

不同空间权重矩阵对我国心血管疾病死亡空间自相关分析的影响

王薇 刘楹宁 殷鹏 王黎君 刘江美 齐金蕾 由金玲 林琳 周脉耕

中国疾病预防控制中心慢性非传染性疾病预防控制中心, 北京 100050

通信作者: 周脉耕, Email: zhoumaigeng@ncncd.chinacdc.cn

【摘要】目的 探索不同空间权重矩阵对我国人群心血管疾病(CVD)死亡空间自相关分析结果的影响及其适用性。**方法** 使用全国人口死亡信息登记管理系统死因监测数据, 构建基于邻接关系的 Rook 矩阵、Queen 矩阵, 以及基于距离关系的 K 最近邻矩阵、距离阈值矩阵, 分别进行 2018 年我国区县水平 CVD 死亡全局及局部空间自相关分析。**结果** 使用 4 类 26 种空间权重矩阵分析我国 CVD 死亡全局自相关均有统计学意义, 全局 Moran's I 统计量在一阶 Rook 矩阵(0.406)、一阶 Queen 矩阵(0.406)、5 个空间单元 K 最近邻矩阵(0.409)以及距离阈值 100 km(0.358)时达到最大。同时, 我国 CVD 死亡呈现局部聚集性分布, 不同空间权重矩阵在探测 CVD 死亡“高-高”“低-低”“高-低”“低-高”空间聚集性方面存在一定差异。**结论** 我国 CVD 死亡存在显著全局及局部自相关性。结合不同的空间权重矩阵进行综合分析, 有助于深入掌握我国区县水平 CVD 死亡空间分布特征, 为有针对性地 CVD 早死区域防控、合理配置资源提供依据。

【关键词】 心血管疾病; 死亡; 空间自相关; 空间权重矩阵; 空间分析

基金项目: 国家重点研发计划(2018YFC1315301)

Influences of using different spatial weight matrices in analyzing spatial autocorrelation of cardiovascular diseases mortality in China

Wang Wei, Liu Yunning, Yin Peng, Wang Lijun, Liu Jiangmei, Qi Jinlei, You Jinling, Lin Lin, Zhou Maigeng
National Center for Chronic and Non-communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 100050, China

Corresponding author: Zhou Maigeng, Email: zhoumaigeng@ncncd.chinacdc.cn

【Abstract】Objective To explore the potential influences and applicability of different spatial weight matrices used in analyzing spatial autocorrelation of cardiovascular disease (CVD) mortality in China. **Methods** Using data from the National Cause-of-death Reporting System, we used adjacency-based Rook and Queen contiguity and distance-based K nearest neighbors/distance threshold. We then conducted global and local spatial autocorrelation analysis of CVD mortality at the county level in China, 2018. **Results** All four categories and 26 types of spatial weight matrices had detected significant global and local spatial autocorrelation of CVD mortality in China. Global Moran's I statistics reached its peak when using first-order Rook (0.406), first-order Queen (0.406), K nearest neighbors including five spatial units (0.409), and distance threshold with 100 kilometers (0.358). Meanwhile, apparent local spatial autocorrelation was found in CVD mortality. Substantial disparities were observed when detecting "High-High clusters", "Low-Low clusters", "High-Low clusters" and "Low-High clusters" of CVD mortality spatial distribution by using different weight matrices. **Conclusions** Using different spatial weight matrices in analyzing the spatial autocorrelation of CVD mortality, we could understand the spatial distribution characteristics of CVD mortality in-depth at the county level in China. In this way, adequate supports could also be provided

DOI: 10.3760/cma.j.cn112338-20201102-01293

收稿日期 2020-11-02 本文编辑 李银鸽

引用本文: 王薇, 刘楹宁, 殷鹏, 等. 不同空间权重矩阵对我国心血管疾病死亡空间自相关分析的影响[J].

中华流行病学杂志, 2021, 42(8): 1437-1444. DOI: 10.3760/cma.j.cn112338-20201102-01293.



on CVD premature death control and rational medical resource allocation regionally.

【Key words】 Cardiovascular disease; Mortality; Spatial autocorrelation; Spatial weight matrix; Spatial analysis

Fund program: National Key Research and Development Program of China (2018YFC1315301)

心血管疾病(cardiovascular diseases, CVD)位居我国城乡居民死亡原因之首^[1]。2017年,我国人群CVD标化死亡率为261.90/10万,死亡约437.80万余人,占全死因死亡的41.89%^[2]。CVD以其发病率高、致残率高、死亡率高、复发率高等特点,已成为影响我国居民健康的重大公共卫生问题^[1]。我国CVD死亡空间异质性较大,掌握CVD死亡空间分布特征可为探索病因和疾病防控提供参考依据^[3]。

空间自相关(spatial autocorrelation)是指空间单元的同属性信息在研究区域内的分布存在一定规律性而并非随机,即存在潜在依赖关系,又称空间依赖(spatial dependence)。空间自相关根植于地理学第一定律,是地理空间信息的一个重要特性,其基本成因是由于空间单元在空间上的邻近,导致其属性值存在关联性^[4-6]。空间自相关分析技术常被运用于对空间单元的同属性信息在研究区域内的分布进行相关性分析,寻找其是否具有聚集性,而疾病的地域聚集性大小可反映出疾病的空间分布规律,为疾病防治和医疗资源合理配置提供理论依据^[7-8]。在空间自相关分析中,通常利用空间权重矩阵(spatial weight matrix)确定并定量测度空间单元地理位置与其对应属性之间的相似性,以此表达研究区域内的空间布局,如拓扑、邻接关系等,并相应地构成了参数检验或模型构建的基础^[6-13]。通过构建良好的空间权重矩阵来反映空间单元实际的联系紧密程度,对准确考察空间自相关特征具有重要意义^[5,12]。

