
第5章

疾病暴发监测系统及其在 时空上的监测方法

陈冬梅

作为一项公共事业,公共健康的实践需要持续收集、分析和解译对其规划、实施和评价具有重要意义的公共健康数据,并将其传达给需要了解相关信息的单位或个人。为了有效地监控公共健康状况,预防和控制疾病的暴发,需要对疾病症状进行监测,并通过一定的系统及时汇总和分析监测数据,反应和报告公共健康中潜在的健康威胁。

疾病症状监测(syndromic surveillance)是指在医生诊断或者实验室检查确诊之前,依靠对患者明显的临床症状的鉴定,对疾病暴发进行早期判定的疾病监测方法(Mandl 2004)。这些监测数据如果能够及时进入一个“症状监测系统”中,则该系统就可以利用实时或近实时的疾病数据进行数据汇总和统计分析,实现对疾病暴发的预警,报告给公共健康部门(Henning, 2004)。通常情况下,医院急诊中心和诊所的医生只对单个病例进行诊断,因此不会关注也无法了解某种疾病在整个区域内的空间分布和时间变化特征。而症状监测系统为一线医生了解区域内疾病的时空变化特征提供了支持(Kamel-Boulos, 2004; Mandl, 2004)。本章重点讨论症状监测系统在探测和预警人群易发传染性疾病暴发以及在探测和预警生物恐怖活动(即由恐怖分子故意释放病原体引起疾病暴发)中的应用。需要指出的是,症状监测系统不仅能够用于监测人群中或动物中潜在的传染性疾病的暴发,还可以应用于公共健康中的其他领域,如伤害预防、环境健康、职业保健、慢性病预防和医疗研究等。

本章首先评述现有的一些电子症状监测系统及其所采用的各种时空异常检测方法,然后介绍加拿大安大略省的一个传染病暴发早期疾病症状监测系统,包括利用从该系统获取的金斯顿(Kingston)地区呼吸道疾病数据,对疾病暴发进

行动态模拟的实例，并据此讨论个人隐私保护对时空异常检测方法敏感度的影响问题。本章最后讨论目前症状监测系统及其所使用的疾病暴发检测方法的一些问题和局限性。

5.1 疾病暴发监测系统

5.1.1 传统的疾病监测系统

传统上，疾病监测依靠的是对医生诊断结果和实验报告数据的人工分析，因而无法对爆发式疾病的暴发情况进行及时有效的监测 (Kamel-Boulos, 2004; Tsui et al. ,2003)。这里有两方面的原因：首先，虽然当短时间内有较多具有相似症状的病例在一地区出现时，这些病例可能预示着一次该疾病的暴发，但是患者会到不同的医院或诊所就诊，而一所医院的医生一般无法与其他医院和诊所的医生共享病例信息，无法了解其他医生对病例的诊断，因此疾病暴发的早期信号往往被忽略。其次，传统上对疾病暴发的诊断往往依靠将人工实验报告与已建立的“金标准”的对比，只有当观测到的疾病发病率和死亡率明显高于根据前一季度的数据计算出的统计预期值时，才能检测出疾病的暴发 (Gestland et al. ,2003; Dafni et al. , 2004)，而这是一个相对缓慢的过程，因为病例数据的获取与分析、生成实验报告以及根据报告对疾病趋势作出判定都需要花费一定的时间。由于存在着这些问题，传统的疾病监测系统在传染性疾病暴发的早期检测中有效性较低。

5.1.2 实时计算机网络疾病症状监测系统

过去的几十年中发生的一些国际性事件引起了许多国家对发展疾病暴发早期预警系统的高度重视 (Lai et al. ,2004; Nordin et al. ,2005)。这在一定程度上可归因于潜在的致命疫情的自然暴发 (如在香港和加拿大出现的 SARS) 和潜在的生物恐怖袭击事件 (如 2001 年在美国出现的炭疽病的传播) 等。为了实现对潜在威胁的早期监测及预警，疾病监测系统需要获取覆盖监测区域多数人口的多源数据，并且拥有对与疾病相关的时间和空间数据进行实时或近实时统计分析的能力 (Gestland et al. ,2003; Nordin et al. ,2005)。

目前，种类繁多的症状监测系统已在不同国家和地区运营。这些系统在数据收集方法、数据特征提取以及疾病暴发分析方法方面各有千秋 (Bravata et al. , 2004)。例如，美国疾病预防与控制中心 (Centers for Disease Control and Prevention) 使用的异常事件早期异常报告系统 (Early Aberration Reporting System,

EARS)^①,通过收集911报警电话数据、急诊中心报告数据(emergency departmental chief complaint data)、医生诊断数据(physician group data)和缺勤数据来对疾病暴发进行早期监测。洛斯阿拉莫斯国家实验室(Los Alamos National Laboratory)使用的快速症状鉴定项目(Rapid Syndrome Validation Project,RSVP)是利用计算机网络收集由医师输入的数据而建立的实时医疗数据库。它通过对症状的分析,而不是疾病诊断结果的分析,对疾病暴发进行追踪和报告(Zelicoff et al.,2001)。美国的社区流行病早期发布电子监测系统(Electronic Surveillance System for the Early Notification of Community Based Epidemics,ESSENSE II)通过收集来自公众和相关社会服务机构的公共健康数据、药品零售店销售额以及电子护士信息,对疾病暴发进行实时监测(Lombardo et al.,2003)。加拿大公共健康局的替代监测预警项目(Alternative Surveillance Alert Project,A.S.A.P)通过收集药品零售店相关药品的零售额来监测传染性胃肠道疾病的暴发疫情。英国健康保护署的国家健康服务直接监督系统则依靠护士帮助热线进行疾病监测(Cooper,2007)。目前,在美国、加拿大、中国台湾等地的多个政府部门和其他机构中运行的开源的实时疾病暴发监测系统(Real-time Outbreak and Disease Surveillance,RODS)则收集和综合医院、诊所、药店和实验室健康数据,进行实时的疾病分类和监测(Tsui et al.,2003;Moore,2004)。在2002年盐湖城冬季奥运会期间,RODS系统被用于检测潜在的生物恐怖事件(Gestland et al.,2003)。

5.1.3 地理信息系统和空间分析在疾病监测系统中的应用

地理信息系统(GIS)在健康领域的应用已经越来越广泛,但目前主要局限于对疾病暴发特征的回顾性研究,例如,对疾病在某段时间内的发病原因进行单因素分析和孤立性研究,而缺少对疾病数据的实时的前瞻性分析,因而在面向未来的健康规划、协调和保护活动方面所发挥的作用仍十分有限(Kamel-Boulos 2004)。然而,美国联邦地理数据委员会(US Federal Geographic Data Committee)的调查表明,在政府机构或者健康部门发布的公共健康数据中,80%~90%的数据将地理位置信息作为关键要素之一。因此,在将这些数据转化为有用信息的过程中,如将这些位置数据与现有的疾病监测数据进行连接、匹配、分析和管理, GIS及其空间分析功能大有可为。

5.2 时空异常疾病暴发检测方法

目前,运行的各种疾病症状监测系统采用了许多不同的疾病暴发检测方法(Dafni et al.,2004;Lombardo and Buckeridge,2007)。这些方法一般采用统计警

