

扫描统计量方法在肺结核发病分布中的应用进展

黄莉^{1,2}, 许琳¹, 李石柱²

1. 云南省疾病预防控制中心, 云南 昆明 650022; 2. 中国疾病预防控制中心寄生虫病预防控制所

摘要: 扫描统计量是空间流行病学中可以用来评价或预测疾病聚集性的一种行之有效的方法,越来越多的研究人员利用扫描统计量方法进行肺结核聚集性及影响因素研究,从空间角度揭示肺结核的分布与传播规律。扫描统计量方法在预测肺结核的暴发方面也有广泛的应用前景,结合结核分枝杆菌基因分型结果还可探测正在发生的区域传播、确定近期传播和识别传播链。目前,在国内仍是值得深入研究的重要课题。本文阐述扫描统计量的基本概念、方法原理和在肺结核领域的应用现状。

关键词: 扫描统计量; 肺结核; 应用

中图分类号: R521 **文献标识码:** A **文章编号:** 1006-3110(2017)06-0766-04 **DOI:** 10.3969/j.issn.1006-3110.2017.06.037

Progress in application of scan statistic in the distribution of pulmonary tuberculosis

HUANG Li*, XU Lin, LI Shi-zhu

* Yunnan Provincial Center for Disease Control and Prevention, Kunming, Yunnan 650022, China

Corresponding author: LI Shi-zhu, E-mail: lysz@chinacdc.cn

Abstract: Scan statistic is an effective method used in spatial epidemiology to evaluate or predict disease aggregation. More and more researchers use this method to study the aggregation of tuberculosis and its influencing factors and reveal the distribution and transmission regularity of tuberculosis from space perspective. Scan statistic has an extensive application prospect in the prediction of tuberculosis outbreak, and combining scan statistic with the genotyping results of *Mycobacterium tuberculosis* can also detect the regional spread, recent transmission and transmission chain. At present, it is still an important subject worthy of further study in China. This paper describes the basic concept, method and principle of scan statistic and the current application status in tuberculosis.

Key words: scan statistic; tuberculosis; application

结核病(tuberculosis, TB)主要是由结核分枝杆菌导致,通常感染并破坏肺(称“肺结核”),是严重危害人类健康的呼吸道传染病,与艾滋病、疟疾一起被世界卫生组织列为全球三大公共卫生问题^[1]。近年来,在流行病学研究领域,空间信息技术已逐步应用于分析疾病的空间分布特征、空间聚集性、绘制疾病风险地图等方面,在疾病防控中发挥了重要作用^[2]。肺结核作为一种传染病,与当地地理环境、经济状况、人口特征等都存在一定的联系,而这种空间相关特点成为空间统计学分析的前提^[3]。

扫描统计量是空间流行病学中可以用来评价或预测疾病聚集性的一种行之有效的方法^[4]。SaTScan 软件为这种空间统计分析提供了很好的技术支撑。目前,越来越多的研究人员利用扫描统计量方法从定量水平探讨肺结核的聚集性、时空分布格局以及与环境、

地理暴露因素间的相关性,预测其流行趋势,从空间角度揭示肺结核的分布与传播规律,进而制定合理有效的预防与控制策略。

1 扫描统计量简介

1.1 扫描统计量的发展历史 扫描统计量的概念由 Naus 于 1965 年首次提出,即用事先选定的时间区间扫描整个观察期所得到的病例数的最大值。1995 年, Kulldorff 和 nagarwalla 针对空间扫描统计量提出一种基于似然比检验的广义数学模型。1998 年, Kulldorff 提出了时空统计量,2001 年提出了前瞻性时空扫描统计量,2005 年进一步提出前瞻性时空重排扫描统计量。

1.2 扫描统计量的应用 扫描统计量是一种广泛应用于通讯、天文、犯罪学、气象、金融、医学等的统计学方法,其目的是探测局部时间和/或空间事件发生数的增加,并检验这种增加是不是由于随机变异造成的。扫描统计量在医学方面的应用主要是实现疾病监测、