目前,我国基于区县水平的人群CVD死亡空间分析研究较少,其空间分布特征尚不明确。本研究基于全国人口死亡信息登记管理系统死因监测数据,描述2018年我国所有区县CVD死亡空间分布特征,识别CVD死亡风险聚集地区,探索不同空间权重矩阵对CVD死亡空间自相关分析结果的影响,以期对我国CVD早死区域防控、合理配置资源提供证据支持。

资料与方法

1. 资料来源:死亡资料来自全国人口死亡信息

登记管理系统,该系统遍布31个省(自治区、直辖市);采用《疾病和有关健康问题的国际统计分类(ICD-10)》编码选择2018年全国所有区县CVD死亡个案,本研究中CVD编码范围:100~199^[14]。人口资料来自中国CDC信息中心2017年底全国所有区县的人口数据。地图资料来自国家基础地理信息系统数据库2010年1:100万中国县界电子地图。

2. 数据质量控制:由于全国人口死亡信息登记管理系统资料漏报难以避免,为保证死因数据质量及可信度,通过数据清洗,对区县数据质量进行比较和判断,将一些认为是漏报严重、有可能影响总体结果的区县进行剔除,形成最终数据库。本次分析中原卫生部死因统计系统和全国疾病监测系统监测点以总死亡率<4.5‰为剔除标准,2013年新增加的监测点以总死亡率<5.0‰为剔除标准,其他区县以总死亡率<3.0‰为剔除标准,共在死因监测点地区剔除93个监测点、非监测点地区剔除312个区县的数据,最终纳入2550个区县的CVD死亡数据进行汇总分析,纳入分析的区县总死亡率最低为3.0‰^[14]。

3. 统计学分析:采用2018年全国2550个区县死亡数据及人口数据计算区县CVD死亡率,并采用2010年全国人口普查数据对CVD死亡率进行标化。本研究对纳入分析的区县CVD标化死亡率空间分布特征进行空间自相关分析,包括全局空间自相关及局部空间自相关;通过构建不同的空间权重矩阵对区县CVD标化死亡率的空间关系进行描述,并探索其对空间自相关分析结果的影响。

(1)全局空间自相关分析:主要用于研究空间单元的属性值在整个研究区域的空间分布特征,能够描述空间单元之间的平均关联程度、空间分布模式及其显著性,即判断研究属性值有无聚集性^[6,15]。全局自相关常用统计量包括:Moran's *I*、Geary's *C*、General *G*等^[16-17]。其中,Moran's *I*是最早提出的、简单且最常用的统计量,在估计聚集区域位于区域边缘时结果更为可靠,故本研究选择Moran's *I*统计量作为全局空间自相关测量指标^[18-19]。公式:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \times \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, i \neq j \quad (1)$$

式中, n 为 2018 年全国区县数量; i 与 j 为空间中两个区县; w_{ij} 表示两个区县空间关系的空间权重矩阵 w 中第 i 行第 j 列的单元值; x_i 与 \bar{x} 分别为第 i 个区县 CVD 标化死亡率和全国平均 CVD 标化死亡率。Moran's I 统计量可以视为观测值变量与其空间滞后变量间的相关系数, 其在 $[-1, 1]$ 范围内取值。取值为正表示相似属性的空间单元全局集聚分布特征, 即存在空间正相关性, 数值相近的属性值在空间分布上互相吸引(邻近), 值越大此类集聚特征越明显; 取值为负表示相异属性的空间单元全局集聚分布特征, 即存在空间负相关性, 数值相近的属性值在空间分布上互相排斥(分离), 值越小此类集聚特征越明显; 等于 0 表示空间随机分布特征, 即不存在空间相关性^[12,17]。

(2) 局部空间自相关分析: 侧重于研究局部范围内空间单元属性值的空间相关性, 即根据每个对象及其领域对象取值判断空间集聚、扩散状态, 计算出聚集地的空间位置和范围, 即热点或冷点^[5-6,17]。局部自相关常用统计量包括: 局部自相关指标 (local indicators of spatial association, LISA), Getis-Ord G_i^* 、局部 Geary's C 等^[5,17,20-21]。本研究使用 LISA 作为局部空间自相关测量指标, 公式:

$$I_i = \frac{(x_i - \bar{x})}{S^2} \sum_{j \neq i} w_{ij} (x_j - \bar{x}) \quad (2)$$

式中, S^2 为 x_i 或 x_j 的方差, 其余与全局 Moran's I 释义相同。 I_i 在 $[-1, 1]$ 范围内取值。取值为正表示相似属性的空间单元局部集聚分布特征, 即存在空间正相关性, 值越大此类集聚特征越明显, 分别存在 2 种局部空间分布模式, 即“高-高”或“低-低”模式; 取值为负表示相异属性的空间单元局部集聚分布特征, 即存在空间负相关性, 值越小此类集聚特征越明显, 分别存在 2 种局部空间分布模式, 即“高-低”或“低-高”模式; 等于 0 表示空间随机分布特征, 即不存在空间相关性^[12,20]。

(3) 构建不同空间权重矩阵对空间自相关分析

结果的影响: 测量空间单元面数据的空间关系一般有 3 种方法, 相应的空间权重矩阵构建方式也有 3 种: ①基于空间单元之间是否有邻接关系的邻接法 (adjacent), 为二元连接矩阵, 指当空间单元的公共边界和/或顶点可明确分辨时, 空间单元相邻接壤即为 1, 不相邻接壤即为 0, 包括 Rook 矩阵、Queen 矩阵等; ②基于空间单元中心距离的距离法 (distance), 主要为地理距离矩阵, 即使用空间单元的质心代表其区位, 不同空间单元距离采用两点间的欧氏距离 (euclidean distance) 进行测量, 且不考虑地表障碍物, 如距离阈值 (distance threshold) 矩阵、K 最近邻 (K nearest neighbors) 矩阵、反距离 (inverse distance) 矩阵等; 此外还有基于区域社会经济综合因素的经济距离矩阵等; ③基于空间单元之间潜在相互影响的交互法 (interaction) 还将空间单元的相对面积纳入了考虑, 如 Cliff-Ord 矩阵 (公式 6)、Decey 矩阵等^[5,8,10,12-13,16,20-28]。

