Analysis of geographical clustering of birth defects in Heshun county, Shanxi province

Wenxue Chi, Jinfeng Wang*, Xinhu Li, Xiaoying Zheng and Yilan Liao

China University of Geosciences, Beijing; Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing; Institute of Urban Environment, Chinese Academy of Sciences, Xiamen; Institute of Population Research, Peking University, Beijing, P.R. China

(Received 18 June 2007; final version received 25 November 2007)

Birth defects remain the leading cause of stillbirth and infant mortality in China, especially in rural areas. The objective of this research is to seek evidence of clustering of cases of birth defects within a highly endemic area and to identify the environmental and landscape characteristics associated with increased risk for birth defects in Heshun county. This study uses four years of data, 1998–2001, to identify clusters of birth defects in Heshun county. A spatial scan statistic was employed to examine the spatial and spatio-temporal clustering of birth defects in the study area. Statistically significant clusters with high relative risk (RR = 2.981, p = 0.008) were identified using a spatial scan statistic. The spatial cluster located in southeast Heshun county has a radius of 7.52 km and includes 27 villages and 27 cases, accounting for 21.25% of the total cases during the study period. The methodology applied in this study was useful for evaluating the spatial distribution of birth defects in Heshun county from 1998–2001. The identified areas may be critical to control birth defects and may provide important direction for further study and targeted interventions.

Keywords: birth defects; China; clustering analysis; spatial scan statistic

Introduction

Birth defects, as defined by the March of Dimes Birth Defects Foundation, refer to any anomaly, functional or structural, that presents in infancy or later in life and is induced by events preceding birth, whether inherited or acquired. Varying from minor cosmetic irregularities to life-threatening disorders, birth defects are the major cause of infant mortality and a leading cause of disability (Carmona 2005). The etiology of birth defects is not well understood, but increased knowledge of the causal factors is prerequisite for applying preventive interventions. A better understanding of the spatial distribution patterns of birth defects would help identify areas and populations at high risk. Shanxi province, a northern region in China, has the highest ratio of neural-tube birth defects in the world (Wu et al. 2004; Li et al. 2006). An understanding of spatial distribution of birth defects in the rural areas of China plays a key role in the successful prevention and medical treatment strategies.

Geographic information systems (GIS) software combined with methods of spatial analysis provide powerful new tools for understanding birth defects and for improving

*Corresponding author. Email: wangjf@lreis.ac.cn
disease prevention and control (Chaput et al. 2002). Several papers were published in a 1999 issue of the *European Journal of Epidemiology* that discussed the history of birth defects cluster investigations, spatial variation analyses and methods for investigation (Williams et al. 2002). These papers provide useful information and examples of birth defects cluster investigations. In this study, the spatial distribution of newly recognized birth defects was investigated for nonrandom patterns and clusters in an area known to be endemic for birth defects. A statistic for concurrently examining spatial and temporal data is available through the use of SaTScan (Kulldorff and Nagarwalla 1995; Kulldorff 1997; Kulldorff et al. 1997, 1998) software. The spatial scan statistic software SaTScan has been used to evaluate the clustering of several other health outcomes including leukemia, (Kulldorff and Nagarwalla 1995; Hjalmars et al. 1996), brain cancer (Kulldorff et al. 1998), prostate cancer (Jemal et al. 2002; Klassen et al. 2005; Abe et al. 2006; Mather et al. 2006; Oliver et al. 2006) and breast cancer (Hsu et al. 2004; Fukuda et al. 2005; Sheehan and DeChello 2005). The objectives of this study are to detect the existence of spatial and temporal clusters of birth defects in Heshun county from 1998–2001. Subsequently, we discuss two possible hypotheses that may explain the pattern of the occurrence of the disease.

**Materials and methods**

**Study site**

The study site is located in the Taihang Mountain region, in eastern Shanxi province (113°03′~113°56′E, 37°03′~37°35′N) at an average 1300 m of altitude. The maximum distance from east to west is about 75 km and the maximum distance from north to south is about 35 km in this county (Figure 1). The area has a total population of 134,522 and encompasses 2250 km². The economic development of the study region is very slow. The impact of low incomes on the economic standard of living of the population is reflected in the levels of direct nourishment and medical treatment.

