

* 预警预测 *

流行性感冒预警方法及其应用

黄智峰¹, 刘晓剑², 杨连朋³, 邹宇华¹, 吴永胜²

摘要： 流行性感冒(流感)是由于人类感染流感病毒而发生的一种急性呼吸道传染病,具有发病率高和迅速蔓延的特点。因此,运用预警方法识别早期流感疫情的异常升高以及流行迹象,使公共卫生部门尽早采取预防措施减少流感暴发或流行显得尤为重要。本研究就近年来常用流感预警方法进行归纳,综述各类预警方法基本原理、优缺点和应用状况,为流感预警方法的应用提供参考。

关键词： 流感; 预警; 预警技术

中图分类号: R18

文献标志码: A

文章编号: 1003-9961(2016)12-0989-06

Methods for influenza early warning and their application HUANG Zhi-feng¹, LIU Xiao-jian², YANG Lian-peng³, ZOU Yu-hua¹, WU Yong-sheng². 1. School of Public Health, Guangdong Pharmaceutical University, Guangzhou 510310, Guangdong, China; 2. Shenzhen Center for Disease Control and Prevention, Shenzhen 518000, Guangdong, China; 3. School of Public Health, Shantou University Medical College, Shantou 515000, Guangdong, China

Corresponding author: WU Yong-sheng, Email: cdc@szcdc.net

Abstract: Influenza is an acute respiratory infectious disease caused by human infection of influenza virus with the characteristics of high morbidity and rapid spread. Therefore, it is important to conduct early warning to timely detect the outbreak of influenza. This paper summarizes the ultimate principles, merits/demerits and application of several common influenza early warning methods for the purpose of providing reference for the influenza early warning.

Key words: Influenza; Early warning; Early warning method

This study was supported by the fund for National High Technology Research Development Program of China (No. 2015AA020109)

流行性感冒(流感)是人类感染流感病毒而发生的一种急性呼吸道传染病,能够通过空气、飞沫或与被污染的物品接触进行传播,具有发病率高和迅速蔓延的特点。20 世纪以来,全球性的流感大流行最少发生过 4 次,分别为 1918 年的“西班牙流感”,1957 年的“亚洲流感”,1968 年的“香港流感”以及 2009 年 4 月始发于墨西哥的甲型 H1N1 流感^[1],均给人类带来巨大的灾难。据世界卫生组织(WHO)估计,每年全球有 300 万~500 万严重病例与流感相关,致使 25 万~50 万人的超额死亡^[2]。因此,流感已成为严重的全球公共卫生问题,迫切需要早期预警并采取有效措施进行预防。如何探索有效的预警方法早期识别流感疫情的异常升高

以及流行迹象,使公共卫生部门尽早采取预防措施,并有效降低流感所带来的经济损失和健康损害成为研究者面临的首要问题。

1 预警的含义

预警是指在目标事件发生之前,提前发出警报,是在缺乏确定的因果关系或充分的数量-反应关系证据的情况下,促进调整预防行为并提出相关应急措施建议的一种方法^[3]。随着计算机和软件技术的发展,一些过去因条件限制难以实现的技术方法开始得到广泛应用,预警方法取得较快发展。但从当前情况看,每种预警方法都有其优越性和局限性。本研究对近年来常用流感预警方法进行归纳分析,综述各方法的特点和应用情况,为流感预警方法的运用提供参考。

2 预警方法的分类

2.1 基于时间维度的预警方法

2.1.1 基于控制图的预警方法 简单控制图法由 Shewhart^[4]于 1924 年首先提出,最早用于质量控制。其基本原理是利用历史监测资料,计算出监测

基金项目:国家高技术研究发展计划(No. 2015AA020109)

作者单位: 1. 广东药学院公共卫生学院,广东 广州 510310; 2. 深圳市疾病预防控制中心,广东 深圳 518000; 3. 汕头大学公共卫生学院,广东 汕头 515000

