

采用两维图论聚类法分析中国直肠肛门闭锁/狭窄的空间分布

乔良 朱军 袁萍 代礼 王艳萍 周光宣 韩颖 刘潇霞 张驯 曹奕

【摘要】目的 探讨中国直肠肛门闭锁/狭窄的地理空间分布规律,为病因研究提供线索。**方法** 利用2001—2005年中国出生缺陷监测网以医院为基础在全国对孕28周至产后7 d的围产儿进行监测的数据,采用两维图论聚类法对全国监测点按直肠肛门闭锁/狭窄的发生率进行空间归类。**结果** 中国直肠肛门闭锁/狭窄总发生率为3.17/万,东部地区发生率高于中、西部,差异有统计学意义(z 值分别为2.50、3.69; P 值分别为0.012、 <0.001)。全国监测点可聚为6类:第Ⅰ类监测点主要分布于黑龙江、吉林、辽宁;第Ⅱ类主要分布于福建、广东、海南、广西以及湖南和江西的南部;第Ⅲ类主要分布于北京、天津、河北、山东、江苏和安徽北部;第Ⅳ类主要分布于浙江、上海、安徽、江苏的南部和湖南、江西的北部及湖北、河南、山西、内蒙古;第Ⅴ类主要分布于宁夏、甘肃、青海;第Ⅵ类主要分布于陕西、四川、重庆、云南、贵州、新疆、西藏。**结论** 聚类结果明确了中国直肠肛门闭锁/狭窄的空间分布特征及其规律,对于进一步分析直肠肛门闭锁/狭窄的环境影响因素和区域化监测具有重要的指导意义。

【关键词】 直肠肛门闭锁/狭窄;两维图论聚类

Using two-dimensional graphic cluster method to study the geographic distribution pattern of patients with anorectal atresia/stenosis, in China QIAO Liang*, ZHU Jun, YUAN Ping, DAI Li, WANG Yan-ping, ZHOU Guang-xuan, HAN Ying, LIU Xiao-xia, ZHANG Xun, CAO Yi. *Department of Epidemiology, West China School of Public Health, Sichuan University, Chengdu 610041, China
Corresponding author: ZHU Jun, National Center for Birth Defects Monitoring, the Second West China Hospital, Sichuan University, Chengdu 610041, China Email: zhu jun1@163.com; YUAN Ping, Department of Epidemiology, West China School of Public Health, Sichuan University, Chengdu 610041, China Email: yuanp1117@126.com

【Abstract】 Objective To investigate the geographic distribution pattern of patients with anorectal atresia/stenosis in China, in order to provide clue for research on its etiology. **Methods** Data were collected from Chinese Birth Defects Monitoring Network (CBDMN), which was a hospital-based congenital malformations registry system. From 2001 to 2005, all fetuses with more than 28 weeks of gestation and neonates up to 7 days of age, were monitored. Two-dimensional graphic cluster method was used to divide monitoring stations into different classes with the incidence rates of anorectal atresia/stenosis. **Results** The overall incidence of anorectal atresia/stenosis was 3.17 per 10 000 during 2001 to 2005. The incidence was higher in Eastern than that in Mid or Western parts of China and the difference was statistically significant ($z=2.50, 3.69; P=0.012, <0.001$). The monitoring stations were grouped into 6 classes. Class I was with Helongjiang, Jilin and Liaoning; Class II was with Fujian, Guangdong, Hainan, Guangxi, and South Hunan and Jiangxi; Class III was with Beijing, Tianjin, Hebei, Shandong, and North Jiangsu and Anhui; Class IV was with Zhejiang, Shanghai, South Anhui and Jiangsu, North Hunan and Jiangxi, Hubei, Henan, Shanxi and Inner Mongolia, Class V was with Ningxia, Gansu and Qinghai; and Class VI was with Shaanxi, Sichuan, Chongqing, Yunnan, Guizhou, Xinjiang and Tibet. **Conclusion** Our findings discovered the geographic distribution patterns of patients with anorectal atresia/stenosis in China. It is important to further analyze the relevant environmental factors attached to it so a better regional monitoring system for anorectal atresia/stenosis can be operated.