**Cases**

There are 322 administrative villages and one town in the study area. Records of birth defect cases and births for four years (1998–2001) were acquired based on hospital registers and also investigations in villages. These investigations were necessary because some pregnant women chose private or community general practice clinics or home births rather than hospital births. Trained healthcare workers conducted face-to-face interviews with mothers. Interviewers used a structured questionnaire and the same interviewers and questionnaire were used for all mothers. The case information included the reporting date and residential location. For the four-year period, the average annual incidence of birth defects in the 322-village area was 29.8 cases per 10,000 persons. The total amount of birth defects in the 322-village area was 135 cases during 1998–2001. Of the 135 cases, 127 had verifiable and integrated addresses. Eight cases (5.93%) were not used due to missing residential address information data. Thus, 127 cases of birth defects were included in the analysis. The data were organized into a case-file, a control file and a geographical coordinate-file and used as input data in the spatio-temporal scanning software SaTScan (URL: http://www.satscan.org). Because the main object of this study was to detect the clustering of birth defects in these poor rural villages, the town was not included, as the urban environmental factors and impact of migration there are somewhat complex. The GIS for spatial analysis determined the locations of the 322 villages. All cases with
address information were geocoded to a point location, where possible, with the ArcView GIS software. The latitude and longitude in decimal degrees for each village centroid were used. Because there were no boundaries defined for the villages, we drew them for each village using a Voronoi chart (Wu et al. 2004).

Methods

Spatial scan statistic

A “cluster” in this context is a real or perceived aggregation of more than the expected number of birth defects cases in a population over a specified time period (Williams et al. 2002). A spatial scan statistic (SaTScan) was applied to identify clusters of birth defects. The spatial analysis in SaTScan imposes a circular window on the map (Green et al. 2003). This window moved over the study region and centered on the centroid of each village. For each location, the radius of the window varies in size from 0–50% of the total population at risk (Vigre et al. 2005). Possible clusters were tested within the variable window around the centroid of each village group (Fang et al. 2006). To identify statistically significant spatial clusters the likelihood ratio test was used. The likelihood function was computed for each specific window and the one with the maximum likelihood constitute the most likely cluster. Details of how the likelihood function is maximized over all windows under the Poisson assumption have been described elsewhere (Kulldorff et al. 1997; Mostashari et al. 2003). For each circle the likelihood of finding the observed
number of cases inside and outside the circle was calculated. The circle with the rate least likely to have occurred by chance is then determined and secondary clusters with excesses are also located. Circles with statistically significant ($p < 0.05$) elevations in birth defect rates were determined through Monte Carlo hypothesis testing (Forand et al. 2002).

For this study, birth defects were defined as cases and the number of births was the control. The Poisson probability model was chosen as the statistical test for this study as recommended in the SaTScan™ User Guide for version 6.1 by Kulldorff in March 2006. Cases and controls were aggregated to the village level. The maximum cluster size was set to 50% of the total cases and controls to scan from small to large clusters (Green et al. 2003). To ensure sufficient statistical power the number of Monte Carlo replications was set to 999. Scanning was also set to search only for areas with high proportions of birth defects. The default of no geographic overlap was used so that secondary clusters would not overlap the most significant cluster. For each of the most likely clusters the SaTScan output includes a listing of the geographic subdivisions, the numbers of observed and expected cases, the population, the relative risk (RR), log likelihood ratio (LLR), and the $p$-value (Abe et al. 2006). Using ArcView GIS software we spatially joined the dbase file exported by SaTScan to the study area administrative region layer and produced a map displaying their physical location of birth defects clusters. In this study, we used three different models to analyze the data – a purely spatial, a purely temporal and a space–time scan statistic.

**Results**

**Purely spatial scan results**

Analysis of confirmed cases of birth defects identified in 1998–2001 in a 322-village area in Heshun county showed there exists spatial variation of birth defects. Using the maximum spatial cluster size of < 50% of the total population, the spatial cluster analysis identified a single cluster consisting of 27 villages located in the southeast regions of Heshun county. The relative risk within the cluster was 2.981, with 27 observed cases compared with 10.55 expected cases. This elevated risk within a nonrandom pattern of birth defects distribution was significant ($p = 0.008$). The geographic distributions of clusters of the spatial scan statistic are shown in Figure 2. For the cluster identified, we list the observed versus expected number of birth defect cases, the radius, the relative risk, log likelihood ratio and the $p$ value. Table 1 shows the results generated by the spatial scan method when the maximum cluster size was set to 50% of the total population.