作者简介:黄智峰,男,广东省肇庆市人,硕士在读,主要从事流行病与卫生统计学工作

通信作者:吴永胜, Tel: 0755-25621102, Email: cdc@szcdc.net

收稿日期:2016-05-19

指标中的均值和标准差,以均值加减若干倍标准差作为预警参考值范围,但根据专业意义通常只取参考值上限为预警上限。该方法是所有类型控制图法的基础,其优点是方法简单,指标容易获取,在疾病监测中可作为一种比较有效的预警方法^[5]。随着控制图法在传染病预警中不断被完善,目前应用较为普遍的控制图有移动平均法(moving average method, MA)、指数权重移动平均法(exponentially weighted moving-average, EWMA)、累积和控制图法(cumulative sum, CUSUM),特别是后两种,已逐渐成为症状监测预警的标准算法^[6]。

EWMA 的基本思想是使用历史数据建立预警数据库,在使用数据时采用指数加权策略,对近期的数据赋予更高的权重,而对远期数据给予较低的权重,使模型能够更好地反映数据影响因素的变化趋势^[7]。当时间序列包含整体趋势和季节性变化时,可建立 Holt-Winters 指数平滑模型,在简单指数平滑模型中加入趋势参数 γ , 季节参数 δ 和趋势性衰减参数 π 。

Williamson 和 Hudson^[8] 认为与 Shewhart 控制图相比,采用 EWMA 进行传染病监测显得更为灵活。此法的优点在于模型考虑了过去资料和现在资料的相关性,其相关性会随着历史资料的久远而下降,呈指数平滑的下降关系。而缺点是需要对基线水平进行估计并确定阈值,不能很好地发现过程中的突发变化,可能会产生一定的假阳性^[9]。

应用案例:Steiner 等^[10] 收集经实验室确认的流感数据,运用 EWMA 能够探测出流感流行季节的开始和结束,同时也可以快速地探测出非流行期的流感暴发。王昕等^[11] 收集 2006—2008 年深圳市流感监测数据,运用 EWMA 拟合流感样病例(influenza-like illness, ILI)的变化趋势,并对 2009 年 ILI% 进行预测,提出了模型的中短期预测效果较好的结论。

CUSUM 是为了提高控制图法的灵敏性而被提出的。其基本原理是通过对观察值与根据历史数据计算出来的期望值之间差值的累积,将过程中的小偏移累加起来,达到放大的效果。该方法能灵敏地探测到数据异常的波动。CUSUM 的计算公式: $S_t = \max[0, S_{t-1} + (x_t - \bar{x})/s - k]$ 。初始值 $S_0 = 0$, t 表示当前时间, $t-1$ 为前一个时间单位, x_t 表示当前的观察发病数, $\bar{x}$ 和 s 表示历史同期发病数的均值和标准差, k 表示最少偏移量, H 为预警阈值。若 $S_t < H$, 表示数据波动正常, S_t 将保持累加,若 $S_t > H$, 即提示有暴发出现,可发出预警信号。当预警信号

发出之后, S_t 需要进行归零重新开始累加。

CUSUM 的优点是只需一年的历史基线资料便能进行预警预测,其次是灵敏性高。所以比较适用于暴发的实时监测,比如流感。然而,它同样受报告的效率和报告过程中的人为因素影响,因此将该应用于监测资料可能缺乏稳健性。

应用案例:1982 年 Tillett 和 Spencer^[12] 将 CUSUM 应用于英国流感大流行探测。而美国的异常事件早期报告系统(the early aberration reporting system, EARS)则设定了轻、中、重 3 个级别的 CUSUM 统计量,作为每日资料异常监测的参考依据^[13]。在国内,杨鹏等^[14] 应用 EARS 系统对 2007—2008 年北京市流感监测数据分析,并与季节性流感病原学监测结果进行比较,发现 CUSUM 模型能高效准确地对流感高峰进行预测和预警。

