

风险监测

时空扫描统计量在细菌性痢疾监测数据分析中的应用

肖辉¹,肖革新²

(1. 国家食品安全风险评估中心,北京 100022; 2. 中国疾病预防控制中心,北京 102206)

摘要:目的 探讨时空扫描统计量在细菌性痢疾监测数据分析中的应用,掌握疾病发病的时空格局与变化趋势。方法 以2008—2009年全国细菌性痢疾发病数据为例,在描述性分析基础上采用SatScan软件对全国2 910个区县逐月进行时空扫描统计分析,扫描结果利用ArcGIS 10.0软件可视化展示。结果 2008年聚集时间为5~10月,聚集空间分为7个聚类,涉及区县数量分别为1 113、10、8、186、6、100、6;2009年聚集时间为8月,聚集空间分为2个聚类,区县数量分别为1 111和485。结论 时空扫描统计量是一种比经典统计学更有效的用来分析食源性疾病时空格局的模型,结合地理信息系统,能够更加直观、全面地展示发病热点聚集区域,有助于为类似研究提供借鉴和参考。

关键词:时空扫描统计量;食源性疾病;似然对数比;相对危险度

中图分类号:R155;R195 文献标识码:A 文章编号:1004-8456(2014)01-0083-05

Application of space-time permutation scan statistics in bacillary dysentery surveillance

XIAO Hui, XIAO Ge-xin

(China National Center for Food Safety Risk Assessment, Beijing 100022, China)

Abstract: Objective To explore the application of space-time permutation scan statistics in bacillary dysentery surveillance, and understand the space-time pattern and the trend of disease. **Methods** Based on descriptive analysis, the national bacillary dysentery incidence data of 2 910 counties during 2008-2009 was analyzed by space-time scan statistics using SatScan software. The results were visualized using ArcGIS 10.0 software. **Results** The accumulation time of 2008 was from May to October and accumulation space was divided into 7 clusters, which included 1 113, 10, 8, 186, 6, 100 and 6 counties respectively. The accumulation time of 2009 was August, and the accumulation space was divided into 2 clusters which included 1 111 and 485 counties respectively. **Conclusion** The space-time scan statistics was more effective for space-time pattern analysis of food borne disease than the classical statistics. Combined with geographic information system, it could be more intuitive, comprehensive to display the hot spot, and provide reference for similar research.

Key words: Space-time permutation scan statistic; food-borne disease; likelihood ratio; relative risk

细菌性痢疾不仅给个人带来疾病困扰,也给世界带来了严重的社会保健问题和沉重的经济负担^[1],细菌性痢疾(简称菌痢)是由志贺菌属感染引起的以腹泻为主要症状的一种常见食源性肠道传染病,传播途径主要是粪-口传播,多为食物和水源的污染,主要临床表现为发热、腹痛、腹泻、里急后重、脓血样便等症状。该病仍是全球面临的重要公共卫生问题,全球每年有1.65亿人次发病,约110万病例死亡^[2]。在我国也属于常见病、多发病,发病率高居甲、乙类传染病前5位^[3],严重影响居民健康和生活质量。

对于疾病的预防控制,需要预先评估其发病高低

和区域聚集情况,由于经典统计分析要求研究变量必须是随机、独立的,难以考虑到其存在空间自相关性,在描述空间分布时仅比较不同地区发病率的高低,未能考虑到疾病不同时间和地区高发的偶然性,也不能客观定量地对一个大的区域范围发病情况进行聚集分析,易导致人为选择性偏倚,因此时空扫描统计量与经典统计分析相比更具有客观性。时空扫描统计量是一种探索疾病在时空上是否存在聚集性,检验疾病发生在时间和空间上是否随机分布的研究方法,可以综合描述在一个区域内病例报告数量和发病范围等信息,查找可能的疾病高发热点区域。该方法已成功应用于手足口病^[4-5]、出血热^[6]、肿瘤^[7]等疾病研究。本文通过对2008—2009年全国细菌性痢疾时空扫描分析,探测我国细菌性痢疾在区县级地区时空聚集性和热点区域,为类似食源性疾病数据分析提供借鉴和参考。

