

基于“大数据”的流感预警研究进展

付之鸥^{1,5} 鲍昌俊² 李中杰³ 王丽萍³ 李苑⁴ 冷寒冰⁵ 彭志行¹

¹南京医科大学公共卫生学院 211166; ²江苏省疾病预防控制中心, 南京 210009; ³中国疾病预防控制中心传染病预防控制处, 北京 102206; ⁴深圳市宝安区疾病预防控制中心 518101; ⁵哈尔滨市宾县卫生健康局 150400

通信作者: 彭志行, Email: zhihangpeng@njmu.edu.cn

【摘要】 传统流感预警体系在监测范围、预测准确性和实时性等方面仍存在不足。随着新一轮科技革命的兴起, 现代信息技术日趋成熟, 以大数据分析为基础的流感预警技术已经成为研究焦点。本文以传统流感监测预警体系作为参照, 从互联网、影响因素、时空趋势、风险评估4个角度总结基于大数据的流感预警研究进展, 综述大数据在流感预警中的优势、不足及未来发展的趋势。

【关键词】 大数据; 流感预警; 监测; 预测

基金项目: 国家自然科学基金(81673275); 国家传染病科技重大专项(2017ZX10201101-002-001, 2018ZX10715-002, 2018ZX10713001); 深圳市科技创新委员会(JCYJ20160427155352873); 江苏省优势学科建设项目

DOI: 10.3760/cma.j.cn112338-20190908-00657

Progress of research regarding the influenza early warning system, based on “Big Data”

Fu Zhiou^{1,5}, Bao Changjun², Li Zhongjie³, Wang Liping³, Li Yuan⁴, Leng Hanbing⁵, Peng Zhihang¹

¹School of Public Health, Nanjing Medical University, Nanjing 211166, China; ²Jiangsu Center for Disease Control and Prevention, Nanjing 210009, China; ³Division of Infectious Diseases, Chinese Center for Disease Control and Prevention, Beijing 102206, China; ⁴Bao'an District Center for Disease Control and Prevention, Shenzhen 518101, China; ⁵Binxian Municipal Health Commission, Haerbin 150400, China

Corresponding author: Peng Zhihang, Email: zhihangpeng@njmu.edu.cn

【Abstract】 Shortcomings have been inherited in the traditional influenza early warning system, often expressed through the scope, accuracy on prediction and real-time performance of the monitor related programs. With the new round of scientific and technological revolution and the increasingly maturity of modern information system, related technology on influenza early warning has become the focus of research in this field, based on big data analysis technology. Using the traditional influenza surveillance and early warning system as reference, this paper summarizes the progress of influenza early warning research, based on the Internet, influencing factors, time and space trends, and risk assessment etc., to summarize the trends on the advantages, shortcomings and future development of big data, used in the early warning system on influenza.

【Key words】 Big data; Influenza early warning; Surveillance; Forecasting

Fund programs: National Natural Science Foundation of China (81673275); Major National Science and Technology Program on Infectious Diseases (2017ZX10201101-002-001, 2018ZX10715-002, 2018ZX10713001); Shenzhen Science and Technology Innovation Committee (JCYJ20160427155352873); Priority Academic Program Development of Jiangsu Higher Education Institutions

DOI: 10.3760/cma.j.cn112338-20190908-00657

“大数据”是指数据规模、多样性和复杂性远远超出传统数据类型和数据分析模式, 需要新的架构、技术、算法分析处理, 并从中提取有价值的信息和隐藏知识的数据; “大数据”不仅指海量的数据, 还包括分析和解释这些海量数据的技术和能力^[1]。传统的数据分析方法是最具有代表性的数据获取事物尽可能多的信息, 以验证假设, 而“大数据”将世界数据化, 探寻数据流间的关联, 以预测为核心, 精准全面的认识事物发展变化的规律^[2]。近年来, 随着互联网以及多学科

多领域数据、知识的交互发展, 大数据技术日趋成熟, 与流感相关的海量数据和非结构化数据挖掘及分析技术已得到广泛应用, 基于“大数据”的流感预警在提升疫情的追踪、响应和预测预警能力方面成效显著^[3]。本文从大数据分析的角度, 对流感预警研究进展进行综述。

一、国内外传统流感监测预警体系

基于哨点医院和网络实验室的流感监测体系在西方国家发展应用较早。欧洲各国在自己的流感监测系统基础上

分别建立了由 WHO 管理的欧洲流感网络 (Euro-Flu) 和由欧盟 CDC 主导的跨国流感监测系统 (European Influenza Surveillance Network, EISN)^[4-5]。EISN 以国家为单位, 每周定性定量报告流感监测信息, 包括流感疫情级别、地理分布及变化趋势、哨点样本数、病毒类型以及每 10 万人的流感样病例 (influenza-like illness, ILI) 和急性呼吸道疾病 (acute respiratory illness) 人数等^[6], 为欧洲各国的流感预警防控提供政策及科学依据。美国的流感监测预警系统具有全国性和综合性, 由美国 CDC 和其他部门相互协作建立的, 可同时门诊患者、住院病例、病毒学、死亡率、流感地理分布概况等进行监测。通过对这些数据的监测来判断流感疫情的发展趋势, 同时开展综合预警, 为美国的流感预防和治疗提供信息服务和指导意义^[7]。我国于 1957 年成立了国家流感中心, 2000 年建立了与 WHO 合作的流感监测网络, 通过对监测的数据进行分析, 掌握流感的动态变化规律。2008 年, 中国 CDC 在全国范围内建立了传染病自动预警信息系统, 通过流感的历史病例数对未来流感的暴发进行预警, 2009 年该监测网络实现了对全国所有地 (市) 级地区的覆盖^[8]。国家流感监测系统以周报形式发布流感信息, 包括暴发疫情、ILI 报告和病原学监测, 并按南、北地区进行分析报告。随着国家流感监测系统的不断完善, 监测网络逐步稳定, 其已成为早期发现和控制流感疫情的重要手段。