2.1.2 自回归积分滑动平均模型(auto regressive integrated moving average model, ARIMA) ARIMA 是以时间序列的方式反映社会现象发展过程以及发展规律,并进行延伸外推预测出发展趋势的方法,其前提是假定事物的过去延续到未来。这个假设前提包含两层含义:一是事件不会发生突然的跳跃变化,以相对小的步伐前进;二是过去和当前的现象可能表明现在和将来活动的发展变化趋势。ARIMA 的建模过程一般分 4 步:①序列平稳化:收集历史资料并编成时间序列,分析数据是否平稳,并使不平稳序列通过差分或季节性差分等方法转为平稳序列;②模型识别:利用 ACF 和 PACF 图能够初步确定 ARIMA(p, d, q)(sp, sd, sq) 模型中 p, d, q 和 sp, sd, sq;③模型的参数估计与适应性检验:通过参数的显著性检验,检验模型中未知参数是否有统计学意义,将没有意义的参数从模型中删除,最终得到最简约而又精确的模型;④预测应用:将建立的最优模型运用到流行病预测研究中。

ARIMA 优点是可以考虑时间对病例数的影响,同时可考虑不同的影响因素来预测病例数,并估计预测值置信区间。但 ARIMA 也有一些局限性:①建立一个标准的 ARIMA 需要长时间的基线资料,历史数据最少需要 1 年;②模型预测短期效果较好,但对于长期预测,模型预测的不确定性将会增大^[15];③模型对于小规模暴发可能出现探测失效,而对于大规模的暴发则探测良好^[16]。

应用案例:甘肃省疾病预防控制中心(CDC)孟蕾等^[17] 收集 2006—2010 年甘肃省哨点医院的流感监测数据,建立 ARIMA 对 2011 年 1—44 周 ILI 进

行预测, 研究结果表明模型拟合效果好, 预测结果与实际报告结果基本一致, 预测效果可靠。Soebiyanto 等^[18]利用中国台湾地区香港和马里科帕县(亚利桑那州)的流感病毒分离数据和气象数据分别建立多个 ARIMA, 研究表明纳入气候因素的 ARIMA 与依赖历史数据的 ARIMA 相比, 预测可靠性有所提高。

2.2 基于时间和空间维度的预警方法

2.2.1 时空扫描统计量 时空扫描统计量是一种基于广义似然比检验的数学模型, 通过综合描述在某个区域内病例报告数、报告病例持续时间长短以及发病范围大小等信息来分析病例的时空聚集性。时空扫描统计量的基本原理是同时考虑时间维度和空间维度两个因素建立时空扫描窗口。扫描窗口为柱状体, 扫描窗口的底对应一定的地理位置, 高对应一定的时间区间。随着科学技术和计算机技术的发展, 扫描窗口的形状由圆形最后演变为不规则形状^[19-21]。假定将研究区域分为若干个子区域, 扫描窗口沿着地理单位中心变动, 窗口每变更一次, 将计算窗口内与窗口外区域之间发病率的差异, 通过计算不同窗口下的对数似然比(log-likelihood ratio, LLR)来判断是否存在区域聚集, 当 LLR 的 P 值小于检验水准 $\alpha = 0.05$ 时, 可认为该区域存在聚集。

该模型的优点是突破了传统研究中时间分布和空间分布分离的缺点, 能够同时在时间维度和空间维度上对疾病的聚集进行精确定位, 并定量计算出发病风险, 对疫情暴发的防控工作具有一定意义。此外, 扫描窗口在整个探测过程中是动态变化的, 可以避免选择偏移。但是, 时空扫描统计量扫描窗口的半径和扫描时间长度参数设定缺乏客观标准, 在实际监测中设定最优扫描半径和扫描时间具有一定难度。

应用案例: Zhang 等^[22]收集 2009 年 10 月至 2010 年 3 月天津地区 2 个镇卫生院和 6 个村卫生室每日流感监测数据, 并利用圆形时空扫描统计量探测农村地区流感聚集现象。研究证实时空扫描统计量在农村地区进行流感监测是有可能的, 在公共卫生实践中具有一定应用价值。Costa 和 Kulldorff^[21]提出了两种不规则时空扫描统计量并优化了算法, 应用于新英格兰 3 个州的流感探测, 研究发现, 与传统的“圆柱形”时空扫描统计量相比, 新方法具有快速探测疾病暴发的潜力, 不规则时空扫描统计量解决了圆形扫描窗口不能探测不规则