收稿日期:2013-08-26

作者简介:肖辉 男 高级工程师 研究方向为计算机科学与技术

E-mail: xiaohui@cfsa.net.cn

通信作者:肖革新 男 副研究员 研究方向为信息技术与空间统计

E-mail: xiaogexin@cfsa.net.cn

1 材料与方法

1.1 数据来源

细菌性痢疾病例数据来源于中国 2008—2009 年法定传染病发病与死亡报告,每条记录包括患者的性别、职业、年龄、发病日期、现住址地理编码等信息。人口数据主要来源于国家统计局,空间扫描单位为全国 2 910 个区县,时间扫描单位为月,时间跨度从 2008 年 1 月—2009 年 12 月。用时空 Kulldorff 扫描统计量,用 SaTScanTM 9.1 软件完成分析;地图绘制和结果可视化展示由 ArcGIS 10.0 软件实现。

1.2 研究方法

时空扫描统计量的基本思想:需要同时考虑时间和空间两个维度,即在地理坐标基础上建立一个时空二维圆柱体活动窗口,每个圆柱反应一个可能的聚集区域^[8]。地理位置和区域大小为圆柱底的半径,时间为圆柱高度,通过在地理单位中心连续不断变化半径大小和时间周期,使得整个圆柱窗口随着区域和时间处于动态变化。对于每个窗口,均根据 Poisson 分布计算出理论发病数,通过与时间发病数构造统计量对数似然比(log likelihood ratio, *LLR*),用 *LLR* 来评价扫描窗口内发病数的异常程度^[9],即窗口内和窗口外发病率差异由对数似然比统计量(*LLR*)给出:

$$LLR = \log\left\{ \left(\frac{c}{\mu}\right)^c \left[\frac{(C-c)}{(C-\mu)}\right]^{(C-c)} \right\}$$

这里, *C* 为总病例数, *c* 是在窗内的实际病例数, μ 是在窗内的期望病例数。即风险人口多,发病人数也多;假设细菌性痢疾病例在时间和空间均呈 Poisson 分布,风险人口少,发病人数也少。在空间分布符合 Poisson 分布,即风险人口多,发病人数也多;风险人口少,发病人数也少。实际观测到的发病人数与理论分布比较,如某年发病人数为 *n*,总风险人口为 *N*,按风险人口的多少,随机分配 *n* 到区

县,这样重复模拟分配多次(如 999 次)。窗口半径的变动范围按人口辖区范围划分,半径每变动一次,均产生数量众多的扫描窗口,未来控制假阳性水准,以 *LLR* 值最大值,产生最可能聚集区域(most likely cluster),表明窗口发病数异常程度最高,然后对窗口进行统计学检验,给出相对危险度(relative risk, *RR*)及利用蒙特卡罗法模拟计算 *P* 值,扫描还对其他 *LLR* 较大窗口进行统计学评价,产生次要聚集区域(secondary clusters),从而找出所有可能的发病数据异常的区域。

本次研究利用 SatScan 9.11 软件进行时空扫描统计分析,按照扫描统计分析最大半径为总人口的 30%,时间周期为研究周期的 30%,蒙特卡罗法模拟次数为 999,进行时空扫描分析,当 *LLR* 的 *P* < 0.05,扫描窗口发病数异常程度差异有统计学意义,可以认为该区域具有聚集特征。

本论文数据整理采用 SAS 9.2,时空扫描软件采用 Kulldorff 等开发的 SatScan 9.11 软件(www.satscan.org)进行时空扫描统计分析,扫描结果可视化展示采用 Arc GIS 10.0。

2 结果

2.1 2008—2009 年发病三间分布

2008 和 2009 年发病人数分别共计 313 434 和 272 222 例。

人间分布:发病率男性高于女性,总发病率之比为 1.35;发病人群构成比居前三位分别为散居儿童 32%、农民 24%、学生 12%。年龄分布以 15 岁以下为主,占总发病人数的 45%(141 045/313 434)。