基于传统流感监测体系, 国内外开展了多项关于流感预测预警技术的研究, 如利用医院就诊数据预测流感发生^[9-10], 对实验室确认病例、ILI 和传染病法定报告传染病信息系统数据在开发流感早期预警系统应用中的有效性进行比较分析^[11], 以及运用传统时间序列模型等方法对历史流感监测数据进行数据挖掘及预测预警^[12-13], 均得到良好的效果。

一直以来, 传统流感监测预警体系在各国的流感防控中起到了举足轻重的作用, 但随着全球经济一体化、信息化、气候变暖等进程的不断加剧, 传统流感监测预警体系已经不能适应如此复杂的变化。例如, 欧洲流感监测系统是在各国自选哨点的基础上建立的, 因各国流感监测项目不同, 导致缺乏统一的规范和整合^[8]。美国流感监测系统报告的不同地理区域的 ILI 活动差异不能准确反映 ILI 患病率的真实差异, 而这些差异可以反映不同地区系统的收集和聚集偏倚^[14]。我国的流感监测预警体系也存在着运行成本昂贵、监测网络覆盖率不高、信息上报效率较低、无法充分展示流感疫情的空间差异、过于依赖流感历史数据而缺乏多维度数据支持、数据挖掘及预测预警方法过于简单等问题^[15]。因此, 开发更加高效、准确的即时预警系统, 建立基于“大数据”流感预警体系, 进一步提升流感疫情的预测和预警能力任重道远、势在必行。

二、基于互联网的流感预警方法

随着互联网信息技术的快速发展和网络覆盖率的不断提高, 互联网“大数据”已经渗透到社会的各个行业、领域, 各种社交软件及搜索引擎早已成为人们发布和了解各种信息的主要渠道, 使得利用互联网“大数据”对流感进行预警成为

可能。与传统流感预警相比, 互联网“大数据”开创了不同于传统流感预警的全新模式^[16-17]: ①不是随机样本, 而是全体数据; ②不是精确性, 而是混杂性 (模糊与概率); ③不是因果关系, 而是相关关系; ④不是滞后预警, 而是即时预警; ⑤不是政府机构主导的点状局部数据采集, 而是全网络、开放式、自动化的数据捕捉; ⑥数据信息完全免费且预警系统运行相对高效。

2009 年 Google 首次尝试利用搜索引擎查询数据对流感趋势进行分析, 并设计了流感预测预警系统 (Google Flu Trends, GFT)^[16]。该系统初期取得了良好的预测效果, 比美国 CDC 的流感预警发布可提前 7~14 d, 但 2011 年的研究表明^[18], 传统 ILI 的哨点监测数据可以比 GFT 更准确的反映非季节性流感活动。2013 年研究认为, GFT 在预测某次流感暴发时, 比实际情况要夸大了几乎一倍^[19]。为提高 GFT 的预测准确性, Google 不断微调预测算法, 但每次修补又造成了新的误差^[20]。

2015 年, Davidson 等^[21]提出了结合“大数据”和传统监测改善流感预警的方法, 将传统监测数据的可靠性和准确性与 GFT 预测的普遍性和即时性相结合。研究人员分别使用贝叶斯模型、机器学习方法和动力学模型尝试集成搜索引擎、社交媒体、电子健康记录、气象数据和传统数据来改善流感预警, 发现上述方法均能够充分提取大数据中的有效信息, 拟合包含实时和历史信息的数学模型, 并补偿传统监测数据中固有的观测偏差和漏报, 可显著改善预警效果^[22-25]。近年, 我国学者也对互联网“大数据”的流感预警技术进行了探索^[26-28], 均取得了良好效果。

2015 年, Yang 等^[29]首次提出 ARGO 模型 (AutoRegression with Google search data), 利用公开的 Google 搜索数据估计当前 ILI 的活动水平。ARGO 是以 Google 搜索查询为外生变量的自回归模型, 可以即时的合并 CDC 报告的新信息, 融入流感流行的季节性, 同时捕捉人们在网上搜索行为随时间的变化, 具有灵活性、自校正性、稳定性和可扩展性等特点。研究人员还提供了理论框架, 既通过隐马尔可夫模型将因果关系论证结合起来。2019 年, 在 ARGO 模型基础上改进得到的 ARGO Net 成功将 Google 搜索频率和电子健康记录数据与流感活动的时空趋势相结合, 产生具有更高的相关性和更低的偏倚的预测^[30]。