公式 (3)~(6) 中, w_{ij} 为空间权重矩阵, i, j 为不同空间单元, d_{ij} 为空间单元 i, j 质心距离, D 为距离阈值, k 为正整数个空间单位, a 和 b 分别为外生的距离摩擦系数和边界共享效应系数, β_{ij} 为 i 被 j 共享边界的长度占样本 i 总边界长度的比例, 此时空间单元之间不再为二元连接关系, 而为同时考虑空间距离及边界长度的共享效应^[10,13,22,24-25]。本研究的分析仅基于邻接法 (公式 3) 及距离法 (公式 4、公式 5)。其中, Rook 矩阵将相邻定义为有公共边但没有公共顶点, Queen 矩阵将相邻定义为有公共边或公共顶点 (公式 3); K 最近邻矩阵将相邻定义为 k 个空间单位内为空间相邻关系, k 个空间单位外则不相邻 (公式 4); 距离阈值矩阵将相邻定义为空间单元之间在距离阈值 D 内则为空间相邻关系, 之外则不相邻 (公式 5)。

本研究分别构建 4 类 26 种空间权重矩阵: ①一阶、二阶、三阶 Rook 矩阵 (Rook1、Rook2、Rook3); ②一阶、二阶、三阶 Queen 矩阵 (Queen1、Queen2、

基于邻接关系 $w_{ij} = \begin{cases} 1, & i \text{ 与 } j \text{ 拥有共同边界或/和顶点} \\ 0, & i \text{ 与 } j \text{ 无共同边界或/和顶点, 或 } i \text{ 与 } j \text{ 相等} \end{cases} \quad (3)$

基于距离关系 $w_{ij} = \begin{cases} 1, & j \text{ 是 } i \text{ 的 } k \text{ 个相邻单位之一} \\ 0, & i \text{ 与 } j \text{ 不相邻} \end{cases} \quad (4)$

$$w_{ij} = \begin{cases} 1, & i \text{ 与 } j \text{ 中心距离 } d_{ij} \geq D \\ 0, & i \text{ 与 } j \text{ 中心距离 } 0 < d_{ij} < D \end{cases} \quad (5)$$

基于空间单元之间潜在相互影响的交互法 $w_{ij} = \begin{cases} d_{ij}^a \beta_{ij}^b, & i \text{ 与 } j \text{ 不等} \\ 0, & i \text{ 与 } j \text{ 相等} \end{cases} \quad (6)$

Queen3); ③K 最近邻矩阵:一邻域、二邻域、三邻域、四邻域、五邻域、六邻域、七邻域、八邻域、九邻域 (KNN1、KNN2、KNN3、KNN4、KNN5、KNN6、KNN7、KNN8、KNN9); ④距离阈值矩阵:100、200、300、400、500、600、800、900、1 000、1 200、1 500 km (DS100、DS200、DS300、DS400、DS500、DS600、DS800、DS900、DS1000、DS1200、DS1500)。基于上述空间权重矩阵,分别采用全局 Moran's *I* 及 LISA 对区县 CVD 标化死亡率进行全局空间自相关及局部空间自相关分析,描述我国 CVD 死亡率空间分布特征,并比较不同空间权重矩阵对其结果分析的影响^[26]。本研究采用 GeoDa 1.14.0 及 R 4.0.2 spdep 包进行空间权重矩阵构建及空间自相关分析^[26]。采用双侧检验,检验水准为 $\alpha=0.05$ 。

结 果

1. CVD 死亡率分布一般特征:2018 年全国 2 550 个区县 CVD 死亡率为 291.00/10 万,标化死亡率为 222.04/10 万。其中,CVD 标化死亡率较高的区县分布于内蒙古自治区、黑龙江省、河北省、河南省、甘肃省、云南省和新疆维吾尔自治区,较低的区县分布于江苏省、上海市、浙江省、福建省、广东省、湖南省和四川省。2018 年,我国人群 CVD 死亡空间分布呈现较大差异,总体呈现北部高、中部次之、南部低的梯度递减特征。

2. 不同空间权重矩阵下 CVD 死亡率全局自相关分析:2018 年,在 Rook、Queen、KNN、DS 的 26 种空间权重矩阵下,我国人群 CVD 死亡率全局 Moran's *I* 统计量均为正值,且经检验均达到显著水平,说明我国区县水平 CVD 死亡率的空间分布总体呈现正向全局自相关特征。其中,一阶、二阶、三阶 Rook 及 Queen 矩阵全局自相关统计量相近,且二者一阶效应空间自相关性最强,均为 0.406。KNN 矩阵随邻接空间单元数量增加,统计量呈现先上升后下降趋势,当将 5 个空间单元纳入存在互相影响的空间关系时,全局自相关性最强 (Moran's *I* 统计量为 0.431),且其随邻近空间单元数量的改变差异较小,均在 0.4 以上。DS 矩阵随着空间单元质心距离的增加统计量递减较快,在距离阈值为 100 km 时全局空间相关性最强 (Moran's *I* 统计量为 0.358),1 500 km 时最弱 (Moran's *I* 统计量为 0.027)。见表 1。

3. 不同空间权重矩阵下我国 CVD 死亡率局部

自相关分析:不同空间权重矩阵下,我国人群 CVD 死亡率均存在局部自相关性,所有 2 550 个区县出现“高-高”(7.41%~25.08%,平均 15.00%)及“低-低”(4.20%~33.77%,平均 17.00%)型占比最高,“低-高”(0.54%~11.95%,平均 5.13%)及“高-低”(0.24%~21.83%,平均 4.45%)型较低,不同空间权重矩阵下局部自相关类型仍存在一定差异。见图 1,表 2。

