

传染病暴发早期预警系统评价内容及其指标

赖圣杰 李中杰 金连梅 杨维中

【关键词】 传染病; 预警系统; 评价

The overview of evaluating contents and indicators on early warning system of infectious disease outbreak LAI Sheng-jie, LI Zhong-jie, JIN Lian-mei, YANG Wei-zhong. Chinese Center for Disease Control and Prevention, Beijing 100050, China
Corresponding author: YANG Wei-zhong, Email: yangwz@chinacdc.cn

【Key words】 Infectious disease; Early warning system; Evaluation

为及时识别和应对传染病暴发与流行,近年来全球广泛开展了传染病暴发探测预警技术研究,探索利用不同类型和来源的监测数据进行传染病暴发早期预警,尝试并建立了各类传染病暴发早期预警系统(预警系统)。预警系统作为一类有别于传统疾病监测系统的传染病暴发早期预警工具,有必要对其性能和运行效果开展专门的评价^[1]。本文系统回顾国内外有关文献,对预警系统评价的研究现状进行概述,并简要介绍预警系统主要评价内容及其指标,以期为我国开展相关研究提供参考和借鉴。

一、预警系统评价研究现状

2000年以来,全球已有多个国家探索和建立了各类传染病暴发早期预警系统^[1-3],但目前预警系统的评价体系研究和案例报道较少,只有美国较为系统地开展了相关的研究。2002年美国全国症状监测大会在美国疾病预防控制中心(CDC)编写的“公共卫生监测系统评价指南”的基础上^[4],结合预警系统的特点,率先提出传染病暴发探测预警系统的评价草案^[5]。2003年,美国CDC成立了专门的工作组,专门研制成套的预警系统评价体系,并于2004年发布了“疾病暴发早期探测预警系统评价框架”^[6]。该评价框架更注重评价预警系统的早期发现能力及相关属性,并从系统运行状况、预警能力评价、系统性能评价等方面进行阐述,介绍了及时性、有效性、灵活性、可接受性等评价指标,但目前尚未见该评价框架应用实例的报道。此外,一些研究机构或学者利用常规的疾病监测系统评价指南或方案指导预警系统评价。例如,Jefferson等^[7]利用2001年美国CDC“公共卫生监测系统评价指南”的评价指标,采用半结构化问卷调查的方式,分析和评价传染病症状监测预警系统的数据质量、预警及时性、可接受性、有效性等特性。

近年来我国学者结合现有的传染病监测信息资源,对传染病预警技术和预警系统进行了探索性研究^[8],但预警系统的评价研究亟待开展。例如,中国CDC于2005年启动了重点传染病监测自动预警信息系统建设与应用试点项目^[9],建立起“传染病自动预警信息系统”,并于2008年4月21日起在全国启动传染病自动预警(时间模型)试运行工作,同时在215个县(区)启动传染病自动预警(时空模型)试点工作^[10],目前急需对系统的运行效果开展深入的评价,从而为系统的修改完善和推广应用提供科学依据。

二、预警系统主要评价内容及其指标

预警系统评价主要是在全面了解系统运行现状的基础上,采用及时性、灵敏度等指标,重点评价传染病暴发的探测预警能力,并根据系统运转状况和用户反馈,用灵活性、可接受性、稳定性等指标综合评价系统性能。主要评价内容及其指标简述如下。

1. 预警系统现状:

(1)预警系统的用途:一般包括预警的疾病种类、暴发类型和规模;预期的及时性、灵敏度和特异度;系统是短期使用,还是长期运行等。评价预警系统的准则之一是系统用途的实现程度。评价时应根据不同用途选择针对性的评价指标。

(2)预警系统的用户:一般包括数据提供者(如医护人员等)和预警信息使用者(如CDC人员、卫生行政部门等)。明确系统用户不仅可以了解系统的服务对象和用户需求,而且有助于确定向哪些类型的用户收集有效的评价信息。

(3)采用的预警方法:预警方法是预警系统的核心,应详细阐述预警方法的原理、参数设置及信号推送方式等。预警方法的一般原理是:首先将经过预处理的监测数据按日/周为单位或一定空间单元进行汇总,形成时间序列或空间维度数据;然后将汇总数据与历史数据进行时间和(或)空间维度的分析、比较,计算预警指标值;若该值超过预警阈值,则认为可能存在疾病暴发并及时发出预警信号。

