

应用时间序列法建立学校结核病预警模型研究

刘维华, 朱建良, 郭付爱, 张洁

河北省石家庄市疾病预防控制中心结核病防治所, 050011

【摘要】 目的 建立石家庄地区学校结核病预警模型, 为学校早期甄别和科学控制结核病疫情提供参考。**方法** 从“结核病信息管理系统”收集 2012—2017 年石家庄地区学校每月总体学生报告病例数和非健康检查学生报告病例数, 进行季节性分解, 按照指数平滑法或自回归移动平均模型 (ARIMA) 方法进行最优化模型筛选。利用 2012—2017 年石家庄地区学校结核病聚集性疫情的数据验证模型, 并确定对疫情进行预警的阈值条件。**结果** 对于总体学生发病例数数据拟合度最好的是 ARIMA(0,0,0)(1,1,0), 平稳 R^2 、 R^2 和正态化 BIC 值分别为 0.71, 0.61, 4.70; 对于非健康检查学生发病例数数据, 拟合度最好的是 ARIMA(0,1,1)(1,0,0), 平稳 R^2 、 R^2 和正态化 BIC 值分别为 0.60, 0.54, 4.08。如果当月非健康检查学生病例预测模型观察值高于拟合值, 并且总体学生病例预测模型中, 观察值也高于拟合值, 提示学校结核病聚集性疫情风险增大。对非健康检查学生病例预测模型采用逐一预测模式, 对总体学生病例预测模型采用连续预测模式。**结论** 利用时间序列法建立的预警模型有助于发现学校结核病疫情。

【关键词】 时间; 序列分析; 结核; 学生保健服务

【中图分类号】 R 181.8 R 52 **【文献标识码】** A **【文章编号】** 1000-9817(2018)11-1682-04

Time series analysis for early warning model of tuberculosis in schools/LIU Weihua, ZHU Jianliang, GUO Fuai, ZHANG Jie. Shijiazhuang Center for Disease Control and Prevention, Shijiazhuang (050011), China

【Abstract】 Objective To establish early warning model of tuberculosis (TB) based in schools of Shijiazhuang. **Methods** The monthly reported TB cases in schools in Shijiazhuang from 2012 to 2017 were collected from Tuberculosis Information Management System. Student TB cases per month from routine as well as random monitoring approaches were collected. The optimal model was selected according to the exponential smoothing method or ARIMA method. By comparing data on reported TB cases from schools during 2012–2017, the threshold of early warning for school tuberculosis cluster cases were verified. **Results** The best model for school TB cases was the simple seasonal model with the smooth R^2 , R^2 and normalized BIC at 0.711, 0.608, and 4.70, respectively. For school TB cases from non-health screening approaches, the best model was Winters plus model with smooth R^2 , R^2 , and normalized BIC at 0.597, 0.544, and 4.082, respectively. When the observed value of Winters plus model was higher than the fitted value, and the observed value of simple seasonal model was also higher than the fitted value, it was the threshold of early warning of school TB outbreak which indicated the enhanced risk. One by one prediction and series prediction were used in simple seasonal model and Winters plus model, respectively. **Conclusion** The early warning models established by time series analysis is helpful for the early detection of TB epidemic in schools.

【Key words】 Time; Sequence analysis; Tuberculosis; Student health services

结核是由结核分枝杆菌感染人体后出现的感染性疾病, 以肺结核为主。学校环境人群高度集中, 流动性强, 学生接触密切, 为传染病的暴发流行提供了条件^[1], 加之结核是呼吸道传染病, 极易在学校发生甚至暴发。学生作为特殊群体, 容易受到关注, 特别是学校结核病聚集性疫情的发生, 给学校带来较大的负面影响^[2]。因此, 了解学校结核病发病趋势, 建立敏感性高的发病预测模型, 采取精准、高效的处置技

术策略, 是目前学校结核病防控的重点。本研究通过时间序列分析方法建立石家庄地区学生结核病的预测模型, 初步构建学校学生结核病聚集性疫情预警系统, 为早期甄别和科学防控提供依据。