**Purely temporal scan results**

In the purely temporal model, no real significant temporal clusters were identified during the period from 1998–2001.

**Space–time scan results**

The present study utilized the retrospective space–time analysis for high rates using a Poisson model to calculate expected cases in each village. The spatial scan window setting was set to a maximum cluster size of 50% of the study population, and the temporal scan window was set to a maximum cluster size of 90% of the study period. The result was that no significant clusters were identified.
Discussion

This study describes and analyzes the spatial and temporal patterns of birth defects in Heshun county from 1998–2001. Using GIS and spatial statistics, we investigated the spatial distribution of confirmed cases of birth defects and identified villages within an area highly endemic for birth defects diseases. Kulldorff’s spatial scan test has several advantages over other local cluster detection tests, namely, avoidance of multiple testing and not restricting scanning for clusters of a pre-specified size, thereby avoiding pre-selection bias (Mather et al. 2006). The present study identified villages in southeast Heshun county as having clustering, something not been previously reported. Additionally, the risk estimate provides valuable information about the geographic disparity of birth defects. The highest level of relative risk for this cluster was 2.981. The results of this analysis must be interpreted with caution due to the small number of case infants for specific birth defects (Araneta et al. 2000; Forand et al. 2002). Birth defects are a low probability event; the total amount of birth defects in the small study area during the four years was relatively small. Therefore, as a result of temporal and space–time scan statistic, no significant clusters were identified in Heshun county.

Table 1. Statistically significant spatial cluster of birth defects in Heshun county, 1998–2001.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>No. of villages</th>
<th>Observed&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Expected&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Radius</th>
<th>RR&lt;sup&gt;d&lt;/sup&gt;</th>
<th>LLR&lt;sup&gt;c&lt;/sup&gt;</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary cluster</td>
<td>27</td>
<td>27</td>
<td>10.55</td>
<td>7.52 km</td>
<td>2.981</td>
<td>10.147</td>
<td>0.008</td>
</tr>
</tbody>
</table>

<sup>a</sup>Observed: number of observed cases in a cluster; <sup>b</sup>Expected: number of expected cases in a cluster; <sup>c</sup>LLR: log likelihood ratio; <sup>d</sup>RR: relative risk.

Figure 2. Map of Heshun county with results from the SaTScan analysis showing the geographical location of the statistically significant spatial clusters (during the whole study period) as black circles.
Although much is still unknown about the etiology and development of birth defects, there is sufficient information to argue that birth defects are most likely caused by a complex combination of factors. Rather than a single explanatory risk, it is likely that many factors contribute to its development, from inherited genetic risk, to lifestyle patterns and environmental exposures to a range of protective and detrimental agents (Klassen et al. 2005). Identifying geographic areas and specific populations which are at an elevated risk of birth defects helps in the process of hypothesis generation for possible etiological mechanisms (Nkhoma et al. 2004). Below we discuss several factors that might explain these results.

**The terrain and water quality factors**

As can be seen in Figure 1 it is evident that the west of the study area is higher than the east in terrain and physiognomy. As a result, principal rivers flow from west to east following the terrain of the land. The 27 villages were located in the relatively lower altitude area. Shanxi is the biggest coal provider among the provinces in China and there are many coal mines in the study area, which indicates a peculiarity with respect to geological background. Coalmining operations are considered one of the most significant and widespread types of water pollution and contaminate the basic substances of life. Daniel Peplow and Robert Edmonds (2005) showed that metal-contaminated sediments cause adverse biological effects at all levels of biological organization from cellular to ecosystem-level responses, even where the corresponding surface water meets water-quality-based criteria. In addition, many chemical plants and paper mills are situated on both sides of these rivers. Moreover, although big industries located in the area purify their wastewater, small industries are suspected of discharging residue into the river. Rivers play a major role in assimilating and carrying off industrial and municipal wastewater, manure discharge and runoff from agricultural fields, and run-off from roadways and streets, which are all responsible for river pollution (Vega et al. 1998). We also know that contaminated groundwater may supply people for drinking or agricultural use at a considerable distance from the site and that contaminated food grown close to the site may be eaten just as much by people living far away as by local residents. However, the 27 villages were located at the lower reaches of these rivers. This means that people living near the downstream of these rivers have a higher risk of having a baby with birth defects. Therefore, the distribution of the polluted rivers in this study area should be a significant factor contributing to the presence of this particular clustering of cases (Goujard 1999).