【Key words】 Anorectal atresia/stenosis; Two-dimensional graphic cluster

DOI: 10.3760/cma.j.issn.0254-6450.2009.02.016

基金项目:国家“十一五”科技支撑计划资助项目(2006BAI05A01)

作者单位:610041成都,四川大学华西公共卫生学院流行病学教研室(乔良、袁萍、韩颖、刘潇霞、张驯、曹奕);四川大学华西第二医院中国出生缺陷监测中心(朱军、代礼、王艳萍、周光宣)

通信作者:朱军,Email: zhu jun1@163.com; 袁萍, Email: yuanp1117@126.com

直肠肛门闭锁/狭窄是以肛门缺如、直肠闭锁(狭窄)或直肠肛门闭锁(狭窄),与邻近器官有瘘或无瘘为特征的先天性畸形,其发病率居消化道先天畸形的前列,病因与遗传因素有关外,还与诸多环境因素,尤其是地理环境因素有关。我国从 1986 年开始建立了以医院为基础的全国出生缺陷监测网,对包括直肠肛门闭锁/狭窄在内的多种高发严重畸形进行连续动态的监测。本研究兼顾监测点之间直肠肛门闭锁/狭窄发病率的相似程度和各监测点的毗邻关系,依据两维图论聚类分析原理对全国监测点按发病率进行归类,旨在了解直肠肛门闭锁/狭窄的空间分布特征和规律,为病因学研究提供线索。

资料与方法

1. 研究对象:2001—2005 年全国出生缺陷监测医院出生的妊娠满 28 周至出生后 7 d 内的围产儿,包括活产、死胎、死产和 7 d 内死亡者。

2. 监测医院:中国出生缺陷监测中心采用以医院为基础的监测方案,此法为 WHO 推荐在发展中国家使用较为理想的出生缺陷监测方法。参加监测的县级及以上医院分布在全国 31 个省、市、自治区 138 个市县。

3. 资料收集:监测医院每季度按月统计围产儿基本信息,填写“围产儿数季报”,对诊断的直肠肛门闭锁/狭窄病例,由监测人员调查后填写“出生缺陷儿登记卡”。所有报表每季度上报省级出生缺陷监测办公室,经审查后上报中国出生缺陷监测中心。具体的监测方法和质量控制见文献[1]。

4. 病例纳入排除标准:将 ICD-10 中编码 Q42.1 和 Q42.3 的确诊病例纳入研究范围。根据国际出生缺陷监测信息交换所规定的标准排除异位肛门,轻度狭窄不需要手术矫正者。

5. 资料分析:

(1)空间分布:采用 Excel 软件建立数据库,用 ArcView GIS 3.2 软件制作全国各省直肠肛门闭锁/狭窄发生率专题地图,以 Poisson 回归比较不同地域之间直肠肛门闭锁/狭窄发病率的差别,该过程由 STATA 9.0 软件完成。

(2)两维图论聚类分析:图是空间一些点与连接点的线的集合。根据两维有序样品聚类的基本要求,既要保证类内疾病相关变量的相似性和其外部差异性,又要保证类内各地理单元的连通性,设加权连通图为 $G=(V, E, D)$ 。其中 V 是各地理单元(本研究为各监测点)地理位置的集合; E 是各监测点邻接

