·疾病监测·

应用空间自相关分析研究广西壮族自治区肝癌的空间异质性分布特征

唐咸艳 黄天壬 朱小东 胡茂琼 徐静 周红霞

【摘要】 **目的** 应用空间自相关分析研究广西壮族自治区肝癌的空间分布特征。**方法** 利用全区2000—2007年肝癌资料，求出各市县肝癌8年平均发病率；应用地理信息系统中的空间统计分析模块进行空间自相关分析；应用Map Info 8.0软件绘制疾病专题图。**结果** 2000—2007年广西肝癌年均发病率存在空间自相关，全域空间自相关系数Moran's I=0.34，P<0.01；全域空间自相关系数G=0.77，P<0.01；Moran's I系数图上下波动，有四处隆起，主要在微观尺度（空间间隔为1～3，实际尺度45～135 km）及宏观尺度（空间间隔16～18，实际尺度720～800 km）上存在聚集性分布，但在空间间隔为1.4，实际尺度约60 km时，空间自相关有波峰，即空间分布有最大的自相关性；疾病专题图显示肝癌高发区主要聚集在桂西南及桂南沿海地区，桂北为低发区。**结论** 2000—2007年广西肝癌的空间格局非随机分布，存在明显的聚集区。

【关键词】 肝肿瘤，原发性；空间自相关；空间异质性

Using spatial autocorrelation analysis to study spatial heterogeneity of liver cancer in Guangxi

TANG Xian-yan, HUANG Tian-ren, ZHU Xiao-dong, HU Mao-qiong, XU Jing, ZHOU Hong-xia. *Public Health School of Guangxi Medical University, Nanning* 530021, China

Corresponding author: *ZHOU Hong-xia*, Email: gmuies@163.com

【**Abstract**】 **Objective** To study the spatial distribution characteristics of liver cancer in Guangxi so as to provide evidence for the development of control and prevention on liver cancer. **Methods** The average eight year morbidity was computed, using the rates of liver cancer in 2000—2007. The spatial statistics module of GIS was used to conduct spatial autocorrelation analysis, and the disease mapping was drawn, using the Map Info 8.0 software. **Results** The average morbidity rate was clustered in Guangxi in the past eight years, with Moran's I index as 0.34 and P value below 0.01. G index appeared to be 0.77 and the P value was below 0.01. Moran's I correlogram lifted up in four spaces, specifically, the cluster took place in both macro-scale (one to three spatial intervals, 45 to 135 km real scale) and micro-scale (16 to 18 spatial intervals, 720 to 800 km real scale). When the spatial interval became 14 and real scale was 60 km, the spatial distribution of liver cancer showed the most intensive autocorrelation. Most of the regions with high morbidity would be clustered in the southwest and southern parts, along the coastal areas of Guangxi while the regions with low morbidity clustered in the northern part of Guangxi. **Conclusion** Liver cancer was found un-randomly distributed and geographically clustered in Guangxi in 2000—2007.

【**Key words**】 Hepatocellular carcinoma; Spatial autocorrelation; Spatial heterogeneity

空间自相关是指一个区域单位上的某种属性值（如发病率）与邻近区域单位上的同一属性值的相关程度，其基本度量指标是空间自相关系数，用空间自相关系数来检验区域单位的某一属性值是否高高相邻、低低相邻或者高低间错分布，即有无聚集性。疾病的分布具有明显空间异质性特点，即有明显的高发区和低发区分布，并呈现出各自的聚集性。在流行病学研究中，疾病的空间分布特点体现了疾病病因的分布特点，因而疾病的空间分布规律一直是流行病学关注的重点。空间自相关分析可以客观地揭示疾病空间分布规律，在疾病的空间分布研究中已得到广泛的应用[1-3]。广西壮族自治区是原发性肝癌（肝癌）高发区，但各县发病情况不同，高发县与低发县间的差异很大，其分布有某种明显的规律性。本研究将应用空间自相关分析探