(4)预警系统的工作流程:指预警系统从监测数据采集和处理、历史数据建模、预警分析运算到信号响应处理的整个运行过程。图1清晰地展示了预警系统的一般工作流程^[6],有助于描述系统组成、运行机制和关键环节^[11]。数据流程和信号响应的相关内容:①数据收集:包括采集方式和数据类别;采集方式包括手工录入和自动采集;数据类别包括各种来源的监测数据(如患者症状、临床诊断或药物销售量等)、预警响应数据等。②数据处理:指在运算前对实时和历史监测数据的筛选、分类、汇总、转换等处理过程。③数据规范和标准:指数据采集、处理、分析和输出全过程所需的数据格式、传输标准、储存方式、安全措施等。④预警信号响应:指预

警信号的响应措施及其工作规范,并说明人员和设备的配备情况。

(5) 监测数据质量: 良好的监测数据质量是有有效预警的基础。通过描述监测数据在人群、时间、地区的分布; 尤其是地区分布, 说明采用的监测数据的代表性。同时分析监测数据报告质量的情况, 包括监测报告内容规范、各项监测指标填报完整率、报告及时性等^[4]。

2. 预警能力评价: 预警能力评价是预警系统评价的核心内容, 主要从及时性、灵敏度、特异度、可探测的暴发规模四个维度进行评价^[12], 通过对相应评价指标的分析, 判断各个维度的最佳值, 从而优选出预警系统的数据类型、预警指标、预警方法及预警阈值等^[13]。

(1) 及时性: 是预警系统的最基本特征, 也是预警能力评价最基本的指标, 体现在发生传染病暴发后, 预警系统能否及时探测到暴发“苗头”并发出相应预警。预警及时性一般用预警系统首次发出预警的时间与参比时间的平均差值进行量化^[12]。参比时间指预警系统之外的其他途径(如医院、学校报告等)或方法(如其他预警模型等)首次发现暴发的时间。首次发出预警时间越早, 及时性越好。此外, 将传染病暴发监测、预警、干预过程中的重要环节, 按起止时间依次标注在一条时间轴上, 绘制成时间节点图(图 2), 可以更直观地评价及时性, 发现制约或提高预警及时性的关键环节^[6]。

(2) 灵敏度: 一般应结合及时性来综合评价预警系统的灵敏度, 即在满足特定及时性要求的条件下, 预警系统发现真正传染病暴发的能力。具体可以采用绘制灵敏度——及时性曲线(ST 曲线)的方法, 直观分析和比较不同及时性要求下的预警灵敏度^[12]。一般用函数 $S_{(x)}$ 表示灵敏度, X 表示及时性; 然后计算传染病暴发开始后 X_d 内预警系统识别出的暴发数占全部暴发数的比例, 即 $S_{(x)}$; 再以 X 为横轴、 $S_{(x)}$ 为纵轴绘制 ST 曲线。如图 3 所示, 若要求预警系统在 5 d 内探测出传染病暴发, 则预警方法 1 的灵敏度更好; 若对及时性要求不高, 允许系统 5 d 后发出预警, 方法 2 则具有较好的灵敏度。

(3) 错误预警率和特异度: 错误预警率指在某一时期内实际没有发生传染病暴发, 但被预警系统错判为存在暴发的情况, 即在实际没有发生暴发的时间内预警的天数占总天数的比例(分子和分母的单位可以是天、周、月等)。例如, 错误预警率是 1/20, 指每天平均有 0.05 次错误预警或每 20 天里会产生 1 次错误预警。特异度是指预警系统正确判断未发生传染病暴发的能力。特异度 = 1 - 错误预警率, 即错误预警率

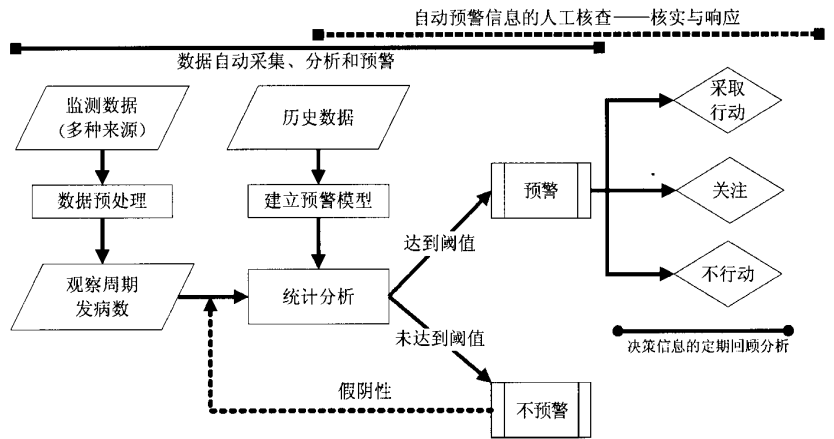
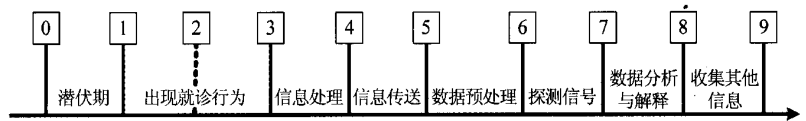