三、基于影响因素的流感预警方法

流感的发生和流行与多种影响因素密切相关, 准确找出疾病影响因素, 探索影响因素与疫情发生间的相互关联, 并将相关影响因素数据集纳入流感预测预警模型, 能够为流感疫情发生风险的分析、评估及预警提供新思路、新方法。

流感的传播通常与气象变量有关。分布滞后非线性模型 (distributed lag non-linear model) 能够在时间尺度上同时兼顾暴露-反应的非线性关系和暴露因素的滞后效应^[31], 已广泛应用在探讨流感与气象因素之间影响效应的研究中^[32]。数据挖掘是一系列数据驱动分析方法, 相较传统回归方法更加灵活, 不需过多的先验信息和参数计算, 并且能够处理大

量的协变量^[33]。越南的研究人员通过回归树(classification and regression tree)探索气候与ILI季节性强度之间的关系,共纳入534个潜在解释变量,发现最低绝对湿度是热带环境中ILI发生的主要驱动因素^[33]。动力学模型也是流感影响因素研究的热点方法,Shaman等^[34]建立SIRS(susceptible-infectious-recovered)模型,探讨绝对湿度是否驱动了温带地区流感传播的季节性变化,发现绝对湿度对流感的季节性和冬季流感暴发均有影响,将异常低的绝对湿度应用到冬季流感暴发的短期预测中是可行的。也有研究通过使用SEIRS(susceptible-exposed infected-removed-susceptible)模型、随机SLIAR(susceptible, latent, infectious symptomatic, infectious asymptomatic, removed)模型等方法,证实流感还会受到人群、文化特征的影响^[35-37]。

危险因素对流感的影响是多种影响因素共同作用的结果。有研究使用增强回归树法和传播动力学模型发现,病毒类型的特异性和气候类型的依赖性共同影响着流感活动^[38],并提示流感病毒与其他病原体存在相互作用^[39]。Chattopadhyay等^[40]分析了10年来1.5亿美国人的疾病史,结合引发流感暴发因素的“大数据集”,模拟流感流行的触发机制及流行过程发现,人口的社会和种群特征,与特定湿度、温度和太阳辐射有关的天气变量,病毒随时间的变异,人口的陆路旅游习惯和流感近时期空动态,如流感波产生的自相关作用等,均是流感的重要影响因素。该模型有着很好的外推预测效果,为制定新的人群水平干预和疫情缓解政策提供了有效依据。

四、基于时空趋势的流感预警方法

流感疫情是一个时空传播的动态过程,相较于单纯的时间模型,综合时间和空间2个维度构建的流感疫情时空风险模型,可以对未来最有可能暴发疫情的时间范围和地理区域进行预测预警,更好地为流感防控提供理论基础和科学依据^[41]。近年来快速发展的空间信息技术将空间分析和可视化表达方法相结合,可以直观反映疾病在时空上的分布和变化特征,为大数据、多维度的流感时空分析提供了技术支撑^[42]。

Kuldorff等^[43-44]于2001和2005年提出前瞻性时空扫描统计量(prospective space-time scan statistics)和时空重排扫描统计量(prospective space-time permutation scan statistics)。2种方法在分析前无需对聚集性的尺寸、位置及规模进行限定,作为探索性分析,可以充分挖掘数据信息,发现疾病存在的异常时空聚集,已经应用于中国流感的预警防控研究中^[45]。时空扫描统计量多使用几何形状(圆形、椭圆形或方形)扫描窗口来发现疾病簇,不适用疾病的发生倾向于聚集在不规则形状的区域时。近年提出的最大连接时空排列扫描统计量(maximum linkage space-time permutation scan statistics)和联合聚类方法(co-clustering approach)则不受扫描形状和大小的限制,可以快速检测具有不规则几何区域的流感暴发,并提供更多有关暴发的空间和时间范围细节^[46-47]。

虽然时空扫描统计量已被应用于多种流行病学研究中,但该方法无法度量与已识别聚类相关的不确定性,且无法解

释协变量信息。贝叶斯方法则可通过在数据中加入额外的协变量和时空动态信息,提供统一的框架分析与每个参数相关的不确定性,并且能够处理多重性问题^[48]。随着集成嵌套拉普拉斯逼近方法(INLA)等统计学方法和软件的提出和应用,贝叶斯方法在传染病发生发展的时空探索上取得了巨大进展^[49]。研究者使用平行区域方法或将狄利克雷过程(Dirichlet process)和粒子过滤器(particle filters)结合,在贝叶斯方法基础上建立新的预测模型,发现这些新模型能够显著提高对某地区2009年流感暴发的预测精度^[50-51]。Osthus等^[52]采用贝叶斯-空时-SIR建模方法来检验流感传播的时空动态,研究结果加强了对流感迁移的理解,并提供预警信息,可指导干预策略。

集成多种时空分析方法,结合相关的气象、社会和经济等影响因素,深入分析流感疫情流行趋势,及时侦测高风险时空区域,是当前研究的热点领域。已有研究发现,基于上述影响因素结合多种预测方法构建的综合预测模型的预测结果比单一预测方法构建的模型更为准确高效^[45,53-54]。

