

☐ 显示缩略图

本文结构

摘要

关键词

Abstract

Keywords

1 资料和方法

1.1 资料来源

1.2 方法

2 结果

2.1 复合季节模型ARIMA(p, d, q)(P, D, Q)s的建立

2.2 模型预测

3 讨论

参考文献

Abstract

Figures

Tables

引用本文

原梅, 张治国, 豆智慧, 王路钦, 张峴, 李卫民, 高基民. 北京市昌平区肺结核发病数ARIMA模型预测[J]. 疾病监测, 2015, 30(12): 1045-1049.

YUAN Mei, ZHANG Zhi-guo, DOU Zhi-hui, WANG Lu-qin, ZHANG Yao, LI Wei-min, GAO Ji-min. Application of ARIMA model in predicting incidence of pulmonary tuberculosis in Changping district, Beijing[J]. Disease Surveillance, 2015, 30(12): 1045-1049. [DOI: 10.3760/cjz151201045](#)

北京市昌平区肺结核发病数ARIMA模型预测

原梅¹, 张治国², 豆智慧³, 王路钦², 张峤³, 李卫民⁴, 高基民¹   

摘要: 目的 探讨时间回归滑动平均混合模型(auto regressive integrated moving average model, ARIMA)模型在北京市昌平区肺结核发病率预测中的应用,阐述建模过程并预测2015年昌平区肺结核发病率数据,为制定防治策略合理配置资源等提供参考。**方法** 采用全国结核病网络专报系统中的2009~2014年现住址为北京市昌平区的肺结核报告发病率数据,通过模型识别、参数估计、检验诊断及模型评价,建立昌平区肺结核发病率数据的ARIMA模型,并预测2015年肺结核发病率。**结果** 现住址为昌平区的肺结核发病率预测模型为ARIMA(0,1,1)(0,1,1)₁₂,预测2015年的新发报告肺结核患者总数为851例,模型2015年第一、二季度(1-6月)预测误差率为1.65%,不到10%模型预测精度较好。**结论** ARIMA模型适用于昌平区肺结核发病率较好的早期预测。

关键词: ARIMA模型 结核病 发病数 预测

Application of ARIMA model in predicting incidence of pulmonary tuberculosis in Changping district, Beijing

YUAN Mei, ZHANG Zhi-guo, DOU Zhi-hui, WANG IU-qin, ZHANG Yao, LI Wei-min, GAO Ji-min

Abstract: Objective To establish an ARIMA model for the prediction of incidence of pulmonary tuberculosis (TB) in Changping district, Beijing in 2015 and provide evidence for the prevention and control of TB and resource allocation. **Methods** The incidence data of TB in local population in Changping during 2009-2014 were collected from the National Tuberculosis Information Management System. An ARIMA model was established by means of model identification, parameter estimation, detection/diagnosis and model evaluation. The incidence of TB in Changping in 2015 was predicted. **Results** With the established model of ARIMA(0,1,1)(0,1,1)₁₂, it was predicted that the pulmonary TB case number would be 851 in local population in Changping in 2015. The error rate of the prediction during January-June 2015 was 1.65%, lower than 10%, indicating the high precision of the model. **Conclusion** This study showed that the ARIMA model is applicable for the prediction of TB incidence in Changping.

Key words: ARIMA model Tuberculosis Incidence Prediction

结核病是一种由结核分枝杆菌感染引起的慢性传染性疾病,可侵及全身各器官,其中以肺部结核感染最为常见。我国每年新发结核病病例130万例,死亡13万例,是全球第二结核病高发国家^[1]。近年来我国结核病疫情有了明显改善,但仍是危害居民健康的重大公共卫生问题^[2]。北京市人口增长速度快、流动人口大及结核病耐药等因素影响,结核病疫情改善,昌平区登记率为10/10万~30/10万,目前北京市结核病机构存在多样性,卫生人力资源在地理分布和人口数量方面的配置均未有很好的均衡性^[3]。昌平区与全市一样,同样面对结核病防治工作的诸多挑战。本研究拟采用密度较高的短期预测模型-自回归滑动平均混合模型(auto regressive integrated moving average model,ARIMA)模型^[4],分析北京市昌平区结核病报告发病数,预测该区未来结核病发病数,为更好地推动该区结核病防治措施的改进及防治资源的配置提供科学依据,并为北京市相关机构预防政策的制定和筹备提供新思路。