DOI: 10.3760/cma.j.issn.0254-6450.2009.02.017

作者单位：530021 南宁，广西医科大学公共卫生学院流行病与卫生统计学教研室（唐咸艳、周红霞）；广西壮族自治区肿瘤防治研究所（黄天壬、朱小东）；湖北黄石理工学院医学院（胡茂琼）；广西医科大学第一附属医院科研部（徐静）

通信作者：周红霞，Email: gmuies@163.com

讨广西壮族自治区肝癌的空间分布，并对分布的规律性进行统计量化，为肝癌防治提供科学依据。

资料与方法

1. 资料来源：2000—2007年各市县肝癌逐年发病数据来源于广西壮族自治区肿瘤防治研究所，入选肝癌病例以病理诊断为准的原发性肝癌；同年各市县总人口数资料来源于自治区统计年鉴；将各市县逐年肝癌发病资料与人口资料在Excel 2003软件中建立数据库，并求出各市县肝癌8年平均发病率。1:25 000 m电子地图由广西壮族自治区测绘局提供。

2. 分析方法：在Map Info 8.0软件中绘制广西壮族自治区肝癌8年平均发病率专题图、局域Getis-Ord得分检验的分布图；空间自相关分析在Arc GIS9.2的空间统计模块中进行；空间自相关系数图采用Excel 2003软件绘制。

3. 空间自相关理论：空间自相关分为全域型空间自相关和局域型空间自相关，常用分析方法有Moran's I、Geary's C、Getis以及空间自相关系数图。

全域空间自相关Moran's I：

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

式中各符号在本研究中的实际意义：n为所研究的疾病空间区域数量；x_i为i区域内的疾病发生率，x_j为j区域内的疾病发生率，$\bar{x}$为所研究区域的疾病平均发生率；w_{ij}为空间权重矩阵。确定方法：

$$w_{ij}=\begin{cases}1 & \text{当区域 } i \text{ 和 } j \text{ 相邻接}\\ 0 & \text{其他}\end{cases}$$

在无效假设下，即研究对象无空间自相关性，此时Moran's I的期望值：

$$E(I)=\frac{-1}{(n-1)}$$

Moran's I的方差有两种假设：一是正态分布假设；另一则是随机分布假设。在随机分布假设下，由于不知道x的理论分布形式，故这种假设常用于疾病发生率的检验[4]。在随机假设下，Moran's I的方差为：

$$Var(I)=\frac{n[(n^2-3n+3)w_1-nw_2+3w_0^2]-k[(n^2-n)w_1-2nw_2+6w_0^2]}{w_0^2(n-1)(n-2)(n-3)}-E(I)^2$$

其中

$$w_0=\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}\qquad w_1=\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij}+w_{ji})^2$$

$$w_2=\sum_{i=1}^{n}(w_{\cdot i}+w_{i\cdot})^2\qquad w_{i\cdot}=\sum_{j=1}^{n}w_{ij}$$

$$k=\frac{[\sum_{i=1}^{n}(x_1-\bar{x})^4/n]}{[\sum_{j=1}^{n}(x_j-\bar{x})^2/n]^2}$$

因此，随机分布假设下Moran's I的Z-score得分检验为：

$$Z_R=\frac{I-E(I)}{\sqrt{\mathrm{var}_R(I)}}$$

当$|Z|>1.96$时，$P<0.05$，拒绝无效假设，存在空间自相关。

至于全域Getis G系数和局域Getis G_i系数，因公式繁琐，本文不作介绍，请参阅文献[4-7]。表1为本研究用到的三种空间自相关分析比较。

表1　几种空间自相关分析方法比较

类型	方法	取值范围	实际意义
全域	Moran's I	$-1\leqslant I\leqslant 1$	$I>0$且有统计学意义，正自相关，聚集分布；$I<0$且有统计学意义，负自相关，均匀分布；$I=0$，无自相关，随机分布
全域	Getis	G为全体实数	$G>0$且有统计学意义，高值聚集分布；$G<0$且有统计学意义，低值聚集分布；$G=0$，随机分布
局域	Getis	G_i为全体实数	G_i实际意义判定标准同全域G