五、基于“流感风险评估”的流感预警方法

控制流感流行的关键是监测系统能否快速、准确发出警报。因此,直观、可靠的流感风险估计方法可以帮助公共卫生机构及时应对即将到来的疾病流行。随着医疗保健和公共卫生机构电子“大数据”信息的迅速增长和完善,监测和预警流感流行季和暴发流行的方法正得到不断的改进。

计算疫情阈值是监测疫情发生的关键之一^[55]。各国现行流感监测系统多采用阈值进行预警,但目前仍没有关于阈值金标准的界定或共识^[56],主要通过对历史数据的检查,以经验的方式指定固定阈值^[57],还可使用简单的统计量如平均值、标准差和中位数等确定阈值^[58-60]。1963年,使用流感非流行期监测的历史数据拟合基线,根据该曲线上某一显著增长确认流行发生阈值的Serfling回归模型被建立后^[61],人们开发了各种Serfling类型的回归模型,用以监测流感疫情的发生和高峰时间^[62-63],并描述流感的季节模式^[64],但这类方法存在下列缺点^[57,65-66]:①需要根据主观标准预先确定流行期和非流行期;②需要长期(通常为 ≥ 5 年)历史数据;③不同流行区域的基线各不相同,缺乏统一标准。

另有研究则通过分析流行曲线的特性来定义流行阈值。Nobre和Stroup^[67]使用指数平滑技术和流感流行曲线的数值导数的特性监测疫情暴发。该方法不需要长期的历史数据,适用于 < 1 年的监测序列,但前提是所选的多项式模型必须与数据良好拟合,并且要进行探索性分析来选择指数平滑模型的参数。瑞典研究者以7d为时间单位对流感监测数据进行线性回归,以线性趋势变化的剧烈程度来确定峰值,再计算正态分布密度函数的导数求出峰值强度^[68]。WHO曾使用移动流行区间法(moving epidemic method, MEM),以确定欧洲流感监测的基线和流行阈值^[69]。MEM通过最大累积速率百分比曲线的斜率是否小于预先设定的标准 δ 来确定流行的开始和结束,但由于区域特异性的原因,预先设定的 δ 难以确定。最近,Cheng等^[70]开发了移动logistic回归方法

(moving logistic regression method, MLRM)来确定中国 30 省份季节性流感流行的阈值,通过 logistic 回归模型来近似累积流行曲线,但 MLRM 的应用局限于“对称流行波”,不适用于非对称或双峰流行波。

虽然目前流感暴发监测的主要方法都是基于阈值的,但也有一些非阈值方法被成功应用。有研究使用分段回归模型(stagewise regression model, SRM)确定流感的流行开始时间^[71-72],通过 SRM 对流行曲线的前半部分(即高峰前的每周 ILI 时间序列)进行拟合,以发病率突然改变的断点作为流行发生的起始点。SRM 不依赖于任何阈值,可以应用于单峰流感季节,且无需较多历史数据,可以根据流行曲线的特性定义流感流行的发生。Cai 等^[73]使用最大曲率法以确定日本特定地区的流感疫情发生和结束的阈值,该方法以流行曲线的最大曲率为基础,利用流行曲线的独特性,保留了非阈值法估计疫情发生的优点。中国深圳市研究中采用综合评分法,对深圳市 ILI% 的基线值和平均值、每周流感病毒检测阳性率和每周 ILI 暴发疫情次数的高低程度评分,计算权重系数,建立“流感指数”判定方程^[74-75],开启了我国将流感监测数据转化为实际应用的新尝试。

六、“大数据”流感预警的前景展望

综上所述,基于“大数据”的流感预警方法和理论已经取得了长足进步,但该领域仍存在诸多困难和挑战^[1,62,76-78]:

①更多的数据意味着更多无效“噪声”干扰,而过于强调数据间的相关性,忽视对疾病流行病学因果关系的解释,则可能使研究沦为“数字游戏”;②在开放的数据环境中,随着互联网环境与舆论导向的快速变化,人们的相关行为也会不断发生变化,这些变化难以及时调整和捕获将造成预警偏差;③多样化、复杂化的“大数据”采集、整合和分析很难像传统“小数据”那样缜密简洁,现阶段数据相关处理技术和架构的落后,导致很多隐藏在大数据中的有效信息难以得到充分挖掘,而在数据采集过程中出现的数据可用性低、质量难以保证等问题同样带来了噪声积累和伪相关等一系列统计学问题;④在高维数据集中,“大数据”分析过程很容易受到“维数诅咒”的影响,稀疏性、多重共线性、模型复杂性、模型拟合的计算成本和模型过度拟合是高维数据集所面临的主要问题;⑤“大数据”越开放越有价值,但数据安全等方面的相关政策与法规并不完善,数据开放和隐私之间难以平衡,“大数据”的利用面临着伦理问题的挑战。

尽管基于“大数据”的流感预警还存在不足,但这并不能否定“大数据”对流感预警的潜在价值^[79]。提高对数据的采集、管理、分析和应用能力,让“大数据”对流感预警产生帮助是当前的关键问题。面对“大数据”引领的“大变革”,随着健康医疗“大数据”体系的持续构建、流感监测体系的覆盖率和可靠性不断增加、互联网和移动设备的进一步普及、数据集集成及挖掘技术的不断创新,未来基于传统流感监测数据,集成互联网、地理气候及社会经济人口等数据源,进行动态多维度可视化的分析与展示,以提供更多更全面的疾病信息,提高流感预测预警的准确性和时效性,将是这一领域的