时间分布特征:细菌性痢疾发病高峰为每年的 5~10 月,发病时间呈现明显的季节性,8 月为最高峰,男性月发病率均高于女性。如图 1。

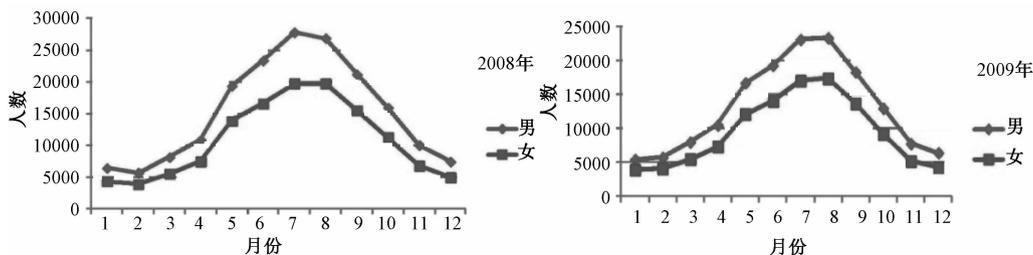


图 1 2008—2009 年发病时间趋势分析

Figure 1 2008-2009 incidence time trend analysis

空间分布:空间分布总体呈现西高东低的发病态势,发病率前 5 位的地区依次为北京、天津、宁夏、新疆、甘肃,发病率分别为 12.73/10 万、9.49/10 万、8.24/10 万、6.06/10 万、5.32/10 万。从空间分

布来看大部分发病率高的省份集中在西部和北方地区,全国发病率空间分布见图 2。

2.2 时空分析结果

对 2008—2009 年间全国 2 910 个区县进行时

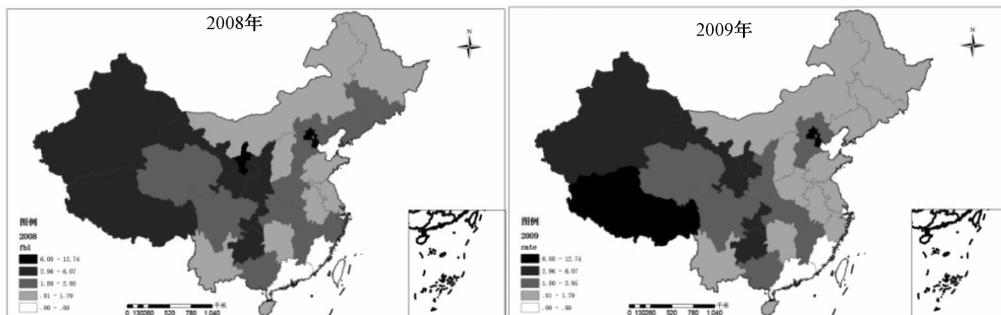


图2 2008—2009年发病率空间分布

Figure 2 The incidence of spatial distribution from 2008 to 2009

空扫描统计分析(以月为单位),结果显示,细菌性痢疾在时间和空间区域上存在明显聚集性。在2008年5月—2008年10月期间共发现7个聚类区

域,2009年8月1日~8月31日期间发现2个聚集区域,具体扫描结果详见表1。

表1 我国2008—2009年区县水平细菌性痢疾时空聚集性分析

Table 1 The level of bacillary dysentery space-time clustering analysis from 2008 to 2009 in China

年份	聚集区域	聚集时间/(年/月/日)	区县个数	RR值	LLR值	P值
2008年	1类	2008/5/1 to 2008/9/30	1 113(新疆、西藏、甘肃、青海、宁夏、内蒙、山西、陕西、四川、北京、河北、河南等)	3.55	45 461.79	<0.001
	2类	2008/5/1 to 2008/10/31	10(浙江)	6.65	4 698.10	<0.001
	3类	2008/5/1 to 2008/10/31	8(湖北)	6.01	3 311.51	<0.001
	4类	2008/5/1 to 2008/10/31	186(贵州、湖南西部、广西)	1.69	1 637.475 8	<0.001
	5类	2008/5/1 to 2008/10/31	6(广东东部)	3.06	1 171.61	<0.001
	6类	2008/6/1 to 2008/10/31	100(江西、湖南东部、福建西部)	1.49	423.98	<0.001
	7类	2008/5/1 to 2008/10/31	6(湖南南部)	2.21	167.97	<0.001
2009年	1类	2009/8/1 to 2009/8/31	1 111(新疆、西藏、甘肃、青海、宁夏、内蒙、山西、陕西、四川、北京、河北、重庆等)	6.32	40 285.04	<0.001
	2类	2009/8/1 to 2009/8/31	485(安徽、浙江、江苏、江西、河南、湖北、湖南等)	2.33	3 614.45	<0.001