^① <http://www.bt.cdc.gov/surveillance/ears/>

报机制,其基本原理是当观察到的疾病发病率高于统计上的预期值时,系统将会发布预警信号。这种统计机制能够检测出人群中明显的疾病暴发信号,即这些信号显著区别于统计上的背景噪声(Tsui et al. ,2003)。传统的公共健康监测方法一般只有在疾病暴发的晚期才能对疾病暴发进行确定,因为只有在疾病暴发的晚期才能获得相关的病例数据。新的时空异常疾病检测方法能对疾病暴发的早期症状进行快速地估计,并及时给出预警,虽然这种估计的准确性需要进一步确认(Tsui et al. ,2003;Wong et al. ,2003)。

所谓观察到的疾病发病率高于统计上的预期值,就是疾病的发病率在时间或者空间上,或者两者均超过正常的期望值,即出现时空集聚,而时空异常检测方法就是要检测这种集聚。时间上的集聚是指疾病的发病率或死亡率在特定时间点异常增高;而空间上的集聚是指在特定的地理区域范围内,疾病的发病率异常高于其他地区(Chen et al. ,2011)。当空间集聚检测方法和时间集聚检测方法能同时运用于传染性疾病暴发的早期监测时,症状监测系统能在感染的早期阶段最大限度地识别新的传染病的暴发。

本章接下来简要介绍和解释目前症状监测系统中所使用的时空集聚检测方法。

5.2.1 时间序列异常检测方法

递推最小二乘法(Recursive Least Squares,RLS)是目前症状监测系统中使用的最简单的方法之一。递推最小二乘法的基本思想是使估计的病例数量与实际的病例数量之间误差的平方和最小(Haykin 2002)。这种方法的优点在于它利用最近的历史数据(过去几天)而不是前一年的数据来检测各类症状(如胃肠道疾病症状、呼吸道疾病症状)中的异常值,因此适用于人口数量发生明显变化的地区(Das et al. ,2003;Gestland et al. ,2003;Moore,2004,Dijk et al. ,2009)。使用RLS方法检测疾病暴发时,当观测到的疾病发病率在95%的置信水平上超过预期的发病率时,监测系统将会发出预警信号。

累积和法(Cumulative Sum,CUSUM)也是目前许多疾病监测系统使用的方法,它计算某变量的一系列观测值与其数学上预期的平均值之间差的累积和。在症状监测系统中,CUSUM将过去几天或过去几个小时观测到的病例平均发病数量与不同时间窗口(过去6个月到过去两年)下的病例平均发病数量进行对比(Das et al. ,2003;Moore et al. ,2008)。这种方法可以对季节变化导致的病例数量的差异进行修正(Mandl,2004)。若某时间段内观测到的病例数量明显高于某一段时期内(如7天)的平均值时,累积和的新增值为正值,累积和会平稳增大。在累积和图中,该时间段的线段会倾斜上升,表示该时段内,疾病的发病率要高于一段时期的平均值。

近期奇异检测法(What is Strange About Recent Events, WSARE)通过建立一些规则,将近期的疾病数据与基准分布的疾病数据进行对比,来检测近期数据中最大的变化(Wong et al., 2005)。该方法充分利用了疾病监测系统收集到的空间、时间、人口和症状等信息,同时也考虑病例记录及病例的其他特征信息(如地理位置),以确定在特定的某天内与历史数据相比,是否存在统计异常值(Wong et al., 2005)。近期奇异检测法融合了许多思想和方法,包括关联规则、贝叶斯网络、假设检验和置换检验等,来检测疾病异常及评价显著性。

除此之外,还有许多其他时间序列检测方法,如小波分析(Wavelet)、脉冲响应(Pulsar)、LOWESS 和 TAD 等,限于篇幅无法在此逐一介绍。这些时间算法的一般描述可以在 Sonesson 和 Brock(2003) 和 Dafni 等(2004) 的研究中找到。除了 WSARE 以外,以上所有监测方法都仅以时间序列统计方法为基础。需要说明的是,单纯的使用时间序列检测方法进行疾病暴发监测,其有效性依赖于监测区域(数据来源区)面积的大小。如果监测区域非常小,病例的数量会很少,这会导致时间序列检测方法对疾病暴发检测的敏感度较低,从而遗漏疾病暴发案例。相反,如果监测区域很大,疾病暴发的早期信号则有可能会被误判为背景噪声。

5.2.2 空间序列和时空(异常)检测方法

疾病暴发可以看作疾病在空间、时间或者两者兼而有之的时空上的集聚。时间序列检测方法一般无法检测疾病的空间集聚,特别是当病例的数量较少而监测的范围较大时。与健康事件有关的集聚统计方法的研究已经进行了几十年,它们一般着眼于解决两个方面的问题:①整体来看,是否存在集聚趋势?②如果有集聚趋势,集聚发生在哪些时段或者哪些空间范围内?第一个问题可以运用全局空间统计方法来解决(Lee and Rogerson, 2007),第二个问题的解决则需要依靠局部统计方法。然而,传统的空间集聚分析方法只能用于分析回顾性(历史)疾病数据(retrospective data),而不能直接用来分析前瞻性的疾病监测数据(prospective surveillance data)。因此,随着疾病监测系统能够获得的前瞻性数据的增加,需要发展新的空间统计方法来检测疾病空间分布的变化(Rogerson 1997; Rogerson and Yamada 2004)。尽管许多传统的空间聚类技术可以用于处理症状监测数据,但接下来我们着重对适用于前瞻性数据的空间制图和空间集聚检测方法进行介绍(表 5.1)。

最早的疾病空间集聚检测方法可以追溯到 1956 年 Bartholomew(1956) 建立的序列检测法。但是由于该方法的复杂性和较多的假设条件,它并没有得到很多关注和应用(Rogerson 1997)。Openshaw 等(1988)提出了地理分析机(geographical analysis machine, GAM)检测法,它能够在给定的显著性水平下,对病例

的空间集聚区进行持续搜索和检测。虽然这种方法最初不是为处理症状监测数据设计的,但是其在疾病空间监测方面的应用取得了良好的效果。它的局限在于,当需要多次检验时,其计算量过大。

Rogerson 是最早将空间分析方法用于疾病症状监测数据的研究者之一。他基于修正的 Tango 统计量提出了一种数据的累积和(CUSUM)统计用于聚类检验以探测空间格局的变化。这种方法的缺点在于它是对疾病全局分布变化而不是对局部分布变化的评估。Rogerson 认为,如果运用局部统计指标(如 Moran's *I*、Getis's *G*)进行累积和分析,可能会达到比使用全局性统计量更好的分析效果(Rogerson and Yamada, 2009)。

2001 年,Rogenson 结合 CUSUM 和 Knox 检验法,发展了一种研究空间点状事件时空互动作用的空间统计方法。利用此方法对乌干达 West Nile 地区 Burkett 淋巴瘤患者的疾病数据的探测表明,淋巴瘤疾病在特定地理位置呈现显著的时空聚类(Rogerson, 2001)。该研究使用 1961—1975 年的数据,因此具有回顾性研究的特点,但是结合 CUSUM 和 Knox 检验组合法,可在疾病新兴空间格局的前瞻性分析中使用。