1 资料和方法

1.1 资料来源

中国结核病管理信息系统又称为结核病网络专报系统, 可供全国所有结核病防治机构实时录入和查询结核患者的报告、转诊追踪、治疗转归等信息, 兼有疫情监测和患者管理功能。本研究采用的时间序列数据均来自于该系统中2009年2014年现住址为昌平区内的肺结核报告发病数据, 该系统登记报告的肺结核发病数按照《中国结核病防治规划实施工作指南(2008年版)》的标准执行^[4]。数据按照月份统计, 见表1。

表 1 2009-2014 年现住址为昌平区的肺结核报告发病数

Table 1 Reported incidence of TB in local population in Changping, 2009-2014

月份	发病数					
	2009年	2010年	2011年	2012年	2013年	2014年
1	23	38	67	44	51	52
2	48	44	53	48	39	54
3	71	70	83	81	65	86
4	101	59	91	70	66	74
5	62	88	87	80	83	92
6	75	77	83	58	48	77
7	62	68	60	65	88	71
8	57	51	60	68	69	79
9	73	66	66	67	87	74
10	51	76	62	69	62	57
11	65	77	55	77	72	69
12	98	97	73	42	58	65
合计	786	811	840	769	788	850

表选项

表选项 ▼

1.2 方法

运用SPSS 19.0软件中的相关模块进行数据处理和分析。本研究是以月份为时间单位的研究序列,故选择复合季节模型^[5,6]:
 $ARIMA(p, d, q)(P, D, Q)s$, 参数 p 为非季节性自回归阶数, d 为一般差分阶数, q 为非季节性滑动平均阶数; 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周期(本研究 $s=12$)。

2 结果

2.1 复合季节模型ARIMA(p, d, q)(P, D, Q)s的建立

模型识别: (1)绘制20092014年现住址为昌平的肺结核报告发病数的时间序列图(图1),显示肺结核报告发病数随时间变化的趋势关系。图中序列显示一定的季节变化,以1年为一个周期,每年发生相似的变化,每年1月或2月,报告发病数相对较少。序列的平均水平和方差波动较大,即均值和方差并不平稳,故该序列不为平稳序列。(2)选择适当的变换或差分使原序列平稳,以初步确定模型ARIMA(p, d, q)(p , d , q 为 P , D , Q 的阶数。 D 值。原序列平均数转乘和1次一般差分。1次季节差分后变为平稳序列(图2),故模型参数确定 $d=1$, $D=1$ 。

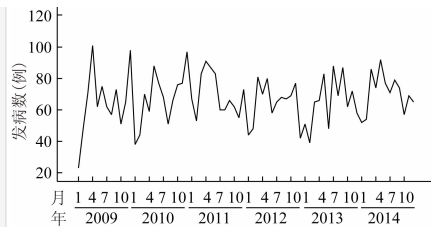


图 1 2009-2014年昌平区肺结核报告发病数时间序列图

Figure 1 Time series of TB cases reported in Changping, 2009-2014

图选项

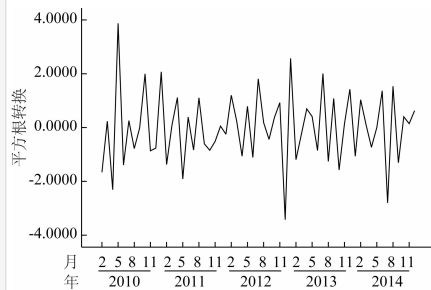


图 2 2010-2014年原序列经平方根转换和差分后时间序列图

Figure 2 Time series of TB cases after square root transform and difference treatment in Changping, 2010-2014

图选项

参数估计：采用最大似然估计识别阶段获得复合季节模型ARIMA(p, d, q)(P, D, Q)s中的p、q、P、Q参数，绘制经平方根变换和差分后序列的自相关系数图(ACF)和偏自相关系数图(PACF)(图 3)。模型中p、q、P、Q参数常取值为0、1、2、3、4等之一，在SPSS 19.0软件Analyze→Forecasting→Create Models→ARIMA过程中，依次取值分别赋给参数p、q、P、Q值，每次赋值后，运行ARIMA过程，以获得模型参数估计值、模型拟合优度指标、模型参数检验结果、模型残差白噪声检验等指标。根据参数有无统计学意义，残差是否为白噪声，模型拟合情况，选定最适模型参数，如此循环，不断改变模型阶数后，拟合优度统计量BIC最小的为最优参数组合。据此，本文时间序列模型参数最后确定为p=0，q=1，p=0，Q=1。