作者简介: 黄莉(1981-),女,硕士在读,主管医师,研究方向:疾病控制及项目管理。

通信作者: 李石柱, E-mail: lysz@chinacdc.cn。

早期预警、聚集性探测,其被广泛应用于传染性疾病和非传染性疾病的空间分布研究,不仅可判断疾病在时间、空间或时空分布上是否存在聚集性,还能对聚集位置进行定位。

1.3 扫描统计模型的选择 SatScan 中的扫描统计模型比较多,选用扫描统计量时需要注意适用条件,在分析时应根据具体资料的特点来判别选用哪种模型。对于离散性数据扫描统计:若是频数数据可选择泊松模型(Poisson model);若是 0/1 二分类数据可选择二项分布模型(Bernoulli model);若仅有病例数则选择时空重排模型(space-time permutation model);若是分类数据则选择多项式模型(multinomial model);若是有序分类数据可选择等级模型(ordinal model);若是生存数据可选择指数模型(exponential model),若为其它类型的数据则选择正态分布模型(normal model)。对于连续性数据扫描统计可选用连续性泊松模型(continuous Poisson model)。这些模型都采用对数似然 LLR (likelihood ratio) 进行假设检验,蒙特卡洛随机化方法模拟 P 值的大小。另外,在分析前应充分熟悉 SaTScan 软件功能和各种模型所需的数据库格式和要求。

2 方法原理

扫描统计量的基本思想是:设定一个扫描窗口,该窗口可在时间和(或)空间移动,窗口的大小和位置均处于动态变化之中。扫描统计量包含时间扫描统计、空间扫描统计和时空扫描统计。时间扫描统计量的扫描窗口为一定的时间长度,空间扫描统计量的扫描窗口为一定的地理区域,时空扫描统计量的扫描窗口为圆柱形,圆柱的底对应地理区域,圆柱的高对应时间。对每一个扫描窗口,根据实际发病数和人口数可计算出预期发病数,然后利用扫描窗口内外的实际发病数和预期发病数构造检验统计量 LLR。在所有扫描窗口中选出 LLR 最大的窗口,这一窗口即为具有最强聚集性的窗口。然后再考虑多重比较问题,在控制假阳性率的前提下,对该窗口的统计学意义进行评价,即采用蒙特卡罗法产生模拟数据集,根据真实数据集在模拟数据集中的位序得到概率 P 值。应用 Kulldorff 开发的软件(<http://www.satscan.org>)实现扫描统计量的参数估计及统计学检验。因 SaTScan 软件不能将分析结果可视化,常借助于 ArcGIS 软件实现分析结果的可视化。

3 应用现状

3.1 肺结核聚集性研究 研究者广泛应用扫描统计量方法研究肺结核的聚集性和发病分布特征,结果表明肺结核的分布不是随机的,具有时间、空间或时空聚集性。这些聚集区域存在较高的结核病负担和结核传播的风险,为结核病控制的重点地区。Nunes^[5]对葡萄牙新发结核病例进行时空扫描,结果确定 2000–2004 年在 3 个地区结核病的发病存在高值聚集,并确定 2002 年是全国性的结核病高发时段。Zaragoza 等^[6]对墨西哥 2006–2010 年结核病进行时空分布研究,结果发现 9 个显著性聚集区。Tadesse 等^[7]对埃塞俄比亚达巴地区 2010 年 10 月–2011 年 9 月确诊的涂阳结核病例进行空间和时空扫描,检测到两个聚集区域。Zhao 等^[8]对 2005–2011 年中国各省肺结核登记进行时空扫描分析,结果表明新涂阳肺结核登记率和结核病总登记率在时空分布上均存在聚集性,一级聚集出现于 2006–2008 年以中国南部为中心的地区,二级聚集位于中国东北部和西部地区。Wang 等^[9]对 2005–2010 年山东省临沂市 180 个乡镇的 35 308 例肺结核病例进行时空扫描研究,发现一个最可能的肺结核高发集群和 9 个次级高发群。Dangisso 等^[10]和 Tiwari 等^[11]的研究也表明肺结核在特定的时间和地理位置发生了聚集,这一发现可用于指导制定针对性的结核病控制策略。

3.2 肺结核空间聚集性影响因素研究 Rakotosamimanana 等^[12]应用空间扫描对马达加斯加岛塔那那利佛的肺结核进行时空分析,进而分析与结核病相关的知识、态度和行为对肺结核分布的影响。Randremana 等^[13]的研究则发现肺结核聚集性的变化与社会经济和病例管理等因素有关。Oren 等^[14]应用 SaTScan 对美国华盛顿 King 县 2004–2008 年的结核病例进行聚集性分析,进而采用多水平模型确定街区水平 SEP(socioeconomic position index)与结核病聚集性之间的关系。Mills 等^[15]的研究发现有呼吸道接触经历的人群高度聚集容易发生结核病聚集和再感染事件。Prussing 等^[16]对美国马里兰州 2004–2010 年培养阳性并进行了基因分型的结核病例进行时空扫描分析,结果发现 2 个结核病聚集区域,这两个区域的共同点是贫穷和拥挤,但是危险人口和行为各异。Li 等^[17]评价新流动人口对北京市结核病空间分布的影响,发现结核病与流动人口和流动人口的增长率存在着明显的相关关系,且新流动人口对结核病的影响要大于当地原有的流动人口。路丽苹等^[18]对上海松江 2006–2011 年新登记肺结核时空分布规律及影响因素进行研究,结果显示松江区 TB 的分布具有一定的时空聚