图 1 预警系统工作流程示意图



注: 0: 发生感染; 1: 出现症状; 2: 就诊; 3: 病例登记; 4: 病例报告; 5: 监测系统发现病例; 6: 运行预警模型/工具; 7: 发出预警信号; 8: 流行病学调查; 9: 流行病学干预

图 2 传染病暴发监测预警时间节点图

越低, 特异度越高, 反之亦然。由于特异度与灵敏度、及时性之间紧密相关, 因而可以用受试者工作特征曲线(ROC 曲线)和动态监测工作特征曲线(AMOC 曲线)分析这些指标间的最佳平衡点^[14]。

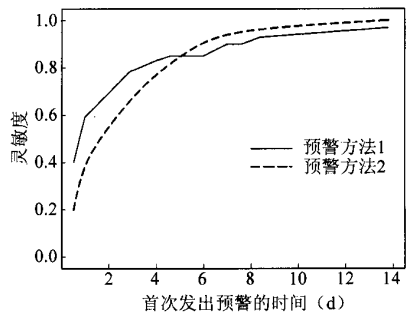


图 3 两种预警方法的 ST 曲线比较

ROC 曲线是用灵敏度(纵轴)和错误预警率(横轴)绘制而成, 曲线上的任意一点代表预警方法的某个预警阈值对应的一对灵敏度和特异度。如图 4 所示, 通常将曲线上最靠近坐标图左上角的点定为灵敏度和特异度最佳的临界值, 此点对应的预警阈值的预警功效可能最大^[12]。国内已有研究采用 ROC 曲线筛选预警阈值并取得较好效果^[8]。此外, 通过计算不同预警方法的 ROC 曲线下的面积, 可用来比较和筛选不同预警方法, 面积越大, 越接近于 1.0, 预警方法的功效越好。

AMOC 曲线则是用预警系统首次发出预警的时间(纵轴)和错误预警率(横轴)绘制的曲线, 表示某种预警方法对应的及时性和特异度之间的关系。如图 5 所示, 提高预警方

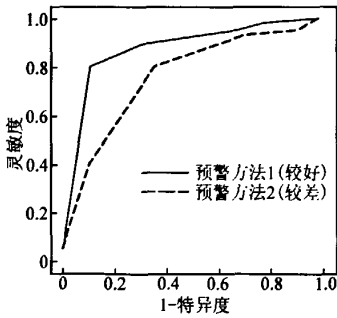


图4 两种预警方法的ROC曲线比较

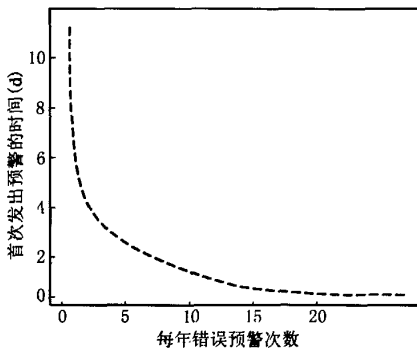


图5 某种预警方法的AMOC曲线

法的及时性,可导致错误预警频次的增加,降低特异度,反之亦然。一般将曲线上最靠近坐标图左下角的点判断为及时性和特异度的理论最佳临界值^[12]。

(4)可探测的暴发规模:预警能力评价的另一个重要指标是预警系统能探测到的最小暴发规模(如发病例数、感染例数、罹患率等),即预警方法的探测能力。用探测所需时间(纵轴)和暴发规模(横轴)绘制及时性——暴发规模曲线图(TOS曲线),可以直观分析不同及时性要求下,系统能够发现的最小的暴发规模,并确定二者之间的最佳平衡点^[12,15]。具体计算方法是在量化暴发规模的基础上,分析预警方法能否探测到不同规模的暴发及所需的时间。如图6所示,TOS曲线可用于比较两种预警方法的暴发探测能力,曲线整体越靠近坐标图的左下角,所代表的预警方法的暴发探测能力越强^[12]。