Rook 及 Queen 矩阵下,我国区县出现“高-高”“低-低”“低-高”“高-低”等不同局部自相关类型的数量接近,且随着阶数的增加而增加。KNN 矩阵随着邻近空间单元数的增多,呈现局部自相关的区县数量也逐步增加,各类局部自相关类型空间分布与 Rook 及 Queen 大致相当。而在 DS 矩阵下,全国大部分区县均存在局部自相关性,占比平均为 60.00% 左右,高于 Rook、Queen 及 KNN 矩阵。除“高-高”及“低-低”型空间分布与 Rook、Queen 及 KNN 相似外,“低-高”及“高-低”型区县数量较大,且随着距离阈值的增加,该 2 种局部自相关区县数量增加明显,尤其是“高-低”型。

表 1 不同空间权重矩阵下我国心血管疾病死亡率全局自相关分析

空间关系	空间权重矩阵	空间权重矩阵参数	全局 Moran's <i>I</i> 统计量	<i>P</i> 值	
基于邻接	Rook	一阶 (Rook1)	0.406	<0.001	
		二阶 (Rook2)	0.367	<0.001	
		三阶 (Rook3)	0.324	<0.001	
	Queen	一阶 (Queen1)	0.406	<0.001	
		二阶 (Queen2)	0.366	<0.001	
		三阶 (Queen3)	0.324	<0.001	
	基于距离	K 最近邻	1 个空间单元 (KNN1)	0.409	<0.001
			2 个空间单元 (KNN2)	0.411	<0.001
			3 个空间单元 (KNN3)	0.415	<0.001
4 个空间单元 (KNN4)			0.418	<0.001	
5 个空间单元 (KNN5)			0.431	<0.001	
6 个空间单元 (KNN6)			0.422	<0.001	
7 个空间单元 (KNN7)			0.420	<0.001	
8 个空间单元 (KNN8)			0.411	<0.001	
9 个空间单元 (KNN9)			0.410	<0.001	
距离阈值		100 km (DS100)	0.358	<0.001	
		200 km (DS200)	0.285	<0.001	
		300 km (DS300)	0.238	<0.001	
		400 km (DS400)	0.197	<0.001	
		500 km (DS500)	0.156	<0.001	
		600 km (DS600)	0.121	<0.001	
		800 km (DS800)	0.073	<0.001	
		900 km (DS900)	0.057	<0.001	
		1 000 km (DS1000)	0.044	<0.001	
		1 200 km (DS1200)	0.031	<0.001	
		1 500 km (DS1500)	0.027	<0.001	

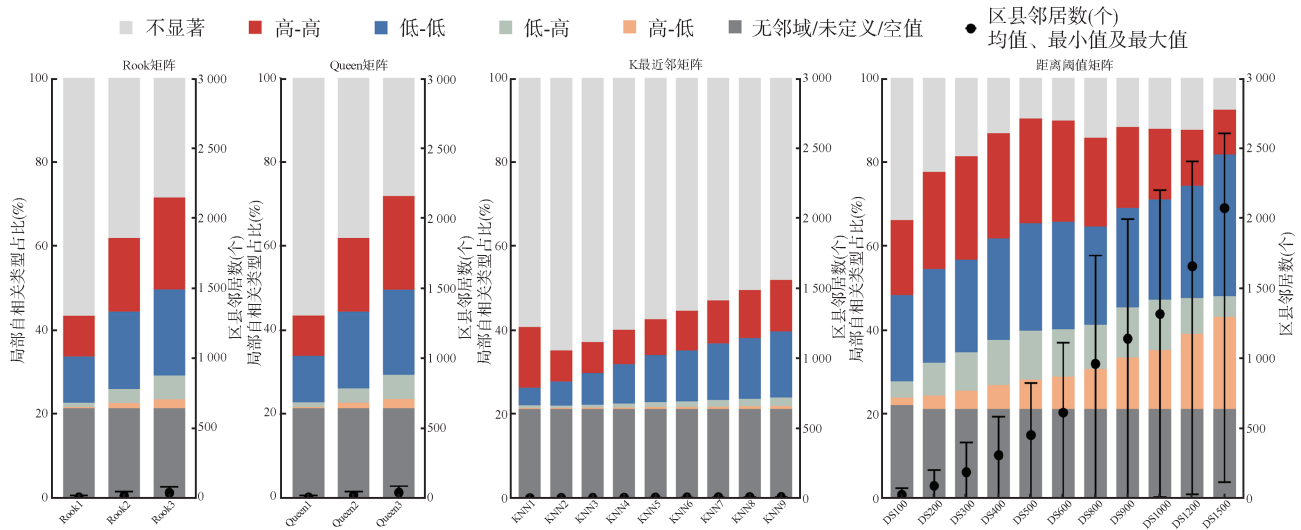


图1 不同空间权重矩阵下我国CVD死亡率局部自相关类型分布情况

就区县邻居数来看,随着Rook和Queen矩阵阶数的增加,区县邻居数也逐步增加,且二者邻居数大致相当(一阶5个、二阶18个、三阶38个);DS矩阵因其基于固定距离阈值,区县邻居数随着地理距离增加而增加迅速,从100 km阈值的24个增加至1 500 km阈值的2 070个;而KNN因其指定了邻居数量,故区县邻居数不变,远小于Rook、Queen以及DS矩阵,因此KNN矩阵下存在局部空间自相关性的区县数量也小于其他3种矩阵。

