

时间序列分析在上海静安区流感样病例预测预警中的应用

沈冰^a, 杨晓明^b, 卑伟慧^c, 沈磊^c, 倪晓芬^c, 孟威^c, 高洁^c

摘要: [目的] 探讨时间序列分析在静安区流感样病例(ILI)预测预警中的应用。[方法] 使用时间序列分析中的自回归求和移动平均模型(ARIMA)分别对静安区成人和儿童2011年7月—2014年7月的每月ILI监测发病数资料进行建模,然后预测2014年8—12月ILI发病水平,并与实际水平对比,以检验模型。[结果] 监测结果显示,儿童和成人ILI发病趋势特点不同,儿童ILI发病未见明显季节性,成人ILI季节性较为明显。儿童监测数据识别最佳模型为ARIMA(0, 1, 1),成人监测数据识别最佳模型为ARIMA(1, 0, 0)(1, 1, 0)₁₂。两个模型参数均具有统计学意义,预测效果良好。[结论] 可使用ARIMA模型对静安区ILI发病趋势进行预测,但需对成人和儿童数据分别分析,根据预测区间可考虑设定预警阈值。

关键词: 时间序列分析; 流感样病例; 预测; 预警; 静安区

Application of Time Series Analysis in Forecasting and Early Warning of Influenza-Like Illness in Jing'an District, Shanghai SHEN Bing^a, YANG Xiao-ming^b, BEI Wei-hui^c, SHEN Lei^c, NI Xiao-fen^c, MENG Wei^c, GAO Jie^c (a.Administrating Office b.Department of Chronic Diseases Control and Prevention c.Department of Infectious Diseases Control and Prevention, Jing'an District Center for Disease Control and Prevention, Shanghai 200041, China). Address correspondence to GAO Jie, E-mail: xiaojusmart@163.com • The authors declare they have no actual or potential competing financial interests.

Abstract: [Objective] To apply time series analysis in forecasting and early warning of influenza-like illness (ILI) occurred in Jing'an District, Shanghai. [Methods] The monthly monitoring data of ILI incidence of adults and children in Jing'an District during July 2011 and July 2014 were modeled with autoregressive integrated moving average model (ARIMA) in time series analysis. Then we predicted the ILI incidence from August to December 2014 by the established model, and tested the model by comparing predicting values and actual values. [Results] The monitoring results showed different trends of ILI between children and adults. There were obvious seasonal variations for adults, but not for children. The optimal model was ARIMA (0, 1, 1) for children and ARIMA (1, 0, 0)(1, 1, 0)₁₂ for adults. The parameters of the two models were both statistically significant, and the result of forecast was ideal. [Conclusion] ARIMA model could be used to predict the incidence trend of ILI in Jing'an District, but the data of children and adults require separate analysis. Warning thresholds could be set according to the range of forecasting results.

Key Words: time series analysis; influenza-like illness; forecast; early warning; Jing'an District

静安区地处上海市中心城区,面积7.62平方公里,常住人口数约25万。城区人口密度较大,但卫生条件佳,多数传染病发病率较低。流感是该区较为常见的影响居民健康的传染病种类,也是引起集体单位聚集性病例发生的常见原因。但由于常规诊疗过程中,医疗机构并不对患者开展流感实验室诊断,流感的发病水平通常以流感样病例(influenza-like illness, ILI)来反映。静安区内设有国家级儿童流感监测点1家,市

级成人流感监测点1家。两家监测点医疗机构每日开展症状监测,收集ILI就诊人数,并通过流感监测网络上报数据。为充分利用监测数据指导疾病防控,提高疾病的预测预警能力,实现通过历史监测数据预估疾病流行态势,并在疾病流行态势超出预计时,及时发现预警信号,本研究拟探讨时间序列分析^[1]中的自回归求和移动平均模型(autoregressive integrated moving average model, ARIMA)在流感监测点症状监测数据,即ILI发病资料的建模和预测预警中的应用。

DOI: 10.13213/j.cnki.jeom.2016.15559

[作者简介] 沈冰(1971—),女,硕士,副主任医师;研究方向:传染病防控;E-mail: sbing002@126.com

[通信作者] 高洁, E-mail: xiaojusmart@163.com

[作者单位] 静安区疾病预防控制中心 a. 行政办公室 b. 慢性病防制科 c. 传染病防制科, 上海 200041

1 材料与方法

1.1 资料来源

收集上海市静安区两家流感监测点医疗机构2011

年7月—2014年7月每日报告的ILI病例资料。根据全国流感监测方案,ILI定义为:发热,体温 $\geq 38^{\circ}\text{C}$,伴有咳嗽或咽痛之一症状,缺乏其之一症状,缺乏其他实验室诊断依据者。由中国疾病预防控制中心信息系统导出监测点医院ILI报告数据。

