

中国出生缺陷区域发生水平的 Kriging 插值模拟研究

武继磊 郑晓瑛

【摘要】 目的 整合文献数据和原有调查数据,绘制中国出生缺陷发生水平阶梯地图,展示中国出生缺陷发生水平的地理差异和区域性集中趋势。方法 利用1987-1988年调查数据为基础,查询1988-2002年间发表文献中关于出生缺陷发生率的数据,在 ARCGIS 软件的支持下,利用 Kriging 插值,分析各个区域的出生缺陷发生水平,绘制出生缺陷发生水平地图。结果 通过文献数据和原调查数据相结合,完成连续平滑的中国出生缺陷发生水平模拟地图,提示中国出生缺陷发生水平的区域特征。结论 利用 Kriging 等空间统计分析技术,可以实现抽样调查数据对全国出生缺陷发生水平的估计,较为准确的反映中国出生缺陷发生水平的分布特点。

【关键词】 出生缺陷; 地理信息系统; Kriging 插值

A simulation on the emerging of birth defects in China with Kriging interpolation method WU Ji-lei, ZHENG Xiao-ying. Institute of Population Research, Peking University, Beijing 100871, China

【Abstract】 Objective To describe the regional patterns of emerging level on birth defects with simulated map by original studied and reference data. **Methods** Based on geographical information system, Kriging interpolation method was used and some related environmental factors were drawn from geographic information system software. **Results** The Kriging method had drawn continual and smooth map on birth defect emerging level and shown some clues with environmental factors. **Conclusion** Kriging method and advanced spatial statistical and analytical methods were helpful in public health studies while simulated map suggested the feasibility of sampled data on national disease surveillance and giving some useful clues to etiological analysis.

【Key words】 Birth defects; Geographic information system; Kriging interpolation

出生缺陷作为影响人口出生素质的重要因素之一,受到越来越多研究者的关注。已有研究表明,出生缺陷发生的原因受到遗传、环境以及遗传与环境共同作用的影响^[1]。在种族集聚、环境特征等人类居住的区域空间位置差异下,人群出生缺陷发生水平也各不相同。这种出生缺陷发生水平在区域位置上表现出的差异,一方面是出生缺陷风险因子在不同空间位置作用的结果;同时,根据区域的差异特征,可以很好的揭示出生缺陷发生的潜在风险因子,为出生缺陷的病因分析提供线索。

我国国土面积广、环境差异显著,人群种族集聚的分布有其特征。因此,研究我国人口出生缺陷的

发生情况,对人群出生缺陷发生水平的区域差异进行分析就显得很有必要。1986年10月至1987年9月,卫生部在全国实施了以医院为基础的出生缺陷监测,选择国内945所医院作为监测点,覆盖除台湾以外的各省,得出了该时期全国各省出生缺陷发生水平^[2]。但从流行病学和病因学角度上看,这次统计是以各省的行政区域为统计单元,未能就出生缺陷发生水平在我国分布模式作深入的分析,可以说是一个遗憾。为此我们以1986年以来上述监测系统中公开发表的区域出生缺陷发生水平等文献数据为基础资料,利用基于地理信息系统(GIS)的空间插值技术,绘制我国出生缺陷发生水平的区域差异地图,为出生缺陷风险评估与干预提供更多的依据。

资料与方法

1. 数据来源: 监测数据来自 1986 年 10 月至

基金项目: 国家重点基础研究发展规划“973”资助项目(2001CB5103); 国家人口和计划生育委员会国家出生缺陷干预工程资助项目[计生科(2000)13号]; 教育部科学技术研究重点资助项目(02185); 北京大学“985”和“211”资助项目(20020903)及中国博士后基金资助项目

作者单位: 100871 北京大学人口研究所/世界卫生组织生殖健康合作中心/北京大学中国人口健康与发展中心

1987 年 9 月卫生部组织全国 29 省(直辖市、自治区)945 所医院同步进行住院分娩的围产儿(含死胎、死产)出生缺陷监测,调查了 1 243 284 例围产儿,其中畸形儿 16 172 例,出生缺陷总发生率为 13‰(图 1)。为了和 1986-1987 年数据相匹配,我们采取了文献回顾法,通过整理国内期刊数据库,共获取 1988-2002 年 12 月前公开发表的期刊文献 256 篇,经过筛选,其中 223 篇文章涉及全国 210 个县级以上地区的出生缺陷发生率(图 1)。地理数据采用中国科学院地理科学与自然资源研究所开发的 1:400 万县级行政数字地图。

