

空间相对危险度估计方法在传染病风险评估中的应用

张业武 郭青 王晓风 于萌 苏雪梅 董言 张春曦

【导读】 介绍可变带宽核密度算法、疾病空间相对危险度估计方法以及空间危险度统计学检验等在疾病风险评估中应用原理,利用2013年云南省鲁甸及周边县区感染性腹泻病网络直报数据中其他感染性腹泻病的空间相对危险度进行估计并绘制疾病图。结果显示其他感染性腹泻病高风险热点区域主要集中于研究地区的东南部大片区域内,表明基于可变带宽的核密度疾病空间危险度估计方法结合疾病制图技术,可为确定重点防控人群和区域提供直观可视化的工具。

【关键词】 空间相对危险度;风险评估;可变带宽核密度估计

Application of spatial relative risk estimation in communicable disease risk evaluation Zhang Yewu¹, Guo Qing¹, Wang Xiaofeng¹, Yu Meng¹, Su Xuemei¹, Dong Yan², Zhang Chunxi¹. 1 Center for Disease Surveillance and Information Services, Chinese Center for Disease Control and Prevention, Beijing 102206, China; 2 Department of Information, Xinjiang Uygur Autonomous Regions Center for Disease Control and Prevention

Corresponding author: Zhang Chunxi, Email: zhangcx@chinacdc.cn

This work was supported by a grant from the National Science and Technology Major Project of China (No. 2013ZX10004218-06-006).

【Introduction】 This paper summarizes the application of adaptive kernel density algorithm in the spatial relative risk estimation of communicable diseases by using the reported data of infectious diarrhea (other than cholera, dysentery, typhoid and paratyphoid) in Ludian county and surrounding area in Yunnan province in 2013. Statistically significant fluctuations in an estimated risk function were identified through the use of asymptotic tolerance contours, and finally these data were visualized through disease mapping. The results of spatial relative risk estimation and disease mapping showed that high risk areas were in southeastern Shaoyang next to Ludian. Therefore, the spatial relative risk estimation of disease by using adaptive kernel density algorithm and disease mapping technique is a powerful method in identifying high risk population and areas.

【Key words】 Spatial relative risk; Risk evaluation; Adaptive kernel density estimation

疾病的发生经常受到当地自然环境和社会生活条件的影响,因此研究疾病空间分布可为其病因及流行因素等提供重要线索^[1-2]。而研究疾病空间分布通常采用标点地图、疾病地区分布图和疾病传播蔓延图等对疾病的发生、发展进行描述,并由此计算出疾病空间分布密度、相对危险度或探测是否存在空间聚集性等^[3]。近年来疾病空间统计和推断方法得到快速发展,而利用核密度平滑相对危险函数估计空间相对危险度是方法之一^[4-6]。为此本研究简

要介绍可变带宽核密度算法在疾病空间相对危险度估计应用原理,以及空间危险度统计学检验方法,并结合实例说明该方法在传染病风险评估研究中的具体应用。

基本原理

为估计疾病风险的空间变化,采用核密度平滑相对危险函数来估计其空间相对危险度。核密度估计方法为非参数估计方法,可根据给定的样本点集合,估计空间随机变量的分布密度函数。由于核密度估计从数据样本本身出发研究数据分布特征,而不必考虑数据分布先验知识,也无需对数据分布附加任何假定,因此在空间统计学理论和应用领域受到高度关注。空间相对危险度估计基本思想是采用可变带宽核密度平滑函数分别计算病例和对照空间

DOI: 10.3760/cma.j.issn.0254-6450.2015.05.025

基金项目: 国家科技重大专项(2013ZX10004218-06-006)

作者单位: 102206 北京, 中国疾病预防控制中心疾病监测与信息服务中心(张业武、郭青、王晓风、于萌、苏雪梅、张春曦); 新疆维吾尔自治区疾病预防控制中心信息科(董言)

通信作者: 张春曦, Email: zhangcx@chinacdc.cn

分布的概率密度,通过比较病例和对照的密度,从而获得疾病空间分布的相对危险度,再对各空间格点的相对危险度进行统计学检验,获取各格点上的统计量及其对应的概率P值,绘制P值等值线,最终获得疾病高风险热点地图^[5]。

