

· 临床论著 ·

我国免费艾滋病抗病毒药物治疗一线人数 ARIMA 模型预测

豆智慧 刘霞 张晓 赵德才 马烨 刘中夫 周义仓 张福杰

【摘要】 目的 探讨 ARIMA 模型在我国成人艾滋病一线治疗人数预测中的应用,阐述建模过程,建立预测模型并预测 2011 年至 2012 年治疗人数,分析预测影响因素,为今后两年艾滋病治疗资源配置提供参考。**方法** 数据来自国家免费艾滋病抗病毒药物治疗信息管理系统 2003 年至 2009 年成人(>15 岁)一线新治疗者按感染途径、基线 CD4 水平分组的季报人数,通过模型识别、参数估计、检验诊断、模型评价,建立艾滋病一线治疗人数 ARIMA 模型,预测 2010 年、2011 年治疗人数。**结果** 我国免费艾滋病一线成人治疗累计人数与在治人数预测模型均为 ARIMA(1,1,0)(0,1,0)₄,预测 2010 年与 2011 年底累计人数分别为 9.9 万、12.1 万,在治人数分别为 7.3 万、8.9 万,模型预测误差率分别为 0.26%、0.67%。我国免费艾滋病一线成人新治人数预测模型为(1,0,0)(0,1,0)₄,预测 2010 年与 2011 年一线新治人数分别为 2.1 万、2.4 万,每季度平均新治人数为 5.2 万,新治人数模型预测误差率为 2.39%。我国艾滋病一线新治人数表现逐年上升趋势和季节变化特点,而且受到感染途径、入选治疗标准等因素显著影响。**结论** ARIMA 模型适用于艾滋病治疗人数早期预测。

【关键词】 自回归移动平均模型模型;获得性免疫缺陷综合症;高效抗逆转录病毒治疗

Application of the ARIMA Model in predicting the number of ART-naïve patients on first-line treatments DOU Zhi-hui, LIU Xia, ZHANG Yao, ZHAO De-cai, MA Ye, LIU Zhong-fu, ZHOU Yi-cang, ZHANG Fu-jie. National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 100050, China

Corresponding author: ZHANG Fu-jie, Email: treatment@chinaaids.cn

【Abstract】 Objective To explore the application of ARIMA model in predicting the number of Chinese patients who will start first line antiretroviral therapy during 2010-2011, through describing the modeling process, building the prediction model

DOI:10.3877/cma.j.issn.1674-1358.2010.02.010

基金项目:国家自然科学基金资助项目(10971163)

作者单位:100050 北京,中国疾病预防控制中心性病艾滋病预防控制中心(刘中夫),治疗与关怀室(豆智慧、刘霞、张晓、赵德才、马烨、张福杰);西安交通大学理学院应用数学系(周义仓)

通信作者:张福杰,Email:treatment@chinaaids.cn

and verification of the model's applicability, with the aim to facilitate the management and planning of Chinese national free ART program. **Methods** Data were extracted from China's National Free ART program database. The quarterly reported numbers of new patients enrolled for ART from 2003 to 2009 were stratified by routes of transmission and CD4 counts. The ARIMA models were built by means of model identification, parameter estimation, inspection and diagnosis and model appraisal. The data of 2003 to 2008 were used to build the time sequence model, with those from 2009 used for model verification. **Results** With $ARIMA(1,1,0)(0,1,0)_4$, the cumulative numbers of Chinese patients given free first-line ART will be 99 000 and 121 000 by the end of 2010 and 2011 respectively, while active patients on ART will be 73 000 and 89 000 by the end of 2010 and 2011 respectively. Error rates for the prediction were 0.26% and 0.67%. With $ARIMA(1,0,0)(0,1,0)_4$ model, the number of new patients enrolled were predicted to be 20 000 in 2010 and 22 000 in 2011. The error rate of this prediction was 2.39%. The number of patients on first-line ART in China have been increasing with clear seasonal variation, and were impacted by routes of HIV transmission routes and change of treatment criteria. **Conclusions** This study shows that the ARIMA model can effectively predict the number of patients starting free ART in China.