2. Kriging 插值^[3]:其原理为假设任意一个测量值被当作一个自由函数(或自由过程、自由场或随机场)的一次实现,并且任何变量的空间变化被表示成三部分之和,即结构分量(指常数均值或多项式趋势)、空间关联分量和白噪音(或称空间不关联的残余项)。Kriging 使用周围样点值来进行这项预报,它是一种最优无偏线性估值方法。对变量在点 x 处的估计值 $Z^*(x)$,可以通过该点影响范围内的 n 个有效观测值 $Z(x_i)$ 的线性组合得到,即:

$$Z^*(x) = \sum_{i=1}^n \lambda_i Z(x_i)$$

式中: λ_i 是赋予观测值 $Z(x_i)$ 的权重,表示各观测值对估计值 $Z^*(x)$ 的贡献,其和为 1,在保证估计无偏性和最优性条件下,由变量半方差函数计算得到。

3. 方法:在地理信息系统软件 ARCGIS(ESRI)环境下,将 1986-1987 年调查的各省市人口出生缺陷总发生率作为该省市下属各县市的出生缺陷平均发生率,如某县市有文献数据提供的出生缺陷发生率,且属于基于住院分娩的围产儿进行统计,则该县市的出生缺陷发生率以较新的文献为准。各县市的地理编码由空间计算的该县市的几何中心点为准,应用 ARCGIS 软件的空间分析模块中的 Kriging 插值功能,生成出生缺陷发生水平的空间插值地图,并叠加全国各省市的行政地图以示空间位置,出生缺陷发生水平以等间距着色标示,得到我国出生缺陷区域发生水平的 Kriging 插值模拟地图(图 2)。

结 果

对比原出生缺陷地图(图 1)和 Kriging 插值模拟地图(图 2)可以看出,以较大行政区划单元统计的出生缺陷发生水平的空间模式具有很大的局限性,如人口的自然分布并不依行政界线的分隔而呈截然

分离的状态,行政界线附件的人口分布呈自然过渡状态,因此其出生缺陷率也不是非此即彼的显著差异;在同一个大的行政区域单元内,人口的分布、自然环境差异也各不相同,我国西部的定居人口分布受自然条件影响,其分布相对集中在绿洲和沿交通干线地带,由于行政单元内广袤的沙漠、戈壁等间隔,人口分布存在显著差异,此时按照较大的行政单元进行统计,则肯定存在较大的偏差。本文以县市为统计单元,地理区划虽然仍稍偏大,但考虑出生缺陷为小概率事件,样本不足会影响监测数据质量,另外通过 Kriging 插值,可以适当弥补因样本不足而产生的缺憾。插值的出生缺陷发生水平模拟地图能够更加详细的反映全国出生缺陷发生水平的区域模式。从图 2 可以看出,出生缺陷高发区主要分布在山西和陕西两省,这与山西省神经管出生缺陷发生率为“珠穆朗玛峰”之称相一致。山东省的出生缺陷发生率存在较大的空间差异,而不是原调查结果描述的显著高发区。

胡焕庸 1935 年发表“中国人口之分布”,编制我国第一幅等值线人口密度图,提出瑗珲-腾冲线,揭示出中国人口 96% 集中于东南半壁,西北省区面积约占全国总面积的一半,而人口只占总人口 4%。经半个世纪的研究,瑗珲-腾冲线走向东北-西南,正是西风带与东南季风的边缘接触带;又与生态脆弱带基本吻合;还与缺硒和大骨节病的分布有关。陈述彭在论述 1954 年其横穿太行山脉的实地考察时,看到晋城附近来自河南黄河沿岸的移民,经过 2~3 代人的定居,往往又受到大骨节病的折磨,降低了生殖能力,重新出现田园荒芜、人烟稀少的景象,他提出跨越瑗珲-腾冲线的移民、屯垦,并不是容易成功的,这条简单的分界线,所揭示的科学内涵非常深刻^[4]。本文通过 Kriging 插值产生出生缺陷发生水平的分布,也不悖这条分界线内涵,沿此线的各个区域,其出生缺陷属于较高的发病区域,广东省一带出生缺陷的高发区主要受基因遗传的地中海贫血缺陷影响,表现与环境关联性较小。