关系的集合,即初始区位联系矩阵 $B^{(0)}$ (公式 1); D 是在图论的树算法中表示各点之间的权值,即初始疾病距离矩阵 $D^{(0)}$ (公式 2)。以加权连通图为 $G=(V, E, D)$ 为基础,用 Kruskal 法构造最小生成树(MST):先选择并连接距离测度最小的两点,再在剩余的点中选择与这两点之一距离测度最小的点与之连接,逐次把剩余的点与已连成的点连接起来,直至把所有的点全部连接完毕。该过程由 DPS7.05 软件完成。

$$B^{(0)} = (b_{ik}^{(0)})_{n \times n}; i, k = 1, 2, \dots, n \quad (1)$$

$$D^{(0)} = (d_{ik}^{(0)})_{n \times n}; i, k = 1, 2, \dots, n \quad (2)$$

式中 $b_{ik}^{(0)}$ 为第 i 与第 k 个监测点之间的区位联系标号, $b_{ik}^{(0)}=1$ 表示两者接壤, $b_{ik}^{(0)}=0$ 表示两者不接壤。 $d_{ik}^{(0)}$ 为第 i 个监测点发生率与第 k 个监测点发生率之间的相似性距离。

用图论中求“颈”的方法分割 MST^[2]。具体步骤:①求“主干”: n 个点由 $(n-1)$ 个边互相连接,其中有两个端点仅有一条边连接,其余点至少有两边连接,因而构成了一条无回路的链,被称为“干”。其中,边数最多的干称为 MST 的“主干”(或“直径”)。②求“子主干”:以 MST 主干上的任一点为起点,在 MST 上找出除主干以外边最多的一条干,称该干为该点的子主干,子主干的边数为该点的“深度”。③找“颈”:规定一个 >1 的整数 a , 在主干上找出深度 $\geq a$ 的点的子主干,诸子主干公共部分中深度为零的各点间的边长为“颈”。④在图中删除颈,使 MST 分割成若干部分,从而实现对各监测点的分类。

通过 ArcView GIS 3.2 软件将最小生成树及其分割后的图层数据叠加至国家基础地理信息系统 1:400 万数据的省界图层中,制成两维图论最小生成树聚类图。并将聚类结果制成聚类地图,以便直观观察。

结 果

2001—2005 年全国共监测围产儿 2 670 367 名,发现直肠肛门闭锁/狭窄病例 846 例,总发生率为 3.17/万。

1. 空间分布:图 1 为 2001—2005 年全国各省(自治区、直辖市)直肠肛门闭锁/狭窄总发生率专题地图。参照中华人民共和国国家统计局 2003 年对我国东、中、西部的划分标准^[3],我国直肠肛门闭锁/狭窄高发区集中在东部地区,而总发生率 $<1.99/$ 万