形状聚集区域的问题, 但模型还需要在日后的研究中进行更多的评估试验。

2.2.2 贝叶斯时空模型 贝叶斯时空模型是在贝叶斯理论的框架下, 为分析时间和空间资料的一种基于概率的数学模型。其基本原理是通过构建贝叶斯时空模型, 根据经验和历史资料指定模型参数的先验分布, 利用观察过程中的数据不断更新参数的先验分布以得到后验分布, 最终基于后验分布作出统计推断。这一过程可以简单理解为利用“样本数据”对“先验信息”进行修正而得到“后验信息”。

贝叶斯时空模型的优势主要表现: ①将任何一个未知参数均看成随机变量, 利用先验分布对其进行描述, 能较好地解决复杂的时间和空间维度中的异质性问题, 揭示空间和时间属性; ②模型修正了经典统计方法在区域人口数极少或发病(或死亡)率极低时所暴露的弱点, 对发病(死亡)率或发病(死亡)危险的估计更加精确; ③能够把与流感有关的影响因素以随机过程的形式引入模型中, 使模型预测更符合实际。但建立该类模型也应注意: 确定参数先验分布范围在贝叶斯时空模型中非常重要, 关系到模型的可识别性问题。这里给出几个确定有先验信息分布时的注意事项: ①注意 0 值。根据贝叶斯定理, 后验分布正比于先验分布和似然函数的乘积。因此, 不管观察数据提供了什么信息, 任何先验分布为 0(或非常小)的参数区域, 其所对应的后验分布也将为 0(或非常小)。②注意认识偏差。因为存在认知偏差, 由专家提取的分布可能会过于狭窄, 无法表达不确定性。除此之外, 由专家提取的分布可能存在系统偏差, 需要分析者进行弥补。③确保与问题相关的论据用于确定先验分布。④确定参数的先验分布时要注意, 有些时候均值在高偏度分布的情形下可能并不具有代表性, 需要利用中位数代替均值确定先验分布^[23]。

应用案例: Chan 等^[24]利用中国台湾台北市 5 所哨点医院 2005 年 9 月至 2008 年 2 月每日 ILI 监测数据, 并将数据分为训练集和测试集。在考虑了气象因素、空间交互作用以及周末和暑期效应等因素对流感活动影响的前提下, 以哨点医院为单位建立层次贝叶斯时空模型。模型以后验概率为 70% 作为预警界值, 并在测试阶段有效地探测出流感峰值, 证明了模型能够提前 1~2 d 探测出 ILI 上升。

2.3 基于回归的预警方法 广义相加模型

(generalized additive model, GAM) 又称广义加性模型, 作为广义线性模型的扩展, 在保留了广义线性模型特点的基础上, 拟合非参数回归, 可增加模型应用的灵活性, 适用于识别自变量和因变量间过于复杂的非线性关系。GAM 主要用于研究某一指标(因变量)与影响因素(自变量)之间的依赖关系, 其预测应用的基本思想是从一组数据出发, 计算出变量间的数学关系式, 确定各因素对因变量的影响程度。利用所求关系式, 根据已有影响因素资料求出未知的因变量, 从而达到预测目的。近年来, 使用 GAM 结合环境因素对传染病进行预警预测得到相对广泛的应用。

该方法的优点为适用于多种分布资料的分析, 模型的构建非常灵活, 并不拘泥于某一种形式的函数。GAM 的直观性较好, 可用效应图识别出变量间非线性关系。此外, 模型还可在控制复杂的混杂作用后, 正确评价自变量序列对因变量序列的影响。但模型应用于时间序列资料时要注意: ①由于大部分时间序列资料通常存在自相关性, 若在使用这一模型时由于残差的自相关性未能得到校正, 在分析结果时会存在偏差^[25]; ②在 GAM 中, 非参数平滑函数的自由度设置十分关键, 可对模型的参数估计和稳定性产生影响。因此, 选择适合的方式确定自由度具有重要意义^[26]。

应用案例: 陈健等^[27]利用 2006—2010 年上海市流感监测数据以及每周气象数据建立 GAM。模型在拟合 ILI 发病数和气象数据方面取得较好表现, 基本符合上海市 ILI 发病情况, 具有一定的预测能力。Soebiyanto 等^[28]对柏林、斯洛文尼亚、卡斯蒂利亚和里昂以及 6 个以色列地区立分别建立 GAM, 研究气象因素与流感活动的关系。证实在热带和亚热带地区, 比湿(specific humidity)对流感活动均有影响, 并利用模型预测出流感活动的开始和结束。