1. 可变带宽核密度平滑法:假设n个二维空间的观察点 x_1, x_2, \dots, x_n 来自未知密度f,采用核密度平滑技术,该密度可通过公式进行估计

$$\hat{f}(z) = \frac{1}{n} \sum_{i=1}^n h_i^{-2} K\left(\frac{z - X_i}{h_i}\right)$$

式中K为核函数,一般选择径向对称的概率密度函数, h_i 为平滑参数或第i个观察值带宽。带宽h选择是估计 $\hat{f}(z)$ 最关键环节,一般有固定带宽(fixed bandwidth)或可变带宽(adaptive bandwidth)两种选择。后者允许每个观测值有不同的带宽,较前者的估计量更具有弹性。本文采用Abramson建议的平方根法则(square-root law)选择局部带宽因子。其优点是对估计量局部偏差收敛的速度快^[5,7]。Abramson可变带宽计算公式

$$h_i = h_0 f(X_i)^{-1/2} \gamma^{-1}$$

式中 h_0 是二阶平滑算子,称之为全局带宽; γ 为 $f(X_i)^{1/2}$ 项的简单几何平均值,用以减轻 h_i 对样本尺度的依赖^[5,7]。

2. 相对危险函数:假设 x_1, x_2, \dots, x_{n1} 和 y_1, y_2, \dots, y_{n2} 分别代表病例和对照的空间坐标点,根据Bithell^[8]的建议,空间格点相对危险度函数等于(未知的)病例核密度函数 $f(z)$ 和对照核密度函数 $g(z)$ 的比值,其公式

$$r(z) = \frac{f(z)}{g(z)}$$

为保证结果的对称性,Kelsall和Diggle^[6]建议采用对数危险度函数 ρ 代替原始值,即令 $\rho = \log(r)$ 以便于计算或进行统计学检验。利用样本数据可获得总体的估计值 $\hat{r} = \hat{f} / \hat{g}$ 。

3. 空间危险度统计学检验和P值等值线:采用Davies和Hazelton^[4]建议的渐近正态法公式,计算各空间格点上的可变带宽危险函数 $\hat{\rho}$ 检验统

计量 $Z(z)$,根据检验假设[如 $H_0: \rho(z) = 0$; $H_A: \rho(z) > 0$],获得各空间格点上统计量 $Z(z)$ 所对应的概率P值。选择合适的P值判断界值(如 $P = 0.05$),绘制P值等值线,即可获得高风险热点区域地图。

4. 统计学分析:使用R软件mapproj包实现空间随机抽样,SAS(版本9.3)软件surveysselect过程实现分层按比例抽样,R软件sprrr包实现空间相对危险度估计^[5]。

实例分析

本研究数据源自中国疾病预防控制中心信息系统传染病报告信息管理系统中2013年度云南省鲁甸及周边地区其他感染性腹泻病例报告信息。其中报告病例822例,报告发病率为31.26/10万,占法定传染病报告发病总数的7.91%。

1. 病例和对照的空间定位:根据病例登记的现住址确定其空间坐标点。采用分层比例方法随机抽取人口样本空间坐标点作为对照的空间位置,即首先以乡镇为单位按人口数采用空间随机抽样方法,生成同等数量的随机人口空间样本点数据库,然后以乡镇为单位按人口比例,从空间样本点数据库中随机抽取与病例数相等的样本点作为对照易感人群样本点。病例和对照的空间分布见图1。

2. 空间相对危险度估计:按照Davies和Hazelton^[4]的方法,将病例和对照设置为相同的全局带宽 h_0 (global h_0),并允许各自选择不同的平滑参数。通过计算确定全局带宽为11 009.09,病例可变