集性。东部工业区是流动 TB 的聚集区域,西南五镇是本地 TB 的聚集区域。对于本地患者,农民职业者及就诊延误与空间聚集区呈独立相关;流动患者中,职业为蓝领工人、年龄小于 60 岁及来源于中国中西部地区是形成空间聚集区的危险因素。

3.3 结核分枝杆菌基因聚集性研究 空间分析和分子生物学技术相结合可发现正在发生的区域传播^[19]。Moonan 等^[20]应用基因分型和地理空间扫描评估美国近期结核分枝杆菌传播,对美国 2005 年 1 月-2009 年 12 月培养阳性结核病例进行基因分型,与空间扫描结果匹配确定近期传播病例,大约 4 个结核病例中有 1 个归因于近期传播。近期传播的危险人群为男性,出生于美国,少数民族或是宗教团体成员,滥用药物及无家可归者。Baker 等^[21]对美国和墨西哥边界结核病基因型聚集特征进行研究,基因型聚集区定义为有 2 例或以上处于同一地理聚集区域的基因分型相同的病例。这种研究方法可以帮助识别整个研究期间可能的传播链。刘媛等^[22]对江苏省阜宁县 681 例活动性肺结核患者分子成簇与空间聚集进行研究,结果 169 株结核分枝杆菌中,有 27 株成簇,分别分布在 12 个簇,提示簇内患者之间存在近期传播。Ricks 等^[23]评估美国移民中的结核病负担。首先用扫描统计量探测出局部聚集区域,并与基因分型结果匹配,如果移民病例不属于聚集区域则判断为复发,结果 5 个移民结核病例中有四个是复发病例,这对美国的结核病消除带来巨大的挑战。

3.4 肺结核预测性研究 Kammerer 等^[24]应用空间扫描统计量等统计学方法,成功探测到了 9 个已知结核暴发中的 8 个,其中 6 个暴发早于当地卫生部门发现之前数月。Althomsons 等^[25]应用 SaTScan 软件确定美国 2006-2010 年间具有特定基因型的结核病例的空间聚集性,进而预测结核病的暴发。国内仅见使用扫描统计量方法进行其它传染病暴发预测的报道。

4 讨论

4.1 空间/时间扫描与时空扫描之间的差异 空间扫描统计量方法有助于监测结核病聚集性的发生,Tiwari 等^[26]应用 GIS 和扫描统计量探测印度 Dehradun 市结核病例的空间分布,结果发现结核病高发率聚集区位于 Dehradun 市的 7 个病房。而时空扫描统计量分析的结果考虑了时间维度,为单纯空间扫描统计量分析的有益补充。如山珂等^[27]利用 SaTScan 软件研究中国 2005-2012 年肺结核发病的时空分布情况,分析显示肺结核发病呈现明显的时空聚集性,其高发区

域存在一定的时空稳定性。殷菲等^[28-29]的研究结果显示,时空聚集性探测比单纯时间聚集性探测更为敏感,在暴发处于局部小范围时就已经发出了预警。其次,时空聚集性探测可对暴发进行精确的时间和空间定位,与单纯时间聚集性分析相比,时空聚集性分析结果可提供更为详尽的信息,从而有利于公共卫生部门及时采取有效的应对措施。此外,在出现传染病疫情暴发时,时空聚集性方法可以确定聚集性最强,需要密切关注甚至优先采取措施的区域,从而可以使有限的资源得到更为合理的配置。可见,时空扫描统计量进行时空联合的交互分析,更适合对肺结核发病时空聚集性的分析。

