

Bootstrap在流感监测预警中的应用

戴启刚, 许可, 黄昊顿, 祁 贤, 王慎骄, 余慧燕, 邓 斐, 霍 翔*

(江苏省疾病预防控制中心急性传染病防制所, 江苏 南京 210009)

[摘要] 目的:应用bootstrap方法估计流感样病例(influenza-like illnesses, ILI)发病水平95%置信区间(95%CI)作为预警阈值,评价该方法在ILI预警中的价值。方法:以2005—2014年(除2006、2009年)的周监测数据为基线,采用历史2年流感病毒最高阳性检出率的30%作为流感流行标准,确定2006年和2009年流感流行期金标准,通过回顾性分析,比较bootstrap法、移动平均法和金标流行期的一致性;并以2012—2015年的周监测数据为基线,探讨bootstrap法在实际预警工作中的应用。结果:江苏省ILI流行曲线有2个高峰,冬春季高峰分布于11月—次年2月,夏季高峰分布于6—8月;bootstrap法、移动平均法和金标流行期一致性均较好;与移动平均法相比,bootstrap法对2009年初的流感流行相对更敏感,与金标流行期的一致性相对更高($\kappa=0.728$)。采用bootstrap法对2016年ILI%开展预警,与金标准结果接近。结论:bootstrap法作为一种不依赖于ILI数据分布的计算密集型方法,可作为传统移动平均控制图预警方法的替代和补充,用于流感流行的监测和预警。

[关键词] 流感样病例;bootstrap;移动平均法;预警

[中图分类号] R181.2

[文献标志码] A

[文章编号] 1007-4368(2017)10-1265-04

doi:10.7655/NYDXBNS20171008

Application of bootstrap method in the early warning for influenza surveillance

Dai Qigang, Xu Ke, Huang Haodi, Qi Xian, Wang Shenjiao, Yu Huiyan, Deng Fei, Huo Xiang*

(Jiangsu Provincial Center for Disease Control and Prevention, Nanjing 210009, China)

[Abstract] **Objective:** To investigate the application of bootstrap method in the early warning for influenza surveillance with 95% CIs of weekly influenza-like illnesses (ILI)% estimated by bootstrap method specified as the early warning thresholds. **Methods:** The influenza surveillance data from 2005 to 2014 (not including 2006 and 2009) were taken as baseline. Thirty Percent of the highest positive rate of influenza viruses in two years was taken as the gold standard of influenza pandemic. Consistency of the early warning results was compared among bootstrap method, moving average method and the gold standard. We also discussed the application of bootstrap in practice in 2016 based on the surveillance data from 2012 to 2015. **Results:** Influenza viruses circulated throughout the year and peaked during cold winter and spring season from November to following February and hot summer from June to August. The result of bootstrap method has high consistency with moving average method and gold standard. The bootstrap method was more sensitive in the early warning of influenza pandemic in 2009 compared with moving average method and more consistent with gold standard($\kappa=0.728$). **Conclusion:** Bootstrap was a compute-intensive method which was not rely on the distribution of ILI% data, and it can be applied as an alternative and complementary method in the early warning for influenza surveillance.

[Key words] influenza-like cases; bootstrap; moving average method; early warning

[Acta Univ Med Nanjing, 2017, 37(10):1265-1268]

流行性感胃是由流感病毒引起的急性发热性呼吸道传染病,流感病毒不断变异,世界各地不断

有流感的散发流行和暴发,一旦有新毒株出现,流行可能迅速波及全球。因此,对流感流行进行监测和预警,已经得到了全球许多国家的重视。传统的疾病监测多为被动监测,依靠病例的临床或实验室诊断结果、描述疾病分布特征、确定危险因素等,是事件发生后的分析,通常难以做到早期预测预警^[1]。而通过对流感症状(发热、咽痛等特定症候群)进行非特异性监测,观察分析监测数据的异常波动,能

[基金项目] 国家卫生计生委科研基金(W201303);重大新发传染病综合防控科技示范工程(BE2015714);江苏省自然科学基金(BK20151595);江苏省青年医学人才(QN-RC2016539);江苏省临床医学科技专项(BL2014081)

*通信作者(Corresponding author),E-mail:huox@foxmail.com

够先于传统的以医生确诊、实验室诊断等为基础的监测系统,有效地早期识别流感暴发的迹象,从而为进一步的疾病病原诊断和防控提供科学依据^[2]。自2005年以来,江苏省逐步建立了完善的流感监测体系,特别是近年来大部分哨点医院完成了基于医院信息系统(hospital information system, HIS)的流感样症状监测的改造,使得流感样病例(influenza-like illnesses, ILI)监测数据的实时性和准确性得到了进一步的提升。本文以2005—2014年的数据为基础,分析本省ILI的流行特征,并基于该历史数据,采用控制图法构建预警线,预警线算法分别采用移动百分位数法P80和bootstrap法上侧95%可信区间(95%CI),探讨2种不同方法在ILI预警应用中的价值。