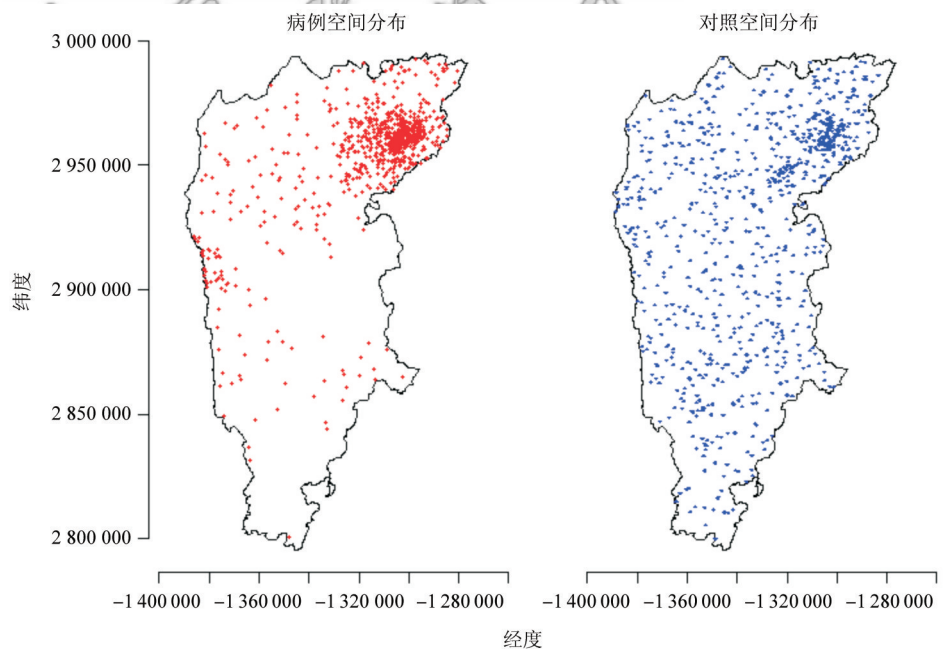


图1 云南省鲁甸及周边地区其他感染性腹泻病例和对照人口的空间分布

平滑参数 h_i (adaptive isotropic smoothing)为2 816.349,对照可变平滑参数 h_i 为1 944.444,以 50×50 空间格点为单位分别计算病例和对照各空间格点上的可变带宽核密度估计值,以此估算空间格点上的对数风险度,并绘制对数风险地图(图2)。可见对数风险值较高的区域主要分布在东北方位邵阳县内,其次在中西部边缘区域,与单纯病例核密度图基本一致。

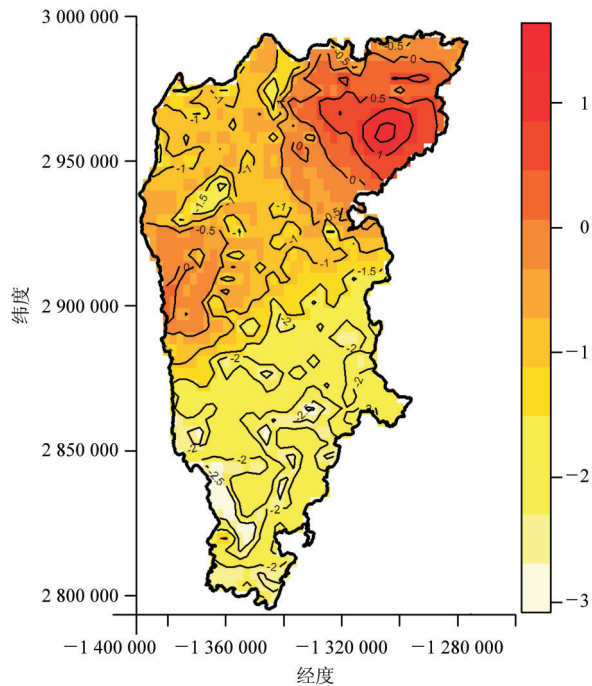


图2 云南省鲁甸及周边地区其他感染性腹泻的可变带宽对数风险分布

对空间各位点的对数风险值 $RR > 0$ 者采用渐近正态法进行统计学检验,获得统计量 $Z(z)$ 对应的概率 P 值,以 $P=0.05$ 作为判断界值,在可变带宽对数风险地图叠加绘制 P 值为0.05等值线地图,即高风险热点地图(图3)。从图3可见,高风险区域仍集中在邵阳县内。