4. 空间自相关系数图：用不同空间间隔下的全域型空间自相关系数Moran's I作纵坐标，空间间隔作横坐标，将不同空间间隔下的全域型Moran's I系数连接成折线图，得到空间自相关系数图，用于分析疾病现象在空间上是否具有阶层性分布。空间自相关系数图能提供以下信息[7]：①空间自相关系数若随空间间隔数增加而依次递减，则表示在该研究区域内某属性值相似的区域呈单核心分布状态，即可能存在单核心聚集区；②空间自相关系数若随空间间隔增加，非依次递减，而呈波浪型曲线，表明在该研究区域内可能存在多处聚集区；③在空间自相关系数大于零的空间间隔区域里存在聚集性，小于零的空间间隔区域内不存在聚集性；系数图波峰值对应的空间间隔处存在最大的空间自相关，即聚集性最强。

本研究拟将广西壮族自治区划分18个空间间隔，每个空间间隔对应的实际尺度约为45 km。

结　果

1. 肝癌年均发病率专题图：如图1所示，广西壮族自治区肝癌年均发病率分布非随机性，存在明显的区域聚集性。高发病区主要聚集在以扶绥县为中心的桂西南及桂南沿海一带；而低发病区主要聚集在桂北以及桂西北地区，其他区域为过渡带。

2. 全域空间自相关分析：以县为基本区域单位进行全域Moran's *I*和全域Getis空间自相关分析。由表2可见：$I=0.34$，$P<0.01$，差异有统计学意义，广西壮族自治区肝癌年均发病率存在正自相关关系，分布具有聚集性；$G=0.77$，$P<0.01$，差异有统计学意义，肝癌年均发病率分布具有聚集性，且具有高值聚集性分布，即发病率高的区域有明显的聚集性分布。

3. 局域Getis*i*得分检验专题图：在Arc GIS9.2中进行局域Getis空间自相关分析，将结果导入到Map Info 8.0中，绘制局域$Z(G_i)$值专题图（图2）：$Z(G_i)$值为正值，且>1.96的区域主要分布在桂西南及桂南沿海一带，这一地区为有统计学意义的高发病率聚集区；$Z(G_i)$值为负值，且<-1.96的区域主要集中在桂北地区，这一带为有统计学意义的低发病率聚集区；其他区域为过渡带。

4. 空间自相关系数图：从图3可得到以下信息：①空间自相关系数图非依次递减，而呈波浪形，提示广西壮族自治区肝癌存在多核心聚集区；②Moran's *I*系数大于零值共有4处，提示肝癌发病率可能存在4个聚集区，在微观尺度（空间间隔1～3，实际尺度45～135 km）和宏观尺度（空间间隔16～18，实际尺度720～800 km）上，聚集现象比较明显，其他中观尺度上聚集现象不明显；③当空间间隔约为1.4，实际尺度约为60 km时，广西壮族自治区肝癌空间分布具有最大的自相关性，存在最强的聚集区。