2004 年,Rogerson 和 Yamada 以美国东北部地区乳腺癌病例为案例,利用模拟数据和县级统计数据,比较了单变量和多变量(CUSUM)方法在监测疾病空间分布中的效果及差异。他们的研究表明,当病例具有较强的空间自相关性时,多变量 CUSUM 在检测病例空间集聚中效果更好;而对于病例的空间自相关作用较弱的地区,单变量 CUSUM 能更好地检测疾病空间分布的变化。

2002 年,Lazarus 等使用了另外一种时空集聚检测的方法。他们以包含患者住址信息的医院就诊病例数据和健康热线数据为数据源,运用广义线性混合模型(GLMM 或 logistic 回归模型,参见本书第 2 章)检测美国马萨诸塞州东部地区与生物恐怖相关的炭疽病例在空间上的集聚。模型的结果表达为基于概率的度量,具有协变量和不可测量特征的可调性,以及多维检验。2004 年,Kleinman 等结合 GLMM 和 CUSUM 检测小尺度地区疾病空间集聚。这种小尺度检测方法在检测由于蓄意释放生物病原体导致的疾病暴发以及其他小尺度地区的疾病暴发(如食物中毒)中,能发挥较好的作用。

与以人口统计区为尺度,进行疾病时空分布分析的研究不同,Olson 等(2005)以患者个人住址数据为基础进行病例时空集聚分析。该研究通过统计分析来比较矩阵判定点与点之间的距离是否存在明显差异。他们的研究表明,即使在病例总数较少的情况下,这种方法在检测病例空间集聚方面仍有较高的敏感度。但是,该方法需要根据疾病发病率的季节变换进行相应的调整。

扫描统计方法也被用来检测和评估疾病的空间、时间或时空集聚(Kulldorff, 1997)。扫描统计的基本方法是通过建立可移动的时间或空间窗口,并比较各窗

口中观测到的病例数量和预期的病例数量来评价疾病的集聚特征。理论上说,扫描窗口既可以是时间段、空间上的圆或椭圆,也可以是能同时表示时间和空间的圆柱或者椭圆柱(Kulldorff, 2001)。蒙特卡罗模拟(Monte Carlo simulation)方法常用来对似然值进行分级,将P值统计量分配给每个聚类,用以评价其统计显著性。空间扫描统计法最初是为对慢性疾病进行回顾性分析而开发的,但后来也被用于传染性疾病的监测(Hefferman et al., 2004)。2001年“9·11”事件后,美国纽约市健康局建立并运行了一个传染性疾病报告系统。该系统以大量稳定的、可利用的健康数据库为基础,结合CUSUM和空间扫描统计检测方法,监测与全国性恐怖事件有关的传染性疾病的暴发(Das et al. 2003)。

Besag 和 Newell 提出了另一种疾病空间集聚检测的方法。他们通过设定一系列包含固定病例数的扫描圆来估计疾病的空间集聚区(Besag and Newell, 1991)。Costa 和 Assuncao 后来对该方法进行了改进,并运用已公布的模拟数据,将改进后的办法与扫描统计方法进行对比。他们的研究表明,总体上看,两种方法在探测疾病聚类的效果上相似,但是当疾病聚类位于人烟稀少的地区时,扫描统计方法的运行效果更好(Costa and Assuncao, 2005)。

表 5.1 空间序列和时空(异常)检测方法概览

方法	检测类型	检测数据	文献
序列检测法	空间	模拟数据	Bartholomew(1956)
地理分析机	空间	回顾性白血病数据	Openshaw 等(1988)
空间扫描统计	空间;时空	回顾性数据;前瞻性数据	Kulldorf (1997); Das 等(2003); Hefferman 等(2004); Kulldorff (2015);
改进的全局累积和(CUSUM)	空间	模拟数据;回顾性淋巴瘤数据	Rogerson(1997)
CUSUM 和 Knox 检验结合法	时空	淋巴瘤回顾性数据	Rogerson(2001)
多变量 CUSUM 和单变量 CUSUM	空间	模拟数据;回顾性乳腺癌数据	Rogerson 和 Yamada (2004)
改进的融合 Moran's I 和 Getis's G 的 CUSUM	空间	回顾性乳腺癌死亡率数据	Rogerson 和 Yamada (2009)
泰森多边形与 CUSUM 结合法	空间	前瞻性医院监测数据	Moore 等(2008)

续表

方法	检测类型	检测数据	文献
广义线型混合模型(GLMM)	空间	前瞻性公共健康监测数据	Lazarus 等(2002); Kleinman 等(2004)
M-Statistic	空间	医院前瞻性监测数据	Olson 等(2005)
近期奇异检测法(WSARE)	时空	前瞻性监测数据	Wong 等(2003,2005)
Besag-Newell 疾病集聚检测法	空间	模拟数据	Besag 和 Newell(1991); Costa 和 Assuncao(2005)

5.2.3 疾病暴发监测方法的问题和挑战

尽管对疾病空间格局变化的监测非常有助于快速探测和及时发现传染病的暴发,但上述介绍的许多空间疾病监测方法并没有在当前运行的疾病症状监测系统中得到广泛应用。多数疾病监测系统仍主要采用时间序列检测法, GIS一般仅用于数据的制图和可视化。由于缺乏更有效空间检测方法,症状监测系统的功能和潜力未能得到充分的发挥(Rogerson and Yamada, 2009)。

时空异常疾病暴发检测的关键在于检测方法的敏感性、针对性和及时性(Hutwagner et al., 2005),它们是症状监测系统中受到持续关注的问题。而空间探测在这些方面的表现严重依赖于可获取的地理数据。对于不准确或不完整的地理编码数据的问题,已有许多研究文献(Elliott and Wartenberg, 2004; Hurley et al., 2003; Moore et al., 2008)。隐私和保密问题一直是影响空间检测和时空检测方法有效性的重要问题(Kamel-Boulous, 2004; Moore et al., 2008)。目前尽管疾病监测系统已经能够获取病例的详细地理信息,但是根据一些联邦、省和相关机构关于隐私保护的法律法规,这些数据常常必须要聚合后到一定区域上才能提供给研究者使用(Gestland et al., 2003; Moore et al., 2008)。当小区域的数据被聚合后,小区域内的疾病暴发信号会被掩盖,同时跨越不同小区域的疾病集聚区也会被稀释,造成空间不确定问题(Jacquez, 2000; Olson et al., 2005)。目前的研究尚未对各种空间检测方法在不同区域合并水平上的时空敏感性做出全面的评估。另外,当不同的疾病传播过程导致相似的疾病空间分布结果时,对疾病空间分布的解释往往具有误导性(Graham et al., 2004; Rogerson and Yamada, 2009)。

症状监测系统往往需要跨越不同行政管辖界限(医院、县、公共健康单元、省或联邦),因此其有效运行往往需要不同辖区之间的相互合作和相互支持(Gestland et al., 2003)。近年来,加拿大、美国和欧洲在这方面已经有不少成功的范例(Moore, 2004; Rolphamre et al., 2004; Lombardo and Buckeridge, 2007)。