讨论

本研究基于全国人口死亡信息登记管理系统数据,描述了2018年我国2 550个区县人群CVD死亡空间分布特征,识别了CVD死亡风险的聚集区域。同时,通过构建不同空间权重矩阵,探索了其对CVD死亡空间自相关分析结果的影响。本研究发现,2018年,在我国区县水平CVD死亡率总体呈现正向全局自相关特征,并且具有较明显的“高-高”(“热点”)及“低-低”(“冷点”)区域聚集性。基于邻接关系的Rook及Queen矩阵,以及基于距离关系的KNN及DS矩阵,均对CVD死亡空间自相关特征分析产生一定影响,尤其是对于局部自相关影响较大。

在4类26种空间权重矩阵中,Rook及Queen矩阵对全局Moran's I统计量的影响几乎一致,KNN矩阵结果稳定,DS矩阵则随距离阈值增加而衰减迅速。Rook矩阵基于空间单元相邻有公共边,Queen矩阵基于空间单元相邻有公共边和公共顶

点,二者原理类似,且均在一阶时Moran's I统计量最大。高阶(二阶及三阶)时逐渐递减,这说明我国区县水平CVD死亡率的空间自相关主要发生在边界直接相邻的区域之间,由邻域向邻域不断扩散的“空间滞后”空间效应向外辐射的现象不明显^[27]。KNN矩阵在5个空间单元时Moran's I统计量最大,即各个区县在距离最近的5个区县为邻近关系时空间聚集性最强,这表明影响CVD死亡率分布的空间风险因素在较小尺度距离内起作用,随着空间尺度的增加,空间效应逐渐减小,但总体变化不大^[29]。基于DS矩阵的CVD死亡全局空间自相关性较Rook、Queen以及KNN矩阵更小,在固定阈值距离为100 km时Moran's I统计量最大,且随着空间单元的距离阈值增加,Moran's I统计量下降迅速,其可能原因为少数区县地理面积大,其质心距离与其他区县质心距离相应较大,而除此之外的其他区县质心距离较小。为了满足具有较大质心距离的区县拥有邻域,设定的距离阈值就会偏大;而偏大的距离阈值使大多数面积较小的区县具有过多邻域,而这些多出来邻域并非真实的CVD死亡空间过程作用区域,从而使得全局空间自相关性减弱^[15]。我国CVD死亡的局部自相关空间分布特征明显,不同空间权重矩阵对探测“高-高”“低-低”“低-高”及“高-低”聚集区有一定差异。其中,DS矩阵下LISA能够探测出的局部区县CVD死亡空间聚集性的可能性最高,KNN矩阵最低;且同类矩阵中,阶数越大或固定距离阈值越小,拥有CVD死亡空间聚集性的区县在全国所有区县中占比越高。这与不同类型空间权重矩阵所定义的具有邻近空

表2 不同空间权重矩阵下我国CVD死亡率局部自相关空间分布情况

空间权重矩阵	局部自相关类型	典型地区	区县数量(个)	占全国所有区县比例(%)
一阶Rook	不显著	-	1 673	56.62
	高-高	内蒙古自治区东北部及东南部、辽宁省西北部、河北省南部、河南省中部、陕西省中部、云南省西南部、新疆维吾尔自治区中西部、西藏自治区南部等地区	286	9.68
	低-低	江苏省中南部、浙江省大部、福建省东南部及西北部、广东省北部、四川省中东部等地区	324	10.96
	低-高	内蒙古自治区东部、山西省、河南省等地区少数分散	30	1.02
	高-低	贵州省、湖南省等地区少数分散	8	0.27
一阶Queen	不显著	-	1 672	56.58
	高-高	内蒙古自治区东北部及东南部、辽宁省西北部、河北省南部、河南省中部、甘肃省西北部、陕西省中部、云南省西南部、新疆维吾尔自治区中西部、西藏自治区南部等地区	283	9.58
	低-低	江苏省中南部、浙江省大部、福建省东南部及西北部、广东省中部及东部、四川省中东部、云南省北部等地区	325	11.00
	低-高	黑龙江省北部、辽宁省中部等地区少数分散	33	1.12
	高-低	湖南省、广西壮族自治区等地区少数分散	8	0.27
K最近邻5个空间单元	不显著	-	1 697	57.43
	高-高	内蒙古自治区东北部及东南部、辽宁省西北部、河北省南部、河南省中部、陕西省中部、云南省西南部、新疆维吾尔自治区中西部、西藏自治区南部等地区	252	8.53
	低-低	江苏省中南部、浙江省大部、福建省东南部及西北部、广东省西部、四川省中东部等地区	330	11.17
	低-高	辽宁省西南部、新疆维吾尔自治区西部、西藏自治区东南部等地区少数分散	34	1.15
	高-低	内蒙古自治区中部、吉林省西部等地区少数分散	13	0.44
距离阈值100 km	不显著	-	1 001	33.87
	高-高	内蒙古自治区北部、黑龙江省西部、辽宁省西北部、河北省南部、河南省中部、陕西省中部、云南省西南部、新疆维吾尔自治区北部及中西部等地区	527	17.83
	低-低	江苏省中南部、浙江省大部、福建省东南部及西北部、广东省北部、四川省中东部等地区	605	20.47
	低-高	辽宁省中东部、湖北省东部、江西省北部等地区	115	3.89
	高-低	安徽省、重庆市、广东省等地区少数分散	51	1.73
距离阈值800 km	不显著	-	420	14.21
	高-高	内蒙古自治区东北部及东南部、黑龙江省北部及东北部、吉林省东北部、辽宁省东部及西北部、河北省中南部、陕西省中部、甘肃省中部、新疆维吾尔自治区西部、西藏自治区南部等地区	625	21.15
	低-低	江苏省中南部、浙江省大部、福建省东南部及西北部、湖北省中南部、云南省东北部、广东省北部等地区	690	23.35
	低-高	黑龙江省中南部、吉林省大部、内蒙古自治区北部及西北部、甘肃省南部、新疆维吾尔自治区西部等地区	311	10.52
	高-低	河南省南部、安徽省西部、湖北省东部、湖南省大部、江西省大部、云南省中西部、广西壮族自治区大部等地区	280	9.48