1 资料来源与方法

1.1 资料来源 根据“结核病信息管理系统”和统计年鉴, 将 2012—2017 年学生资料按照年、月进行归类整理。

1.2 时间序列分析 时间序列分析是指将某一现象发生的数量变化, 依照时间先后顺序排列, 以揭示随时间变化的发展规律, 从而用于预测现象发展数量及

【作者简介】 刘维华 (1977—), 女, 河北宁晋人, 博士, 副主任技师, 主要从事结核病监测与预防控制工作。

DOI: 10.16835/j.cnki.1000-9817.2018.11.022

动向。时间序列法包括自回归移动平均模型(*ARIMA*)、指数平滑法、自回归模型。其中 *ARIMA* 模型已被广泛应用于传染性疾病的发病预测^[3-4]。*ARIMA* 乘积季节模型,即 *ARIMA*(p, d, q)(P, D, Q)模型,可以处理并分析含有季节性成分的时间序列,对季节周期性的时间序列资料的预测更准确^[5]。其中 *AR* 是自回归, *MA* 是移动平均, p, q 是自回归和移动平均阶数, P, Q 是季节性自回归和移动平均阶数, d, D 分别为非季节性和季节性差分次数。

“结核病信息管理系统”中患者来源分为:健康检查、因症就诊、因症推荐、转诊和追踪。根据患者来源,统计学生发病例数,按照年月进行汇总。非健康检查学生病例是指排除健康检查发现的学生病例(包括因症就诊、因症推荐、转诊和追踪的学生病例);全体学生病例包括所有来源病例。根据月发病数制作时间序列图,利用季节性分解显示季节变化,判断季节波动是随时间加剧还是持续稳定存在。

1.3 模型构建及拟合 *ARIMA* 模型建模分 4 步进行^[6]:(1)序列的平稳性,通过差分使原序列满足平稳可逆的要求。(2)模型识别,主要通过自相关图和偏自相关图把握模型的大致方向,为目标序列定阶,提供几个粗模型以便进一步分析。(3)参数估计和模型诊断,利用非线性最小乘法估计模型参数,对模型残差序列进行白噪声检验,判断其适合性,并根据标准

化贝叶斯(*BIC*)准则判断模型的拟合优度,一般 *BIC* 值最小的模型较合适,并计算平均相对误差,平均相对误差较小说明模型拟合较好。(4)预测。利用 SPSS 23.0 软件中“专家建模器”模块,采用条件自动筛选分别建立对非健康检查学生病例和全体学生病例的最优预测模型。确定平稳的 R^2, R^2 及正态化 *BIC* 值。并对 2018 年 1—6 月的数据进行预测,评估连续预测和逐一预测的最佳方式。

收集石家庄地区 2012—2017 年学校结核病聚集性疫情的资料,验证最优模型,初步确定学校结核病聚集性疫情的预警条件。

2 结果

2.1 学生总体发病情况 2012—2017 年石家庄地区共报告学生结核病例 2 338 例。2012—2017 年学生人群报告发病率分别为 20/10 万, 17/10 万, 22/10 万, 21/10 万, 21/10 万, 31/10 万, 差异有统计学意义($\chi^2 = 4.46, P = 0.04$)。

2.2 时间序列图 由于近几年的学生人群人口数没有明显变化,因此使用总体学生发病数制作时间序列图。见图 1。月发病数范围为 8~74 例,平均 32.47 例;月发病数在 2017 年 4 月出现明显上升趋势;波动有一定的季节性且持续存在,表现为 4—5 月和 9—11 月升高, 7—8 月降低, 12 月至次年 2 月降低。