下面我们介绍利用实时疾病症状爆发监测系统(RODS)在加拿大安大略省进行疾病暴发检测模拟的研究,着重阐述利用扫描统计和空间自相关分析进行疾病暴发检测的过程,并探讨个人隐私法对症状监测系统中检测算法的灵敏度的影响。

5.3 研究实例

5.3.1 研究区和数据

在加拿大,实时疾病症状爆发监测系统(RODS)是能够提供最密集、最详细疾病监测数据的监测系统之一。它的开发源于安大略省的一个两年试验计划。该计划从两方面评价 RODS 系统的运行效果:一方面是该系统能否通过监测医院急诊部提供的数据,为早期检测易感传染病或由生物恐怖活动引起的胃肠道疾病和呼吸道疾病的暴发提供帮助;另一方面是该系统是否有助于公共健康服务中的疾病暴发调查和疾病应急响应(Moore et al. ,2008)。RODS 系统最初仅在 Kingston、Frontenac 和 Lennox & Addington (KFL&A) 地区的公共健康部门(KFL&A Public Health Unit)中使用,后来推广到整个安大略省。关于 RODS 系统更详细的介绍可以参考相关文献(Tsui et al. ,2003; Moore et al. ,2008; Dijk et al. ,2009)。

本节介绍的研究以安大略省 RODS 系统提供的 2006 年呼吸道疾病病例数据为基础,对呼吸道疾病爆发的可能性进行情景模拟。该数据包括 2006 全年中到 KFL&A 地区 6 所医院中任意一所医院的急诊部就诊呼吸道疾病症状的所有病例。每一病例的数据都包括就诊时间、就诊者住址邮政编码(六位邮政编码)、症状等信息。为了保护个人隐私,在数据预处理中将就诊者住址邮政编码的最后一位删去,将剩下的五位邮政编码与加拿大统计中心(Canada Statistics)提供的邮政编码分区地图相对应,对就诊者的住址重新进行空间定位。由于我们的研究只针对 KFL&A 地区,邮政编码为非 KFL&A 邮政编码的病例被过滤掉。

图 5.1 为呼吸道疾病单日就诊病例时间变化图 2006 年,呼吸道疾病单日就诊病例的平均值为 29 例,两个峰值均出现在冬季,分别为年底和 3 月初。自 3 月到 8 月底,单日就诊病例的数量逐渐波动下降,8 月后又逐渐上升,到年底时达到峰值。作为该时间变化趋势的响应,RODS 系统对疾病暴发的敏感度在秋季和冬季要低于在春季和夏季的。在秋季和冬季时,如果呼吸道疾病的暴发只导致就诊病例相对于平均值的小幅上升,RODS 系统将很难检测出该暴发。

在试验阶段,安大略省 RODS 系统只使用了时间序列检测方法。为了提高

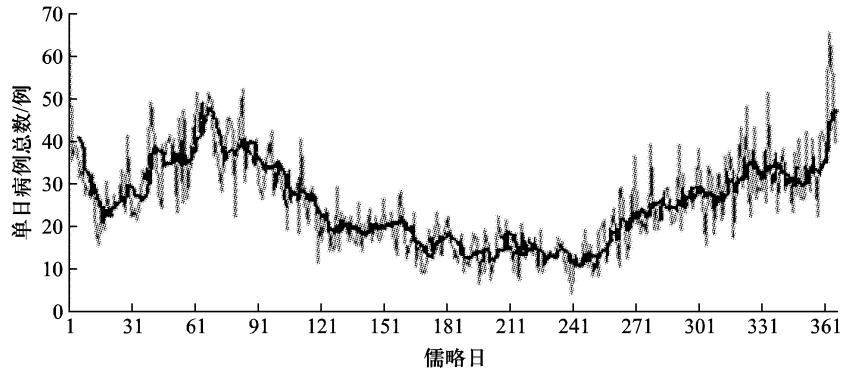


图 5.1 2006 年 KFL&A 地区呼吸道疾病单日就诊病例
时间变化(黑色趋势线为 7 日移动平均线)

RODS 系统的敏感度以及更加迅速地监测局部疾病暴发,我们在皇后大学的监测研究小组将空间集聚检测方法运用到 RODS 系统中并对其有效性进行测试。该研究关注两个方面:①相对于六位邮政编码的定位数据,使用五位邮政编码数据对疾病暴发进行空间检测的有效性如何?②使用空间集聚或者时空集聚检测方法进行疾病暴发监测时,如果要及时发布预警信号,至少需要多少病例数据?

测试疾病暴发检测方法敏感度的最好方式是对其进行重复模拟测试,测试的目的是检验系统从背景噪声中检测出疾病暴发信号的能力。这种测试通过将测试数据插入到真实数据中,进行不同时间长度和不同地理区域范围的疾病暴发模拟,然后通过估计检测方法的敏感度和针对性来评价疾病监测系统的有效性(Feinberg and Shmueli, 2005)。

5.3.2 疾病暴发动态模拟

基于连续三日的呼吸道疾病的病例数据,我们对呼吸道疾病短时间内在 Kingston 市中心快速快播的场景进行了模拟。我们假设由于某些原因,呼吸道疾病在 Kingston 市中心的某一地点暴发,并在第一天结束时影响周边 100 m 范围内的地区(图 5.2a 中的小圆)。相应地,在第一天中,周边 100 m 地区内除正常情况下的病例,新增了 3 例病例到 KFL&A 医院的急诊中心就诊。随着疾病的进一步传播,到第二天结束时,周边 300 m 范围受到影响(图 5.2a 中的中圆)。第二天中,周边 300 m 地区内除正常情况下的病例,新增了 3 例病例到医院就诊。到第三天结束时,疾病的影响范围扩大到周边 500 m 范围(图 5.2a 中的大圆),导致第三天中周边 500 m 范围内除了正常情况下的病例,新增了 3 例病例。需要说明的是,这里的 3 个病例的数量是人为设定的,如果每天在疾病影响区内的新增病例超过 3 例,理论上系统将会检测出更强的病例空间集聚。

首先,我们模拟了3天中每天正常情况下的病例场景。模拟的方法是假设病例随机分布于KFL&A地区的各邮政编码区。对正常情况的模拟须满足两个条件:①模拟病例的总数等于RODS系统提供的该地区的病例数量的平均值。②一个邮政编码区的人口越多,病例分布在该区的机会越大。模拟之后,我们用Poisson模型和Moran's *I*模型对模拟结果进行了检测,结果显示,在正常场景中病例不显示空间集聚现象。

接下来,我们进行了疾病暴发的场景模拟,即将疾病暴发引起的新增病例附加到相应的正常场景中。为了能够测试在RODS中使用五位邮政编码数据与使用六位邮政编码数据的差异,我们将六位邮政编码数据进行合并转换为五位邮政编码数据,然后分别对五位邮政编码的病例数据和六位邮政编码的数据进行暴发场景模拟和病例空间集聚检测。