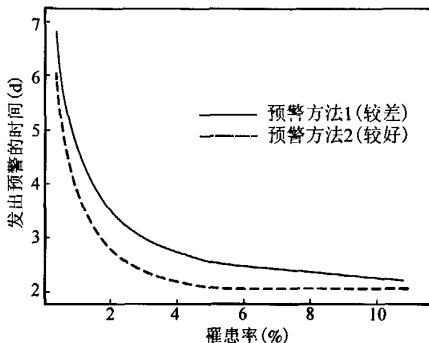


图6 两种预警方法的TOS曲线比较

3. 系统性能评价:预警系统性能评价综合反映了系统的预警方法、运行机制和信息技术的各项特性,以及用户的使用感受,主要评价指标包括实用性、灵活性、可接受性、可推广性、稳定性等^[4,6]。

(1)实用性:是一项综合评价指标,在分析预警系统的应用效果、优于其他方法的特点及用户评价后做出综合评价。若预警系统有利于提高传染病监测预警工作效率,能够对重点传染病暴发实现早期预警,促使及时采取干预措施,则说明该预警系统具备一定的实用性。

(2)灵活性:指预警系统随需求改变而改变的能力。灵活的系统应根据需求随时调整,或只需花费少量的资源和时间进行修改。结构简单、易于操作的系统,往往具有较好的灵活性。该指标可用具体案例和用户使用体验进行说明。

(3)可接受性:反映用户对预警系统的参与意愿和认同程度,可使用用户满意度、报告单位的参与率、数据报告质量和及时性、预警信号响应处理的比例和及时性进行评价。

(4)可推广性:体现预警系统在不同的地区或运行环境中推广应用的难易程度。如果系统实施方案简单可行、系统建设和运行不易受软硬件的限制、预警方法简便而实用,往往推广性较好。该指标可用已有案例与经验进行评价。

(5)稳定性:指预警系统能否持续稳定运行。可使用持续稳定运行的时间,系统硬件、软件、经费、人员等原因中止运行次数和时间、预警信号发送的延迟或丢失率进行评价。

4. 系统成本效益分析:系统成本是指系统建立和维护运转所需的资源,主要包括固定成本和可变成本。固定成本主要包括购买设备、软件及人员培训、工资等所需的费用。可变成本主要是预警响应处理的费用,如诊断、密切接触者的追踪、流行病学调查分析等费用^[6]。成本效益分析可以更好地评价预警系统的价值,但预警系统的效益和错误预警或重复预警造成的负面效应等不易估算,因此成本效益分析方法的应用尚待进一步研究。

5. 分析总结和结果利用:通过收集和整理评估资料,分析各项评价指标结果,对预警系统做出综合评价并进行总结和建议。内容包括系统现状、各项评价指标结果、预警系统的优缺点、取得的经验教训、系统目标的实现情况,并进一步提出改进建议。此外,应跟踪系统评价报告的应用情况。

三、预警系统评价中应注意的问题

1. 评价信息的收集:针对不同指标应采取不同的资料收集方法。例如,灵敏度、特异度等指标应收集监测数据进行定量计算,而实用性、灵活性、可接受性等定性指标则可采用问卷调查、小组讨论、用户访谈等方式收集资料。此外,在时间和资源有限的情况下,可根据评价目的,有针对性地收集核心评价信息,实现快速评估。一般优先收集以下核心信息:预警原理、预警频次、预警阈值等预警方法的信息;数据分析、现场调查的预警信号比例及响应措施内容,反映预警响应情况的信息;暴发探测能力主要收集预警系统探测到的暴发和及时性,未能发现的暴发及其发现途径,以及相应的病原体、宿主人群等信息^[16]。

2. 评价数据的选取:评价数据的选取是评价的重要前提。需要用含有暴发数据的监测资料计算灵敏度,以及不存在暴发的监测数据计算特异度、错误预警率。评价数据主要有三种类型:一是真实的历史或实时监测数据;二是完全模拟数据;三是在真实数据中插入暴发的模拟数据。一般采用历史或实时监测数据评价预警系统识别真实事件的能力;用模拟数据评价预警系统探测生物恐怖等罕见事件的能力。由于模拟数据仍有较多局限性,须开展更多研究以验证暴发假设模型的有效性^[17,18]。此外,传染病暴发不易界定,一般由相关部门结合疾病严重程度、事件报告标准、公众关注情况和既往经验等因素综合判定^[6]。