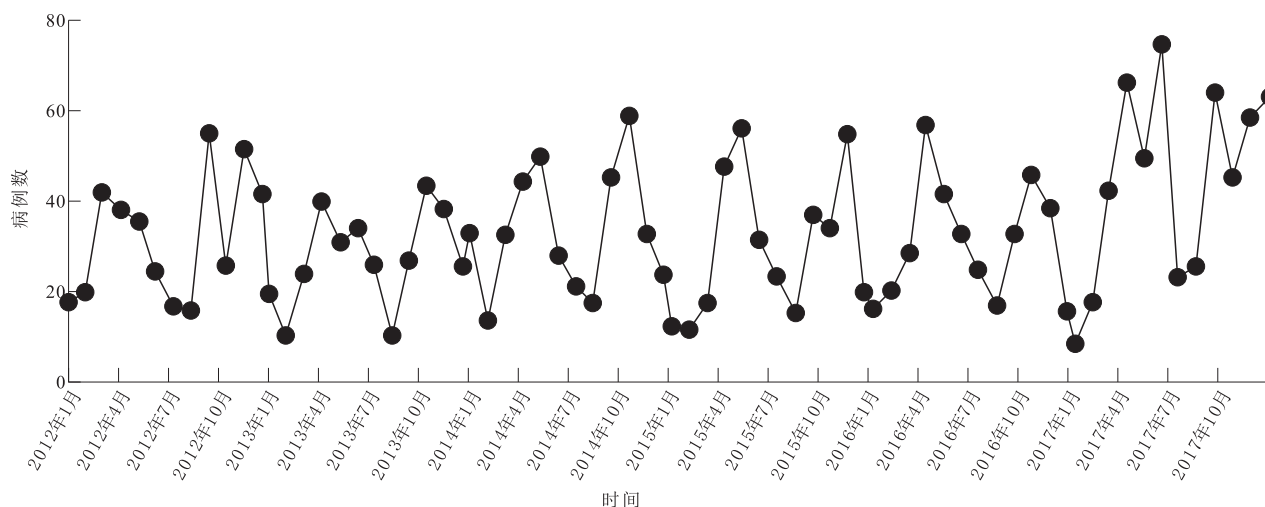


图1 石家庄地区 2012—2017 年学生结核病发病时间

2.3 建立时间序列预测模型 因时间序列图中 2012—2017 年学生月发病数呈持续季节性波动,所以在专家建模器的条件设置中选择的 2 种模型均考虑季节性模型。经过筛选,对于总体学生发病例数数据拟合度最好的是 *ARIMA*(0,0,0)(1,1,0),模型残差序列 Ljung—Box $Q = 11.06, P = 0.81$,其平稳 R^2, R^2 和正态

化 *BIC* 值分别为 0.71, 0.61, 4.70;对于非健康检查学生发病例数数据拟合度最好的是 *ARIMA*(0,1,1)(1,0,0),模型残差序列 Ljung—Box $Q = 9.82, P = 0.83$,其平稳 R^2, R^2 和正态化 *BIC* 值分别为 0.60, 0.54, 4.08。预测模型如图 2~3。