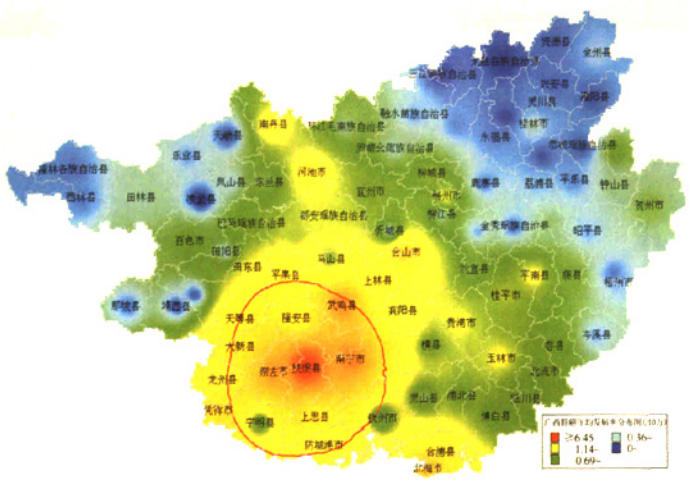

图1　2000－2007年广西壮族自治区肝癌年平均发病率分布

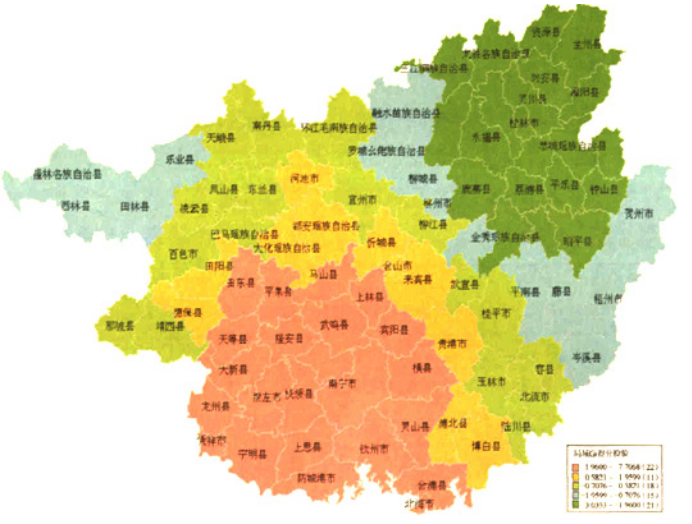

图2　广西壮族自治区局域G_i得分检验分布图

表2　广西壮族自治区肝癌全域空间自相关分析

方法	自相关系数	*Z*-score检验	*P*值
Moran's *I*	0.34	7.81	<0.01
Getis	0.77	4.47	<0.01

注：$P=0.05$对应的*Z*-score界值为1.96；$P=0.01$对应的*Z*-score界值为2.58

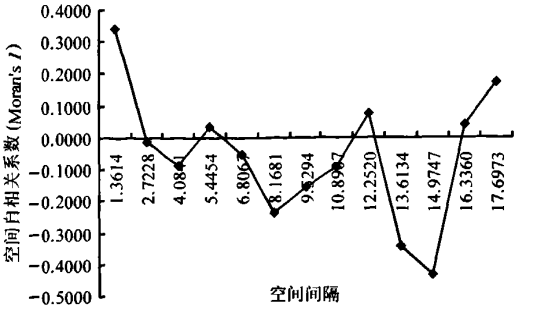

图3　广西壮族自治区肝癌分布空间自相关系数图

讨　论

空间自相关分析作为空间统计学中的主要方法，已在疾病空间分布方面得到很好应用[8-10]。空间自相关分析主要有Moran's I、Geary's C、Getis及空间自相关系数图等方法，其共性之处均可对疾病空间聚集性进行量化研究；不同之处在疾病聚集强度、聚集区域与范围、聚集状态类型等方面提供的信息不同。经统计推断，全域Moran's I系数能说明在研究范围内疾病有无聚集性分布，但是不能说明疾病是什么状态的聚集，即是高值聚集还是低值聚集。全域Getis空间自相关弥补了全域Moran's I的不足；Getis系数大于零且有统计学意义，说明疾病在研究区域内呈高值聚集；反之，疾病在区域内的分布为低值聚集。然而，不论是全域Moran's I还是全域Getis系数，都是从全局角度说明疾病空间分布有无聚集性以及是什么状态的聚集性，都不能确切说明疾病具体聚集在哪些空间区域内。局域Getis系数克服了这种弊端，若某区域单位的局域Getis系数为正值且有统计学意义，说明此区域单位与周边区域单位在某属性值（如发病率）上都是高值，呈现高值簇聚集区；反之，则说明此区域单位与周边区域单位在某属性值（如发病率）上都是低值，呈现低值簇聚集区。