研究方向。相信在不远的将来,基于“大数据”的流感预警将成为趋势,就像今天的天气预报那样准确与平常。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- [1] Lee CH, Yoon HJ. Medical big data: promise and challenges [J]. *Kidney Res Clin Pract*, 2017, 36(1): 3-11. DOI: 10.23876/j.krcp.2017.36.1.3.
- [2] Khoury MJ, Ioannidis JP. Big data meets public health [J]. *Science*, 2014, 346(6213): 1054-1055. DOI: 10.1126/science.aaa2709.
- [3] McGowan CJ, Biggerstaff M, Johansson M, et al. Collaborative efforts to forecast seasonal influenza in the United States, 2015-2016 [J]. *Sci Rep*, 2019, 9(1): 683. DOI: 10.1038/s41598-018-36361-9.
- [4] WHO/Europe influenza surveillance (Euro-Flu. org) [EB/OL]. [2019-10-22]. <http://www.euroflu.org>.
- [5] European Centre for Disease Prevention and Control. European Influenza Surveillance Network (EISN) [EB/OL]. [2019-10-22]. <https://www.ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/eisn>.
- [6] Flu News Europe. Week 41/2019(7-13 October 2019) [ER/OL]. [2019-10-22]. <http://flunews europe.org>.
- [7] USCDC. U.S. influenza surveillance system: purpose and methods [ER/OL]. [2019-10-22]. <http://www.cdc.gov/flu/weekly/overview.htm>.
- [8] 任玉环,姚建义,闫磊,等.国内外流感监测系统的综合调查研究[J]. *疾病监测*, 2010, 25(10): 833-838, 841. DOI: 10.3784/j.issn.1003-9961.2010.10.020.
- [9] Ren YH, Yao JY, Yan L, et al. Research on influenza surveillance systems in China and abroad [J]. *Dis Surveill*, 2010, 25(10): 833-838, 841. DOI: 10.3784/j.issn.1003-9961.2010.10.020.
- [10] Gerke AK, Yang M, Tang F, et al. Association of hospitalizations for asthma with seasonal and pandemic influenza [J]. *Respirology*, 2014, 19(1): 116-121. DOI: 10.1111/resp.12165.
- [11] Gerke AK, Tang F, Yang M, et al. Predicting chronic obstructive pulmonary disease hospitalizations based on concurrent influenza activity [J]. *COPD J Chron Obstruct Pulmon Dis*, 2013, 10(5): 573-580. DOI: 10.3109/15412555.2013.777400.
- [12] Yang XT, Liu DP, Wei KF, et al. Comparing the similarity and difference of three influenza surveillance systems in China [J]. *Sci Rep*, 2018, 8(1): 2840. DOI: 10.1038/s41598-018-21059-9.
- [13] 杨召,叶中辉,赵磊,等. ARIMA-BPNN 组合预测模型在流感发病率预测中的应用 [J]. *中国卫生统计*, 2014, 31(1): 16-18.
- [14] Yang Z, Ye ZH, Zhao L, et al. Application of ARIMA-BPNN Hybrid Model to Predict the Incidence of Influenza in Henan province [J]. *Chin J Health Stat*, 2014, 31(1): 16-18.
- [15] 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. DOI: 10.11847/zgggws2014-30-02-32.
- [16] Meng L, Wang XH, Bai YN, et al. Prediction of ARIMA model of influenza-like cases in sentinel hospital of Gansu Province [J]. *Chin J Public Health*, 2014, 30(2): 228-230. DOI: 10.11847/zgggws2014-30-02-32.
- [17] Baltrusaitis K, Vespignani A, Rosenfeld R, et al. Differences in regional patterns of influenza activity across surveillance systems in the united states: comparative evaluation [J]. *JMIR Public Health Surveill*, 2019, 5(4): e13403. DOI: 10.2196/13403.
- [18] 吴洁,刘丽红,邹惠英,等. 流感监测网络现状及展望 [J]. *现代预防医学*, 2016, 43(10): 1880-1883.
- [19] Wu J, Liu LH, Zou HY, et al. The status and Outlook of Influenza surveillance network system [J]. *Mod Prev Med*, 2016,