【Key words】 ARIMA model; AIDS; HAART

我国免费艾滋病治疗人数未来变化趋势及其预测,愈加受到关注。本文采用一种精确度较高的短期预测模型——ARIMA 模型,早期预测我国艾滋病治疗人数,以及时发现问题,采取应对措施。

资料与方法

一、一般资料

本文时间序列数据采自国家免费艾滋病抗病毒药物治疗信息系统^[1]中2003年至2009年成人(>15岁)一线新开始治疗者(78 824例)感染途径、治疗基线CD4水平的季报人数及其累计、在治季报人数(见表1),其中2003年至2008年季报人数用于建立时间序列模型,2009年季报人数用于验证所建立的模型。

二、方法

ARIMA 模型全称为自回归移动平均模型(Autoregressive integrated moving average model, ARIMA),由 Box 和 Jenkins 于 20 世纪 70 年代初提出的时间序列预测方法,所以称为 box-Jenkins 模型^[2],其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,用一定的数学模型来近似描述这个序列。本文研究序列是以季度为时间单位的序列,故选择复合季节模型^[3]: $ARIMA(p,d,q)(P,D,Q)_s$,其中参数 p 为非季节性自回归阶数, d 为一般差分阶数, q 为非季节

性移动平均阶数;P 为季节性自回归阶数,D 为季节差分阶数,Q 为季节性移动平均阶数,s 为季节模型的时间单位相应的周期(本研究 s = 4)。统计软件采用 SPSS 17.0。研究步骤^[2,3]如下。

表 1 我国免费艾滋病抗病毒药物治疗一线人数的时间序列

年/季	累计治疗人数	正在治疗人数	开始治疗人数	感染途径 ^a			基线 CD4(个/μL) ^a	
				采供血	静脉吸毒	性途径及其他	< 200	≥ 200
2003/1	823	822	823	818	0	5	61	125
2003/2	1836	1731	1013	1005	0	8	55	101
2003/3	3160	2942	1324	1285	1	38	230	417
2003/4	4682	4333	1522	1456	9	57	267	431
2004/1	6362	5799	1680	1599	12	69	337	526
2004/2	8226	7399	1864	1762	10	92	446	501
2004/3	10804	9574	2578	2384	27	167	907	887
2004/4	13858	12266	3054	2613	84	357	1257	1128
2005/1	15603	13594	1745	1357	63	325	765	599
2005/2	17913	15423	2310	1635	156	519	1138	754
2005/3	20195	17110	2282	1374	269	639	1292	610
2005/4	22342	18691	2147	1181	323	643	1256	565
2006/1	24111	19831	1769	855	269	645	1097	401
2006/2	26435	21501	2324	962	439	923	1570	470
2006/3	28859	23188	2424	933	447	1044	1694	400
2006/4	31108	24706	2249	751	469	1029	1534	453
2007/1	33453	26282	2345	734	490	1121	1577	498
2007/2	36608	28604	3155	917	669	1569	2109	685
2007/3	39989	31042	3381	879	758	1744	2293	793
2007/4	43208	33290	3219	761	786	1672	2146	819
2008/1	46221	35307	3013	650	719	1644	1956	806
2008/2	50591	38569	4370	870	988	2512	2886	1230
2008/3	55238	41960	4647	771	1095	2781	3045	1366
2008/4	59595	45013	4357	827	963	2567	2786	1360
2009/1	63556	47703	3961	570	905	2486	2573	1216
2009/2	69179	51965	5623	867	1066	3690	3567	1865
2009/3	74940	56323	5761	906	1033	3822	3590	1998
2009/4	78824	58949	3884	423	736	2725	2374	1411