2.4 基于互联网搜索数据的预警方法 随着搜索引擎逐渐成为人们查询信息的主要渠道, 使利用互联网搜索数据对流感进行预警成为可能。由于搜索数据中包含着用户查询流感的相关信息, 通过对信息的挖掘与分析, 就可以了解到流感的流行状态。从目前来讲, 基于互联网搜索的新型监测数据可以作为传统监测数据的有益补充, 其原因: ①网络使用人口越来越多; ②群众利用网络进行健康相关知识搜索越来越频繁; ③互联网监测时效性高, 能迅速反映出疾病流行现况。作为美国互联网搜

索市场占有率最高的 Google 在流感监测工作方面也不甘落后, 2009 年推出了基于自身搜索数据的流感预警系统 Google Flu Trends(GFT), 引起了全球各国流行病监测部门的重视, 其发表在 *Nature* 的论文至今已被引用 2000 多次^[29]。

初期 GFT 的基本原理非常简单^[30]: 假设一个人患了流感, 那么他可能会通过互联网查询关于自身病情的信息, 通过追踪与流感相关的搜索关键词使用频率, 找出与流感流行季节高度相关的检索词并计算其检索频率, 就有可能估算出流感流行情况。

GFT 的优点主要是反应迅速, 依靠搜索数据的实时性, 预警系统给出的流感流行报告时间与美国 CDC 官方报告相比提前 7~14 d, 通过及时预警有利于公共卫生部门调拨应急物资, 制定应急方案, 降低疫情扩散的风险。而目前 GFT 还处于探索阶段, 其预测过程存在局限性: ①GFT 的搜索词是基于用户已有的认知进行归纳的, 当新型流感病毒出现时, 用户搜索的范围仍停留于过往的认识中, 这种情况可能会造成 GFT 预测的偏差; ②Google 作为一家以商业盈利运作的公司, 为满足其日常业务的需求, Google 搜索引擎的算法是会不断调整的, 而搜索算法的改变和用户的搜索行为将会影响 GFT 的预测结果^[31]; ③认为搜索数据可以完全替代传统监测数据, 而非作为后者的补充。

在新甲型流感暴发后, GFT 的预测问题开始相继出现, 但同时也经历着 2.0 和 3.0 的改进^[32-33]。2014 年 10 月 31 日, Google 的高级工程师在博客上宣布谷歌对 GFT 进行了新的升级^[34], 改良后的 GFT 纳入了美国 CDC 流感监测数据, 而不只是用监测数据校正模型和建立基线^[35]。尽管基于互联网搜索数据的预测方法目前遇到一些问题, 但随着科技进步以及对互联网数据的了解, 基于互联网搜索数据进行流感预警的尝试依然值得关注。

3 小结

本研究主要针对近年来的常用预警和热点技术进行归纳, 对其基本原理、优缺点和应用进行综述。总的来讲, 预警主要是以计算期望值的方式来预测流感流行趋势, 只是对于不同预警方法其计算方法有所不同。随着社会的发展和科学的进步, 应用于流感监测的资料类型将会逐渐丰富, 应用于流感的预警方法也会越来越多。在实际运用中, 应考虑研究目的、资料类型以及可行性等因素, 选择适合的预警方法, 才能取得更好的效果。