2.3 可视化展示

利用 ArcGIS 对上述结果进行可视化展示,如图 3

所示。2008年探测到7个时空聚集区,2009年探测到2个时空聚集区。

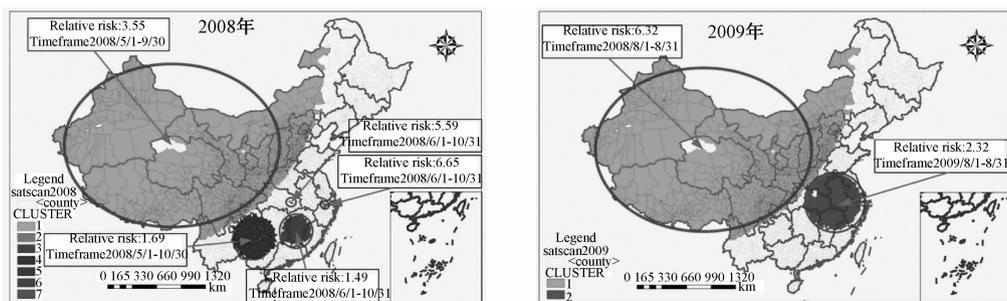


图3 2008年1月—2009年12月时空扫描聚集区域

Figure 3 Time space scanning analysis of gathering area from 2008 January to 2009 December

3 讨论

从三间分布分析来看,发病率男性高于女性,这可能与男性活动范围广,从事社交、生产活动接触病原菌机会多有关。发病年龄分布以5岁以下儿童居多,这与儿童身体发育不成熟,机体免疫力较低,在受到痢疾感染以后容易发病,且儿童防范意识差,容易暴露被感染有关;青壮年发病率较高,可能与活动范围广及接触病原体机会较多有关^[10]。

从发病时间来看主要集中在5~10月,与夏、秋季苍蝇孳生及细菌繁殖,且人们喜食生冷食物有关^[11]。从发病区域来看西部高于中东部,这可能是西部卫生设施水平较低有关^[12]。

利用时空扫描统计分析探索细菌性痢疾聚集性,由于事先未对聚集性的大小、位置、规模作任何假定,避免了选择偏倚,可以最大程度地进行数据挖掘,发现时空聚集性的存在。通过对

2008—2009年全国细菌性痢疾病例进行时空扫描统计分析,在不考虑地区因素情况下,按月份进行扫描分析发现2008年5~10月、2009年8~9月这段时间全国细菌性痢疾高发月份,与报告发病率高发月份、高发季节均一致,能客观地展示细菌性痢疾发病的时间规律,较好地评估细菌性痢疾不同时间周期发病异常增高现象。另外,在空间扫描和时空扫描两种分析方法上均发现多个最可能聚集区域。由于两种统计分析方法在统计结果上差别不大,只是时空扫描方法考虑到时间周期变化因素,分析得到的个别聚集区域有所变化,以及原无统计意义的区县转换为有统计意义,这些情况可能反映出一种亚聚集状态,提示高危因素的存在及变化,在时间预防控制中应给予重视。而分析结果中明确存在聚集现象的区域,在整体上发病较为接近,表明内部特征更为近似,不同于其他地区,提示来自同一总体的概率较大,存在传染病发病异常增多,明显聚集发病现象。