为了更好的描述 Kriging 插值模拟地图的结果,在原有数字地图的基础上,附上我国数字高程图、能源矿产图和土壤类型图,以更好的论述结果。对照图 2~5,可以看出,人群的出生缺陷发生水平与其所处的海拔高度落差、地质状况以及相应的土壤状况有关,我国出生缺陷高发区一般具有如下特征:

1. 由图 3 可见,出生缺陷发生水平较高的地区

基本分布在海拔高度 1000~2000 m 的范围内,特别是这个海拔高度具有较大落差,如黄土高原上的山西和陕西两省,受水土流失的侵扰,人们的居住环境多为沟壑遍地中较为平坦地区,但这些平坦的地区又多为冲刷沉积区,若岩层中重金属等有害物质超标,则这些沉积区势必成为有害物质的富集区,类似地区还有山东省鲁中丘陵区、云贵高原的横断山区等。较大的落差说明该地区存在地质断层。由此,不同时期岩石或土壤母质均暴露在同一环境下,从而导致地质环境更为复杂。

2. 从图 4 可以看出,在出生缺陷高发区除地质环境表现的较为复杂和较大落差的地质断层特点外,还可以发现其具有分布较多能源矿产的特点。其矿产一般为质量较低的烟煤、褐煤等,这些矿产的开采和利用更加容易对环境造成污染。如山西省的煤炭矿产资源居全国之首,其神经管出生缺陷发生率也是世界上高发区之一。据调查,该省的 8 个主要煤炭矿区的出生缺陷率均高于非矿区,且吕梁山区的烟煤矿区出生缺陷发生情况要高于太行山区的质优无烟煤矿区;此外,云贵高原的烟煤矿区出产的煤多为高硫、高氟煤,其利用对环境产生二次污染,引起酸雨危及环境及人口健康,此外当地居民也多有氟中毒的报道。

3. 图 5 显示,在高发地区的土壤类型上,出生缺陷高发区的土壤一般为黏土、棕壤以及褐土等类型,这类土壤中容易含有重金属离子及其络合物,不易淋溶或降解^[5]。黄土高原区,由于降水的缺乏,有害物质的淋溶和降解较为困难,而云贵高原区,其降雨虽多,但黏土质阻碍了重金属离子的降解;山东半岛北部的高发区,为黄河入海口,其土壤环境特征为盐碱土,土壤盐碱化严重,海水内浸,易将底层重金属离子提到土壤浅层。

讨 论

近年来信息技术特别是 GIS 技术以及基于 GIS 的空间统计和分析等相关学科的发展,促进了空间数据的分析和应用。Kriging 插值正是利用了相邻区域属性值相似性较大进行分析^[6]。在既定区域内,

利用研究事物的空间属性,通过已知的区域属性值推测未知区域属性值的过程成为空间插值(interpolation)。空间插值一般是将离散的采样点数据转换为连续的曲面数据,从而可以进行多尺度的数据分析。除 Kriging 插值外,还有基于地理尺度的反向距离加权插值、全局多项式插值等方法,其中 Kriging 方法可以评估预测值的不确定性,采取不同的参数和算法,得到局域的最优、无偏估计,从而得到了更加广泛的应用^[7]。

插值模拟我国出生缺陷发生率水平的区域差异目的在于揭示其分布规律,由这些规律提供的病因学线索进行进一步的分析。由于 Kriging 方法具有光滑作用,即将极大值偏小估计,而极小值偏大估