参考文献

- [1] Zeng XX, Li KS. The Pandemic history of influenza: review and revelation of the influenza pandemic in last century [J]. *Medicine and Society*, 2010, 23(11): 4–6. (in Chinese)
曾祥兴, 李康生. 流感百年: 20 世纪流感大流行的回顾与启示 [J]. *医学与社会*, 2010, 23(11): 4–6.
- [2] World Health Organization. Factsheet: influenza (seasonal) [EB/OL]. (2014–03) [2016–04–22]. <http://www.who.int/mediacentre/factsheets/fs211/en/>.
- [3] Hu P, Zhao SW, Zheng KQ, et al. Theory and application of public health emergency monitoring and early warning system [J]. *Occupation and Health*, 2004, 20(8): 5–7. (in Chinese)
胡培, 赵世文, 郑克勤, 等. 突发公共卫生事件监测与预警系统理论概念及应用 [J]. *职业与健康*, 2004, 20(8): 5–7.
- [4] Shewhart WA. Some applications of statistical methods to the analysis of physical and engineering data [J]. *Bell Labs Tech J*, 1924, 3(1): 43–87.
- [5] Yin ZY, Fang CF. Discussion on early warning methods of infectious diseases [J]. *Chinese Journal of Health Statistic*, 2010, 27(2): 218–220. (in Chinese)
尹志英, 方春福. 传染病预警预测方法探讨 [J]. *中国卫生统计*, 2010, 27(2): 218–220.
- [6] Yang WZ. Early warning of infectious disease theory and practice [M]. Beijing: *People's Medical Publishing House*, 2012: 57–58. (in Chinese)
杨维中. 传染病预警理论与实践 [M]. 北京: 人民卫生出版社, 2012: 57–58.
- [7] Hunter JS. The exponentially weighted moving average [J]. *J Qual Technol*, 1986, 18(4): 203–210.
- [8] Williamson GD, Hudson GW. A monitoring system for detecting aberrations in public health surveillance reports [J]. *Stat Med*, 1999, 18(23): 3283–3298.
- [9] Wang XL, Wang QY, Luan RS, et al. The main model of early warning of infectious diseases [J]. *Modern Preventive Medicine*, 2008, 35(22): 4339–4341. (in Chinese)
王小莉, 王全意, 栾荣生, 等. 传染病疫情早期预警的主要模型 [J]. *现代预防医学*, 2008, 35(22): 4339–4341.
- [10] Steiner SH, Grant K, Coory M, et al. Detecting the start of an influenza outbreak using exponentially weighted moving average charts [J]. *BMC Med Inform Decis Mak*, 2010, 10: 37.
- [11] Wang X, Cheng XW, Fang SS, et al. Application of exponential smoothing model in forecasting influenza incidence [J]. *China Tropical Medicine*, 2011, 11(8): 938–939. (in Chinese)
王昕, 程小雯, 房师松, 等. 指数平滑模型在流感样病例预测中的应用 [J]. *中国热带医学*, 2011, 11(8): 938–939.
- [12] Tillett HE, Spencer IL. Influenza surveillance in England and Wales using routine statistics: development of 'csum' graphs to compare 12 previous winters and to monitor the 1980/81 winter [J]. *J Hyg*, 1982, 88(1): 83–94.
- [13] Hutwagner L, Thompson W, Seeman GM, et al. The bioterrorism preparedness and response early aberration reporting system (EARS) [J]. *J Urban Health*, 2003, 80 Suppl 1: i89–96.
- [14] Yang P, Wang QY, Shi WX, et al. Pre-warning analysis of influenza surveillance in Beijing 2007–2008 [J]. *Chinese Journal of Disease Control and Prevention*, 2010, 14(2): 108–111. (in Chinese)
杨鹏, 王全意, 石伟先, 等. 北京市 2007–2008 年流感监测预警分析 [J]. *中华疾病控制杂志*, 2010, 14(2): 108–111.
- [15] Wang X, Cheng XW, Fang SS, et al. Prediction of influenza incidence by using ARIMA [J]. *China Tropical Medicine*, 2011, 11(6): 659–661. (in Chinese)
王昕, 程小雯, 房师松, 等. 用 ARIMA 模型预测流感的变化趋势 [J]. *中国热带医学*, 2011, 11(6): 659–661.
- [16] Reis BY, Mandl KD. Time series modeling for syndromic surveillance [J]. *BMC Med Inform Decis Mak*, 2003, 3(1): 2.
- [17] Meng L, Wang XH, Bai YN, et al. Prediction of influenza-like illness by using ARIMA in Gansu province [J]. *Chinese Journal of Public Health*, 2014, 30(2): 228–230. (in Chinese)
孟蕾, 王新华, 白亚娜, 等. 甘肃省哨点医院流感样病例 ARIMA 模型预测 [J]. *中国公共卫生*, 2014, 30(2): 228–230.
- [18] Soebiyanto RP, Adimi F, Kiang RK. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters [J]. *PLoS One*, 2010, 5(3): e9450.
- [19] Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic [J]. *J Roy Stat Soc*, 2001, 164(1): 61–72.
- [20] Takahashi K, Kulldorff M, Tango T, et al. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring [J]. *Int J Health Geogr*, 2008, 7(1): 14.
- [21] Costa MA, Kulldorff M. Maximum linkage space-time permutation scan statistics for disease outbreak detection [J]. *Int J Health Geogr*, 2014, 13(1): 20.
- [22] Zhang Y, Li L, Dong XC, et al. Influenza surveillance and incidence in a rural area in China during the 2009/2010 influenza pandemic [J]. *PLoS One*, 2014, 9(12): e115347.
- [23] Siu NO, Kelly DL. Bayesian parameter estimation in probabilistic risk assessment [J]. *Reliab Eng System Saf*, 1998, 62(1/2): 89–116.
- [24] Chan TC, King CC, Yen MY, et al. Probabilistic daily ILI syndromic surveillance with a spatio-temporal Bayesian hierarchical model [J]. *PLoS One*, 2010, 5(7): e11626.
- [25] Yu SL, Peng XW. A method for removing residual autocorrelation of time series in generalized additive models [J]. *Chinese Journal of Health Statistics*, 2010, 27(5): 450–454. (in Chinese)
余松林, 彭晓武. 广义加性模型配合时间序列资料时消除残差自相关性的一种方法 [J]. *中国卫生统计*, 2010, 27(5): 450–454.
- [26] Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality [J]. *J Roy Stat Soc*, 2006, 169(2): 179–203.
- [27] Chen J, Zhang L, Lu S, et al. Prediction of influenza-like illness in Shanghai based on the generalized additive method [J]. *Chinese Journal of Epidemiology*, 2013, 34(4): 404–408. (in Chinese)