5.3.3 模拟结果

根据模拟结果,呼吸道疾病病例数在第一天、第二天和第三天分别有33例、32例和33例,RODS系统本身所提供的RLS和CUMUS等时间序列检测方法没有对这样的情形发出疾病预警信号。由此可以得知,当初始病例数很小时,现有RODS系统中的时间序列检测方法并不能提供有效的疾病暴发检测及预警。

图5.2显示了使用空间集聚法进行疾病暴发模拟的检测结果。图5.2a中的同心圆代表了3天中疾病影响区的扩大。图5.2b显示的是使用六位邮政编码数据模拟时的检测结果,小圆、中圆和大圆分别代表第一天、前两天和连续三天的检测结果。图5.2b显示这种动态检测法提供了清晰的病例空间集聚和扩张信息,尽管该信息不是十分精确。作为对比,图5.2c显示了使用五位邮政编码数据模拟时的检测结果,图中的圆表代表连续3天的检测结果。

我们用Poisson模型对单日模拟场景分别进行了检测。结果显示,第一天中,当3例新增病例在周边100m地区出现时,病例的空间分布呈现很强的集聚性($P=0.001$);到第二天时,当3例新增病例在周边300m地区出现时,病例的空间集聚性相对减弱($P=0.090$);第三天时,当3例新增病例在周边500m地区出现时,病例的空间集聚性进一步减弱($P=0.125$)。模拟结果显示,若使用六位邮政编码数据进行场景模拟,为了能够持续检测出明显的病例空间集聚现象,至少需要两例或3例新增病例出现在Kingston市中心周边100m地区,4例或更多病例出现在周边300m地区。

与单日检测不同,对3日数据的时空变化检测显示,病例集聚区的范围逐渐扩大,而且集聚性一直较强。当选择一个覆盖两天的时间窗口时,可以检测出一个明显的病例空间集聚区($P=0.001$);而当选择一个覆盖3天的时间窗口时,同样可以检测出明显的病例空间集聚区,且其集聚性并未减弱($P=0.001$)。因

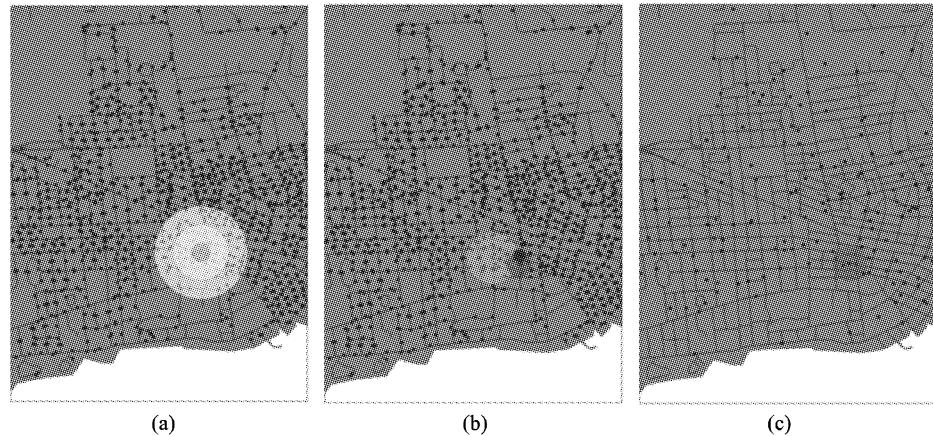


图 5.2 模拟呼吸道疾爆发场景及集聚检测结果:(a)模拟的增长集聚;(b)使用六位邮政编码数据模拟的检测结果;(c)使用六位邮政编码数据模拟的检测结果(见书末彩插)

此,即使在病例总数随时间变化很小的情况下,这种时空变化检测方法仍然能够动态地检测出病例空间集聚范围的扩大。因此,疾病危险性增强的事实不会被病例总数相对稳定的表象所掩盖。

进一步的研究表明,当把六位地理编码数据合并转化为五位地理编码数据时,检测结果无法反映病例空间集聚范围的扩张。当使用五位地理编码数据进行疾病暴发场景模拟时,在第一天的模拟结果中无法看到病例空间集聚现象,而在前两天和连续 3 天的模拟中,只能检测出相同的病例空间集聚区(图 5.2c)。当使用五位地理编码数据进行疾病暴发场景模拟时,若仅有 3 例或更少的新增病例分布在疾病暴发地点周边 100 m 或 300 m 范围内,系统将无法检测出明显的病例空间集聚现象。

5.4 讨 论

5.4.1 症状监测系统中的隐私保护

当症状监测系统使用关于个人的数据,特别是当这些数据包含身份认证信息并通过互联网进行传输时,个人隐私和保密问题必须要受到高度重视。一些身份认证信息和地理位置信息,如住址信息和邮政编码等,需要严格的管理和处理。个人数据保密是一个复杂的问题。例如,当一个受保护的数据库与其他数据库相连接用于 GIS 分析时,新的数据库可能并没有采取相应的保护措施,这可

能导致一些个人信息的识别和泄露。在这种情况下,错误的信息识别和正确的信息识别都会导致个人利益的受损(Kamel-Boulos,2004)。但是,在很多情况下存在着这样一种矛盾,即一方面必须要保护个人信息,另一方面又需要使用个人信息服务于公共利益,症状监测系统的运行就是一个例子。因此,在多数情况下,在系统运行之前,联邦、省或相关机构需要制定相关的隐私保护的法律法规,例如,2002年冬奥会期间针对RODS系统运行的法律保护措施(Gestland et al., 2003)。

有一些监测系统采取尽量避免使用或传输个人身份信息的策略。这些监测系统所依赖的是大量的群体性信息。由于这些信息并不包括个人身份信息,因此较少涉及隐私问题。这种系统的优点在于能够覆盖范围更广的地区且系统运行的成本较低,需要的工作人员较少。但是其缺点也很明显:由于个人数据和地理信息数据的缺失,这种系统无法检测疾病的空间集聚,因此疾病监测的敏感度和针对性较低,更容易发布错误的预警和集聚区预测。另外,使用群体性信息代替个人信息还会产生空间不确定问题,这是因为群体性信息将一定空间单元内的信息合并表达,这样一方面空间单元内部地域分异及其所包含的疾病暴发信号会被掩盖;另一方面,当疾病的集聚区跨越不同空间单元的边界时,其空间集聚也会被掩盖。

为了保护个人隐私,在症状监测数据库中,可以用一些数字代码代替个人姓名和住址信息来与空间位置链接。另外,对数据进行聚合也是保护患者隐私的常用方法。例如,在安大略省RODS系统的试验运行中,在城市地区将六位邮政编码聚合为五位。但是,模拟结果显示,这种数据聚合导致了疾病时空异常预警的延迟。这种延迟所造成的后果的严重性取决于疾病潜伏期的长短和传播速度的快慢。如果一种疾病具有高传染性且传播速度很快,这种延迟可能会导致公共健康部门错过最佳的疾病控制时间。在乡村地区,试验保留了六位邮政编码,因为编码的最后一位总是为0。但是在乡村地区,一个邮政编码往往覆盖很大范围的地区,因此邮局的位置对该地区居民的实际居住位置并不具有很大的代表性。当传染病在城市地区暴发时,很可能分布在城市不同地方的居民与同一个患病者有接触,而这些不同地区的患者会到不同的医院就诊,在这种情况下邮政编码水平上的数据会将疾病发生状况稀释,不会显示出空间集聚现象,监测系统很难对疾病最初的暴发地点和疾病的传播进行空间定位和检测。而在乡村地区,当一些随机的病例出现在同一个邮政编码区时,则会造成空间异常预警的偏差,因而为了在乡村地区正确检测出空间异常区,要求病例在更大的空间范围表现出集聚。这种乡村和城市地区对病例空间集聚范围的差异给疾病检测系统带来了挑战,因为目前多数疾病监测系统所使用的空间异常检测方法并不能自动适应不同地区对分析尺度不同的要求。