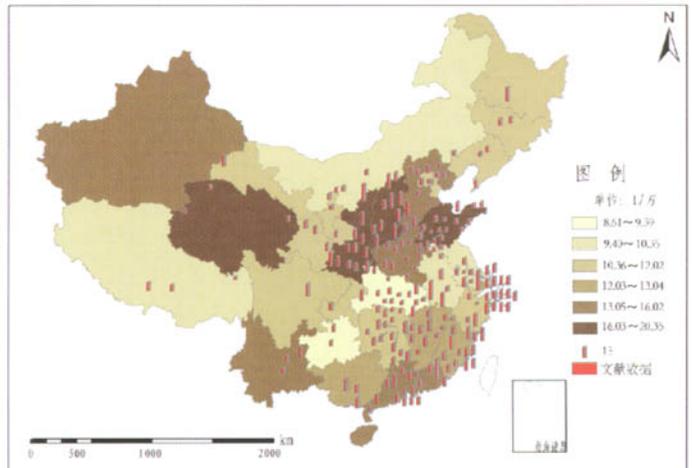


图1 1986-1987 年出生缺陷发生率和文献数据的出生缺陷发生率

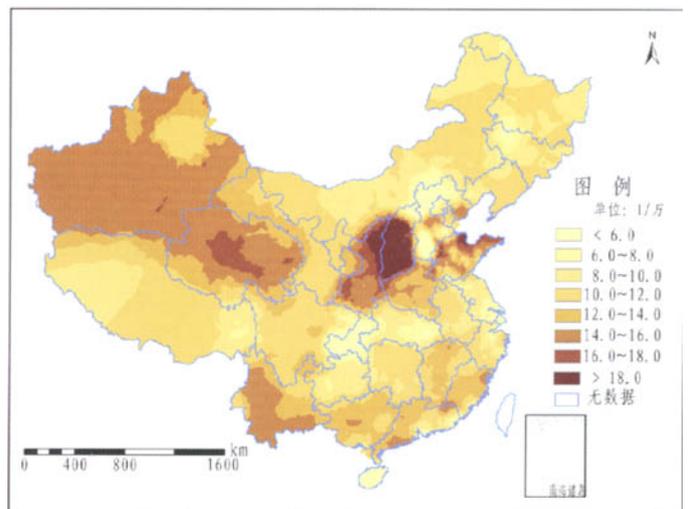


图2 Kriging 插值生成的出生缺陷发生水平模拟图

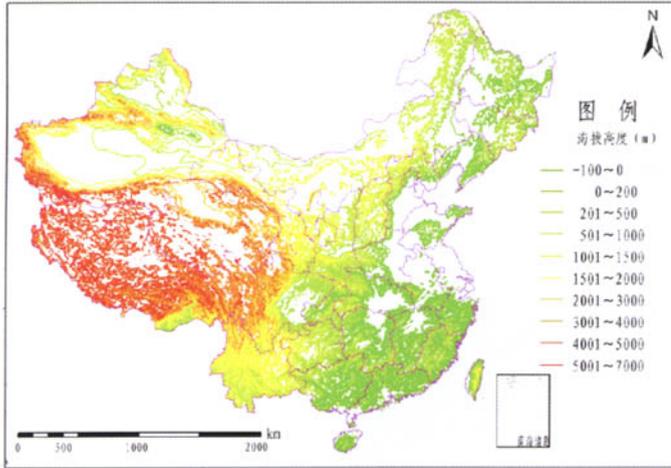


图3 我国海拔高度等值线图

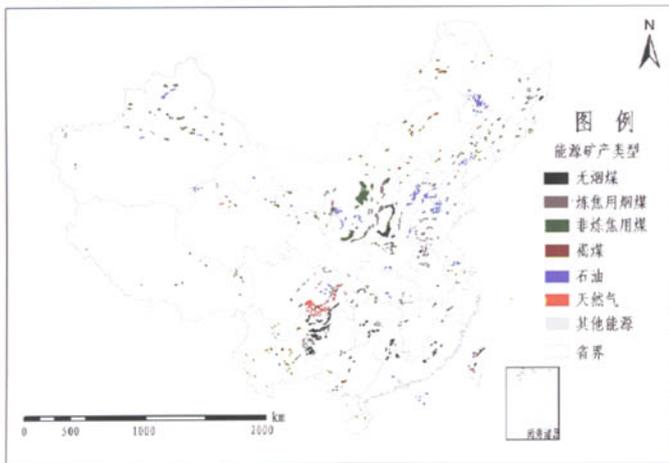


图4 我国能源矿产分布图

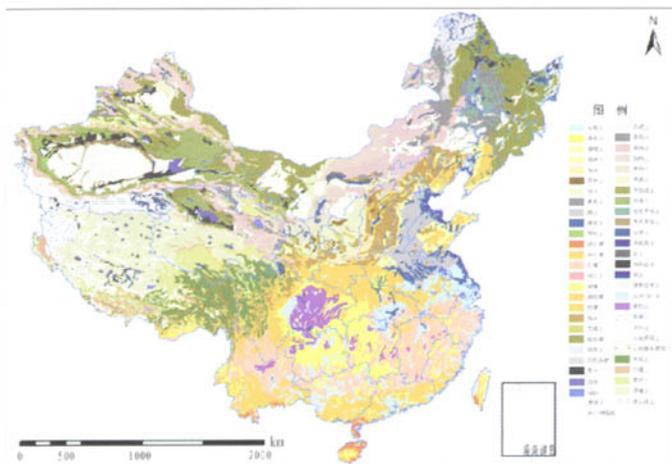


图5 我国土壤类型分布图

计,因此整幅图像趋于平缓。这对于出生缺陷来说,作为小概率事件,其监测结果受到随机因素的影响,Kriging 的平滑恰好能够消除出生缺陷监测数据中的异常值^[8]。因出生缺陷的发生,受到多种风险因子影响,这种地理意义上的平滑也可能忽视了人本身的影响因素如基因突变等非地理因素,但作为基因突变等风险因子的载体人依然具有空间分布特征,故而当研究不追求局域疾病发生率的准确性,而在于刻画整个区域的疾病发生模式时,如本次出生缺陷发生水平的插值模拟,能够揭示出生缺陷发生水平的区域特征,从而为病因学研究提供有效的线索。

当然,本文也存在诸多局限,主要表现在:①本模拟结果受文献数据和调查设置地点的影响,在我国西北部,其文献发表数量明显偏低,其中新疆、青海、内蒙古三个区域没有关于出生缺陷资料的公开发表数据,而其模拟的参考点主要依据是1986-1987年的调查结果,可以看出新疆、青海等区域呈现大片的出生缺陷发病水平相似的高发区,这与其人口依绿洲、交通干线分布存在明显的差异。