- 43(10):1880-1883.
- [16] Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data[J]. *Nature*, 2009, 457(7232):1012-1014. DOI:10.1038/nature07634.
- [17] Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet [J]. *Can Med Associat J*, 2009, 180(8):829-831. DOI:10.1503/cmaj.090215.
- [18] Ortiz JR, Zhou H, Shay DK, et al. Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends [J]. *PLoS One*, 2011, 6(4):e18687. DOI:10.1371/journal.pone.0018687.
- [19] Butler D. When Google got flu wrong [J]. *Nature*, 2013, 494(7436):155-156. DOI:10.1038/494155a.
- [20] Lazer D, Kennedy R, King G, et al. Google flu trends still appears sick: an evaluation of the 2013-2014 flu season [J]. *Social Science Electronic Publishing*, 2014, 40(3):165-172. DOI:10.2139/ssrn.2408560.
- [21] Davidson MW, Haim DA, Radin JM. Using networks to combine "big data" and traditional surveillance to improve influenza predictions[J]. *Sci Rep*, 2015, 5:8154. DOI:10.1038/srep08154.
- [22] Santillana M, Nguyen AT, Dredze M, et al. Combining search, social media, and traditional data sources to improve influenza surveillance [J]. *PLoS Comput Biol*, 2015, 11(10):e1004513. DOI:10.1371/journal.pcbi.1004513.
- [23] Yang W, Lipsitch M, Shaman J. Inference of seasonal and pandemic influenza transmission dynamics [J]. *Proc Natl Acad Sci USA*, 2015, 112(9):2723-2728. DOI:10.1073/pnas.1415012112.
- [24] Ertem Z, Raymond D, Meyers LA. Optimal multi-source forecasting of seasonal influenza [J]. *PLoS Comput Biol*, 2018, 14(9):e1006236. DOI:10.1371/journal.pcbi.1006236.
- [25] Zimmer C, Leuba SI, Yaesoubi R, et al. Use of daily Internet search query data improves real-time projections of influenza epidemics [J]. *J R Soc Interface*, 2018, 15(147):20180220. DOI:10.1098/rsif.2018.0220.
- [26] Guo P, Zhang JJ, Wang L, et al. Monitoring seasonal influenza epidemics by using internet search data with an ensemble penalized regression model [J]. *Sci Rep*, 2017, 7:46469. DOI:10.1038/srep46469.
- [27] Xu QN, Gel YR, Ramirez LLR, et al. Forecasting influenza in Hong Kong with Google search queries and statistical model fusion [J]. *PLoS One*, 2017, 12(5):e0176690. DOI:10.1371/journal.pone.0176690.
- [28] Liang F, Guan P, Wu W, et al. Forecasting influenza epidemics by integrating internet search queries and traditional surveillance data with the support vector machine regression model in Liaoning, from 2011 to 2015 [J]. *Peer J*, 2018, 6(7):e5134. DOI:10.7717/peerj.5134.
- [29] Yang SH, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO [J]. *Proc Natl Acad Sci USA*, 2015, 112(47):14473-14478. DOI:10.1073/pnas.1515373112.
- [30] Lu FS, Hattab MW, Clemente CL, et al. Improved state-level influenza nowcasting in the United States leveraging Internet-based data and network approaches [J]. *Nat Commun*, 2019, 10:147. DOI:10.1038/s41467-018-08082-0.
- [31] Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models [J]. *Stat Med*, 2010, 29(21):2224-2234. DOI:10.1002/sim.3940.
- [32] Lau SYF, Wang XX, Wang M, et al. Identification of meteorological factors associated with human infection with avian influenza A H7N9 virus in Zhejiang Province, China [J]. *Sci Total Environ*, 2018, 644:696-709. DOI:10.1016/j.scitotenv.2018.06.390.
- [33] Thai PQ, Choisy M, Duong TN, et al. Seasonality of absolute humidity explains seasonality of influenza-like illness in Vietnam [J]. *Epidemics*, 2015, 13:65-73. DOI:10.1016/j.epidem.2015.06.002.
- [34] Shaman J, Pitzer VE, Viboud C, et al. Absolute humidity and the seasonal onset of influenza in the continental United States [J]. *PLoS Biol*, 2010, 8(2):e1000316. DOI:10.1371/journal.pbio.1000316.
- [35] Dalziel BD, Kissler S, Gog JR, et al. Urbanization and humidity shape the intensity of influenza epidemics in U.S. Cities [J]. *Science*, 2018, 362(6410):75-79. DOI:10.1126/science.aat6030.
- [36] Kucharski AJ, Kwok KO, Wei VWI, et al. The contribution of social behaviour to the transmission of influenza A in a human population [J]. *PLoS Pathog*, 2014, 10(6):e1004206. DOI:10.1371/journal.ppat.1004206.
- [37] Litvinova M, Liu QH, Kulikov ES, et al. Reactive school closure weakens the network of social interactions and reduces the spread of influenza [J]. *Proc Natl Acad Sci USA*, 2019, 116(27):13174-13181. DOI:10.1073/pnas.1821298116.
- [38] Chong KC, Lee TC, Bialasiewicz S, et al. Association between meteorological variations and activities of influenza A and B across different climate zones: A multi-region modelling analysis across the globe [J]. *J Infect*, 2019, 80(1):84-98. DOI:10.1016/j.jinf.2019.09.013.
- [39] Opatowski L, Baguelin M, Eggo RM. Influenza interaction with cocirculating pathogens and its impact on surveillance, pathogenesis, and epidemic profile: A key role for mathematical modelling [J]. *PLoS Pathog*, 2018, 14(2):e1006770. DOI:10.1371/journal.ppat.1006770.
- [40] Chattopadhyay I, Kiciman E, Elliott JW, et al. Conjunction of factors triggering waves of seasonal influenza [J]. *eLife*, 2018, 7:e30756. DOI:10.7554/eLife.30756.
- [41] 赖圣杰, 廖一兰, 张洪龙, 等. 2011-2013年国家传染病自动预警系统中时间模型和时空模型应用效果比较 [J]. *中华预防医学杂志*, 2014, 48(4):259-264. DOI:10.3760/cma.j.issn.0253-9624.2014.04.005.
- [42] Lai SJ, Liao YL, Zhang HL, et al. Comparing the performance of temporal model and temporal-spatial model for outbreak detection in China infectious diseases automated-alert and response system, 2011-2013, China [J]. *Chin J Prev Med*, 2014, 48(4):259-264. DOI:10.3760/cma.j.issn.0253-9624.2014.04.005.
- [43] Li XL, Yang Y, Sun Y, et al. Risk distribution of human infections with avian influenza H7N9 and H5N1 virus in China [J]. *Sci Rep*, 2015, 5:18610. DOI:10.1038/srep18610.
- [44] Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic [J]. *J R Stat Soc Ser A Stat Soc*, 2001, 164(1):61-72. DOI:10.1111/1467-985X.00186.
- [45] Kulldorff M, Heffernan R, Hartman J, et al. A space-time permutation scan statistic for disease outbreak detection [J]. *PLoS Med*, 2005, 2(3):e59. DOI:10.1371/journal.pmed.0020059.
- [46] Dong W, Yang K, Xu QL, et al. Spatio-temporal pattern analysis for evaluation of the spread of human infections with avian influenza A (H7N9) virus in China, 2013-2014 [J]. *BMC Infect Dis*, 2017, 17:704. DOI:10.1186/s12879-017-2781-2.
- [47] Costa MA, Kulldorff M. Maximum linkage space-time permutation scan statistics for disease outbreak detection [J]. *Int J Health Geogr*, 2014, 13:20. DOI:10.1186/1476-072X-13-20.
- [48] Ullah S, Daud H, Dass SC, et al. Detecting space-time disease clusters with arbitrary shapes and sizes using a co-clustering approach [J]. *Geospat Health*, 2017, 12(2):567. DOI:10.4081/gh.2017.567.
- [49] Zou J, Karr AF, Datta G, et al. A Bayesian spatio-temporal

