

Research article

Open Access

Exploratory spatial data analysis for the identification of risk factors to birth defects

Jilei Wu^{1,2}, Jinfeng Wang¹, Bin Meng¹, Gong Chen², Lihua Pang², Xinming Song², Keli Zhang³, Ting Zhang⁴ and Xiaoying Zheng^{*2}

Address: ¹Institute of Geographical Sciences and Nature Resources Research, CAS, Beijing, 100101, P. R. China, ²Institute of Population Research, Peking University, Beijing, 100871, P. R. China, ³Department of Resources and Environment, Peking Normal University, Beijing, 100875, P. R. China and ⁴Capital Institute of Pediatrics, Beijing, 100020, P. R. China

Email: Jilei Wu - wujl@reis.ac.cn; Jinfeng Wang - wangjf@reis.ac.cn; Bin Meng - mengb@reis.ac.cn; Gong Chen - chengong@pku.edu.cn; Lihua Pang - pang@pku.edu.cn; Xinming Song - xmsong@pku.edu.cn; Keli Zhang - kunqing@bnu.edu.cn; Ting Zhang - zhangting@263.net.cn; Xiaoying Zheng* - xzheng@pku.edu.cn

* Corresponding author

Published: 18 June 2004

Received: 09 October 2003

BMC Public Health 2004, 4:23

Accepted: 18 June 2004

This article is available from: <http://www.biomedcentral.com/1471-2458/4/23>

© 2004 Wu et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Birth defects, which are the major cause of infant mortality and a leading cause of disability, refer to "Any anomaly, functional or structural, that presents in infancy or later in life and is caused by events preceding birth, whether inherited, or acquired (ICBDMS)". However, the risk factors associated with heredity and/or environment are very difficult to filter out accurately. This study selected an area with the highest ratio of neural-tube birth defect (NTBD) occurrences worldwide to identify the scale of environmental risk factors for birth defects using exploratory spatial data analysis methods.

Methods: By birth defect registers based on hospital records and investigation in villages, the number of birth defects cases within a four-year period was acquired and classified by organ system. The neural-tube birth defect ratio was calculated according to the number of births planned for each village in the study area, as the family planning policy is strictly adhered to in China. The Bayesian modeling method was used to estimate the ratio in order to remove the dependence of variance caused by different populations in each village. A recently developed statistical spatial method for detecting hotspots, Getis's G_i^* [7], was used to detect the high-risk regions for neural-tube birth defects in the study area.

Results: After the Bayesian modeling method was used to calculate the ratio of neural-tube birth defects occurrences, Getis's G_i^* statistics method was used in different distance scales. Two typical clustering phenomena were present in the study area. One was related to socioeconomic activities, and the other was related to soil type distributions.

Conclusion: The fact that there were two typical hotspot clustering phenomena provides evidence that the risk for neural-tube birth defect exists on two different scales (a socioeconomic scale at 6.84 km and a soil type scale at 22.8 km) for the area studied. Although our study has limited spatial exploratory data for the analysis of the neural-tube birth defect occurrence ratio and for finding clues to risk factors, this result provides effective clues for further physical, chemical and even more molecular laboratory testing according to these two spatial scales.

Background

Birth defects, formally 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 caused 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. However, they can be prevented and early intervention is important to ameliorate their consequences [1]. But this requires an accurate understanding of the causes and risk factors in advance.

According to results of birth defects research, the probability of birth defects caused by genetic factors may be similar in various regions. However, environmental risk factors, such as chemicals, toxins, and environmental pollution account for different ratios of birth defect occurrences in different regions. Those environmental risk factors, including socioeconomic status and geographical elements, often have spatial associations as well as various patterns.

As long as diseases have been recognized, it has been apparent that many of them are manifested in clusters. Interest in those clusters resides not so much in the mere aggregation of cases but rather in populations that have a high rate of disease. Experience in epidemiology should remind us that clustering can also be observed for variables that are not causes but serve as markers for the causes. The scientific reason to study disease clusters is to learn about clustering of the causes [2] (Rothman, 1990). Exploratory spatial data analysis methods, which aim at testing hypotheses of spatially distributed object analysis, can serve as a tool in identifying risk factors for birth defects.

Based on research, the ratio of birth defect occurrences is estimated to be about 40–50‰ in P. R. China. Shanxi province, a northern region in China, has the highest ratio of neural-tube birth defects in the world. In order to reach prepotency, we selected Heshun, one county of Shanxi, as an experimental region for study (Figure 1). This county lies in the Taihang Mountain region and forms a relatively closed area. Most of the people in this county are farmers and seldom change their living environment. Furthermore, there have been no large-scale movements of people in the history of this region. The inherited and congenital causes of birth defects are similar among the people in this region, and those causes explain only a small fraction of all neural-tube birth defects cases. Following the CDC's guidelines [3] for investigating clusters of health events, exploratory spatial data analysis methods were introduced to study the relationship between environmental risk factors and neural-tube birth defects.