4.2 扫描统计量方法与其它空间分析方法的对比 扫描统计量方法可纳入时间因素,因此可以准确地判断高发病区域随着时间的变化趋势。Randremanana 等^[13]分 2 个时期对塔那那利佛肺结核病例进行空间聚集性分析,结果在第二个时期发现出现于第一时期的东部聚集区消失了,取而代之的是位于南部的聚集区($RR=12.73$, $P=0.001$),说明空间聚集区由东部向南部移动了。Touray 等^[30]采用 GIS 和 SaTScan 分析冈比亚 GreaterBanjul 地区 2007-2008 年 1 145 例肺结核的聚集性,两种方法结果一致,均显示存在 2 个区域的高发和低发区。金瑾^[31]和 Hassarangsee 等^[32]应用空间自相关方法和扫描统计量方法分别进行肺结核空间流行特征研究,发现单纯空间扫描统计量的分析结果与局部空间自相关的分析结果比较一致,但又有明显的差别,究其原因,是由于所使用统计量的基本原理不同。

5 展望

扫描统计量方法不仅能动态分析肺结核的时间与空间分布特征,而且可以从全新的角度和方式来研究和认识肺结核,并从其发生和流行的立体环境观察其时空传播的规律。国外广泛应用扫描统计量方法对肺结核在时间、空间或时空方面表现出来的聚集性及聚集性产生的原因进行研究,研究提示肺结核在时间、空间上具有特定的分布模式,其分布特征随环境与社会等因素的变化而变化。国内越来越多的研究人员应用扫描统计量方法进行肺结核发病分布的研究,但对肺结核空间差异产生的原因还较少研究,今后可以从区域水平及个体水平收集肺结核发生的可能危险因素,比较研究中确定的肺结核发病聚集区域与非聚集区域的结核病疫情、防控措施的实施情况、地理环境、经济状况、人口特征、生产生活方式等,进一步探讨肺结核

聚集区形成的原因,从而改进肺结核防控策略,降低肺结核疫情。

扫描统计量方法在预测肺结核的暴发方面也有广泛的应用前景,结合结核分枝杆菌基因分型结果还可探测正在发生的区域传播、确定近期传播和识别传播链。目前,在国内仍是值得深入研究的重要课题。随着学科研究的进一步深入及 GIS、GPS 和 RS 的发展和完善,扫描统计量方法也会得到进一步完善,应用的领域也将更加广泛。

参考文献

- [1] Murray CJ, Ortblad KF, Guinovart C, et al. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013[J]. *Lancet*, 2014,384(9947):1005–1070.
- [2] 康万里, 郑素华. 空间扫描统计在中国菌阳结核病分布中的应用[J]. *中国卫生统计*, 2012,29(4):487–489.
- [3] 唐益, 龚德华, 白丽琼. 湖南省 2003–2011 年活动性肺结核患者登记的空间分析[J]. *中国防痨杂志*, 2012,34(12):764–767.
- [4] 李秀央, 陈坤. 扫描统计量的理论及其在空间流行病学中的应用[J]. *中华流行病学杂志*, 2008,29(8):828–831.
- [5] Nunes C. Tuberculosis incidence in Portugal: spatiotemporal clustering[J]. *Int J Health Geogr*, 2007,6:30.
- [6] Zaragoza BA, Hernandez TM, Bustamante ML, et al. Spatial and temporal distribution of tuberculosis in the State of Mexico, Mexico[J]. *Sci World J*, 2012,2012:570278.
- [7] Tadesse T, Demissie M, Berhane Y, et al. The clustering of smear-positive tuberculosis in Dabat, Ethiopia: a population based cross sectional study[J]. *PLoS One*, 2013,8(5):e65022.
- [8] Zhao F, Cheng S, He G, et al. Space-time clustering characteristics of tuberculosis in China, 2005–2011[J]. *PLoS One*, 2013,8(12):e83605.
- [9] Wang T, Xue F, Chen Y, et al. The spatial epidemiology of tuberculosis in Linyi City, China, 2005–2010[J]. *BMC Public Health*, 2012,12:885.
- [10] Dangisso MH, Datiko DG, Lindtjorn B. Spatio-temporal analysis of smear-positive tuberculosis in the Sidama Zone, southern Ethiopia[J]. *PLoS One*, 2015,10(6):e126369.
- [11] Tiwari N, Adhikari CM, Tewari A, et al. Investigation of geo-spatial hotspots for the occurrence of tuberculosis in Almora district, India, using GIS and spatial scan statistic[J]. *Int J Health Geogr*, 2006,5:33.
- [12] Rakotosamimanana S, Mandrosovololona V, Rakotonirina J, et al. Spatial analysis of pulmonary tuberculosis in Antananarivo Madagascar: tuberculosis-related knowledge, attitude and practice[J]. *PLoS One*, 2014,9(11):e110471.
- [13] Randremananana RV, Sabatier P, Rakotomanana F, et al. Spatial clustering of pulmonary tuberculosis and impact of the care factors in Antananarivo City[J]. *TM & IH*, 2009,14(4):429–437.
- [14] Oren E, Narita M, Nolan C, et al. Neighborhood socioeconomic position and tuberculosis transmission: a retrospective cohort study[J]. *BMC Infect Dis*, 2014,14:227.
- [15] Mills HL, Cohen T, Colijn C. Modelling the performance of isoniazid