的省(自治区、直辖市)基本位于西部地区。

疆、西藏(图3)。

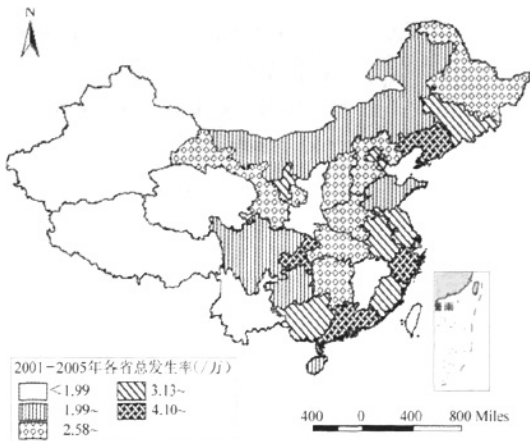


图1 2001—2005年全国各省直肠肛门闭锁/狭窄总发生率分布

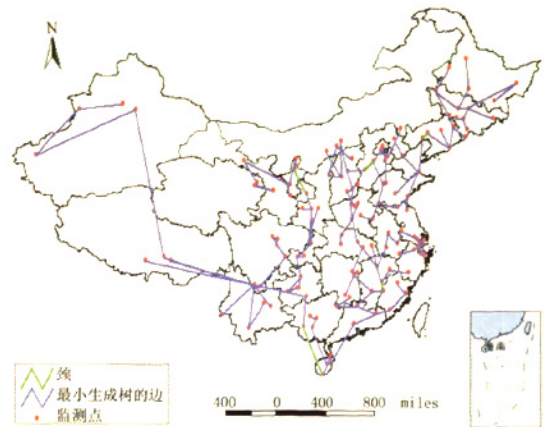


图2 2001—2005年全国直肠肛门闭锁/狭窄监测点二维图论最小生成树聚类图

2001—2005年我国东、中、西部地区直肠肛门闭锁/狭窄总发生率分别为3.59/万、2.94/万和2.54/万。Poisson回归结果显示,东部地区的发生率高于中、西部,而中部和西部地区发生率的差异无统计学意义($z = 1.37, P = 0.170$)(表1)。

表1 2001—2005年全国不同地区直肠肛门闭锁/狭窄空间特征别发病率比较

地区	出生人数	畸形例数	发生率(/万)
东部	1 299 956	467	3.59 ^{a,b}
中部	780 155	229	2.94
西部	590 256	150	2.54
合计	2 670 367	846	3.17

注:^a东部与中部地区比较, $z = 2.50, P = 0.012 (P < 0.05)$; ^b东部与西部地区比较, $z = 3.69, P < 0.001$

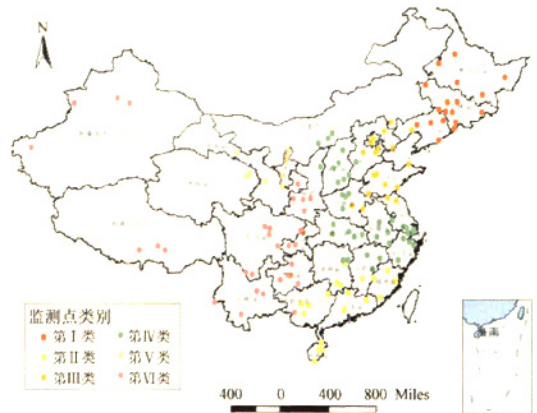


图3 2001—2005年全国直肠肛门闭锁/狭窄监测点聚类图

2. 二维图论聚类分析: 兼顾各监测点直肠肛门闭锁/狭窄发生率的相似性和各监测点间的空间相关性, 构建最小生成树(图2)。根据图论中求“颈”的方法, 定义正常数 $a = 2$, 可将监测点聚为6类, 图2中绿色线段即为类间分割线(即“颈”)。

按照2001—2005年直肠肛门闭锁/狭窄总发生率, 将聚成的6类结果由高到低分为I~VI类。第I类监测点主要分布于黑龙江、吉林、辽宁; 第II类主要分布于福建、广东、海南、广西以及湖南和江西的南部; 第III类主要分布于北京、天津、河北、山东、江苏和安徽的北部; 第IV类主要分布于浙江、上海、安徽和江苏的南部、湖南和江西的北部、湖北、河南、山西、内蒙古; 第V类主要分布于宁夏、甘肃、青海; 第VI类主要分布于陕西、四川、重庆、云南、贵州、新

讨论

直肠肛门闭锁/狭窄是儿童最常见的消化道畸形之一。由于肛门和直肠的病理改变, 1/3的患儿术后仍存在不同程度的排便功能障碍, 多数需要终生治疗, 严重地影响患儿的生活质量和心理发育, 给患儿、家长和社会带来沉重的负担^[4]。有研究表明^[6,9], 孕期接触过多环境污染物, 可能增加出生缺陷的发病风险。本研究显示, 我国东部地区直肠肛门闭锁/狭窄的发生率高于中部和西部地区。东部地区经济水平相对较高, 工农业基础较为雄厚, 主要的工业和海洋水产业的发达地区也大多分布于东部, 工业污染以及河流域污染在这些地区较为严重, 由此推测该地区孕期妇女可能接触更为严重的各种物理、化学等有害物质, 从而易导致直肠肛门闭锁/狭窄的发生。