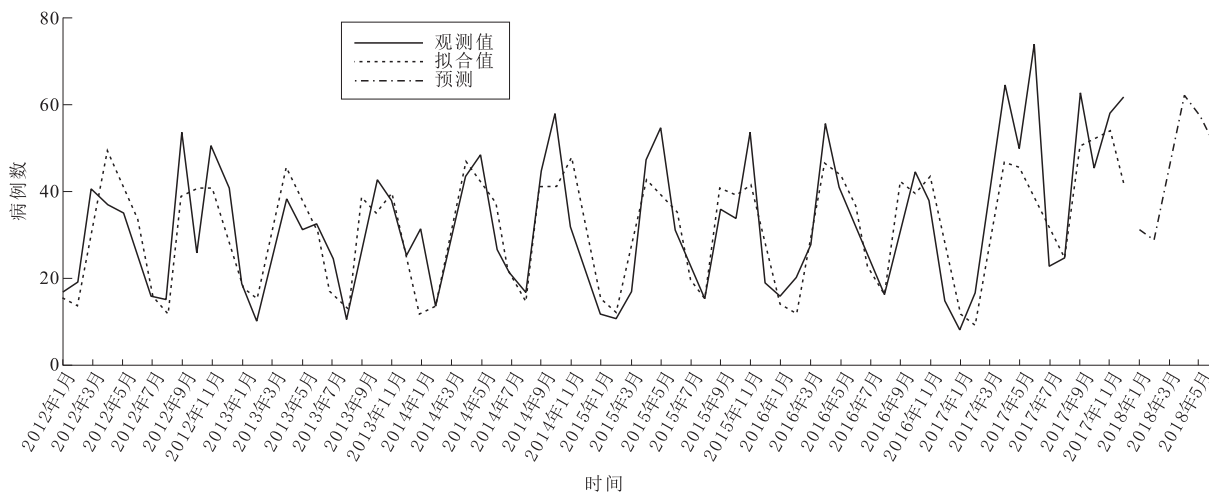


图2 总体学生结核病例拟合度最佳的预测模型

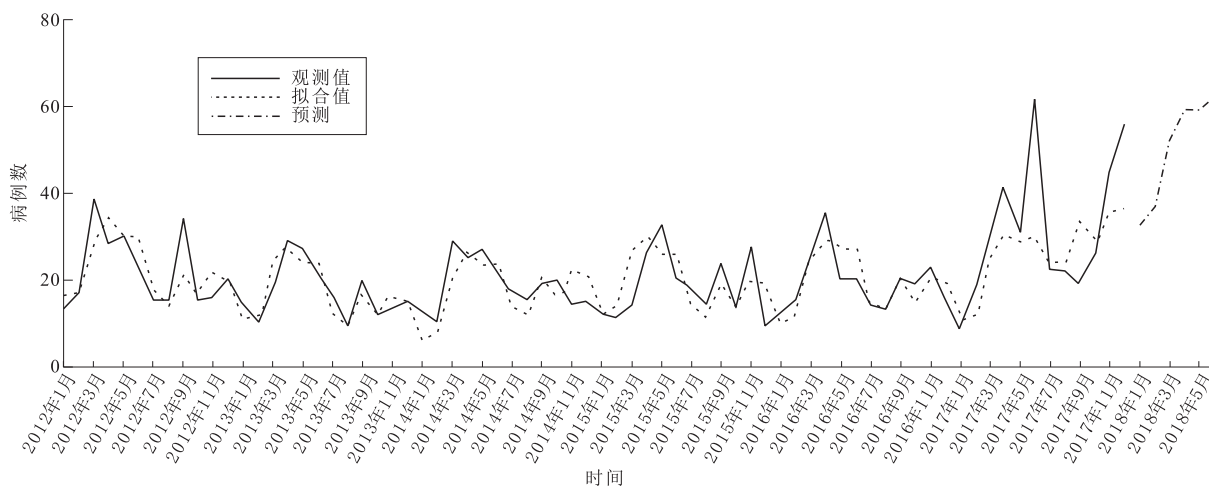


图3 非健康检查学生结核病例拟合度最佳的预测模型

2.4 学校结核病聚集性疫情与预测模型的验证
2012—2017 年石家庄地区发生的学校结核病聚集性疫情共 7 起。见表 1。

表 1 石家庄地区 2012—2017 年学校结核病聚集性疫情

序号	时间	聚集性疫情学校
1	2012 年 11 月	河北体育学院
2	2013 年 10 月	河北职业技术学校
3	2014 年 1 月	精英中学
4	2015 年 5 月	河北美术学校
5	2015 年 11 月	平山实验中学
6	2016 年 4 月	辛集中学
7	2017 年 6 月	行唐中学
8	2018 年 1 月	石家庄 35 中

验证发现,当非健康检查学生病例预测模型,观察值高于拟和值时,并且当月总体学生病例预测模型中观察值也高于拟合值,会在当月或后 2 月出现 1 次疫情。

2.5 序列预测值 利用时间序列分析,采用连续预测模式,利用 2012—2017 年的数据预测 2018 年 1—6 月数据;采用逐一预测模式,即利用 2012—2017 年的数

据预测 2018 年 1 月数据,然后利用 2012—2018 年 1 月数据预测 2018 年 2 月数据;之后依次预测。见表 2。

表 2 学校结核病病例时间序列分析预测值比较

时间	总体学生病例			非健康检查学生病例		
	实际	连续	逐一	实际	连续	逐一
	观察值	预测值	预测值	观察值	预测值	预测值
2018 年 1 月	65	31.49	31.49	46	32.26	32.26
2018 年 2 月	26	29.16	34.42	20	36.44	38.18
2018 年 3 月	45	44.66	47.89	32	51.95	51.26
2018 年 4 月	68	62.16	64.36	40	59.47	43.87
2018 年 5 月	62	57.49	59.81	36	58.98	39.65
2018 年 6 月	50	50.99	53.18	34	61.99	26.66

总体学生病例预测模型中,连续预测值低于逐一预测值,更接近于实际观察值;在非健康检查学生病例预测模型中,逐一预测值低于连续预测值,更接近于实际观察值。其中,2017 年 12 月中观察值 60 远高于拟合值 38,推测可能会有聚集性疫情,在 2018 年 1 月石家庄 35 中出现疫情,并且构成了突发公共卫生事件。

3 讨论

学校具有人群高度聚集、学生接触密切的特点,加之高度紧张的学习状态使学生抵抗力降低,容易造成传染病流行^[7]。因此,运用敏感性高的工具和方法,早期侦测和识别聚集性病例至关重要^[8]。本研究分析了 2012—2017 年石家庄地区学校结核病发病趋势,初步建立了一种敏感的发病预测模型,以期对学校结核病防控起到预警作用。