1.2 方法

1.2.1 数据平稳化 对2011年7月—2014年7月静安区成人和儿童ILI每月发病数的时间序列分布图和自相关函数进行观察,判断其是否为平稳序列。如不是,则进一步尝试通过差分或对数变换等方式,使不平稳的序列平稳化。

1.2.2 模型的识别和定阶 采用逐步试探的方法,摸索出BIC值最小的最优模型。如数据为平稳序列,则将模型标记为 $\text{ARIMA}(p, d, q)$,其中 p 是自回归的阶, d 是差分次数, q 是移动平均的阶。如数据为不平稳序列,需同时考虑周期性,则模型标记为: $\text{ARIMA}(p, d, q)(P, D, Q)s$,其中 p, d, q 的含义与之前相同, P, D, Q 分别代表季节周期的自回归的阶、差分次数和移动平均的阶, s 指循环长度。

1.2.3 模型的诊断性检验 模型建立后,模型的拟合值和实际值之间的差距即为残差。对2011年7月—2014年7月儿童和成人ILI发病数的模型拟合值和实际值之间的残差进行白噪声检验,计算Ljung-Box Q统计量,以判断残差是否为随机性误差。如残差序列为白噪声序列,则模型达到较好的预测效果;反之,则提示所建模型不是最终模型,仍需进一步对模型进行拟合,直到实现残差为白噪声序列方可。

1.2.4 模型的预测分析 对使用2011年7月—2014年7月3年历史数据建立且通过检验的模型进行预测分析,利用所建立的模型预测2014年8月—2014年12月的静安区成人和儿童ILI发病数值及其95%可信区间,并与实际发生值进行对比。如实际值落入预测值的可信区间内,则进一步检验了模型的良好和精确预测效果,模型可应用于实际工作;若实际值不在可信区间内,应结合实际情况进一步分析是否有其他可解释的原因改变发病报告水平,并判断是否需继续改变模型参数。

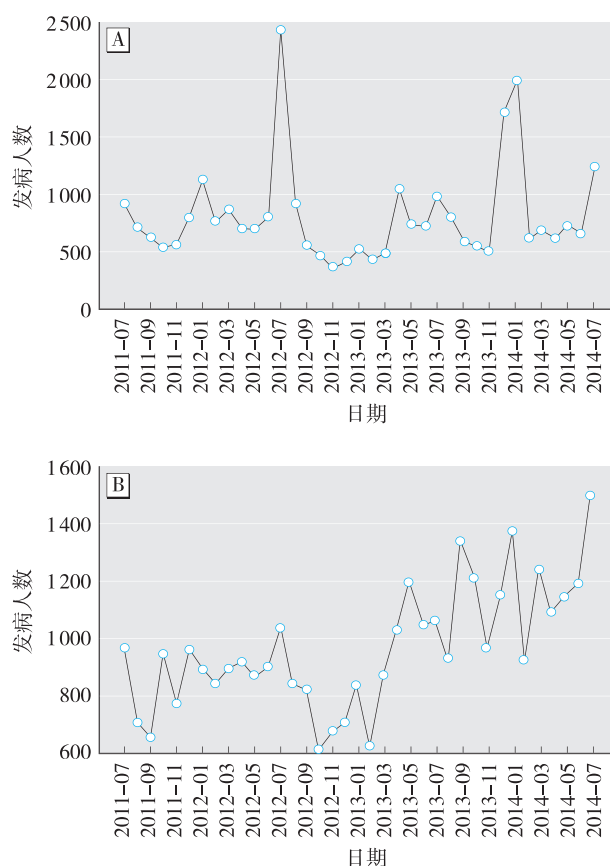
1.3 统计学分析

导出数据,使用Excel 2007建立数据库,SPSS 17.0进行模型构建、诊断和预测^[2]。

2 结果

2.1 ILI发病趋势分析

监测数据显示,2011年7月—2014年7月静安区成人ILI发病趋势均具有较为明显的季节性,季节性波动周期约为12个月,数据不平稳;儿童ILI发病趋势则未见明显的季节性。成人和儿童的波动趋势和幅度均存在较大差异。见图1。



[注]A:成人;B:儿童。

图1 2011年7月—2014年7月静安区ILI发病时间序列分布

2.2 模型建立

2.2.1 模型识别 对成人和儿童的ILI监测数据分别进行模型识别。儿童数据未见明显季节性,数据不平稳、波动明显,考虑进行一阶差分,结合对原始数据的对数变换,减少原始数据的方差波动,使时间序列平稳,直接构建ARIMA模型,即 $\text{ARIMA}(0, 1, 1)$,经与其他多个模型比较, $\text{ARIMA}(0, 1, 1)$ 的正态化BIC值最小,为10.456,识别为最佳模型。成人数据季节性较明显,首先对原始数据进行对数变换减少方差波动后,再进行一次季节性差分,周期长度12个月。经过数据预处理后,根据自相关和偏自相关函数结果提示,模型应为1的自回归过程,故首先构建季节性ARIMA模