间关系的区县数量有关,区县邻居数越多,局部区域中空间范围越大,拥有局部空间自相关性的区县就越多^[27]。除局部空间自相关外,区县邻居数的定义可能还对全局空间自相关结果造成一定影响:对于Rook和Queen矩阵,随着阶数的增加区县邻居数也逐渐增加,因此拥有局部空间自相关区县数量增加,全局空间自相关减弱;对于DS矩阵,区县邻居数随着距离阈值增加而迅速增加,全局空间自相关减弱也较快;而在本研究中,KNN矩阵明确指定了相邻的空间单元数量(1~9个),故区县邻居数变化范围不大,且远小于Rook、Queen、DS矩阵,因此

KNN矩阵下存在局部空间自相关性的区县数量也小于其他3种矩阵,全局空间自相关同样保持在相对较稳定水平。就LISA探测出的“高-高”及“低-低”聚集区来说,不同空间权重矩阵定义下虽有一定不同,但差别不大,我国CVD死亡风险较高区域多集中于东北地区如黑龙江、辽宁、内蒙古部分地区,中部地区如河北、河南、陕西、甘肃部分地区,西北部地区如新疆部分地区等;而CVD死亡风险较低区域则集中于东部地区如江苏、浙江、福建部分地区等。然而,在探测“低-高”及“高-低”聚集区时,DS矩阵明显优于Rook、Queen及KNN矩阵,

这可能与其基于地理距离的性质有关,从而能够基于阈值范围纳入更多区县进行局部相关性分析。

本研究发现,不同空间权重矩阵下的 CVD 死亡全局自相关及局部自相关结果可能互为关联:对于同一种矩阵,全局 Moran's I 统计量下降,而 LISA 探测出具有空间聚集性的区县数量上升。尽管通过定义更多的区县邻居数能够有助于探测 CVD 死亡局部风险,但仍需考虑空间单元的总体分布情况。总体来说,Rook 和 Queen 矩阵具有拓扑不变性,对空间单元间的公共边界或顶点可明确分辨资料时具有较好探测效果,但难以体现地理距离对空间单元的影响;KNN 矩阵对于空间点在空间均匀分布时具有较好效能,然而可能无法反映稀疏空间单元之间实际情况;DS 矩阵与 KNN 矩阵类似,然而其阈值确定的主观性较大,且无法区分阈值范围内不同距离的影响^[27]。就我国区县水平 CVD 死亡空间自相关分析来看,西部地区区县面积较大,东部地区区县面积相对较小,确定合适的距离阈值是采用 DS 矩阵描述空间单元关系的一大难点:较大的距离阈值将高估局部空间自相关特征,较低的距离阈值则无法探测 CVD 死亡的风险地区。因此,采用空间单元邻居数量来构建空间权重矩阵,可能是稳健表达我国区县水平空间关系的方法。考虑到区县水平 CVD 死亡“空间滞后”效应并不明显,空间相关性在较小尺度距离内起作用,一阶 Rook、一阶 Queen 矩阵以及纳入 5 个空间单元的 K 最近邻矩阵在一定程度上反映了我国 CVD 死亡空间自相关特征的空间权重矩阵,三者结果一致且相对稳定,能较为合理反映我国区县空间分布情况,为 CVD 死亡空间自相关分析提供思路。我国 CVD 死亡的空间聚集性分布特点,或与其影响因素的空间分布差异有关,如地理因素、饮食行为、生活方式、环境气候等^[1,3,30]。

本研究存在不足。从数据质量来看,各地死因监测完整性仍是数据质量的重要挑战之一,漏报问题仍然存在^[31-32]。所以,CVD 死亡的空间特征描述尤其是局部自相关分析结果,将会受到漏报问题的影响,如“冷点”地区可能由于存在漏报现象的区县 CVD 死亡水平较低所致。而因数据质量控制所致的约 14% 区县无数据的情况也在一定程度上影响了分析结果。从分析方法来看,本研究仅使用了空间关系中较为常见的邻接关系和距离关系,这两类矩阵均基于地理学第一定律来构建和描述空间单元之间的物理关联^[23,27]。然而,CVD 死亡水平是一

个综合而系统的人群健康指标,生活习惯及行为方式作为 CVD 死亡的重要危险因素,将很大程度上受到个人所处区域的社会经济发展程度、受教育水平等因素的影响^[33-34]。有研究发现,仅使用事物之间的地理位置来反映不同空间单元及其关联强度相对粗糙,地理距离矩阵不足以描述空间单元间复杂的经济社会关系及其交互影响^[27]。因此,本研究未纳入经济距离空间权重矩阵可能会对准确描述我国 CVD 死亡空间分布特征带来一定影响^[35]。此外,由于本研究纳入了几乎全国所有区县,样本量较大,所有空间权重矩阵类型均能探测到 CVD 死亡存在显著空间自相关特征,但在公共卫生实践中未必有指导意义,使用时应持审慎态度。

不同的空间权重矩阵对空间邻近的定义不同,导致空间单元邻域的空间分布状态、范围及数量有所差别,从而影响空间自相关性的分析结果^[15,36]。空间权重矩阵并不存在固定的模式,需根据分析的具体内容和性质来生成合适的度量指标以反映出空间单元属性资料之间的关系^[8]。实际上,任何空间权重矩阵都只能是近似表示出其空间的关系^[37]。随着我国社会经济的发展,城镇化和人口老龄化进程不断加快,以 CVD 为代表的慢性病的影响因素及不同地区的疾病空间关系,将较以往更加复杂化和多样化。因此,在充分考虑社会经济、人口结构、环境气象等因素对空间关系的影响条件下,构建一个能够更为准确的刻画我国人群 CVD 死亡空间特征的空间权重矩阵,将是一个重要的研究方向。