王婷^[9]研究显示,2011—2016 年大连的学校结核病发病率由 40.86/10 万下降至 29.95/10 万;广西 2011—2015 年学生的发病率由 13.46/10 万下降至 12.14/10 万^[10];2012—2017 年石家庄地区学校学生发病率平均为 20.9/10 万,高于广西,低于大连,但 2017 年的发病率比往年有明显增高,提示石家庄地区学校结核病聚集性疫情可能会增多,学校应加强结核病防控工作。在本研究中,总体学生病例呈季节性波动,4,5,9—11 月发病率较高,和王婷^[9]报道的结果一致,但和全国报道的结果不同^[11]。因为石家庄地区执行学校开学结核病体检筛查较好,因此开学的 3,4 月中,结核病管理系统中录入的健康检查学生病例多,9,10 月有秋季学生体检。因此,在预测模型中排除了健康检查病例。

目前,国家预警体系预警多是针对急性传染病^[12],2018 年 7 月刚开始在“结核病信息管理系统”中对学生结核病进行单一病例预警。结核病是一种慢性传染病,国家预警覆盖面过大,且单一地区的预警线不同,因此有必要建立适合于本地区的局部预警模式。在本次预测模型中,建立了总体学生病例预测模型和非健康检查学生病例预测模型,并根据 2 个模型的对比,推测学校结核病聚集性疫情可能性大小。在非健康检查学生病例模型中,当学生病例的观察值高于拟合值时,往往提示学校聚集性疫情风险增大。结合总体学生病例预测模型,可以看出学校疫情可能发生在当月,也可能发生在后月。目前在 2012—2018 年 1 月的数据中,预测准确率达到 100%,但该模型的预测准确率还需要经过更长时间验证。

本研究的预警指标是非健康体检学生病例模型中观察值高于拟合值,总体学生病例模型中,同期观察值高于拟合值,则当月或后 2 月本地区出现学校结核病聚集性疫情的风险增高。非健康检查学生病例是散发的门诊接诊病例数据,利用往年连续的发病数据建立模型。通过连续预测和逐一预测方式,与 2018

年实际观察值进行对比发现,对于总体学生病例的预测采用连续预测,数据更接近真实值;而对于非健康检查学生病例预测采用逐一预测,数据更接近真实值。

良好的监测数据来源和质量是有效预警的基础,采用法定报告传染病数据作为暴发探测数据来源具有明显优势,但要求基层门诊医生在录入时确保学生来源准确,严格按照健康检查、转诊、追踪等正确信息录入。本研究的模型是整个石家庄地区的数据模型,范围较大,可以考虑各县区根据自己的数据进行模型构建,创建更小范围的模型,提高模型预警的敏感性。除学生发病病例数据外,还应积极探索其他的预测数据,如学校缺勤监测、医疗机构的症状监测、药店的药品监测等,将大数据联合利用起来,以提高预警的及时性和准确性。

4 参考文献

- [1] 张颖,曾光.2004 年全国学校突发公共卫生事件分析[J]. 中国学校卫生,2007,28(1):81-82.
- [2] 王黎霞,成诗明,陈伟. 学校结核病防治工作手册[M]. 北京:军事医学科学出版社,2012:8-9.
- [3] 陆波,闵恩韬,闵红星,等. 应用 ARIMA 模型预测麻疹发病率的可行性研究[J]. 中国卫生统计,2015,32(1):106-107.
- [4] 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.
- [5] WANG Y W, SHEN Z Z, JIANG Y. Comparison of ARIMA and GM (1,1) models for prediction of hepatitis B in China[J]. PLoS One, 2018, 13(9):e0201987, doi: 10.1371/journal.pone.0201987. eCollection.
- [6] 姚英,沈毅.手足口病发病趋势的 ARIMA 模型预测[J]. 浙江预防医学,2015,27(2):147-149.
- [7] 肖占沛,马雅婷,王长双,等. 2004—2013 年河南省学校传染病突发公共卫生事件分析[J]. 现代预防医学,2015,42(12):2245-2248.
- [8] YIN Z Y, FANG C F. Study on early warning methods of infectious diseases[J]. Chin J Health Stat, 2010, 27(2):218-220.
- [9] 王婷. 2011—2016 年某市学校结核病疫情特征分析[D]. 大连:大连医科大学, 2017.
- [10] 崔哲哲,林定文,林玫. 广西学生肺结核发病趋势分析及预警系统建立[J]. 中国学校卫生,2017, 38(1): 144-147.
- [11] 杜昕,陈伟,黄飞,等. 2004~2008 年全国学生肺结核报告发病特征分析[J]. 中国健康教育,2009,25(11):803-810.
- [12] 张洪龙,曾令佳,赖圣杰,等. 2016 年国家传染病自动预警信息系统运行情况分析[J]. 疾病监测,2018, 33(2):159-167.

收稿日期:2018-08-10;修回日期:2018-10-09