注: ^a 指感染途径、基线 CD4 为开始治疗者分组,基线 CD4 人数不等于开始治疗人数是因为未检测。

1. 模型识别 : (1) 绘制序列图和分析时间序列自相关图(autocorrelation chart function, ACF) 和偏自相关图(partial autocorrelation chart function, PACF)。我国免费艾滋病一线抗病毒药物治疗人数时间序列(见表 1) 的序列图均显示上升趋势、季节变化,其 ACF、PACF 图均为拖尾,即非平稳序列。(2) 选择适当的变换或差分使原序列平稳,初步确定模型 ARIMA(p,d,q) (P,D,Q)s 中 d、D 值。①差分:

在表1序列中,累计人数、在治人数这2个序列在1次一般差分和1次季节差分($d = 1, D = 1$)之后均为平稳序列,新治人数相关6个序列在1次季节差分($d = 0, D = 1$)之后均为平稳序列,因为上述各序列ACF、PACF图显示自相关系数和偏自相关系数均在95%可信区间限内。②变换:采供血新治疗人数时间序列图显示2004年为一陡增高峰,故采用2003年与2005年各季度相应治疗人数的平均值代替2004年实际治疗人数。基线CD4 < 200个/ μL 新治疗人数时间序列图显示序列方差大而采用平方根变换。以变换后的新序列建模。

2. 参数估计:采用最大似然估计识别阶段获得模型ARIMA($p, 1, q$)($P, 1, Q$)₄、ARIMA($p, 0, q$)($P, 1, Q$)₄中 p, q, P, Q 参数。 p, q, P, Q 参数常取值为0、1、2、3、4……之一。在SPSS17.0软件Analyze→Forecasting→Create Models→ARIMA过程中,先从0~4中由小到大取值分别赋给参数 p, q, P, Q 值(如:1、0、0、0或0、1、0、0或……),每次赋值之后,运行ARIMA过程,获得模型参数估计值、参数检验结果、模型参差白噪声检验、模型拟合优度等指标。筛选参数有统计学意义、模型参差为白噪声的模型,选定模型参数。如此循环,本研究各时间序列模型参数最后确定为 $p = 1, q = 0, P = 0, Q = 0$,即1阶自回归模型,记作AR(1)。当 $p = 1$ 时,检验各时间序列即AR(1)均具有统计学意义($P < 0.01$)。

3. 模型检验:绘制各序列模型残差ACF和PACF图,判断参差自(偏)相关系数是否在95% CI之内;从ARIMA过程中,获得模型检验统计量(Ljung-Box Q),判断模型残差服从以0为均数的平稳随机过程,即白噪声。

4. 选择模型:通过上述过程,本研究最终确定累计人数与在治人数时间序列预测模型均为ARIMA(1,1,0)(0,1,0)₄,模型预测误差率(2009年预测数减2009年实际数之差除以2009年实际数 $\times 100\%$)分别为0.26%、0.67%;开始治疗人数各时间序列(见表1)预测模型均为ARIMA(1,0,0)(0,1,0)₄,各模型预测误差依次为2.39%、15.50%、18.90%、0.02%、16.30%和2.40%。

结 果

一、我国免费艾滋病抗病毒药物治疗一线人数预测结果

1. 累计、在治人数2010年至2011年预测:根据累计、在治人数模型ARIMA(1,1,0)(0,1,0)₄预测,2010年累积、在治人数分别为9.9万(95% CI 9.5~10.3)、7.3万(95% CI 7.0~7.7),2011年累积、在治人数分别为12.1万(95% CI 11.2~13.1)、8.9万(95% CI 8.0~9.8)(见图1)。可见,2010年至2011年、在治人数呈上升趋势,平均每季度分别增加 4.8×10^3 (95% CI 3.7~5.9)、 3.5×10^3 (95% CI 2.5~4.5)。