Methods

Spatial position and expression

There were 322 villages and one town in the study area. Since the main object of this study was the relationship between environmental risk factors and neural-tube birth defects, the town was not included, as the environmental factors there are somewhat complex. And birth defects registers in the town were removed from the study as well. The locations of the 322 villages were determined by the Geographical Information Systems for spatial analysis. As there were no boundaries defined for the villages, we drew them for each village using a Voronoi chart (Figure 2).

Bayesian modeling of neural-tube birth defects data

As one of the areas with the highest ratio of neural-tube birth defects, inspection branches were well organized in this county. Records of birth defect cases for four years (1998~2001) were acquired based on hospital registers and investigation in villages. These cases were divided into neural-tube birth defects and other birth defects by organ system. Neural-tube birth defects include anencephaly, spina bifida, encephalocele, holoprosencephaly and hydrocephalus, among others. Different birth defects may be caused by different risk factors, and we limited our research to neural-tube birth defects.

However, there are seldom full records for normal births. For the total number of birth records, we used the number of births planned every year for each village. As the family planning policy has been strictly carried out, this number reflects the real births in maximum. Furthermore, because birth defects are low probability events, four years' neural-tube birth defects cases were added together and considered as one year's cases for the calculation of the occurrence ratio.

When the occurrence ratio was calculated, the number of births in each village was different because of population differences, which would cause a bias in the ratio acquired by simply dividing the number of birth defects by the number of all births. Generally, villages with smaller populations, which correspondingly have fewer births, will have larger variances for calculating the ratio of birth defects occurrence. Simply dividing the number of birth defects by the number of all births may cause an error in our spatial analysis. [4]

In order to remove the dependence of the sampling variance on population size and calculate the neural-tube birth defects ratio, Bayesian modeling methods were used through WinBUGS software. The observed number of cases in each village was treated as a binomial random variable with parameter P_i in our analysis. P_i is the probability of a live birth in village i having a birth defect. The standard rate (observed birth defects cases divided by all

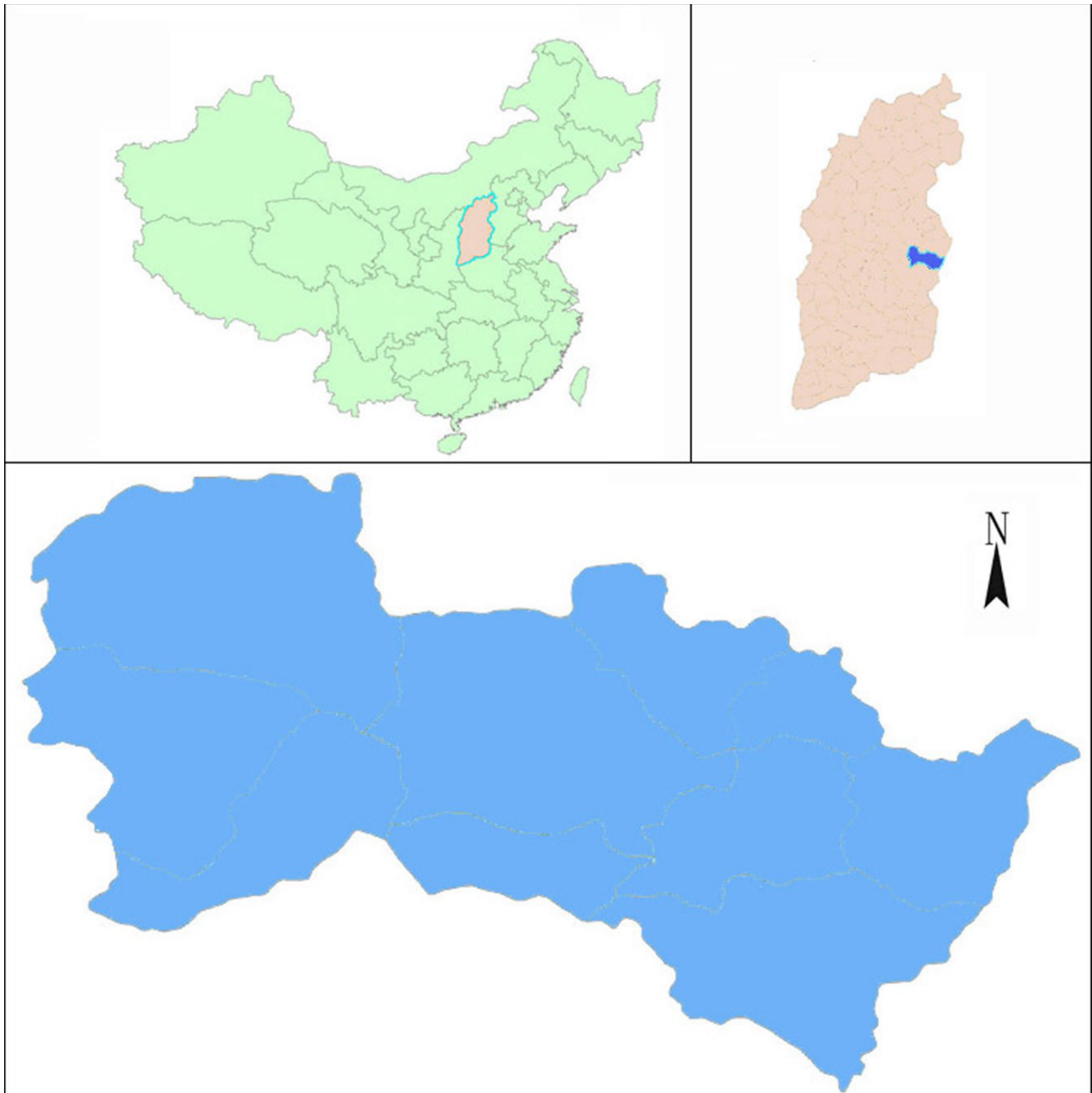
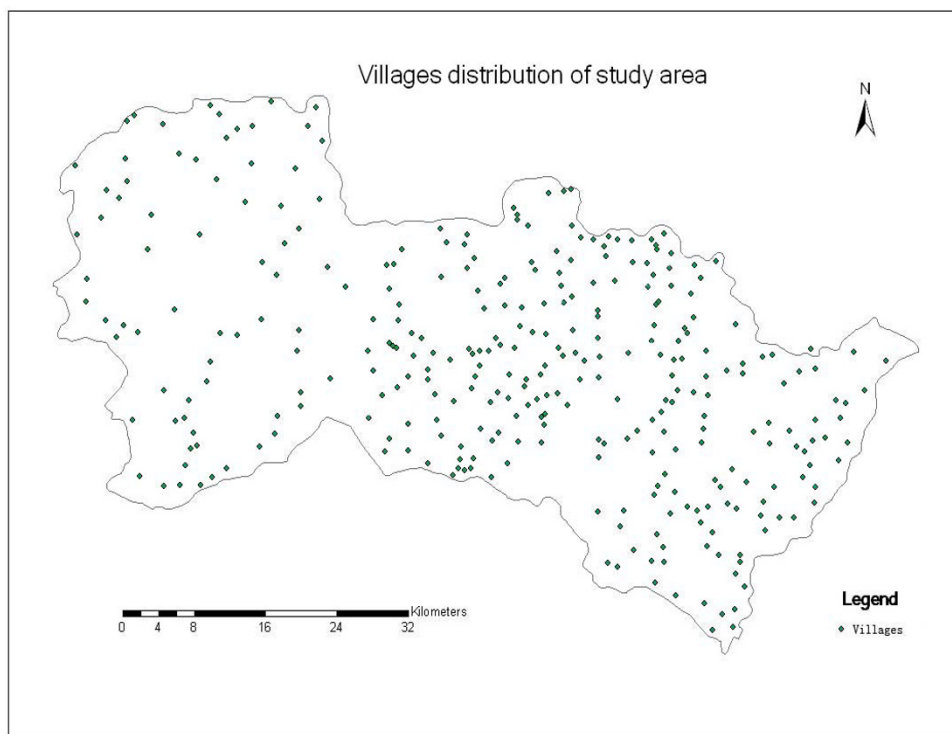