②本文模拟所采用的出生缺陷发生数据主要与1986-1987年监测方法类似或者可比的情况下文献发表的数据,主要是基于医院的监测资料,受监测方法和种类的影响,其监测种类也主要是出生时所能看到的病种。根据WHO关于出生缺陷发生情况的冰山示意分布图^[9],若需分析5岁以下儿童的出生缺陷发生情况,则需要重新作调查和分析。③本文采取的文献数据来源于1988-2002年公开发表的文章和资料,其监测时间跨度较大,同一地点的多次监测结果按照平均值处理,但这期间出生缺陷发生水平的变化会影响到该地点的出生缺陷发生水平,从而对模拟结果产生影响。

总之,本研究为根据现有的数据和信息进行出生缺陷发生水平的空间分布

特征分析提供一个方法学上研究的新思路,从而为今后的出生缺陷水平监测或者类似流行病学调查提供借鉴。特别是卫生部即将在全国正在开始实施以人群为基础出生缺陷监测,这项由以医院为基础的出生缺陷监测转向以人群为基础的全局大规模监测将在今后几年内获取覆盖面更大、更为可靠的全国监测数据,如采取本研究论述的方法,可以对产生的数据再次充分利用,获取更可靠、准确的不同地区的出生缺陷发生水平,一方面提升对全国监测数据的分析水平;另一方面也可以从中挖掘传统流行病学方法不能发现的新的地区差异上的原因,为出生缺陷病因研究假设提供更多的依据。

参 考 文 献

[1] 安笑兰,符绍莲. 环境优生学. 北京:北京医科大学 中国协和医科大学联合出版社,1995.