5.4.2 数据来源与数据质量

这里“数据质量”是指来自于不同公共健康系统的数据的准确性。为了评价数据质量,我们通常将所收集的数据与一些“金标准”进行对比,以确定其相应的敏感度和针对性,并利用其实现对疾病病例的准确分类(Mandl 2004)。一般认为,在症状监测系统中,获得完全准确无误的数据往往是不可能也是不必要的。如上文所讨论的,多数症状检测系统会依据统计上的噪声数据发布早期检测预警。然而研究表明,在 RODS 运行过程中,恰恰是系统组织上的问题(如数据共享与隐私保护)而不是技术上(如具体算法和参数选择)的问题,是导致疾病预警延迟的主要原因。这种现象再次表明,基于 GIS 的疾病监测系统的成功运行需要其受益者的紧密合作,其中包括公共健康系统、私人企业和研究机构。

症状监测系统不仅受到数据获取及时性和数据处理方法的影响,也受到数据来源的影响(Buehler et al. 2003)。RODS 症状监测系统主要依靠医院急诊部、诊所和家庭医生提供的病例数据,并假定患病者在症状出现初期即到这些机构就诊。但是,如果患病者不是在症状出现初期即就诊,尤其是一些潜伏期较短的疾病,如炭疽,就会导致数据出现质量问题。针对这个问题,可以收集一些非传统数据作为系统的补充数据源,包括 911 电话(急诊车)、健康咨询电话、在线咨询、处方药销售额(如抗病毒药物和抗生素)、非处方药销售额(如感冒药)、零售店某些商品的销售额(如餐巾纸、果汁)等(Lombardo et al. ,2003; Bravata et al. ,2004; Moore, 2004)。零售额也叫作柜台销售额(over-the-counter sales, OCS),某些商品零售额的增长可以指示疾病的發生和疾病的类别。另外,缺勤记录也可以作为数据源,但是其可靠性往往不强,因为缺勤的原因往往没有记录,或者记录较晚,因此无法确定其是否是疾病原因导致的(Feinberg and Shmueli, 2005)。

流行病学家也常常根据病例地理分布的特殊性来诊断疾病发病区。例如,SARS 暴发不久,专家就发现发病者一般最近去过香港或者与从香港来的人员接触过(Lai et al. ,2004)。在最近的猪流感暴发中,加拿大出现的多数病例都直接或间接地与来自墨西哥的人员有过接触。另外,通过 911 电话和健康咨询电话的拨打者的拨打时间和地址信息,可以对病例进行时间和空间定位。同时,零售点也能提供一些病例信息,因为一些顾客会持有固定零售店的打折卡或航空里程卡,或者自愿向零售点提供他们的邮编或电话信息(Feinberg and Shmueli, 2005)。

通过结合以上不同的数据,疾病监测系统从背景噪声中检测出疾病暴发信号的能力将会大大增强。但是,当处理大量来自于不同数据源的数据时,疾病检测方法也会面临新的挑战。首先,目前疾病监测系统所使用的检测方法一般只适用于传统数据,而不能直接用于处理这些非传统数据(Wong et al. ,2005)。另

外,这些非传统数据本身存在许多非确定性,而目前疾病监测系统所使用的检测方法不能对这些非确定性进行有效的处理。一些非传统方法,如机器学习中的人工神经网络分析、序列预警的多层次系统等,能够更有效地从多源数据库中提取信息,可以成为未来疾病监测系统发展多源数据处理能力的技术方向。

当个体位置信息无法得到时,常用邮政编码区多边形的质心来代表区内居民的位置,这种不准确和不完全的地理编码常常会导致数据的空间定位偏差和错误分类。有研究者已经着眼于解决这类地理编码的局限性(Hurley et al., 2003; Elliot and Wartenburg, 2004)。

除了地理编码问题外,疾病编码错误、医生的错误诊断以及由于聚合监测区域导致的基准人口数量的增大都是导致数据质量问题的原因。其他一些原因,如连锁药店对某种药物的促销以及医疗系统引进新的药物等,也会造成数据的错误分类,进而导致疾病监测系统发布错误的预警信号。

当疾病暴发或生物恐怖袭击发生时,人为地放大或压制疾病信号都会影响疾病数据的质量(Pavlin 2003)。疾病监测系统并不能代替流行病学家对疾病暴发和诊断的实地考察。流行病学家必须对疾病监测系统发布的每一个信号进行仔细调查,评价数据的有效性并判定是否存在数据的错误分类。这种调查往往需要对预警信号所涉及的疾病集聚区中的单个病例进行检验,在这方面,系统内的人口数据和地理数据也能够起到辅助判定潜在的数据错误分类的作用。

5.5 结 论

当传染性疾病暴发时,疾病症状监测系统能够帮助利益相关者(stakeholder)实时或近实时地评估疾病暴发的严重程度及其空间分布(Sosin,2003)。现有的疾病症状监测系统中最常用的检测方法一般仍局限于对病例时间序列的分析。由于基于一定暴发源的疾病传播本身是一个空间扩散的过程,在分析中缺失空间信息就无法对一种疾病传播过程是否时“暴发”以及暴发源在哪里做出准确和全面的评判。另一方面,做到对疾病空间分布的正确解释也非易事,因为不同的传播过程可能会导致类似的疾病空间分布模式,往往对疾病空间分布的解释具有误导性。现实中,疾病暴发后的传播过程会在不同的时间和空间尺度上呈现错综复杂的分布特征,只有全面考虑这一过程的时空耦合特征才能更有效地探测“感染性病因”(Graham et al., 2004)。地理信息系统技术为疾病症状监测系统在时间和空间数据的无缝集成和综合分析方面提供了技术基础。

检测方法的敏感性和针对性是症状监测系统中的关键问题。当疾病的传播每天仅导致少量的病例产生时,目前的监测系统可能无法有效地检测其暴发。

同时,如何能在保护患者个人隐私的前提下,充分利用多种数据源,对潜在的疾病暴发进行更加有效的监测,也是一个持续受到关注的问题。随着症状监测系统检测范围的扩大和其所使用数据源的增多,需要发展能够整合和分析多源数据的更有效的时空检测方法,以便可以从多个不同数据源集成信息。一些已经在其他领域运用的非传统时空分析方法,如机器学习中的人工神经网络分析、贝叶斯时空方法(Bayesian spatial-temporal approach)、基于先验知识的数据挖掘和数据融合技术(knowledge-based data mining and data-fusion approach)等可以作为未来研究的方向。