研究表明,2008—2009年西部地区出现聚集性提示发病与卫生条件状况及西部特有的自然环境有关,西部由于经济发展水平比较低,卫生状况比东部差,由细菌性痢疾污染食物导致食源性疾病发病增多,呈现聚集分布,将成为重点防控区域。中部部分区县出现聚集性区域并且呈现移动变化的趋势,有可能与菌株的毒力因子、食物和水源污染、环境因素、气候及抗生素的使用等多种因素有关^[13],并且将继续影响一些局部地区高发。

加强细菌性痢疾流行病学分析,特别是结合地理信息系统进行病例的空间分布、聚集发病特征分析,可以为疫情预测、防控策略的选择和重点防控地区的确定提供依据。本研究通过利用时空扫描统计分析,发现了细菌性痢疾不是随机分布的,存在明显时空聚集性,可以为今后采取针对性的预防控制措施,控制食源性肠道病高发热点区域提供参考依据。时空扫描统计分析方法弥补了单纯流行病学发病率比较的不足,也避免了对食源性疾病高发区域的人为判断,较经典统计分析更有说服力,同时结合地理信息系统,更加直观、全面地展示了发病聚集区域,为以后采取针对性的预防控制措施,提供了科学参考依据。

本研究采用时空扫描统计分析了2008—2009年细菌性痢疾监测数据,在此基础上,需进一步收集人口流动、气候、环境污染、社会经济学等方面的资料,采用空间插值、空间自回归、空间地理加权回归模型、地理风险探测器^[14]等空间统计方法来探讨细菌性痢疾与这些相关因素的关系,并建立时空预

测模型对疾病危害因素进行预测。例如 Wang 等^[15-16]采用地理风险探测器探讨食物、营养及自然环境与出生缺陷影响因素的关系, HU^[17]等采用地理加权回归分析我国手足口病与气象的关系。此外食物环境污染物空间格局分析也是食源性疾病研究的重要组成部分, Cattle^[18]等采用指示克里金技术基于已知抽样点污染物浓度去估计非抽样点污染物浓度。 Asmari^[19]等采用泊松克里金插值对2003—2007年336个县食道癌发病数据进行分析,得出了 Ardebil、Mazandaran 和 Kordestan 三省与其他省相比有较高的风险。 Adhikar^[20]等人采用指标和概率克里格方法描述印度德里市 Najafgarh 街区地下水铜、铁、锰污染情况,有助于为食品安全风险评估获得基础性资料。综上所述,空间统计技术在疾病研究中的应用还大有潜力。

参考文献

- [1] Lanzas C, LU Z, Grohn Y T. Mathematical modeling the transmission and control of foodborne pathogens and antimicrobial resistance at preharvest[J]. *Foodborne Pathog Dis*, 2011, 8(1):1-10.
- [2] Sharma A, Singh S K, Kopi L. Molecular epidemiological characteristics of shigellasp plisolated from river narmada during 2005-2006[J]. *Environ Health*, 2009, 71(6):61-66.
- [3] ZHONG W, CHANG ZH, ZHANG J. Analysis on bacillary dysentery surveillance data collected from the national surveillance system in 2007[J]. *Chinese Journal of Epidemiolog*, 2010, 31(3):304-307.
- [4] DENG T, HUANG Y, YU S, et al. Spatial-temporal clusters and risk factors of hand, foot, and mouth disease at the district level in Guangdong Province, China. *PLoS One*. 2013; 8(2): e56943. DOI: 10.1371/journal.pone.0056943.
- [5] 肖革新,胡跃华,马家奇,等.中国2008—2011年手足口病空间聚集性及变化趋势研究[J]. *中华流行病学杂志*, 2012, 33(8):808-812.
- [6] LIN H L, LIU Q Y, GUO J Q, et al. Analysis of geographic distribution of HFRS in Liaoning Province between 2000 and 2005[J]. *BMC Public Health* 2007, 7:207.
- [7] Kulldorff M, Athas W F, Feurer E J, et al. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico[J]. *American Journal of Public Health*, 1998, 88(9):1377-1380.
- [8] Kulldorff M, Heffernan R, Hartman J, et al. A space-time permutation scan statistic for disease outbreak detection[J]. *PLoS Medicine*, 2005, 2(3):e59.
- [9] Mostashari F, Kulldorff M, Hartman J J, et al. Dead bird clusters as an early warning system for West Nile virus activity[J]. *Emerging Infectious Diseases*, 2003, 9(6):641-646.
- [10] 李永红.细菌性痢疾流行病学特征的变迁[J]. *内科*, 2013, 8(1):58-59.
- [11] 关鹏,曲波,何苗,等.关联规则及其在细菌性痢疾风险预测中的应用[J]. *现代预防医学*, 2007, 34(11):2007-2008.
- [12] 张华一.青海省2005—2009年细菌性痢疾监测分析[D].太