3. 评价时应考虑的其他因素:预警系统的能力受监测数据质量、报告及时性、传染病特点和预警方法的适用条件等多种因素的影响或制约。因此,应尽量在相同条件下比较不同预警方法,排除或减少其他因素的干扰;在设计评价方案时,应充分考虑不同类型用户的需求,细化评价指标和评价标准;在计算灵敏度、特异度和及时性等指标的最佳平衡点后,还应考虑疾病的特点、暴发的危害和资源可及性等因素,才能得出科学的评价结论。此外,应分析预警系统对不同地区、不同传染病和不同流行水平的适用性,从而为预警系统的推广应用提供科学依据。

综上所述,目前预警系统评价研究已取得一定进展,但如何将各项评价内容及其指标进行综合分析平衡,并有效地应用于我国现有的传染病暴发早期预警系统的评价,尚需进行深入的实践与探索。

参 考 文 献

- [1] Henning KJ. What is syndromic surveillance. *MMWR*, 2004, 53 Suppl:S5-11.
- [2] Tsui FC, Espino JU, Dato VM, et al. Technical description of RODS: a real-time public health surveillance system. *J Am Med Inform Assoc*, 2003, 10(5):399-408.
- [3] Valenciano M, Bergeri I, Jankovic D, et al. Strengthening early warning function of surveillance in the Republic of Serbia: lessons learned after a year of implementation. *Euro Surveill*, 2004, 9(5):24-26.
- [4] German RR, Lee LM, Horan JM, et al. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. *MMWR Recomm Rep*, 2001, 50 (RR-13):1-351.
- [5] Sosin DM. Draft framework for evaluating syndromic surveillance systems. *J Urban Health*, 2003, 80 (2 Suppl 1): S8-13.
- [6] Buehler JW, Hopkins RS, Overhage JM, et al. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. *MMWR Recomm Rep*, 2004, 53(RR-5):1-11.
- [7] Jefferson H, Dupuy B, Chaudet H, et al. Evaluation of a syndromic surveillance for the early detection of outbreaks among military personnel in a tropical country. *J Public Health (Oxf)*, 2008, 30(4):375-383.
- [8] 杨维中, 邢慧娴, 王汉章, 等. 七种传染病控制图法预警技术研究. *中华流行病学杂志*, 2004, 25(12):1039-1041.
- [9] 金连梅, 杨维中. 我国传染病预警工作研究现状分析. *中国公共卫生*, 2008, 24(7):548-549.
- [10] 中国疾病预防控制中心. 全国传染病自动预警(时间模型)试运行和传染病自动预警(时空模型)试点工作. 北京:中国疾病预防控制中心, 2008.
- [11] Buehler JW, Sosin DM, Platt R. Evaluation of surveillance systems for early epidemic detection // M'ikanatha NM, Lynfield R, Beneden CA, eds. *Infectious Disease Surveillance*. 1st ed. Massachusetts: Blackwell, 2007:432-442.
- [12] Wagner MM, Wallstrom G. Methods for Algorithm Evaluation // Wagner MM, Moore AW, Aryel RM. *Hand book of Biosurveillance*. Burlington: Elsevier, 2006:301-310.
- [13] Hashimoto S, Murakami Y, Taniguchi K, et al. Detection of epidemics in their early stage through infectious disease surveillance. *Int J Epidemiol*, 2000, 29(5):905-910.
- [14] Fawcett T, Provost F. Activity monitoring: noticing interesting changes in behavior. *The Fifth ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1999. San Diego (CA): ACM Press, 1999:53-62.
- [15] Straetemans M, Altmann D, Eckmanns T, et al. Automatic outbreak detection algorithm versus electronic reporting system. *Emerg Infect Dis*, 2008, 14(10):1610-1612.
- [16] Sosin DM, Dethomasis J. Evaluation challenges for syndromic surveillance - making incremental progress. *MMWR*, 2004, 53 Suppl:S125-129.
- [17] Kulldorff M, Zhang Z, Hartman J, et al. Benchmark data and power calculations for evaluating disease outbreak detection methods. *MMWR*, 2004, 53 Suppl:S144-151.
- [18] Buckeridge DL, Burkom H, Moore A, et al. Evaluation of syndromic surveillance systems-design of an epidemic simulation model. *MMWR*, 2004, 53 Suppl:S137-143.

(收稿日期:2008-11-14)

(本文编辑:尹廉)