目前,虽然在全球范围内已经发展了许多症状监测系统,但是由于数据质量和检测方法的局限性等问题,尚没有明确的证据表明这些监测系统能对传染性疾病的暴发进行有效准确的早期监测。然而相关测试表明,疾病监测系统确实能够有效地模拟和预测一些年周期性疾病(如流感)的暴发(Platt et al., 2003; Lombardo and Buckeridge, 2007)。通过不同时间框架下一系列的空间集聚检测,症状监测系统的确能够反映出疾病的集聚特征。因此,随着症状监测系统逐渐覆盖更多的医院和更大范围的监测区域,空间集聚检测方法在监测系统中的作用也将更加明显。

总体而言,虽然基于疾病传播是个时空现象这个本质, GIS 和空间分析理应在流行病学研究中发挥重要作用,然而现实情况是,空间分析方法在流行病研究中的应用远没有达到其应有的程度。这个问题的形成有多方面的原因,解决起来也相对复杂。在技术层面,流行病学研究中所使用的许多统计分析方法需要一个独立性假设,这与空间统计分析技术中所依赖的空间相关性相矛盾,导致人们对空间分析技术存在偏见。在知识结构层面,流行病学研究者还没有广泛采纳空间思维方式和使用空间分析技术。这些问题导致了开发适用于流行病研究的 GIS 软件的商业驱动力不足(Elliott and Wartenberg, 2004)。我们在实际工作中发现,虽然一些参与症状监测系统开发和应用的流行病学家已经关注到了 GIS 的空间分析优势,他们也只能有限地进行一些与疾病传播有关的空间位置、接触性(exposure)、共变性(covariate)分析和生物过程建模(biological disease process modelling)。流行病学研究者往往局限于这些功能,仅仅是因为这些分析和建模在当前的 GIS 软件中能够方便地实现(Jacquez 2000)。

本章所介绍的疾病暴发检测方法多是采集和分析病例的时间和空间信息,并基于其统计特征进行建模,这些方法并不能对疾病爆发和疾病传播过程的生物医学机制做出详细的分析。若要对复杂的疾病传播真正做到时空预测,需要在统计模型之外进一步建立疾病传播的机理模型,而在目前的疾病监测中,这一环节尚未受到关注(Jacquez, 2000)。

症状监测系统的长期运行也面临着挑战。有效的疾病监测系统的开发,一

方面需要发展新的疾病监测和检测技术,另一方面也要使监测系统在日常公共健康监测中表现出实用性。在实用性方面,为症状监测系统设定合适的预警门槛是一个非常重要的问题。疾病预警发布后,需要花费大量的人力物力进行进一步的疾病调查,因此必须充分考虑疾病检测方法的敏感性和针对性,在有效的疾病暴发检测和发布错误的预警信号之间需求最佳的平衡点(Stoto et al., 2004)。针对这一问题,一种思路是根据对某种传染性疾病威胁程度的预期,对系统检测方法的敏感性和针对性做出相应的调整。另外需要指出的是,虽然疾病监测系统可以有效地辅助流行病学家和公共健康部门对疾病暴发的评估,但是它只能作为医生诊断、实验室检测和其他公共健康监测手段的辅助工具,而不能完全代替这些传统手段的功能,即疾病监测系统并不能独立完成对疾病暴发的全部调查工作。

加拿大安大略省的实时症状监控系统(RODS)试验计划是政府、公共健康部门和高校研究机构相互合作开发和运行基于地理信息系统的疾病症状监测系统的一个范例。在发展中国家,基于地理信息和公共健康信息的疾病监测系统尚未广泛使用。但是近年来,一些国际性事件,如全球性流感暴发、许多国家中的持续性恐怖袭击、禽流感的暴发等,已经使公众意识到了对灾难性的流行性疾病暴发进行监测的重要性。因此,疾病监测系统的发展和改进应当受到各个国家政府部门的高度重视。同时,基于空间的流行病研究技术的发展也应受到足够的重视,从而实现对疾病传播的空间和时间结果的更有效的分析。

致谢

本研究得到了加拿大自然科学和工程基金积及 Geomatic for Informed Decision(GEOIDE):National Center of Excellence 项目的支持。本章部分内容的英文版曾发表于 2011 年的 Annals of Geographic Information System。感谢 KFL&A 地区的公共健康部门提供数据,感谢田杰的数据模拟帮助,以及宋明洁和刘文宝的中文翻译。

参 考 文 献

- Bartholomew, D. J. 1956. A sequential test of randomness for events occurring in time or space. *Biometrika*, 43, 64–78.
- Besag, J. and Newell, J. 1991. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society A*, (154), 154, 143–55.
- Bravata, D. M. ,K. M. McDonald, W. M. Smith. 2004. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. *Annals of Internal Medicine* 140(11):910–922.
- Buehler, J. W. ,R. L. Berkelman, D. M. Hartley, C. J. Peters. 2003. Syndromic surveillance and biot-

- errorism-related epidemics. *Emerging Infectious Disease* 9(10) :1197 – 1204.
- Chen, D. J. Cunningham, K. Moore, J. Tian. 2011. Spatial and temporal aberration detection for disease outbreak in syndromic surveillance. *Annals of Geographic Information System*. 17(4) :211 – 220
- Cooper, D. 2007. Case study: use of tele-health data for syndromic surveillance in England and Wales. *Disease surveillance: a public health informatics approach*. J. S. Lombardo and D. L. Buckridge, Wiley-Interscience :335 – 362.
- Costa, M. A and R. M. Assuncao. 2005. A fair comparison between the spatial scan and the Besag- Newell Disease clustering tests. *Environmental and Ecological Statistics*, 12 :301 – 319.
- Dafni, U. G. ,S. Tsiodras, D. Panagiotakos, K. Gkolfinopoulou, G. Kouvatseas, Z. Tsourtzi, and G. Saroglou. 2004. Algorithm for statistical detection of peaks-Syndromic surveillance system for the Athens 2004 Olympic games. *Morbidity and Mortality Weekly*.
- Das, D. ,D. Weiss, F. Mostashari, and T. Treadwell. 2003. Enhanced drop-in syndromic surveillance in New York City following September 11, 2001. *Journal of Urban Health* 80 :i76 – i88.
- Elliot, P. ,and D. Wartenberg. 2004. Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives* 9 :998 – 1006.
- Feinberg, S. E. ,and G. Shmueli. 2005. Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. *Statistics in Medicine* 24 :513 – 529.
- Gestland, P. H. ,R. M. Gardner, F. -C. Tsu, J. U. Espino, R. T. Rolfs, B. C. James, W. W. Chapman, A. W. Moore, and M. M. Wagner. 2003. Automated syndromic surveillance for the 2002 winter Olympics. *Journal of the American Medical Informatics Association* 10(6) :547 – 554.
- Graham, A. J. ,P. M. Atkinson, and F. M. Danson. 2004. Spatial analysis for epidemiology. *Acta Tropica* 91 :219 – 225.
- Haykin, S. 2002. *Adaptive Filter Theory*, Prentice Hall.
- Henning, K. J. 2004. Overview of syndromic surveillance. *What is syndromic surveillance? Morbidity and Mortality Weekly*(53) :5 – 11.
- Hefferman, R. ,F. Mostashari, D. Das, A. Karpati, M. Kulldorff, and D. Weiss. 2004. Syndromic surveillance in public health practice, New York City. *Emerging Infectious Disease*, 10 :858 – 864.
- Hurley, S. E. ,T. M. Saunders, R. Nivas, A. Hertz, and P. Reynolds. 2003. Post office box addresses: a challenge for geographic information systems-based studies. *Epidemiology* 14(4) :386 – 391.
- Hutwagner, L. ,T. Browne, G. M. Seeman, and A. T. Fleischauer. 2005. Comparing Aberration Detection Methods with Simulated Data. *Emerging Infectious Disease* 11(2) :314 – 316.
- Jacquez, G. M. 2000. Spatial analysis in epidemiology; nascent science or a failure of GIS? *Journal of Geographical Systems* 2 :91 – 97.
- Kamel-Boulos, M. N. 2004. Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom. *International Journal of Health Graphics* 3(1).
- Kleinman, K. ,R. Lazarus, and R. Platt. 2004. A general linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American*