型,初步选择模型ARIMA(1,0,0)(1,1,0)₁₂。经与其他多个模型比较,ARIMA(1,0,0)(1,1,0)₁₂的正态化BIC值最小,为12.637,故识别为最佳模型。

2.2.2 模型参数估计 对成人ILI监测数据的ARIMA(1,0,0)(1,1,0)₁₂模型计算结果表明,模型的两个自回归系数,包括季节性和非季节性自回归系数,均有统计学意义。对儿童数据的ARIMA模型计算结果表明,自回归系数有统计学意义。见表1。

表1 2011年7月—2014年7月静安区每月ILI发病数时间序列分析模型构建系数及其相关统计量

模型	常数项	标准误	T-Ratio	P
成人监测点, ARIMA(1,0,0)(1,1,0) ₁₂				
ARI	0.495	0.191	2.592	0.017
SARI	-0.758	0.166	-4.587	<0.001
儿童监测点, ARIMA(0,1,1)				
ARI	0.628	0.142	4.428	0.001

2.2.3 模型诊断试验 成人和儿童ILI预测模型构建后,分别对两个监测数据构建的模型进行白噪声残差分析。成人模型残差序列自相关函数的Box-Ljung Q统计量最小值为17.42, $P>0.1$,残差为白噪声序列,ARIMA(1,0,0)(1,1,0)₁₂可用于成人ILI发病的预测。儿童监测数据的残差序列自相关函数的Box-Ljung Q统计量最小值为10.456, $P>0.1$,残差也为白噪声序列,故ARIMA(0,1,1)也可用于儿童ILI发病预测。

2.2.4 模型的预测 利用2011年7月—2014年7月监测数据建立的模型对2014年8月—2014年12月静安区ILI发病情况进行预测,预测值与实际值较为接近,见表2。

表2 2014年8月—12月静安区每月ILI发病人数模型预测值与实际值对比

时间	实际病例数	预测病例数	95%下限	95%上限
成人监测点				
2014-08	1219	754	364	1404
2014-09	703	533	235	1060
2014-10	507	496	214	1002
2014-11	435	408	175	828
2014-12	689	615	263	1250
儿童监测点				
2014-08	1037	1334	923	1870
2014-09	938	1354	913	1940
2014-10	1047	1375	905	2010
2014-11	1086	1396	898	2080
2014-12	1169	1418	892	2150

3 讨论

传染病发病通常受到错综复杂的因素影响,很难识别出特定的发病规律,这使得预测传染病的发病难以通过结构式的因果模型实现。时间序列分析方法是分析预测对象的历史数据随时间的变化规律,建立数学模型,并将模型外推,预测未来时间疾病的发生情况,近年来广泛用于公共卫生和传染病相关的医学研究领域。ARIMA模型是时间序列分析方法中一种较为常见的模型,预测精度较高,对样本容量和概率分布没有特别严格的要求,可以用于较为复杂的,难以识别典型特征的数据。近年来,许多研究尝试将ARIMA模型应用于传染病预测预警领域,取得了较为理想的效果^[3-8]。

本研究对静安区的儿童监测点ILI病例数据和成人监测点的ILI数据分别尝试建立ARIMA模型。监测数据显示儿童和成人的ILI发病趋势存在明显不同,所适用的模型存在很大差别。儿童ILI数据经过一阶差分后的ARIMA(0,1,1)模型即达到较好的效果;成人ILI数据则需要季节性差分后建立ARIMA(1,0,0)(1,1,0)₁₂,才达到较好的效果。这进一步提示,ARIMA模型可应用于ILI预警,但由于不同人群的流行规律存在不同,在建模过程中,应考虑对不同人群分别构建,才能更为准确地预测ILI发病趋势,降低误差。

时间序列分析方法更适用于研究跨度时间较长及重复性测量的资料。但本研究所使用的数据仅为2011年7月—2014年7月的ILI监测数据。这主要是考虑到在2011年7月以前,静安区儿童ILI监测方法与现在不同,医疗机构未纳入急诊数据。为确保监测方法的一致性和监测数据的稳定性,本研究选用2011年7月以后的监测数据进行分析。用于模型构建的历史数据仅为3年,在一定程度上影响了模型的精确度,这可能也是模型预测可信区间范围较宽的原因之一。随着监测的继续开展,在未来研究中,可以尝试使用更长时间的监测数据来构建模型。