**The chemical fertilizer factor**

Comparing Figures 3 and 4 with Figure 1, we find obvious evidence that the north area of the 27 villages had a higher usage of chemical fertilizers. Several studies have shown that chemical materials can cause health problem such as birth defects, nerve damage and cancer (Amler 1991; Scialli et al. 1997; Czeizel et al. 2004). As a result of the terrain, the rivers flowed to the 27 villages over an area with a high usage of chemical fertilizers. The usage of chemical pesticides and fertilizers impacted on the quality of the water villagers drank. The majority of people living in the study area are farmers, whose lifestyles have remained unchanged for a long time. Although these rivers are polluted, many local residents still live very near. Due to economic factors, some women perform agricultural work during pregnancy. Therefore, opportunities for pregnant women to come in contact
with chemical substances are increased. Li et al. (2006) reported that environmental factors are strongly associated with the risks for neural tube defects in Shanxi province. Chen et al. (2004) reported the chemical element content analysis of the water in study area. In her paper, direct measurement of soil chemicals and residents’ exposure to them helped to
assess the risk factors. In her conclusion, the chemical element content in these 27 villages area is obviously higher than in other village. Wu et al. (2004) also reported that the soil type and chemical elements in the lithosphere have an influence on human health in the study area. Evidence from these past studies has given our findings strong support.

Spatial analysis combined with analysis of environmental factors can increase understanding of birth defects within a highly disease endemic area. The authors believe that some mechanism besides these reasons was at work in distributing birth defects around these 27 villages. In fact, prenatal screening would have been extremely helpful. Unfortunately, there was very poor living and working conditions, hence the local people had no money to screen during the maternal pregnant period. Most people in rural villages are not covered by health care insurance. Although villagers understood that the county hospital provided high quality services, they were reluctant to go there, mainly because of the direct and indirect costs such as high medical costs and long traveling times to the hospital.

During the past 10 years, governments have been taking measures to prevent the occurrence of birth defects, which has indeed reduced the prevalence to some extent. However, because the underlying causes of many cases still remain unknown, the ratio of birth defects is still high (Li et al. 2005). Both natural and artificial environmental exposure and genetic predisposition are thought to contribute to birth defects, but very few tools are available to study their relative importance (Cordero 2005). As mapping and spatial analysis software become easier to use, public health professionals will increasingly employ these tools to aid in disease surveillance and cluster investigations (Forand et al. 2002). The spatial scan statistic method approach illustrated in this article would seem to be a useful starting point for birth defects analysis of clustering patterns. This study demonstrates the utility of geographic information systems and spatial analysis tools to clarify the distribution of birth defects. The current study should be viewed as a preliminary study of the spatio-temporal patterns of birth defects in Heshun county. Knowing specific areas with high rates of birth defects would help health policymakers to focus the scope of prevention programs and health care delivery, thus providing for efficient allocation of public health resources. In an area in which birth defects are highly endemic, aiming prevention strategies at areas of highest risk can potentially increase a program’s effectiveness (Himman et al. 2006). Persons at highest risk should be informed of that risk and of the options for risk reduction. Funds spent on programs might be better spent on areas where cost-effectiveness can be maximized (Chaput et al. 2002). This paper can assist state and local health departments or birth defects surveillance systems in the investigation of birth defects clusters by providing definitions and recommendations to aid in the development of a standardized strategy.

**Conclusion**

This study analyzed the spatial and temporal distribution of birth defects in Heshun county from 1998–2001 by the spatial scan statistic method. SaTScan identified a geographic area in southeastern Heshun county as the most likely cluster, which had not been previously reported. Our results provide evidence that environmental factors including terrain, water quality, usage of chemical fertilizers, socio-economic status and lifestyle behaviors may be contributing causes of the original clusters. The findings identified in this study may help other researchers conduct additional etiological studies and devise more comprehensive strategies for the prevention of birth defects in China. Furthermore, the use of spatial statistics to analyze birth defects monitoring data which
has a spatial component can be an efficient method of contributing to the knowledge of the epidemiology of important diseases of infant in developing countries.

References


Li ZW, Ren AG, Zhang L, Guo ZY, Li Z. 2006. A population-based case-control study of risk factors for neural tube defects in four high-prevalence areas of Shanxi province, China. Paediat Perinatal Epidemiol. 20:43–53.