1 对象和方法

1.1 对象

发热(体温 $\geq 38\text{ }^{\circ}\text{C}$)伴咳嗽或咽痛之一,而缺乏其他的实验室确定诊断依据的病例,即为ILI;ILI%即ILI占同期门急诊就诊总人数的比例。

截至2014年底,江苏省共设立29个国家级流感监测哨点医院,4个省级流感监测哨点医院,分布于全省13个省辖市,医院类型主要为综合性医院和儿童医院。通过HIS系统,每日收集医院内科/儿科门诊、发热门诊和急诊就诊的ILI人次及当日就诊的总人次,计算ILI%,按周汇总并通过中国流感监测信息系统进行信息报告。本文主要利用2005—2014年的ILI%监测数据结果,以历年监测的周ILI%数据为基线,以流感流行年的ILI%为测试对象,分析不同控制图预警算法在ILI早期预警中的实际意义。

1.2 方法

流感的发病一般表现为散发,发病率按照既往的变化规律如季节性、周期性等发生变化^[3]。经典流行病学中,流行的定义是指发病率显著超过历年散发发病率的水平^[4],该定义未从定量的角度精确阐述流行的概念,有部分学者采用“ $\bar{x}+2s$ ”作为流行标准^[5],由于ILI%数据分布不服从正态分布,“ $\bar{x}+2s$ ”的流行标准亦不适用于ILI%的判断。本文参考Zheng、Cowling、杨筱婷等^[6-8]的研究,采用流感病毒实验室检出率作为金标准;以历史2年的核酸阳性率高值中位数的30%作为流感流行金标准。

以2005—2014年(除2006、2009年)周监测数

据为基线,以2006和2009年的ILI%为例,比较不同控制图预警算法的一致性及其在ILI早期预警中的实际意义。一致性的度量采用 κ 系数,当 $\kappa > 0.75$ 时,可以认为一致性较好^[9]。

1.2.1 移动百分位数法

多数哨点医院ILI以周为单位报告,因此以1周作为1个观察周期,参比历年同期监测数据,并前后摆动2个周期,以P80为预警界值。

1.2.2 Bootstrap法

从基线ILI%数据中,等比例、有放回地重复抽样(在程序编写阶段,本文尝试抽样50、100、1 000、5 000次,考虑到计算结果的稳定性和计算机运算效率,最终将抽样次数定为1 000^[10]),计算每1个抽样样本的ILI%中位数,取1 000个中位数的 $Q_{2.5}$ 和 $Q_{97.5}$ 作为95%CI的上下限,以上侧95%CI作为周ILI%的预警线。

为探讨bootstrap方法在实际监测工作中的应用,本文以2012—2015年的数据为历史基线,评价bootstrap法在2016年ILI%监测预警中的效果。

2 结果

2005—2014年江苏省周ILI%波动于2.78%~7.89%(不含2009年流感大流行),中位数为4.20%,ILI每年有2个高峰,冬春季高峰分布于11月—次年2月,夏季高峰分布于6—8月。根据历史基线2年的核酸阳性率高值中位数的30%,判定江苏省2006年第9~16周和第24~34周,2009年第32~52周为流感流行,设为本研究的金标流行期(图1)。

比较bootstrap法、移动平均法和上述金标流行期的一致性,由图2可见,bootstrap法和移动平均法确定的阈值较接近,流感流行期阈值约5%,非流行期阈值约4.5%;由于利用了历年前后2周的数据,移动平均法求得的阈值曲线较bootstrap法更为平滑;两种方法判定的2006年和2009年流感流行期基本一致。

3种方法两两对比的结果:2006年bootstrap法和移动平均法确定的流行期完全一致($\kappa=1.000, P<0.001$),2009年的52周中有6周不一致($\kappa=0.768, P<0.001$);bootstrap法确定的流行期和金标流行期相比,2006年有11周不一致($\kappa=0.577, P<0.001$),2009年有8周不一致($\kappa=0.699, P<0.001$);移动平均法确定的流行期和金标流行期相比,2006年有11周不一致($\kappa=0.577, P<0.001$),2009年有6周不一致($\kappa=0.771, P<0.001$),详见表1。