讨论

空间相对危险度估计为非参数空间统计方法,属于空间探索性分析范畴^[9-11]。近年来空间相对危险度估计方法在流行病学研究中的应用逐渐增多^[8,11-13]。本文将可变带宽的核密度估计方法应用于空间相对危险度估计中,并对估计的风险区域进行统计学检验,为确定疾病热点高发地区提供了客观依据。

在空间统计研究领域中,如何估计密度函数至关重要。对于核密度估计一般采用固定带宽或可变

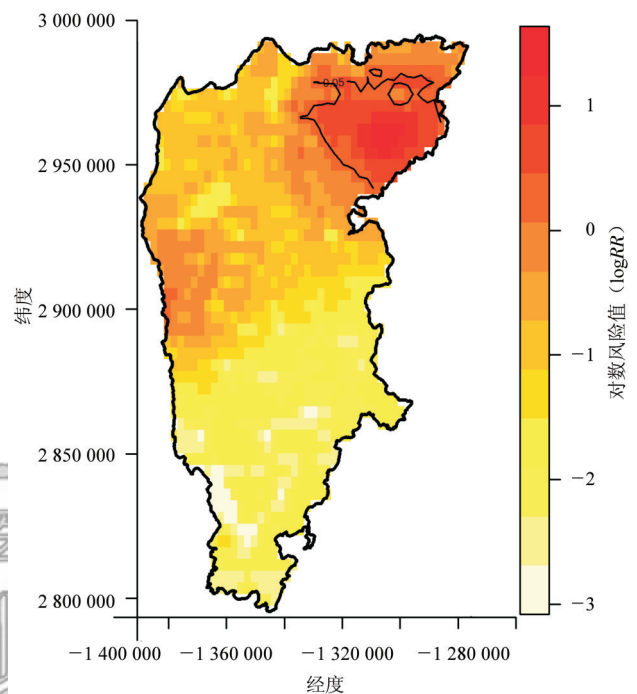


图3 云南省鲁甸及周边地区其他感染性腹泻的对数风险热点及其合并概率等值线($P=0.05$)

带宽。一般而言可变带宽较固定带宽的估计量更具有弹性。在数据密集的区域,可选取较小的带宽以降低估计的偏差;在数据稀疏的区域,可选取较大的带宽以降低估计的方差^[4]。在实际应用中,Benschop等^[12]研究发现与固定带宽比较,可变带宽方法可矫正低密度区域中过度平滑和高密度区域中平滑不足问题。Zhang等^[11]通过比较发现,可变带宽方法在描述空间异质性方面能力更强,但对自定义的平滑参数改变过于敏感,在可变带宽的算法中,带宽对危险估计值和危险显著性检验 P 值的影响大,而对于危险估计值的标准差变化影响小。

与其他非空间流行病学研究类似,病例在空间分布直接受到易感人群在空间分布的影响,而空间相对危险度被认为可有效避免因易感人群空间分布不均衡所造成的混杂因素的影响。因此,本文介绍的相对危险函数已成为空间流行病学研究中描述疾病分析空间变化较为常用的工具^[6,14]。

本文实例源自我国地震灾害后疾病风险评估研究。2014年8月云南省鲁甸县发生6.5级地震,为此本研究利用2013年度鲁甸及周边地区法定传染病病例报告的历史信息,并获得地震灾区各相关传染病高风险区域。高风险区域意味着某种病因或高风险的流行因素的存在。这些危险因素可能会因为自然灾害的原因而进一步恶化,因此本文结果提示应

在所确定的高风险区域有针对性开展现场调查,找出可能影响疾病暴发或流行的主要危险因素,有的放矢地开展防控措施。

参 考 文 献

[1] Rothman KJ. Epidemiology: an introduction[M]. 2nd ed. Oxford University Press, 2012.

[2] Li LM. Epidemiology[M]. 6nd ed. Beijing: People's Medical Publishing House, 2007. (in Chinese)
李立明. 流行病学[M]. 6版. 北京: 人民卫生出版社, 2007.

[3] Waller LA, Gotway CA. Applied spatial statistics for public health data[M]. Hoboken, New Jersey: John Wiley & Sons, 2004.

[4] Davies TM, Hazelton ML. Adaptive kernel estimation of spatial relative risk[J]. Stat Med, 2010, 29(23): 2423-2437.

