处理后再进行单位根检验,结果显示该序列是不含截距项和时间趋势项的平稳随机序列(见图 5)。

2.4.2 时间序列模型的识别和参数估计 时间序列通过 1 阶非季节性差分后平稳,根据最小二乘法原理初步拟定模型参数,分别建立 ARIMA(p,d,q)模型,其中 $d=1$, p 、 q 分别取值 0、1、2 和 1、2、0,共 8 种模型,并进行残差序列相关检验,选出的最优模型^[7-8]为 ARIMA(2,1,2),8 种模型的 AIC 值详见表 2。

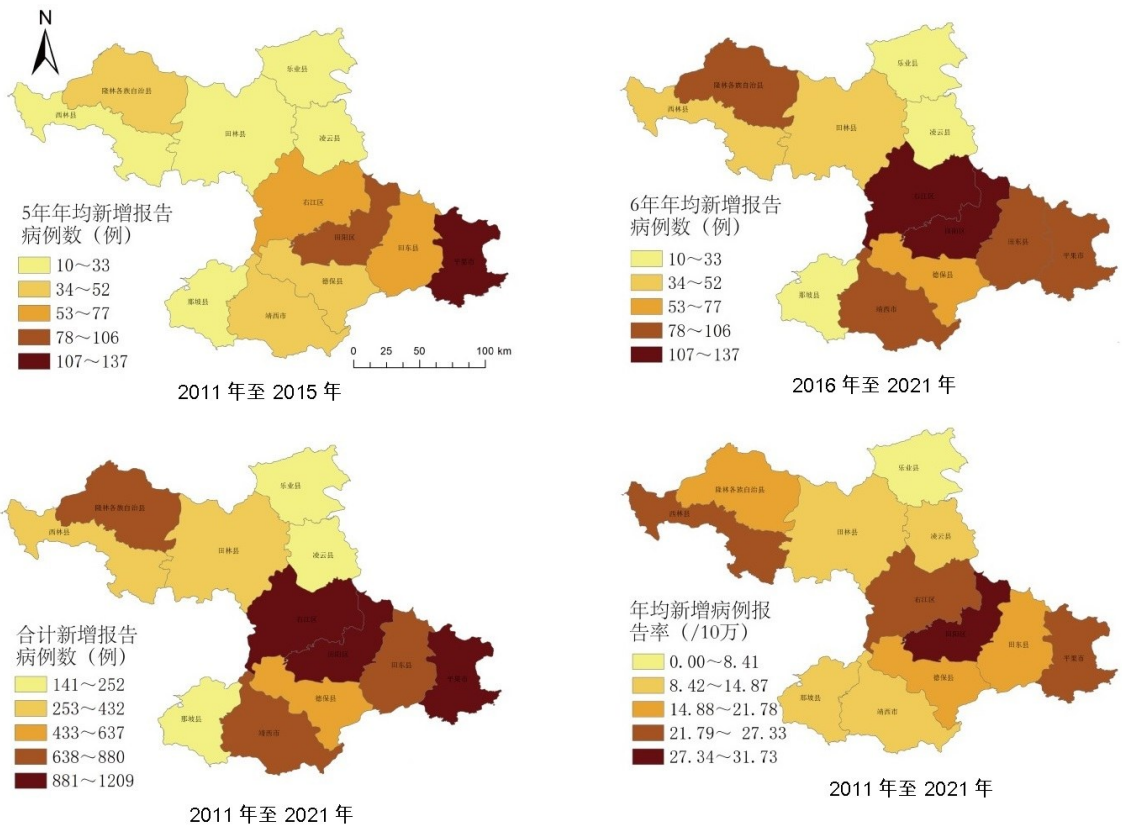


图 2 2011 年至 2021 年百色市艾滋病报告病例地区分布

2.4.3 模型的预测 根据建立的 ARIMA (2,1,2)模型,对百色市 2021 年 7 月至 2021 年 12 月艾滋病报告病例数进行预测,并将预测值和真实值进行比较,以评估该模型的预测精度,数据的平均拟合准确度为

85.29%,平均相对误差为 14.71%,说明该模型的拟合效果良好。根据上述模型进一步进行预测,获得 2022 年 1~6 月艾滋病报告病例预测值。见表 3 和图 6。