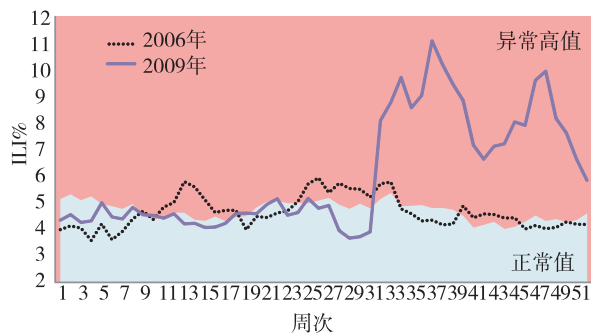


图 1 Bootstrap 控制图法 95% 预警阈值和 2006、2009 年 ILI% 流行曲线

Figure 1 The 95% CIs of weekly ILI% estimated by bootstrap method and the epidemic curve of ILI% in 2006 and 2009

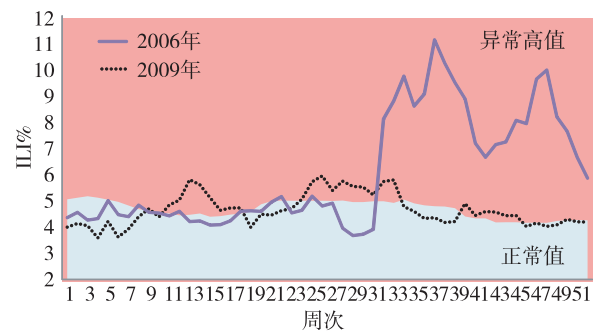


图 2 移动平均控制图法 95% 预警阈值和 2006、2009 年 ILI% 流行曲线

Figure 2 The 95% CIs of weekly ILI% estimated by moving average method and the epidemic curve of ILI% in 2006 and 2009

表 1 Bootstrap 法、移动平均法、金标流行期的一致性

Table 1 The consistency of bootstrap method, moving average method and gold standard

Bootstrap 法	2006 年移动平均法确定流行期				2009 年移动平均法确定流行期			
	非流行	流行	κ 值	P 值	非流行	流行	κ 值	P 值
非流行	26	0	1.000	<0.001	21	2	0.768	<0.001
流行	0	26			4	25		

金标流行	2006 年 bootstrap 法确定流行期				2009 年 bootstrap 法确定流行期			
	非流行	流行	κ 值	P 值	非流行	流行	κ 值	P 值
非流行	24	9	0.577	<0.001	20	4	0.728	<0.001
流行	2	17			3	25		

金标流行	2006 年移动平均法确定流行期				2009 年移动平均法确定流行期			
	非流行	流行	κ 值	P 值	非流行	流行	κ 值	P 值
非流行	24	9	0.577	<0.001	19	5	0.576	<0.001
流行	2	17			6	22		

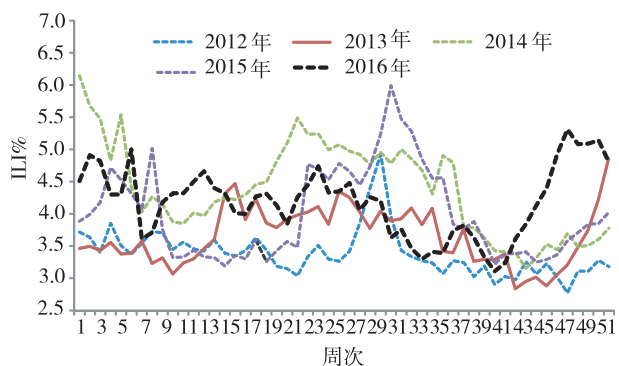


图 3 2012—2016 年分周 ILI% 流行曲线

Figure 3 The weekly epidemic curve of ILI% from 2012 to 2016

与历史基线数据相比,2016 年 ILI% 夏季高峰略高,但未超出历史基线水平,而冬春季高峰从 40 周开始上升,高峰开始时间与历年相比大幅提前(图 3)。通过 bootstrap 法预警结果可见,44 周 ILI% 超出预警阈值(图 4)。按 2014—2015 年最高核酸阳性率 30% 作为金标准阈值,45 周达到流感流行。

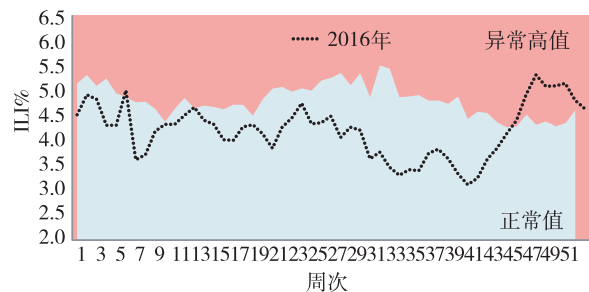


图 4 Bootstrap 控制图法 95% 预警阈值和 2016 年 ILI% 流行曲线

Figure 4 The 95% CIs of weekly ILI% estimated by bootstrap method and the epidemic curve of ILI% in 2016