- Journal of Epidemiology*, 159, 217 – 224.
- Kulldorff, M. 1997. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26, 1481 – 1496.
- Kulldorff, M. 2001. Prospective Time Periodic Geographical Disease Surveillance Using a Scan Statistic. *Journal of the Royal Statistical Society, Series A*, 164, 61 – 72.
- Kulldorff, M. 2015. *SATScan User Guide V9.4*. <http://www.satscan.org>
- Lai, P. C. , C. M. Wong, A. J. Hedley, S. V. Lo, P. Y. Leung, J. Kong, and G. M. Leung. 2004. Understanding spatial clustering of severe acute respiratory syndrome (SARS) in Hong Kong. *Environmental Health Perspectives* 112(15) :1550 – 1556.
- Lazarus, R. , K. Kleinman, I. Dashevsky, C. Adams, P. Kludt, A. DeMaria, and R. Platt. 2002. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. *Emerging Infectious Disease*. 8(8) :753 – 760.
- Lee, G. , and P. A. Rogerson. 2007. Monitoring Global Spatial Statistics. *Journal of Stochastic Environmental Research and Risk Assessment*, 21, 545 – 553.
- Lombardo, J. , H. Burkom, E. Elbert, S. Magruder, S. H. Lewis, W. Loschen, J. Sari, C. Sniegoski, R. Wojcik, and J. Pavlin. 2003. A systems overview of the electronic surveillance system for the early notification of community-based epidemics (ESSENCE II). *Journal of Urban Health* 80(2) :i32 – i42.
- Lombardo, J. S. , and D. L. Buckeridge, Eds. 2007. *Disease surveillance: a public health informatics approach*, Wiley-Interscience.
- Mandl, K. D. 2004. Implementing syndromic surveillance: a practical guide informed by early experience. *Journal of the American Medical Informatics Association* 11(2) :141 – 150.
- Moore, K. 2004. Real-time syndromic surveillance in Ontario, Canada: the potential use of emergency departments and telehealth. *European Journal of Emergency Medicine* 11(1).
- Dijk, A. V. , J. Aramini, G. Edge, and K. Moore. 2009. *Real-time surveillance for Respiratory disease outbreaks, Ontario, Canada*. *Emerging infectious diseases*. 15(5) :799 – 801.
- Moore, K. M. , B. L. Edgar, and D. McGuinness. 2008. Implementation of an automated, real-time public health surveillance system linking emergency departments and health units: rationale and methodology. *Canadian Journal of Emergency Medicine*, 10, 114 – 119.
- Nordin, J. D. , M. J. Goodman, M. Kulldorff, D. P. Ritzwoller, A. M. Abrams, K. Kleinman, M. J. Levitt, J. Donahue, R. Platt. 2005. Simulated anthrax attacks and syndromic surveillance. *Emerging Infectious Disease* 11(9) :1394 – 1398.
- Olson, K. L. , M. Bonetti, M. Pagano, and K. D. Mandl. 2005. Real time spatial cluster detection using interpoint distances among precise patient locations. *Medical Informatics and Decision Making* 5.
- Openshaw, S. M. Charlton, A. W. Craft, J. M. Birch. 1988. An investigation of leukaemia clusters by use of a geographical analysis machine. *The Lancet* 1;272 – 73.
- Platt, R. , C. Boccino, B. Caldwell, R. Harmon, K. Kleinman, R. Lazarus, A. F. Nelson, J. D. Nordin, D. P. Ritzwoller. 2003. Syndromic surveillance using minimum transfer of identifiable data: the example of the national bioterrorism syndromic surveillance demonstration program. *Journal of Ur-*

- ban Health 80(2) :i25 – i31.
- Rogerson, P. A. 1997. Surveillance systems for monitoring the development of spatial patterns. *Statistics in Medicine*, 16, 2081 – 2093.
- Rogerson, P. A. 2001. Monitoring point patterns for the development of space-time clusters. *Journal of the Royal Statistical Society*, 164, 87 – 96.
- Rogerson, P. A. , and I. Yamada. 2004. Monitoring change in spatial patterns of disease: comparing univariate and multivariate cumulative sum approaches. *Statistics in Medicine*, 23, 2195 – 2214.
- Rogerson, P. A. , and I. Yamada. 2009. *Statistical Detection and Surveillance of Geographic Clusters*. Boca Raton: Chapman & Hall/CRC.
- Rolfhamre, P. , K. Grabowska, K. Ekdahl. 2004. " Implementing a public web based GIS service for feedback of surveillance data on communicable diseases in Sweden. " *Infectious Diseases* 4:1471 – 2334.
- Sonesson, C. , D. Bock. 2003. " A review and discussion of prospective statistical surveillance in public health. " *Journal of the Royal Statistical Society, Series A*:166;5 – 21.
- Sosin, D. M. 2003. Draft framework for evaluating syndromic surveillance systems. *Journal of Urban Health* 80(2) :i8 – i13.
- Stoto, M. A. , M. Schonlau, and L. T. Mariano. 2004. Syndromic surveillance: Is it worth the effort? *Chance* 17 (1) :19 – 24.
- Tsui, F. , J. U. Espino, and V. M. Dato. 2003. Technical description of RODS: A real-time public health surveillance system. *Journal of the American Medical Informatics Association* 10: 399 – 408.
- Wong, W. , A. Moore, G. Cooper, and M. Wagner. 2003. WSARE: what's strange about recent events? *Journal of Urban Health* 80(2) :i66 – i75.
- Wong, W. , A. Moore, G. Cooper, and M. Wagner. 2005. What's strange about recent events (WSARE) : An algorithm for early detection of disease outbreaks. *Journal of Machine Learning Research* 6:1961 – 1998.
- Zelicoff, A. , Brillman, J. , and D. Forslund. 2001. The rapid syndrome validation project (RSVP). *Proceeding of American Medical Informatics Association Symposium*, 2001 ;771 – 775.