Chinese)

陈健,张磊,陆帅,等. 采用广义可加模型分析预测上海市流感样病例发病情况[J]. 中华流行病学杂志,2013,34(4):404-408.

[28] Soebiyanto RP, Gross D, Jorgensen P, et al. Associations between meteorological parameters and influenza activity in Berlin (Germany), Ljubljana (Slovenia), Castile and León (Spain) and Israeli districts [J]. *PLoS One*, 2015, 10(8):e0134701.

[29] Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data[J]. *Nature*, 2009, 457(7232):1012-1014.

[30] Zou XH, Zhu WF, Yang L, et al. Google Flu Trends—the initial application of big data in public health [J]. *Chinese Journal of Preventive Medicine*, 2015, 49(6):581-584. (in Chinese)

邹晓辉,朱闻斐,杨磊,等. 谷歌流感预测——大数据在公共卫生领域的尝试[J]. 中华预防医学杂志,2015,49(6):581-584.

[31] Lazer D, Kennedy R, King G, et al. The parable of Google Flu: traps in big data analysis [J]. *Science*, 2014, 343(6176):1203-1205.

[32] Cook S, Conrad C, Fowlkes AL, et al. Assessing Google Flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic [J]. *PLoS One*, 2011, 6(8):e23610.

[33] Lazer D, Kennedy R, King G, et al. Google Flu Trends still appears sick: an evaluation of the 2013-2014 flu season [R]. SSRN 2408560, 2014.

[34] Stefansen C. Google Flu Trends gets a brand new engine [EB/OL]. (2014-11-03) [2016-04-25]. <http://googledata.org/google-research/google-flu-trends-gets-a-brand-new-engine/>.

[35] Harris D. In the name of accuracy, Google retools its flu trends model [EB/OL]. (2014-10-31) [2016-04-25]. <https://gigaom.com/2014/10/31/in-the-name-of-accuracy-google-retools-its-flu-trends-model/>.



黄智峰

ORCID:0000-0001-7853-473X

作者贡献:

黄智峰:文章主要撰写人,负责搜集文献并撰写文章

刘晓剑:修改文章指导者

杨连朋:课题组成员,参与讨论文章撰写

邹宇华:学校指导老师,负责论文审阅并提出修改建议

吴永胜:指导文章撰写全过程