2. 新治人数2010年、2011年预测:根据新治人数模型ARIMA(1,0,0)(0,1,0)₄预测,2010年、2011年新治人数分别为2.0万(95% CI 1.6~2.4)、2.2万(95% CI 1.6~2.8)(见图2a)。可见,2010年至2011年新治人数呈上升趋势,平均每季度增加 5.2×10^3 (95% CI 4.0~6.4);每年新治人数呈现季节波动,即冬春季低、夏秋季高。

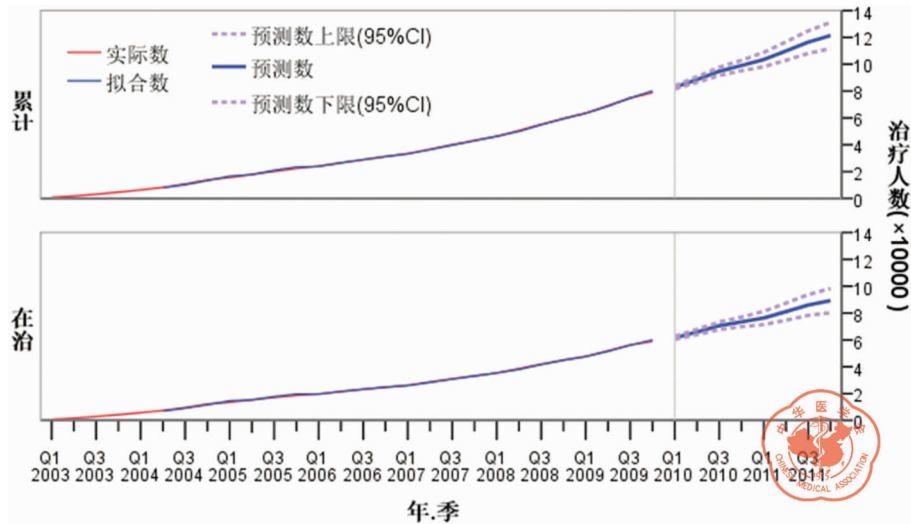


图1 我国免费艾滋病一线治疗中累计和在治实际人数与预测人数的关系

3. 新治人数预测的影响因素:(1)感染途径:将新治疗人数按感染途径分组建模预测显示(见图2b):①性途径(含不详途径)感染者2010年至2011年新治人数分别为1.4万(95% CI 1.2~1.6)、1.6万(95% CI 1.1~2.0),平均每季度新治人数为 3.7×10^3 (95% CI 2.9~4.5),表现快速上升趋势与季节波动。静脉吸毒感染者2010年至2011年新治人数分别为 3.0×10^3 (95% CI 2.1~4.1)、 2.7×10^3 (95% CI 0.7~4.7),平均每季度新治疗人数 0.7×10^3 (95% CI 0.4~1.1),表现略降趋势与季节波动。采供血感染者2010年至2011年平均每季度新治人数 0.4×10^3 (95% CI 0~1.1),表现低水平波动。②在2010年至2011年新治人数中,性途径(含不详途径)、静脉吸毒和采供血感染者构成率为77%、15%和8%。(2)治疗入选标准:将新治疗人数按治疗基线CD4分组建模预测显示(见图2c):①基线CD4 < 200个/ μL 者2010年至2011年新治人数分别为1.3万(95% CI 0.9~1.8)、1.6万(95% CI 0.9~2.5),平均每季度新治人数为 3.7×10^3 (95% CI 2.3~5.3),表现继续上升趋势与季节波动。基线CD4 ≥ 200 个/ μL 者2010年至2011年新治人数分别为 7.1×10^3 (95% CI 4.6~9.6)、 7.9×10^3 (95% CI 3.7~12.1),平均每季度新治疗人数 1.9×10^3 (95% CI 1.0~2.7),表现上升趋势与季节波动。②2010年至2011年新治人数中基线CD4 < 200个/ μL 、 ≥ 200 个/ μL 者构成率分别为66%和34%。