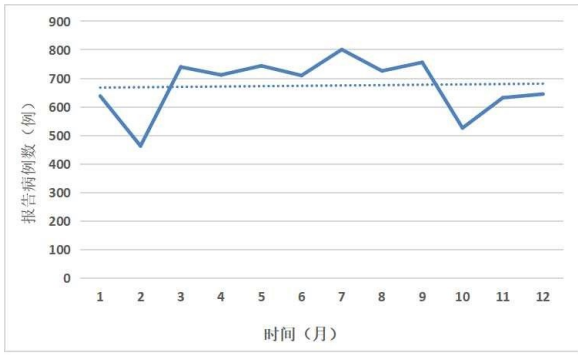


图 3 2011 年至 2021 年百色市艾滋病报告病例月份分布

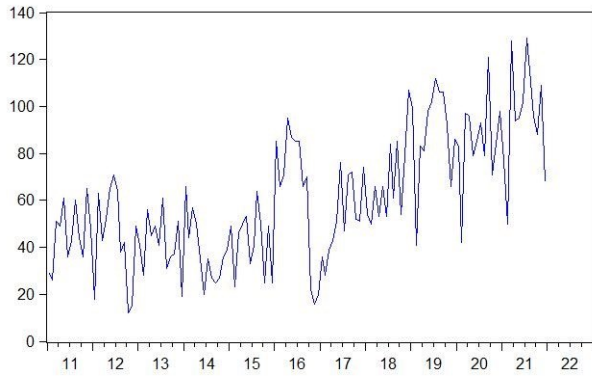


图 4 2011 年至 2021 年百色市艾滋病报告病例原始序列图

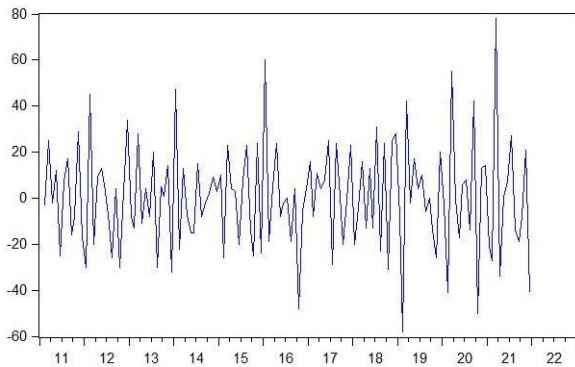


图 5 2011 年至 2021 年百色市艾滋病报告病例经非季节性 1 阶差后分序列图

表 2 ARIMA 各模型的 AIC 值

AR	MA	AIC
0	1	0.975226
0	2	0.955559
1	0	1.060218
1	1	0.931893
1	2	0.915586
2	0	0.915196
2	1	1.013586
2	2	0.899245

表 3 2021 年 7 月至 2022 年 6 月艾滋病报告病例数的拟合和预测

时间	实际值/人	预测值/人	绝对误差/人	相对误差/%
2021 年 7 月	129	102	27	20.93
2021 年 8 月	115	92	23	20.00
2021 年 9 月	96	99	3	3.13
2021 年 10 月	88	91	3	3.41
2021 年 11 月	109	100	9	8.26
2021 年 12 月	68	92	24	35.29
合计	605		89	14.71
2022 年 1 月		93		
2022 年 2 月		90		
2022 年 3 月		101		
2022 年 4 月		93		
2022 年 5 月		103		
2022 年 6 月		94		

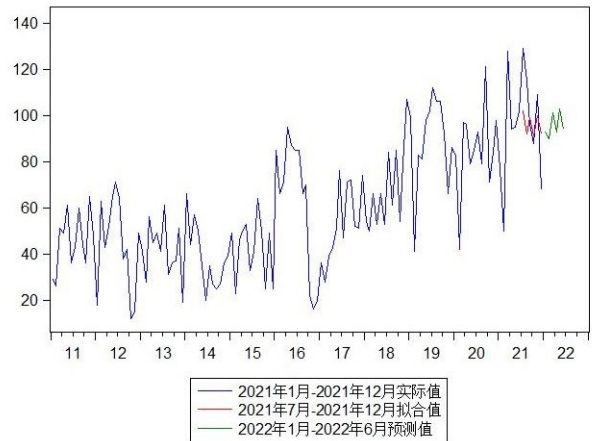


图 6 百色市 2011 年 1 月至 2022 年 6 月艾滋病报告病例数实际值、拟合值和预测值

3 讨论

本研究结果显示,2011 年至 2021 年百色市艾滋病报告病例数呈逐年上升趋势,艾滋病防控形势仍然严峻。这与广西艾滋病当前疫情一致。目前广西艾滋病疫情尚处于广泛流行期,由卖淫、嫖娼、吸毒等高危人群向普通人群进行扩散^[2]。百色市疫情上升可能与广西艾滋病感染的新特点有关,近年来,广西每年新报告病例经性传播的比例超过 90%^[9],其中农村 50 岁及以上中老年人群、男男性行为人群、高校青年学生等艾滋病感染率呈逐年上升^[10-15]。可能与老年人艾滋病相关知识缺乏、自我保护意识差、夫妻关系不和谐/离异/丧偶、经济条件改善,不安全商业性行为等因素有关^[11,14-15]。另外,近年来随着新媒体的兴起,年轻人利用微信、QQ 等社交软件网络交友更加便捷^[14],高校学生 HIV 感染率也明显上升,且以同性传播为主^[12-13]。

此外,广西 2010 年至 2020 年先后实施了两轮(第一轮 2010 年至 2014 年,第二轮 2015 年至 2020 年)的防治艾滋病攻坚工程,通过扩大宣传教育覆盖率和艾滋病检测范围、推行一站式抗病毒治疗服务等,病例的发现率大幅提高,HIV 感染者/AIDS 发现率从 2015 年底的 68.9% 提高到 2021 年的 82.5%^[16]。