- approach for real-time detection of disease outbreaks: a case study [J]. *BMC Med Inform Decis Mak*, 2014, 14: 108. DOI: 10.1186/s12911-014-0108-4.
- [49] Blangiardo M, Cameletti M, Baio G, et al. Spatial and spatio-temporal models with R-INLA [J]. *Spat Spatio-Temporal Epidemiol*, 2013, 7: 39–55. DOI: 10.1016/j.sste.2013.07.003.
- [50] Birrell PJ, Pebody RG, Charlett A, et al. Real-time modelling of a pandemic influenza outbreak [J]. *Health Technol Assess*, 2017, 21(58): 1–118. DOI: 10.3310/hta21580.
- [51] Zou J, Zhang ZQ, Yan H. A hybrid hierarchical Bayesian model for spatiotemporal surveillance data [J]. *Stat Med*, 2018, 37(28): 4216–4233. DOI: 10.1002/sim.7909.
- [52] Osthus D, Hickmann KS, Caragea PC, et al. Forecasting seasonal influenza with a state-space SIR model [J]. *Ann Appl Stat*, 2017, 11(1): 202–224. DOI: 10.1214/16-AOAS1000.
- [53] Yamana TK, Kandula S, Shaman J. Individual versus superensemble forecasts of seasonal influenza outbreaks in the United States [J]. *PLoS Comput Biol*, 2017, 13(11): e1005801. DOI: 10.1371/journal.pcbi.1005801.
- [54] Pei S, Kandula S, Yang W, et al. Forecasting the spatial transmission of influenza in the United States [J]. *Proc Natl Acad Sci USA*, 2018, 115(11): 2752–2757. DOI: 10.1073/pnas.1708856115.
- [55] Paranthaman K, Conlon CP, Parker C, et al. Resource allocation during an influenza pandemic [J]. *Emerg Infect Dis*, 2008, 14(3): 520–522. DOI: 10.3201/eid1403.071275.
- [56] Wang RP, Jiang YG, Michael E, et al. How to select a proper early warning threshold to detect infectious disease outbreaks based on the China infectious disease automated alert and response system (CIDARS) [J]. *BMC Public Health*, 2017, 17(1): 570. DOI: 10.1186/s12889-017-4488-0.
- [57] Cowling BJ, Wong IOL, Ho LM, et al. Methods for monitoring influenza surveillance data [J]. *Int J Epidemiol*, 2006, 35(5): 1314–1321. DOI: 10.1093/ije/dyl162.
- [58] Baumeister E, Duque J, Varela T, et al. Timing of respiratory syncytial virus and influenza epidemic activity in five regions of Argentina, 2007–2016 [J]. *Influenza Other Respir Viruses*, 2019, 13(1): 10–17. DOI: 10.1111/irv.12596.
- [59] Baumgartner EA, Dao CN, Nasreen S, et al. Seasonality, timing, and climate drivers of influenza activity worldwide [J]. *J Infect Dis*, 2012, 206(6): 838–846. DOI: 10.1093/infdis/jis467.
- [60] Ly S, Arashiro T, Ieng V, et al. Establishing seasonal and alert influenza thresholds in Cambodia using the WHO method: implications for effective utilization of influenza surveillance in the tropics and subtropics [J]. *Western Pac Surveill Response J*, 2017, 8(1): 22–32. DOI: 10.5365/WPSAR.2017.8.1.002.
- [61] Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths [J]. *Public Health Rep*, 1963, 78(6): 494–506. DOI: 10.2307/4591848.
- [62] Olson DR, Konty KJ, Paladini M, et al. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales [J]. *PLoS Comput Biol*, 2013, 9(10): e1003256. DOI: 10.1371/journal.pcbi.1003256.
- [63] Wang XL, Wu SS, MacIntyre CR, et al. Using an adjusted Serfling regression model to improve the early warning at the arrival of peak timing of influenza in Beijing [J]. *PLoS One*, 2015, 10(3): e0119923. DOI: 10.1371/journal.pone.0119923.