原:山西医科大学,2012.

- [13] 贾蕾,黎新宇,刘桂荣,等.北京气象因素与细菌性痢疾发病关系的研究[J].现代预防医学,2007,34(13):2470-2480.
- [14] WANG J F, HU Y. Environmental health risk detection with GeogDetector [J]. Environmental Modelling & Software, 2012, 33:114-115.
- [15] WU J L, WANG J F, MENG B, et al. Exploratory spatial data analysis for the identification of risk factors to birth defects [J]. BMC Public Health, 2004, 4:23.
- [16] WANG J F, LI X H, Christakos G, et al. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China [J]. International Journal of Geographical Information Science, 2010, 24(1):107-127.
- [17] HU M G, LI Z J, WANG J F, et al. Determinants of the incidence

of hand, foot and mouth disease in China using geographically weighted regression models [J]. PLoS ONE, 2012, 7(6):e38978.

- [18] Cattle J A, McBratney A B, Minasny B. Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination [J]. Journal of Environmental Quality, 2002, 31(5):1576-1588.
- [19] Asmarian N S, Ruzitalab A, Amir K, et al. Area-to-Area Poisson Kriging analysis of mapping of county-level esophageal cancer incidence rates in Iran [J]. Asian Pacific Journal of Cancer Prevention, 2013, 14(1):11-13.
- [20] Adhikary P P, Dash Ch J, Bej R, et al. Indicator and probability kriging methods for delineating Cu, Fe, and Mn contamination in groundwater of Najafgarh Block, Delhi, India [J]. Environmental Monitoring and Assessment, 2011, 176(1-4):663-676.

《中国食品卫生杂志》2014 年征稿征订启事

《中国食品卫生杂志》创刊于 1989 年,由卫生部主管,中华预防医学会、中国疾病预防控制中心营养与食品安全所主办,国内公开发行。是中国食品卫生专业唯一的中文核心期刊、中国科技核心期刊。被中国知网(CNKI)中国期刊全文数据库收录。

期刊影响力:该杂志于 2003—2004、2005—2006 和 2007—2008 年度连续获得中华预防医学会优秀期刊一等奖;获得卫生部首届医药卫生优秀期刊三等奖。在预防医学领域的影响因子排名:中国知网多年排在第 6 名,万方数据库统计多次排在第 3 名;2013 年版影响因子(中国知网) 0.968,87 种预防医学期刊中排第 8 名。

所设栏目有:专家述评、论著、研究报告、实验技术与方法、监督管理、调查研究、食品安全标准、风险评估、风险交流、食物中毒、综述等,及时报道食品卫生领域的最新科研动向、食品安全监管、事件处理以及国内、国际有关食品卫生的政策、法律法规和标准等最新信息。

刊发周期:审稿通过后一般 3 个月内刊出;省部级(含)以上基金资助论文、具有创新性的优秀论文可加急审稿、优先发表。

欢迎投稿、欢迎订阅

在线投稿:<http://www.zgspws.com>

订阅:由北京报刊发行局发行,邮发代号为 82-450。双月刊,逢单月末出版。2014 年每期定价 28 元,全年 168 元。如漏订可汇款到编辑部订阅过刊(免费邮寄,挂号加收 3 元)。

地址:北京市朝阳区潘家园南里 7 号 行政楼 208 室

电话/传真:010-67791542 **邮政编码:**100021 **E-mail:**spws462@163.com