preventive therapy for reducing tuberculosis in HIV endemic settings: the effects of network structure[J]. *J R Soc Interface*, 2011,8(63):1510–1520.

- [16] Prussing C, Castillo-Salgado C, Baruch N, et al. Geo-epidemiologic and molecular characterization to identify social, cultural, and economic factors where targeted tuberculosis control activities can reduce incidence in Maryland, 2004–2010[J]. *Public Health Rep*, 2013,128(Suppl 3):104–114.
- [17] Li T, He XX, Chang ZR, et al. Impact of new migrant populations on the spatial distribution of tuberculosis in Beijing[J]. *Int J Tuberc Lung Dis*, 2011,15(2):163–168.
- [18] 路丽苹, 洪建军, 高谦, 等. 上海松江 2006–2011 年新登记肺结核时空分布规律及影响因素[J]. *中华疾病控制杂志*, 2014,18(8):709–714.
- [19] Haase I, Olson S, Behr MA, et al. Use of geographic and genotyping tools to characterise tuberculosis transmission in Montreal[J]. *Int J Tuberc Lung Dis*, 2007,11(6):632–638.
- [20] Moonan PK, Ghosh S, Oeltmann JE, et al. Using genotyping and geo-spatial scanning to estimate recent mycobacterium tuberculosis transmission, United States[J]. *Emerg Infect Dis*, 2012,18(3):458–465.
- [21] Baker BJ, Moonan PK. Characterizing tuberculosis genotype clusters along the United States–Mexico border [Short communication][J]. *Int J Tuberc Lung Dis*, 2014,18(3):289–291.
- [22] 刘媛, 王学高, 刘美伶. 江苏省阜宁县活动性肺结核病患者分子成簇与空间聚集研究[J]. *中华传染病杂志*, 2013,31(8):449–453.
- [23] Ricks PM, Cain KP, Oeltmann JE, et al. Estimating the burden of tuberculosis among foreign-born persons acquired prior to entering the U. S., 2005–2009[J]. *PLoS One*, 2011,6(11):e27405.
- [24] Kammerer JS, Shang N, Althomsons SP, et al. Using statistical methods and genotyping to detect tuberculosis outbreaks[J]. *Int J Health Geogr*, 2013,12(1):15.
- [25] Althomsons SP, Kammerer JS, Shang N, et al. Using routinely reported tuberculosis genotyping and surveillance data to predict tuberculosis outbreaks[J]. *PLoS One*, 2012,7(11):e48754.
- [26] Tiwari N, Kandpal V, Tewari A. Investigation of tuberculosis clusters in Dehradun city of India[J]. *Asian Pac J Trop Med*, 2010;486–490.
- [27] 山珂. 肺结核发病空间聚集分布及影响因素研究[D]. 济南:山东大学, 2014.
- [28] 殷菲, 李晓松, 冯子健, 等. 前瞻性时空扫描统计量在传染病早期预警中的应用[J]. *中华预防医学杂志*, 2007,41(增刊):155–158.
- [29] 殷菲, 冯子健, 李晓松, 等. 基于前瞻性时空重排扫描统计量的传染病早期预警系统[J]. *卫生研究*, 2007,36(4):455–458.
- [30] Touray K, Adetifa I, Jallow A. Spatial analysis of tuberculosis in an urban west African setting: is there evidence of clustering[J]. *Trop Med Int Health*, 2010,15(6):664–672.
- [31] 金瑾. 2010–2012 年济宁市肺结核空间流行病学研究[D]. 济南:山东大学, 2014.
- [32] Hassarangsee S, Tripathi N, Souris M. Spatial pattern detection of tuberculosis: a case study of Si Sa Ket Province, Thailand[J]. *Int J Environ Res Public Health*, 2015,12(12):16005–16018.

收稿日期:2016–11–07