- [64] Liu XX, Li YH, Zhu YB, et al. Seasonal pattern of influenza activity in a subtropical city, China, 2010–2015 [J]. *Sci Rep*, 2017, 7(1): 17534. DOI: 10.1038/s41598-017-17806-z.
- [65] Amorós R, Conesa D, Martínez-Beneito MA, et al. Statistical methods for detecting the onset of influenza outbreaks: a review [J]. *Revstat Stat J*, 2015, 13(1): 41–62.
- [66] Costagliola D, Flahault A, Galinec D, et al. A routine tool for detection and assessment of epidemics of influenza-like syndromes in France [J]. *Am J Public Health*, 1991, 81(1): 97–99. DOI: 10.2105/ajph.81.1.97.
- [67] Nobre FF, Stroup DF. A monitoring system to detect changes in public health surveillance data [J]. *Int J Epidemiol*, 1994, 23(2): 408–418. DOI: 10.1093/ije/23.2.408.
- [68] Spreco A, Eriksson O, Dahlström O, et al. Integrated detection and prediction of influenza activity for real-time surveillance: algorithm design [J]. *J Med Internet Res*, 2017, 19(6): e211. DOI: 10.2196/jmir.7101.
- [69] Vega T, Lozano JE, Meerhoff T, et al. Influenza surveillance in Europe: establishing epidemic thresholds by the moving epidemic method [J]. *Influenza Other Respir Viruses*, 2012, 7(4): 546–558. DOI: 10.1111/j.1750-2659.2012.00422.x.
- [70] Cheng XJ, Chen T, Yang Y, et al. Using an innovative method to develop the threshold of seasonal influenza epidemic in China [J]. *PLoS One*, 2018, 13(8): e0202880. DOI: 10.1371/journal.pone.0202880.
- [71] Charu V, Zeger S, Gog J, et al. Human mobility and the spatial transmission of influenza in the United States [J]. *PLoS Comput Biol*, 2017, 13(2): e1005382. DOI: 10.1371/journal.pcbi.1005382.
- [72] Geoghegan JL, Saavedra AF, Duchêne S, et al. Continental synchronicity of human influenza virus epidemics despite climatic variation [J]. *PLoS Pathog*, 2018, 14(1): e1006780. DOI: 10.1371/journal.ppat.1006780.
- [73] Cai J, Zhang B, Xu B, et al. A maximum curvature method for estimating epidemic onset of seasonal influenza in Japan [J]. *BMC Infect Dis*, 2019, 19(1): 181. DOI: 10.1186/s12879-019-3777-x.
- [74] 逯建华, 何建凡, 房师松, 等. 深圳市流感指数的制定及应用 [J]. *实用预防医学*, 2016, 23(5): 628–629. DOI: 10.3969/j.issn.1006-3110.2016.05.039.
- Lu JH, He JF, Fang SS, et al. Development and application of the flu index in Shenzhen City [J]. *Pract Prev Med*, 2016, 23(5): 628–629. DOI: 10.3969/j.issn.1006-3110.2016.05.039.
- [75] 逯建华, 何建凡, 许舒乐, 等. 深圳市流感指数运行结果调查 [J]. *中国公共卫生管理*, 2016, 32(4): 567–568. DOI: 10.19568/j.cnki.23-1318.2016.04.055.
- Lu JH, He JF, Xu SL, et al. Survey of influenza index in Shenzhen [J]. *Chin J Pub Health Manag*, 2016, 32(4): 567–568. DOI: 10.19568/j.cnki.23-1318.2016.04.055.
- [76] 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. DOI: 10.1126/science.1248506.
- [77] Santillana M, Zhang DW, Althouse BM, et al. What can digital disease detection learn from (an external revision to) Google Flu Trends? [J]. *Am J Prev Med*, 2014, 47(3): 341–347. DOI: 10.1016/j.amepre.2014.05.020.
- [78] Dinov ID. Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data [J]. *Giga Science*, 2016, 5(1): 12. DOI: 10.1186/s13742-016-0117-6.
- [79] Milinovich GJ, Williams GM, Clements ACA, et al. Internet-based surveillance systems for monitoring emerging infectious diseases [J]. *Lancet Infect Dis*, 2014, 14(2): 160–168. DOI: 10.1016/S1473-3099(13)70244-5.

(收稿日期: 2019-09-08)

(本文编辑: 李银鸽)