总体上,针对静安区儿童ILI和成人ILI构建的ARIMA模型都起到了较好的预测效果,2014年8—12月实际发病数据均在预测数据的区间内。但是实际上,由于影响传染病发病的因素众多,常常超出人们的预计范围,即使十分精确的数学模型也难以完全准确地预计传染病的发病。我们可以尝试将预测数据的95%上限作为预警线,用于常规工作中的预警,一旦

监测报告的实际ILI发病数据超出这一限值,则提示有其他不同于常规的因素出现,使得疾病的发病高于其常规状态,应启动相应调查来确认是否出现了突发流感疫情或不明原因疾病的流行;同时根据实际工作效果,进一步调整优化预警限制。对于ILI监测数据,也希望在进一步的研究中,尝试探索适合ILI就诊数据的预测预警模型,以寻找最为适用的预警阈值,指导实际的流感和流感样病例防控工作。

·作者声明本文无实际或潜在的利益冲突。

参考文献

- [1]何书元.应用时间序列分析[M].北京:北京大学出版社,2004:185-225.
- [2]张文彤.SPSS 11 统计分析教程(高级篇)[M].北京:北京希望电子出版社,2002:250-289.
- [3]李廷杰,陈秀山,李燕芬,等.应用时间序列统计方法分析广东省1984—1993年乙型肝炎季节性分布特性[J].中华流行病学杂志,1998,19(2):103-106.
- [4]李来英,张小平.起伏型时间序列分析方法在流行性出血热预测中的应用[J].中国卫生统计,1997,14(3):64.
- [5]赛晓勇,张治英,徐德忠,等.不同时间序列分析法在洞庭湖区血吸虫病发病预测中的比较[J].中华流行病学杂志,2004,25(10):863-866.
- [6]吴家兵,叶临湘,尤尔科,等.时间序列模型在传染病发病率预测中的应用[J].中国卫生统计,2006,23(3):276-276.
- [7]李燕婷,张宏伟,任宏,等.上海市流感样病例发病趋势的时间序列分析和预测模型研究[J].中华预防医学杂志,2007,41(6):496-498.
- [8]彭志行,陶红,贾成梅,等.时间序列分析在麻疹疫情预测预警中的应用研究[J].中国卫生统计,2010,27(5):459-463.

(收稿日期:2015-09-21)

(英文编辑:汪源;编辑:王晓宇;校对:汪源)

【告知栏】

新书介绍:《现代环境毒理学》

《现代环境毒理学》是由孟紫强教授主编,北京大学、复旦大学等10所著名大学21位本专业顶尖专家参编的大型环境毒理学专著,该书被评为《“十二五”国家重点图书出版规划项目》,已于2015年12月由中国环境出版社隆重出版。

本书共四篇三十七章,计150万字,力求从各个领域全面介绍现代环境毒理学基础理论、研究方法和实际应用技术。第一篇(第一章至第六章)主要介绍现代环境毒理学的基础理论和学科历史,对环境污染危害健康的基本规律、环境化学物毒性作用和致癌变、致畸变、致突变的基础理论和最新发展进行了论述。第二篇(第七章至第十五章)论述了9个主要环境毒理学分支学科的内容和最新进展,包括大气环境毒理学、水环境毒理学、土壤环境毒理学、工业环境毒理学、室内环境毒理学、纳米环境毒理学、食品环境毒理学、生物环境毒理学和现代环境分子毒理学等,极大拓宽了本学科的研究和应用领域。第三篇(第十六章至第三十章)对16类环境污染物(或因素)的毒性及其防护进行阐述,包括大气颗粒物、大气气体污染物、重金属、微量元素、农药与肥料、环境致癌物、持久性有机污染物、内分泌干扰物、石油、有机溶剂、电离辐射、电磁辐射、光污染、噪声污染以及环境生物污染等。第四篇(第三十一章至第三十七章)介绍环境毒理学主要研究方法和实际应用技术。包括对环境化学物一般毒性和特殊毒性的测试方法、大气细颗粒物(PM_{2.5})毒性研究方法、环境流行病学实用方法与技术、环境化学物的安全性和健康危险度评价方法、环境化学污染突发事件应急处理方法以及环境评价方面的环境影响医学评价与环境污染健康影响评价方法等。

本书可供从事环境保护、环境科学与工程、环境医学与毒理学、公共卫生与预防医学以及地方病学等方面教学、科研、管理的人员阅读参考,也可作为相应专业研究生教材或参考书,还可供进行生命科学、农业科学、物理学、化学、地质地理学、宇宙科学、气象学等与健康科学交叉研究的人员学习参考。