通过构建不同的空间权重矩阵,能够深入对我国人群 CVD 死亡的空间自相关特征进行分析,探测 CVD 死亡的高危和低危地区,加深对 CVD 死亡区域差异规律的认识与理解,从而有助于因地制宜的为我国 CVD 早死防制重点地区提供证据支持。本研究提示,我国应继续加强对 CVD 早死防控工作,减少“热点”区域的“扩散效应”,同时采取有力措施维持“冷点”区域 CVD 低死亡水平状态。

利益冲突 所有作者均声明不存在利益冲突

志谢 感谢泰康溢彩公共卫生及流行病防治专项基金的支持

参 考 文 献

- [1] 国家心血管病中心. 中国心血管病报告-2018[M]. 北京:中国大百科全书出版社, 2019. National Center for Cardiovascular Disease, China. Report on cardiovascular diseases in China 2018[M]. Beijing: Encyclopedia of China Publishing House, 2019.
- [2] Institution of Health Metrics and Evaluation. GBD results tool[EB/OL]. [2020-10-08]. <http://ghdx.healthdata.org/gbd-results-tool>.
- [3] Liu SW, Li YC, Zeng XY, et al. Burden of cardiovascular diseases in China, 1990-2016: findings from the 2016 Global Burden of disease study[J]. JAMA Cardiol, 2019,