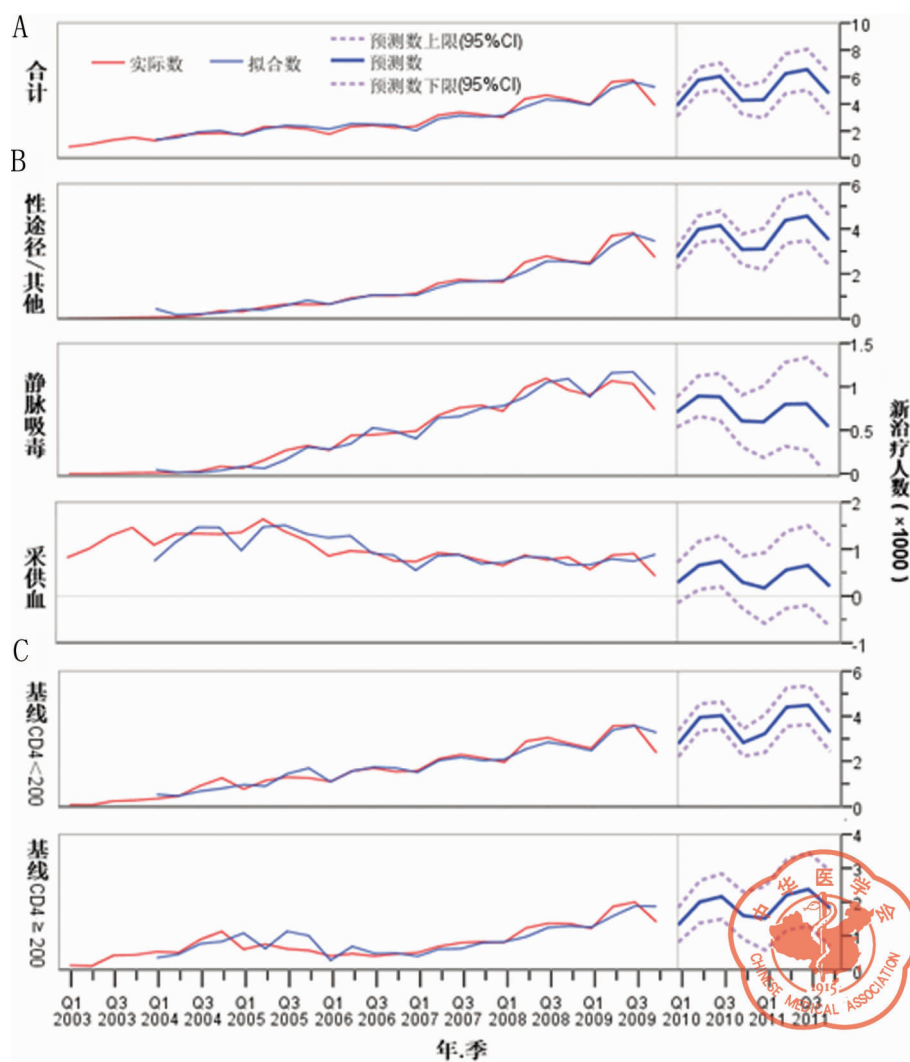


图2 我国免费艾滋病一线新治疗者实际人数与预测人数关系

讨 论

一、主要预测结果及其解释

本研究发现我国免费艾滋病抗病毒药物治疗一线成人累计、在治人数时间序列预测模型均为 $ARIMA(1,1,0)(0,1,0)_4$, 而新开始治疗人数各时间序列预测模型均为 $ARIMA(1,0,0)(0,1,0)_4$ 。各模型2003年至2009年实际数与拟合数较好吻合, 实际数均在拟合数95% CI可信限内, 且各模型预测误差率小于3%, 故本文预测模型基本合理。