[2] 中国出生缺陷监测协作组. 中国出生缺陷地图集. 成都:成都地图出版社,1992.
 [3] Liebhold AM, Rossi RE, Kemp WP, et al. Geostatistics and geographic information systems in applied insect ecology. *Ann Rev Entomol*, 1993, 38:303-327.
 [4] 陈述彭. 人口统计的时空分析. *中国人口·资源与环境*, 2002, 12(4):3-7.
 [5] 张辉. 土壤环境学. 北京:化学工业出版社,2006.
 [6] Harvey JM. Tobler's first law and spatial analysis. *Ann Assoc Am Geographers*, 2004, 94:284-289.
 [7] Carrat F, Valleron AJ. Epidemiologic mapping using the "kriging" method: application to an influenza-like epidemic in France. *Am J Epidemiol*, 1992, 135:1293-1300.
 [8] Oliver MA, Muir KR, Webster R, et al. A geo-statistical approach to the analysis of pattern in rare disease. *J Public Health Med*, 1992, 14:280-289.
 [9] WHO. Services for the prevention and management of genetic disorders and birth defects in developing countries. WHO Publication, WHO/HGN/GL/WAOPBD/99.1. Geneva, 1999.

(收稿日期:2006-06-02)

(本文编辑:张林东)

· 疾病控制 ·

浙江省流动人口肺结核病疫情状况分析

陈松华 王晓萌

资料来源于浙江省 2005 年结核病登记季报疫情资料以及疾病监测信息管理系统的肺结核病资料。流动人口肺结核病登记报告:主要是指以人户分离,户籍仍在原地,在外地居住 3 个月以上以打工为生的人群;流动人口肺结核病例控制要求:以县为单位按“浙江省外来流动人口肺结核病人管理办法(试行)”要求对流动人口中的肺结核病例进行登记报告和管理。资料的汇总、统计分析由 SPSS 13.0 和 Excel 软件完成。

结果与分析:2005 年累计发现流动人口活动性肺结核病例 7965 例,占全省活动性肺结核病例的 22.48%;其中流动人口涂阳肺结核病例 4096 例,占全省涂阳肺结核病例的 27.59%。在浙江省非流动人口肺结核病例为 27 462 例;涂阳病例分别为 4096 和 10 752 例,非流动人口肺结核病例的涂阳比例(51.42%)比流动人口肺结核病例(39.15%)的涂阳比例高,差异有统计学意义($\chi^2 = 381.98, P = 0.00, RR = 1.31$);由于涂阳肺结核病例具有明显的传染性,所以流动因素起到了传播的作用,是非流动人口病例的 1.31 倍($AR = 23.86$)。在省内各地涂阳肺结核病例中流动比例各地不平衡,高的地市流动人口涂阳肺结核病例占全体涂阳肺结核病例的 48.5%,低的仅占 5.2%,基本上是浙江省沿海地市流动人口涂阳肺结核病例比例较高。利用方差分析 SNK 多重比

较($\alpha = 0.05$),可以认为嘉兴、宁波、温州、杭州、金华市的涂阳流动比例较高,地区间无差异($P = 0.06$);湖州、台州、绍兴市次之,地区间无差异($P = 0.12$);舟山、衢州、丽水市较低,地区间无差异($P = 0.22$)。地区构成比显示流动人口肺结核病例基本上在沿海的杭州、宁波、温州、嘉兴、台州等市分布,占总体流动人口肺结核病例的 72.02%;活动性肺结核病例 1~4 季度流动比例分别为 18.71%、23.64%、23.91% 和 23.32% ($\chi^2 = 89.52, P = 0.00$),流动病例构成比中以 2 季度最高,为 29.99%;涂阳肺结核 1~4 季度流动比例分别为 22.90%、29.35%、29.10% 和 27.94% ($\chi^2 = 47.03, P = 0.00$),利用方差分析 SNK 多重比较,1 季度比另外 3 个季度的涂阳流动比例低,3 个季度间差异无统计学意义($P = 0.88$)。流动人口肺结核病例季节性波动比较大,人口外流的方向是由经济相对落后地区流向发达地区,病例的地理分布各地不平衡,主要倾向于省内经济相对发达的沿海地区。流动人群的总体传染性比非流动人群要高,而涂阳流动肺结核病例占全省涂阳病例的比例大,这对我们的结核病控制工作提出了严峻的挑战,流动肺结核病例的防治已成为现阶段浙江省结核病控制工作的重要问题之一,在结核病控制工作中应该进一步加强。

(收稿日期:2006-12-01)

(本文编辑:尹廉)