- 4(4):342-352. DOI:10.1001/jamacardio.2019.0295.
- [4] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB/Z 33451-2016 地理信息 空间抽样与统计推断[S]. 北京: 中国标准出版社, 2017. General Administration of Quality Supervision, Inspection and Quarantine, Standardization Administration. GB/Z 33451-2016 Geographic information-Spatial sampling and statistical inference[S]. Beijing: China Standards Press, 2017.
- [5] 王振声. 基于空间模型的小地域疾病制图研究[D]. 武汉: 武汉大学, 2014. Wang ZS. Research of disease mapping over small-area based on spatial model[D]. Wuhan: Wuhan University, 2014.
- [6] 陈雅淑. 局部空间自相关指标的适用性研究[D]. 上海: 华东师范大学, 2009. Chen YS. Investigation of local indexes of spatial autocorrelation application: take sex ratio of Jiangsu for example[D]. Shanghai: East China Normal University, 2009.
- [7] 林静静, 张铁威, 李秀央. 疾病时空聚集分析的研究与进展[J]. 中华流行病学杂志, 2020, 41(7): 1165-1170. DOI: 10.3760/cma.j.cn112338-20190806-00582. Lin JJ, Zhang TW, Li XY. Research progress on spatiotemporal clustering of disease[J]. Chin J Epidemiol, 2020, 41(7): 1165-1170. DOI: 10.3760/cma.j.cn112338-20190806-00582.
- [8] 苏茜, 冯子健, 蒋敏, 等. 不同空间权重矩阵在疟疾空间分布模式分析中的探讨[J]. 中华疾病控制杂志, 2010, 14(5): 419-422. Su Q, Feng ZJ, Jiang M, et al. An analysis of spatial distribution pattern of malaria using different spatial weight matrixes[J]. Chin J Dis Control Prev, 2010, 14(5): 419-422.
- [9] Anselin L. Spatial econometrics: methods and models[M]. Netherlands: Springer, 1988.
- [10] Cliff AD, Ord JK. Spatial processes: models and applications [M]. London: Pion Limited, 1981.
- [11] Cressie NAC. Statistics for Spatial Data[M]. New York: Wiley, 1993.
- [12] 刘畅. 中国区域经济增长与集聚的实证研究[D]. 长春: 吉林大学, 2012. Liu C. The empirical study on China's regional economic growth and agglomeration[D]. Changchun: Jilin University, 2012.
- [13] 朱文康. 空间计量模型的权重矩阵构造与分析[D]. 上海: 上海交通大学, 2014. Zhu WK. Constructing and analyzing spatial weight matrix of spatial econometric model[D]. Shanghai: Shanghai Jiaotong University, 2014.
- [14] 中国疾病预防控制中心慢性非传染性疾病预防控制中心, 国家卫生健康委统计信息中心. 中国死因监测数据集-2018[M]. 北京: 中国科学技术出版社, 2019. National Center for Chronic and Non-communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention; National Health Statistics Information Center, National Health Commission. China mortality surveillance yearbook, 2018[M]. Beijing: Science and Technology Press of China, 2019.
- [15] 冯昕, 杜世宏, 舒红. 空间权重矩阵对空间自相关的影响分析——以我国肾综合征出血热疾病为例[J]. 武汉大学学报: 信息科学版, 2011, 36(12): 1410-1413. DOI: 10.13203/j.whugis2011.12.013. Feng X, Du SH, Shu H. Influences of spatial weight matrices on spatial autocorrelation: a case study of HFRS in China[J]. Geomatics and Information Science of Wuhan University, 2011, 36(12): 1410-1413. DOI: 10.13203/j.whugis2011.12.013.
- [16] 周晓农. 空间流行病学[M]. 北京: 科学出版社, 2009. Zhou XN. Spatial epidemiology[M]. Beijing: Science Press, 2009.
- [17] 王劲峰, 廖一兰, 刘鑫. 空间数据分析教程[M]. 2版. 北京: 科学出版社, 2019. Wang JF, Liao YL, Liu X. Spatial data analysis[M]. 2nd ed. Beijing: Science Press, 2019.
- [18] 吴珣, 杨婕, 张红. 不同空间权重定义下中国人口分布空间自相关特征分析[J]. 地理信息世界, 2017, 24(2): 32-38. DOI: 10.3969/j.issn.1672-1586.2017.02.006. Wu X, Yang J, Zhang H. Analyzing spatial autocorrelation of population distribution in different spatial weights: A case of China[J]. Geomat World, 2017, 24(2): 32-38. DOI: 10.3969/j.issn.1672-1586.2017.02.006.
- [19] Cliff AD, Haggett P, Ord JK, et al. Elements of spatial structure: a quantitative approach[M]. Cambridge: Cambridge University Press, 1975.
- [20] 高乐. 空间计量模型中空间权重矩阵的构造方法[D]. 长春: 东北师范大学, 2019. Gao L. Construction method of spatial weight matrix in spatial econometric model[D]. Changchun: Northeast Normal University, 2019.
- [21] Getis A, Aldstadt J. Constructing the spatial weights matrix using a local statistic[J]. Geograph Anal, 2004, 36(2): 90-104. DOI: 10.1111/j.1538-4632.2004.tb01127.x.
- [22] 王守坤. 空间计量模型中权重矩阵的类型与选择[J]. 经济数学, 2013, 30(3): 57-63. DOI: 10.3969/j.issn.1007-1660.2013.03.011. Wang SK. Several types of weights matrix and their extended logic[J]. J Quant Econom, 2013, 30(3): 57-63. DOI: 10.3969/j.issn.1007-1660.2013.03.011.
- [23] 吴燕. 空间计量经济学模型及其应用[D]. 武汉: 华中科技大学, 2017. Wu Y. Spatial econometric model applied research[D]. Wuhan: Huazhong University of Science and Technology, 2017.
- [24] Cliff AD, Ord JK. Spatial autocorrelation[M]. London: Pion Limited, 1973.
- [25] 徐彬. 空间权重矩阵对 Moran's I 指数影响的模拟分析[D]. 南京: 南京师范大学, 2007. Xu B. Simulation analysis on the influences of different spatial weight matrices on Moran's I [D]. Nanjing: Nanjing Normal University, 2007.
- [26] 龚德鑫, 刘涛, 朱志华, 等. 空间自回归分析及其在 R 软件中的实现[J]. 华南预防医学, 2020, 46(4): 450-453. DOI: 10.12183/j.scjpm.2020.0450. Gong DX, Liu T, Zhu ZH, et al. Spatial autocorrelation analysis and its application in R[J]. South China J Prev Med, 2020, 46(4): 450-453. DOI: 10.12183/j.scjpm.2020.0450.
- [27] 张可云, 王裕瑾, 王婧. 空间权重矩阵的设定方法研究[J]. 区域经济评论, 2017(1): 19-25. DOI: 10.14017/j.cnki.2095-5766.2017.0003. Zhang KY, Wang YJ, Wang J. Research on the specification methods of spatial weight matrix[J]. Reg Econom Rev, 2017(1): 19-25. DOI: 10.14017/j.cnki.2095-5766.2017.0003.
- [28] Lam C, Souza PCL. Estimation and selection of spatial weight matrix in a spatial lag model[J]. J Busin Econom Stat, 2020, 38(3): 693-710. DOI: 10.1080/07350015.2019.1569526.
- [29] 林冠文, 马金香. 不同空间权重矩阵在白血病空间统计分析中的探讨[J]. 中华疾病控制杂志, 2013, 17(7): 617-620. Lin GW, Ma JX. Study on the spatial statistical analysis of leukemia using different spatial weight matrixes[J]. Chin J Dis Control Prev, 2013, 17(7): 617-620.
- [30] 李镛冲. 我国高血压患病、知晓、治疗和控制的水平空间分析[D]. 北京: 中国疾病预防控制中心, 2016. Li YC. Multilevel spatial analysis of hypertension prevalence, awareness, treatment and control in China [D]. Beijing: Chinese Center for Disease Control and Prevention, 2016.
- [31] Guo K, Yin P, Wang LJ, et al. Propensity score weighting for addressing under-reporting in mortality surveillance: a proof-of-concept study using the nationally representative mortality data in China[J]. Popul Health Metr, 2015, 13: 16. DOI: 10.1186/s12963-015-0051-3.
- [32] Zeng XY, Adair T, Wang LJ, et al. Measuring the completeness of death registration in 2 844 Chinese counties in 2018[J]. BMC Med, 2020, 18(1): 176. DOI: 10.1186/s12916-020-01632-8.
- [33] GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017[J]. Lancet, 2018, 392(10159): 1923-1994. DOI: 10.1016/S0140-6736(18)32225-6.
- [34] 何柳, 施小明, 胡永华. 地区社会经济水平与心血管疾病研究进展[J]. 中国公共卫生, 2014, 30(7): 936-939. DOI: 10.11847/zgggws2014-30-07-27. He L, Shi XM, Hu YH. Progress of research on regional social-economic level and cardiovascular disease[J]. Chin J Public Health, 2014, 30(7): 936-939. DOI: 10.11847/zgggws2014-30-07-27.
- [35] 林光平, 龙志和, 吴梅. 我国地区经济收敛的空间计量实证分析: 1978-2002 年[J]. 经济科学, 2005, 4 增刊: 1: 67-82. DOI: 10.13821/j.cnki.ceq.2005.S1.005. Lin GP, Long ZH, Wu M. A spatial analysis of regional economic convergence in China: 1978-2002[J]. China Econom Quart, 2005, 4 Suppl 1: 67-82. DOI: 10.13821/j.cnki.ceq.2005.S1.005.
- [36] 李霖, 周玉杰, 靳凤撰. 基于不同空间权重的总和生育率自相关分析[J]. 测绘地理信息, 2015, 40(5): 62-64. DOI: 10.14188/j.2095-6045.2015.05.018. Li L, Zhou YJ, Jin FZ. Spatial autocorrelation analysis of total fertility rate based on different spatial weight matrixes[J]. J Geomat, 2015, 40(5): 62-64. DOI: 10.14188/j.2095-6045.2015.05.018.
- [37] Odland J. Spatial autocorrelation[M]. California: Sage Publications Inc, 1988.