本研究预测2010年至2011年我国艾滋病一线成人治疗人数呈现快速上升趋势。提示管理者对治疗管理、药物供应、服务流程等应做出调整 and 安排, 统筹医疗资源, 增强医护力量, 拟定详细的应对计划, 迎战面临的治疗压力。艾滋病一线成人治疗人数在2010年至2011年上升的原因^[4,5]: 近7年来, 我国艾滋病综合防治工作广泛深入地开展和加强, HIV抗体阳性筛查率、适治患者发现率、治疗覆盖率逐年提高, 以及免费艾滋病医疗救治和药物保障体系逐年完善的结果。

二、治疗人数预测的影响因素

1. 受季节气候影响:每年1、4季度新治疗人数比其他季度人数少,可能因为12、1、2月份是全年最冷的时间以及元旦、春节等节假日影响,即艾滋病是一种慢性进展性感染疾病,可择期治疗^[4]。

2. 受重大事件影响:2003年至2004年国家对于河南及其周边省份献血员人群开展了HIV抗体阳性“大筛查”^[6,7],造成2004年新治人数陡增,本文对此矫正后建模以减少预测结果影响。

3. 感染途径影响:本研究显示在2010年2011年一线成人新治疗者中,平均每季度新治人数分别为:性途径感染者为3700(77%),表现上升趋势;静脉吸毒感染者为700(15%),表现略降趋势;采供血感染者为400(8%),表现低水平波动。这与以前研究结果以及我国HIV感染者以性途径为主的特点相符^[4,7]。

4. 入选标准影响:2008年初我国免费艾滋病抗病毒药物治疗入选标准放宽,即将CD4在200~350个/ μ L患者由推荐治疗改为要求治疗^[1],这将增大新开始治疗人数。因此本文按基线CD4 \geq 200个/ μ L建模显示,2008年以来该人群治疗人数表现上升趋势,而且预测2010年至2011年每季度平均新治人数为1900(34%)。

三、本研究局限性

ARIMA模型是一种由结果到结果的预测,非因果模型,不能直接反映HIV抗体阳性筛查率、适治患者发现率、治疗覆盖率、治疗入选标准变化、政策干预强度等影响治疗人数的决定因素,只是这些因素的一种近似的综合反映。本文2010年至2011年预测治疗人数是假设未来两年与2003年至2009年我国艾滋病治疗管理政策等因素和变化相似条件下的近似估计。如果2010年至2011年上述假设条件发生重大变化,则可能发生预测数偏差。

志谢 中山大学公共卫生学院医学统计与流行病学系张晋昕主任统计指导

参 考 文 献

- 1 张福杰,曹韵贞,赵红心,等. 国家免费艾滋病抗病毒药物治疗手册. 第2版. 北京:人民卫生出版社,2008.
- 2 张文彤主编. SPSS11 统计分析教程. 北京:希望电子出版社. 2002;250-287.
- 3 杨帆,秦银河,刘丽华. ARIMA模型在门诊人次预测中的应用. 中华医院管理杂志,2009,25;28-31.
- 4 Zhang F, Dou Z, Ma Y, et al. Five-year outcomes of the China National Free Antiretroviral Treatment Program. Ann Intern Med, 2009,151:242-251.
- 5 Zhang F, Haberer JE, Wang Y, et al. The Chinese Free ART Program: challenges and responses. AIDS,2007,21:S143-S148.
- 6 Zhang F, Dou Z, Yu L, et al. The effect of highly active antiretroviral therapy on mortality in HIV-infected former plasma donors in China. Clin Infect Dis,2008,47:825-833.
- 7 冯丹,曹秀堂,董军,等. 综合性医院收容量预测的ARIMA模型构建研究. 解放军医院管理杂志,2007,14:101-103.

(收稿日期:2010-01-16)

(本文编辑:孙荣华)

豆智慧,刘霞,张晓,等. 我国免费艾滋病抗病毒药物治疗一线人数ARIMA模型预测[J/CD]. 中华实验和临床感染病杂志:电子版,2010,4(2):171-177.