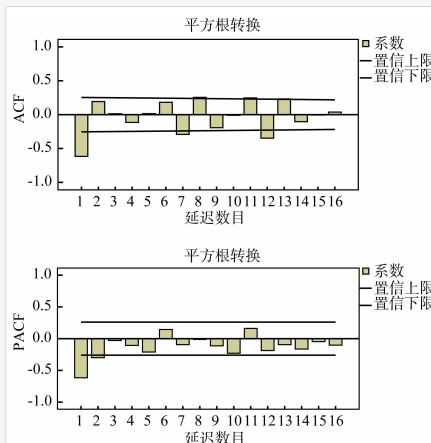


图 3 原序列经平方根变换和差分后ACF和PACF图

Figure 3 ACF and PACF of time series of TB cases after square root transform and difference treatment in Changping, 2009-2014

图选项

模型检验：绘制时间序列模型的残差ACF和PACF图(图 4)，判断残差自相关系数和偏自相关系数均在95%CI内。从ARIMA过程中，获得模型检验统计量Ljung-Box Q的P值>0.05，该统计量无显著性，故判断该模型残差服从以0为均数的平稳随机过程，即白噪声。选择模型：通过上述过程，本文时间序列复合季节ARIMA模型确定为ARIMA(0, 1, 1)(0, 1, 1)₁₂，该模型及各项参数差异均有统计学意义(P < 0.01)，见表 2。

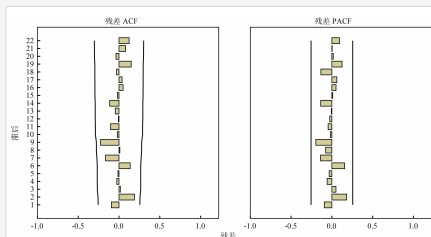


图 4 残差ACF和PACF图

Figure 4 Residual ACF and PACF

图选项

表 2 ARIMA 模型参数

Table 2 ARIMA model parameter

发病数-模型	SE	值	Sig
平方根差分1			
MA 滞后1	0.083	10.171	0.000
季节性差分1			
MA，季节性滞后1	0.144	3.401	0.001

表选项

2.2 模型预测

根据以上拟合得到的模型预测2015年现住址为昌平区肺结核发病数，见表 3。2015年现住址为昌平区的新发报告肺结核患者为851例，其中2015年第一、二季度(16月)预测误差率为1.65%，不到10%，模型预测精度较好，预测图及可信区间见图 5。

表 3 ARIMA模型预测2015年昌平区肺结核发病数

Table 3 Incidence of TB predicted with ARIMA model in Changping, 2015

月份	报告数	ARIMA(0, 1, 1)(0, 1, 1) ₁₂		
		预测值	LCL	UCL
1	59	53	30	82
2	51	52	28	80
3	65	82	51	118
4	83	76	46	110
5	86	90	57	129
6	72	70	41	104
7		76	45	112
8		75	44	112
9		78	46	116
10		63	35	97
11		72	42	109
12		64	35	100

表选项

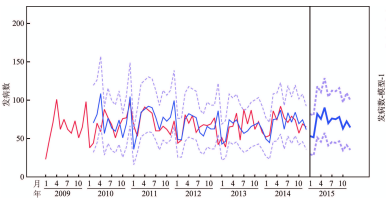


图 5 2009-2015年昌平区肺结核实际报告发病数与预测数关系

Figure 5 Relationship between actual case number and predicted case number of TB in Changping, 2009-2015

3 讨论

在医学领域，ARIMA模型是一种重要的时间序列分析定量预测模型，且为精密度较高的短期预测模型，是由Box等^[1]于20世纪70年代初提出的时间序列预测方法，故又称Box-Jenkins模型，其将预测对象随时间推移形成的数据序列视为一个随机序列，用一定的数学模型近似描述这个序列，进而从时间序列的过去值及现在值预测未来值。本研究采用全国结核病网络专报系统中2009-2014年现住址为昌平区的肺结核报告发病数建立ARIMA预测模型，确定适合的模型为ARIMA(0, 1, 1)(0, 1, 1)₁₂。模型2009-2014年的实际数与拟合数基本吻合，实际数均在拟合数95%可信区间内，预测得到的2015年每月肺结核发病数中，比较第一、二季度实际报告发病数与预测数可知模型预测精度较好，说明本研究建立的模型基本合理。