在地区分布上,百色市艾滋病报告病例主要集中在东南区域,重点分布在右江区、田阳区、平果市、田东县、靖西市等地,具有较为明显的地区聚集性。可能由于这些县(市、区)为百色市重要的交通枢纽中心和经济金融中心,人口密集,人员流动性强,外来务工人员数量多而杂,在这些单身离家打工潮中可能出现“临时夫妻”,或从事性服务工作、艾滋病防控意识薄弱、流动性强,自我防护意识薄弱等因素有关,但具体原因有待进一步研究。

目前,时间序列 ARIMA 模型被广泛用于传染病的预测和分析^[17]。本文利用时间序列 ARIMA(2,1,2)模型对百色市 2021 年 7~12 月艾滋病月报告病例数进行预测,数据的平均拟合准确度为 85.29%,平均相对误差为 14.71%,说明该模型的拟合效果良好,可以用于艾滋病的短期预测。ARIMA 模型能利用时间序列分解出疾病的各种变化情况,并用序列图形式直观地展示疾病的变化趋势^[18],但需要注意的是,时间序列模型是在不考虑艾滋病影响因素的前提下,对未来流行趋势进行预测,由于艾滋病的流行受各种复杂因素影响,如流动人口、宣传力度、干预效果、抗病毒治疗覆盖率等,所以本研究仅进行短期预测。

参考文献:

[1] 央视财经央视新闻客户端徐丹. 警惕! 这种病,我国感染者达 114 万例! [EB/OL]. [2022-10-27]. <https://baijiahao.baidu.com/s?id=1717958282406055760&wfr=spider&for=pc>.

[2] 蓝光华,沈智勇,葛宪民,等. 广西艾滋病综合防控与管理 工作回顾与展望[J]. 中国临床新医学,2021,14(10):951-955.

[3] PORTET S. A primer on model selection using the Akaike Information Criterion[J]. Infect Dis Model,2020,5: 111-128.

[4] CHEN Y G. Spatial autocorrelation approaches to testing residuals from least squares regression[J]. PLoS One, 2016,11(1):e146865.

[5] 张孟媛,张强,罗佳伟,等. 重庆市艾滋病发病人数的 ARIMA 时间序列分析[J]. 中国卫生统计,2018,35(5):650-654.

[6] (美)乔治 E. P. 博克斯(George E. P. Box),(英)格威利姆 M. 詹金斯(Gwilym M. Jenkins),(美)格雷戈里 C. 莱因泽尔(Gregory C. Reinsel)//王成章,尤梅芳,郝杨. 时间序列分析[M]. 北京:机械工业出版社,2011:350-360.

[7] VAN RCM,JACKSON D. A new justification of the Hartung-Knapp method for random-effects meta-analysis based on weighted least squares regression[J]. Res Syn Meth,2019,10(4):515-527.

[8] ESTRADA E,CAPEROS J M,PARDO A. Change in the center of the distribution and in the individual scores:Relation with heteroskedastic pre- and post-test distributions [J]. Psicothema,2020,32(3):410-419.

[9] 葛宪民,杨文敏,朱秋映,等. 广西壮族自治区 2010—2017 年艾滋病流行病学特征分析[J]. 中华流行病学杂志, 2019,40(3):315-321.

[10] 孟琴,蓝光华,沈智勇,等. 广西 2006—2016 年男男性行为者 HIV 感染流行病学特征及趋势分析[J]. 中国艾滋病性病,2018,24(2):155-158.

[11] 王木,裴容,谭川,等. 2020 年达州市 50 岁及以上 HIV/AIDS 患者艾滋病感染相关因素定性分析[J]. 预防医学情报杂志,2022,38(6):740-744.

[12] 孟琴,朱秋映,刘玄华,等. 2011—2019 年广西新报告学生 HIV/AIDS 特征分析[J]. 应用预防医学,2021,27(4):295-298.

[13] 麻小龙,朱鑫,李亚静,等. 2014—2019 年河南省洛阳市 所高校学生 HIV 感染状况分析[J]. 河南预防医学杂志, 2022,33(3):212-214.

[14] 惠珊,许艳,王璐,等. 部分省市艾滋病异性传播方式 构成及特征分析[J]. 疾病监测,2011,26(6):458-462.

[15] 农丽萍,何波,农全兴,等. 中老年艾滋病高危人群及其 高危行为影响因素分析[J]. 中国皮肤性病杂志,2017, 31(6):641-644,655.

[16] 人民资讯. 广西第二轮防治艾滋病攻坚工程见成效[EB/ OL]. [2022-12-20]. <https://baijiahao.baidu.com/s?id=1723151643405312077&wfr=spider&for=pc>.

[17] 高雅,王伶,吴伟,等. 辽宁省手足口病疫情季节性 ARIMA 模型预测效果评价[J]. 中国公共卫生,2017,33(10):1482-1484.

[18] 吴尊友. 中国特色的艾滋病防治策略[J]. 中华疾病控制 杂志,2019,23(8):885-889.

收稿日期:2022-10-28;修回日期:2022-12-31