3 讨论

对 ILI 的监测和早期预警,设定合理的预警阈值,可有效早期识别流感流行或暴发的迹象,有利于提前做好应对流感流行的物资,提高疾病控制工作的运行效率。本文采用 bootstrap 再抽样技术,以

bootstrap 样本的上侧 95%CI 为流感流行的预警阈值,并与传统预警流行的方法相比较,判定 bootstrap 法应用在流感流行预警中的实际意义。

江苏省于 2005 年以后,逐步建立和完善流感监测网络,近年大部分流感哨点医院报告采用 HIS 系统使得 ILI%更加趋于规范和稳定,本研究将 2005—2014 年的 ILI%纳入分析,由于流感发病有明显的季节性和周期性,流行程度受到气候、循环优势毒株型别、流感疫苗接种水平等诸多因素的影响,反映到 ILI%数据上,则表现为数据分布不呈正态分布,不同时间点上历年 ILI%的离散程度不一致。基于上述原因,传统的“ $\bar{x} \pm 2s$ ”方法并不适用于 ILI%的数据。本文结合 Zheng, Cowling、杨筱婷等^[6-8]的研究,采用历史 2 年流感病毒最高阳性检出率的 30% 作为流感流行标准,确定 2006 年和 2009 年流感流行期金标准。Bootstrap 再抽样技术作为一种计算密集型统计方法,是以原始现有样本资料为基础,借助经验分布理论进行有放回的重复抽样,进而估计统计量的分布特征,该方法适用于本研究中 ILI%数据总体分布未明确的情况。为判定 bootstrap 法的实际应用意义,本研究将其与金标准以及同属非参数法的移动平均法相比较。

使用 bootstrap 法上侧 95%CI 作为阈值和移动平均法 P80 比较,两种方法确定的 2006 年流行期完全一致, κ 值为 1;2009 年初(5~12 周)实际观测到的 ILI%与两种方法的阈值均比较接近,由于两种方法的阈值有细微差别,造成共计 6 周预警信号的不一致,但是整体 κ 值达 0.768,说明两种方法的一致性较好。与移动平均法相比,bootstrap 法对 2009 年初的流感流行相对更敏感,与金标流行期的一致性相对更高($\kappa=0.728$)。为检验 bootstrap 法预警的外推性和在实际监测工作中的使用价值,本文基于历史数据,采用 bootstrap 法对 2016 年的 ILI%数据开展预警工作,观测到的 ILI%数据于第 44 周超出 bootstrap 法的预警阈值,与流感核酸检出金标准的流行周次(45 周)较为接近,提示采用 bootstrap 法开

展 ILI%预警具有较高的敏感性。

综上所述,bootstrap 法作为一种不依赖于 ILI 数据的分布的计算密集型方法,与主流的移动平均法相比,预警信号的一致性较好,且敏感度略优于移动平均的方法,可作为传统移动平均控制图预警方法的替代和补充,用于流感流行的监测和预警。

[参考文献]

- [1] 宁芳,段玮,高培,等. 流感样病例监测系统数据异常波动与预警分析[J]. 中国公共卫生, 2007, 23(10): 1210-1211
- [2] 霍翔,李亮,祖荣强,等. 流感样病例监测在甲型 H₁N₁ 流感大流行中的早期预警效果分析[J]. 江苏预防医学, 2010, 21(1): 8-11
- [3] 史继新,张文增,冀国强,等. Arima 模型在流感样病例预测预警中的应用[J]. 首都公共卫生, 2010, 4(1): 12-16
- [4] 齐秀英. 流行病学 [M]. 8 版. 北京: 人民卫生出版社, 2013: 15
- [5] 杨维中,兰亚佳,李中杰,等. 国家传染病自动预警系统的设计与应用 [J]. 中华流行病学杂志, 2010, 31(11): 1240-1244
- [6] Zheng W, Aitken R, Muscatello DJ, et al. Potential for early warning of viral influenza activity in the community by monitoring clinical diagnoses of influenza in hospital emergency departments[J]. BMC Public Health, 2007, 7: 250
- [7] Cowling BJ, Wong IO, Ho LM, et al. Methods for monitoring influenza surveillance data [J]. Int J Epidemiol, 2006, 35(5): 1314-1321
- [8] 杨筱婷,刘新风,孟蕾,等. 甘肃省流感流行预警方法探研[J]. 中华流行病学杂志, 2016, 37(3): 430-433
- [9] 王洁贞,赵跃进. Kappa 统计量及其应用[J]. 中国卫生统计, 1995, 12(6): 49-50
- [10] Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians[J]. Stat Med, 2000, 19(9): 1141-1164

[收稿日期] 2016-11-17

本刊邮发代号 28-61

网址: <http://jnmu.njmu.edu.cn>