聚类分析是按“物以类聚”原则研究事物分类的方法,其基本思想是将每一个样本视为多维空间中一个矢量的端点,求出各样本之间的相似程度,按相似程度的大小,把样本归类。该方法具有分类明确的优点,因而在许多学科领域被广泛应用。然而,传统的聚类分析方法在分析疾病空间结构时所得结果不易做出合理的解释。在疾病空间结构聚类分析中,不仅要注重疫源地之间的相似程度,还应考虑取样地理单元,即不同取样地理单元的毗邻关系。两维图论聚类分析模型将二维有约束空间系统聚类思想与图论中的最小生成树方法有机结合,利用地理信息系统(GIS)的空间分析手段与图论的树算法相结合的方法进行地理区划研究,既考虑到各区内部的相似性,又保持了空间上的连通性,对于分析各地理单元之间的相似性、展示疾病空间结构状态、识别疾病空间结构异质性界限等具有重要意义。罗罡辉等^[7]将主成分分析与两维图论聚类分析有机结合,确立了土地整理优先度的评价方法。曹阳等^[8]基于两维图论聚类分析对中国储粮区域进行了划分,为区域性储粮技术应用和制定宏观区域性储粮技术规程提供了参考依据。然而,目前将两维图论聚类应用于出生缺陷的研究尚鲜见报道。

本研究采用两维图论聚类对全国直肠肛门闭锁/狭窄的发生率范围进行了划分,研究结果对于进一步分析和监测直肠肛门闭锁/狭窄的环境影响因素具有重要的指导意义。一方面,先天畸形的发生除与遗传因素有关外,还与地理环境、气候、经济、文化水平等诸多因素有关,按照两维图论聚类分析的结果挖掘具有地区差异性的环境特点,有利于深入分析直肠肛门闭锁/畸形的相关环境影响因素;另一

方面,尽管我国直肠肛门闭锁/狭窄的发生率呈现出东高西低的分布特点,然而,各地区内部仍然具有其各自多层次的经济、自然环境特点及人群特征,就监测能力而言,如此大范围的监测难以实现相关影响因素的细化,而本研究的分类结果为在各地实现直肠肛门闭锁/狭窄的环境影响因素的可及性区域监测提供了思路,通过区域划分,可在同一类地区内定期详细收集区域内各地理单元的相关环境危险因素,实现监测的区域化。

此次分类考虑的是各监测点之间而非各省之间的地理毗邻关系,使得分类结果更能体现地理区划的要求。但是,如果在同一省内的不同监测点经聚类分析后,分属于不同类别时,将会增大从各省行政区域层面进行监测的难度。

参 考 文 献

[1] 代礼,朱军,周光董,等.1996—2000年全国神经管缺陷的动态监测.中华预防医学杂志,2002,36(6):402-405.
 [2] 方开泰,潘恩沛.聚类分析.北京:地质出版社,1982.
 [3] http://www.stats.gov.cn/tjzs/t20030812_402369584.htm.
 [4] 王维林.先天性肛门直肠畸形基础与临床研究.继续医学教育,2006,20(18):22-24.
 [5] 白亚娜,屈燕,胡晓斌,等.出生缺陷致病因素的多因素分析.中国妇幼保健,2004,19:44-46.
 [6] 李思涛,肖昕,刘秀香.中国围产儿出生缺陷危险因素的Meta分析.临床儿科杂志,2008,26(4):350-353.
 [7] 罗罡辉,吴次芳,徐保根.土地整理优先度评价方法及其应用研究.浙江大学学报(农业与生命科学版),2004,30(3):347-352.
 [8] 曹阳,卞科,陈春刚,等.基于两维图论聚类分析的中国储粮区域划分.中国粮油学报,2005,20(4):122-124.

(收稿日期:2008-10-20)

(本文编辑:张林东)