分析本研究时间序列的序列图，现住址为昌平区的肺结核报告发病数整体表现出稳定波动趋势。2014年与前几年相比，报告发病数略有增加，这可能与早期患者的发现增加所致。近年随着公共卫生服务项目的增多，新型农村合作医疗及医疗机构转诊工作的加强以及诊断设备的不断改进，使更多的人得到早期发现和诊断。本研究没有发现明显的肺结核发病周期性规律(即多见于冬春季)，这可能与现在各相关机构重视和加强肺结核防治措施，致其发病规律不再典型。本研究发现每年的1月或2月现住址为昌平区的肺结核报告发病数相对偏少，可能因为节假日医疗机构服务时间相对减少，或是节假日家人、亲朋好友欢聚，人们不愿去医院就医，不能及时发现和诊断。因此，不断提高发现结核病患者能力，统筹医疗资源，增强医疗力量仍是需要坚持的重要策略，也是控制和降低结核病疫情的必要措施。相关机构可以根据肺结核发病数预测情况，制定年度工作重点，合理配置人力和资源；同时加强结核病干预措施，如加强健康教育、提高知晓率，加强医疗机构督导检查等。居民自身也应注重健康，不能因节日而延迟就医，延误病情。本研究使用建立的模型预测2015年现住址为昌平区的肺结核发病数，不仅有利于该区相关机构掌握一定时间内结核病发病的变化趋势，评估已采取防控措施的实施效果，还可根据未来发病趋势的预测结果，为相关机构今后防控工作的人力、财力、物力安排和部署提供参考依据。

本研究拟合得到的模型，尽管本身及各参数都有统计学意义，但也存在一定缺陷。ARIMA模型随着预测时间的延长，预测误差会越来越大，但总体来说其预测准确性还是比较高的；同时预测得到的2015年发病数是在假设未来一年与前几年的预防管理政策等因素和变化相似条件下的近似估计值，如果2015年上述假设发生重大变化，则可能发生预测偏差。

参考文献

[1] World Health Organization. Global tuberculosis control 2011[R]. Geneva:WHO,2011.

[2] Zhang YY, Li P, Ran QZ, et al. Introduction to Modern Clinical Medicine[M]. 2nd ed. Beijing:Science Press,2012.(in Chinese) 张燕燕,李萍,冉启志,等.现代临床医学概论[M]. 2版.北京:科学出版社,2012.

[3] Chen X, Wang YL, He XX, et al. Human resources survey and staffing assessment of Beijing district level tuberculosis institutions[J]. Chinese Journal of Antituberculosis,2012,34(9):580-584.(in Chinese) 陈曦,王延莉,贺晓新,等.北京市区(县)结核病防治机构人力资源现状调查及配置评价[J].中国防痨杂志,2012,34(9):580-584.

[4] Department of Health Bureau of Disease Prevention and Control of the People's Republic of China, Ministry of Health Department of the People's Republic of China, Chinese Center for Disease Control and Prevention. China Tuberculosis Control Program Implementation Guide[M]. The 2008 ed. Beijing:Beijing Union Medical University press,2008.(in Chinese) 卫生部疾病预防控制中心,中国疾病预防控制中心.中国结核病防治规划实施工作指南[M]. 2008年版.北京:中国协和医科大学出版社,2008.

[5] Dou ZH, Liu X, Zhang Y, et al. Application of the ARIMA Model in predicting the number of ART-naïve patients on first-line treatments[J/CD]. Chinese Journal of Experimental and Clinical Infectious Diseases:Electronic Version,2010,4(2):171-177.(in Chinese) 豆智慧,刘霞,张峻,等.我国免费艾滋病抗病毒药物一线人数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.

[6] Yang F, Qin YH, Liu LH. Application of ARMA model in prediction of outpatient headcount[J]. Chinese Journal of Hospital Administration,2009,25(1):28-31.(in Chinese) 杨帆,秦银河,刘丽华.ARIMA模型在门诊人次预测中的应用[J].中华医院管理杂志,2009,25(1):28-31.

[7] Box GEP, Jenkins GM, Reinsel GC. Time Series Analysis, Forecasting and Control[M]. 3rd ed. Englewood Cliffs, NJ:Prentice-Hall,1994.