[5] Davies TM, Hazelton ML, Marshall JC. Sparr: Analyzing spatial relative risk using fixed and adaptive kernel density estimation in R[J]. J Stat Softw, 2011, 39(1): 1-14.

[6] Kelsall JE, Diggle PJ. Non-parametric estimation of spatial variation in relative risk [J]. Stat Med, 1995, 14 (21/22): 2335-2342.

[7] Abramson IS. On bandwidth variation in kernel estimates-a square root law[J]. Ann Statist, 1982, 10(4): 1217-1223.

[8] Bithell JF. An application of density estimation to geographical epidemiology[J]. Stat Med, 1990, 9(6): 691-701.

[9] Harpole JK, Woods CM, Rodebaugh TL, et al. How bandwidth selection algorithms impact exploratory data analysis using kernel density estimation[J]. Psychol Meth, 2014, 19(3): 428-443.

[10] Cislighi C, Biggeri A, Braga M, et al. Exploratory tools for disease mapping in geographical epidemiology [J]. Stat Med, 1995, 14 (21/22): 2363-2381.

[11] Zhang ZJ, Davies TM, Gao J, et al. Identification of high-risk regions for schistosomiasis in the Guichi region of China: an adaptive kernel density estimation-based approach [J]. Parasitology, 2013, 140(7): 868-875.

[12] Benschop J, Hazelton ML, Stevenson MA, et al. Descriptive spatial epidemiology of subclinical salmonella infection in finisher pig herds: application of a novel method of spatially adaptive smoothing[J]. Vet Res, 2008, 39(1): 2.

[13] Carlos HA, Shi X, Sargent J, et al. Density estimation and adaptive bandwidths: a primer for public health practitioners[J]. Int J Health Geogr, 2010, 9: 39.

[14] Clough HE, Fenton SE, French NP, et al. Evidence from the UK Zoonoses Action Plan in favour of localised anomalies of *Salmonella* infection on United Kingdom pig farms[J]. Prev Vet Med, 2009, 89(1/2): 67-74.

(收稿日期: 2014-09-01)

(本文编辑: 张林东)

读者·作者·编者

本刊常用医学词汇缩略语

本刊对以下较为熟悉的一些常用医学词汇将允许直接用缩写,即在文章中第一次出现时,可以不标注中文和英文全称。

A值	吸光度值	HBcAg	乙型肝炎核心抗原	PBS	磷酸盐缓冲液
AIDS	艾滋病	HBsAg	乙型肝炎e抗原	PCR	聚合酶链式反应
ALT	丙氨酸氨基转移酶	HBsAg	乙型肝炎表面抗原	RR	相对危险度
AST	天冬氨酸氨基转移酶	Hb	血红蛋白	RT-PCR	反转录聚合酶链式反应
BMI	体重指数	HC	臀围	SARS	严重急性呼吸综合征
CHD	冠心病	HDL-C	高密度脂蛋白胆固醇	SBP	收缩压
CI	可信区间	HFRS	肾综合征出血热	SCr	血清肌酐
COPD	慢性阻塞性肺疾病	HI	血凝抑制试验	T2DM	2型糖尿病
CT	计算机断层扫描技术	HIV	人类免疫缺陷病毒	TC	总胆固醇
DBP	舒张压	HPV	人乳头瘤病毒	TG	甘油三酯
DNA	脱氧核糖核酸	ICU	重症监护病房	UA	尿酸
ELISA	酶联免疫吸附试验	IDD	碘缺乏病	WBC	白细胞
FPG	空腹血糖	IFG	空腹血糖受损	WC	腰围
GMT	几何平均滴度	LDL-C	低密度脂蛋白胆固醇	WHR	腰臀围比值
HAV	甲型肝炎病毒	M值	中位数	WHtR	腰围身高比
HBV	乙型肝炎病毒	MRI	磁共振成像	WHO	世界卫生组织
HCV	丙型肝炎病毒	MS	代谢综合征	抗-HBs	乙型肝炎表面抗体
HDV	丁型肝炎病毒	MSM	男男性行为者	抗-HBc	乙型肝炎核心抗体
HEV	戊型肝炎病毒	OR	比值比	抗-HBe	乙型肝炎e抗体