流行病学研究显示，广西壮族自治区肝癌的空间分布具有明显的区域异质性[11]，但对异质性缺少量化描述。本研究应用空间统计方法，将广西壮族自治区肝癌发病率作为一种疾病区域化变量，制作疾病专题图（图1）和自相关统计量分析，比较客观地展示了肝癌发病率在全区内分布不均匀性，即存在空间异质性分布特点：桂西南及桂南沿海地区为肝癌高发区，主要以扶绥县为中心，半径约60 km的红黄色区域内；空间自相关系数（图3）也说明在空间间隔约为1.4，实际尺度约为60 km时，广西壮族自治区肝癌空间分布具有最大的自相关性，存在最强的聚集区，两者结果相符。疾病专题图虽然较好地展示了肝癌分布的聚集性现象，但是否真有聚集性需经空间自相关分析加以客观量化和验证。全域Moran's I=0.34，P<0.01，说明广西壮族自治区肝癌分布存在聚集性；全域Getis=0.77，P<0.01，说明肝癌存在高值聚集性，即发病率高的区域聚集簇。为进一步探讨肝癌哪些地区是高发病率聚集区，哪些地区是低发病率聚集区，本研究进一步做局域Getis空间自相关分析，并将各市县的局域$Z(G_i)$值在广西壮族自治区地图上直观显示（图2）：高发病区[$Z(G_i)$>1.96]集中在桂西南及桂南沿海，共22个市县；低发病区[$Z(G_i)$<-1.96]聚集在桂西北，共21个市县；其他市县为肝癌发病中间带。可见，广西壮族自治区肝癌发病率存在明显的高发区和低发区，应因地制宜，加强综合防治。

需要指出的是本研究存在一定的偏倚。广西壮族自治区肿瘤防治研究所位于南宁市，是全区最高水平的肿瘤防治研究机构，对就诊的每例肿瘤患者都有详细的诊疗记录，但由于我国还缺乏法定的肿瘤注册制度，位于桂北区域的肿瘤患者可能因为距离原因就近治疗，而存在入院偏倚。此外由于空间统计分析技术本身是在地理学研究中发展而来，对疾病的描述还存在缺陷，如何发展这一方法并应用于疾病的流行病学研究是值得关注的问题。特别突出的一点是疾病的量化（或属性值）目前都基于行政划分单元，如发病率等，是按某一行政区域的人口数计算而来，而疾病分布则更重要的是基于各种自然和人文环境的地理分布，这一矛盾突出地影响了疾病空间流行病学分布的研究。对疾病的测量还需要发展更为接近地理环境属性的指标。

参考文献

[1] Fang LQ, Yan L, Liang S, et al.Spatial analysis of hemorrhagic fever with renal syndrom in China. BMC Infect Dis, 2006, 6: 77.

[2] Lai PC, Wong CM, Hedley AJ, et al. Understanding the spatial clustering of Severe Acute Respiratory Syndrome (SARS) in Hong Kong.Environ Health Perspect, 2004, 112(15): 1550-1556.

[3] Zhang ZY, Xu DZ, Zhou Y, et al. Spatial analysis of snail distribution in Jiangning county.J Med Coll PLA, 2002, 17(2): 88-91.

[4] 王劲峰. 空间分析. 北京：科学出版社，2006：77-91.

[5] 周国法，徐汝梅. 生物地理统计学——生物种群时空分析的方法及其应用. 北京：科学出版社，1998：1-42.

[6] 王政权. 地统计学及其在生态学中的应用. 北京：科学出版社，1998：62-94.

[7] 刘湘南，黄方，王平，等. GIS空间分析原理与方法. 北京：科学出版社，2005：189-194.

[8] 陈炳为，许碧云，李德云，等. 应用区域型空间自相关系数分析疾病的聚集性.中国公共卫生，2006，22(9)：1146-1147.

[9] 陈炳为，李德云，倪宗瓒.四川省碘缺乏病的空间自相关性.现代预防医学，2003，30(2)：158-159.

[10] 方立群，曹春香，陈国胜，等. 地理信息系统应用于中国大陆高致病性禽流感的空间分布及环境因素分析.中华流行病学杂志，2005，26(11)：839-842.

[11] 黄天壬，余家华，张振权，等. 广西肝癌流行特征和流行趋势分析.广西医学，2000，22(4)：677-679.

（收稿日期：2008-08-25）

（本文编辑：张林东）