Figure 1
Location of the study area

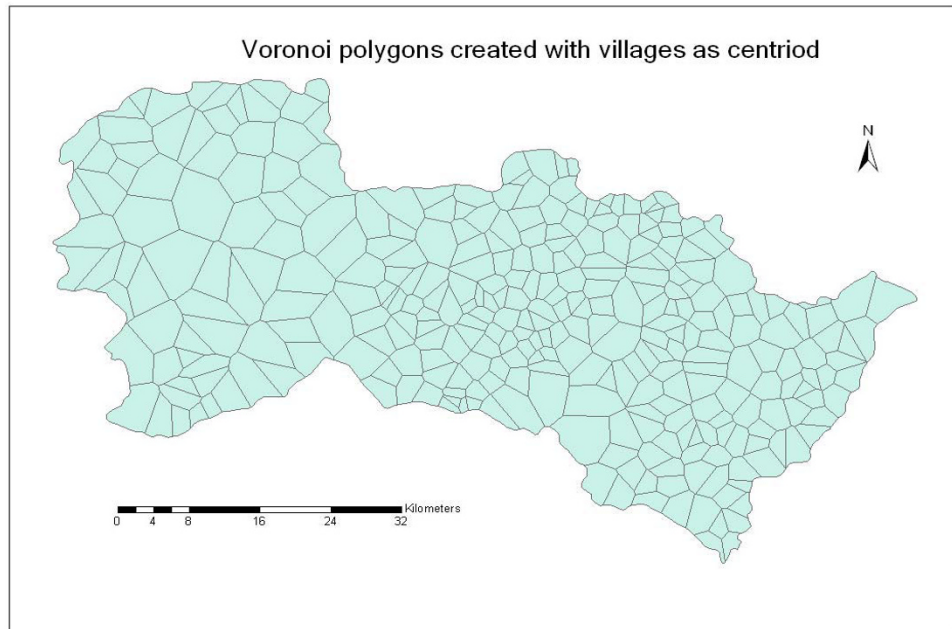
births) is the maximum likelihood estimate of P_i . As the environmental and socioeconomic status are similar, the P_i is assumed to be constant within the same village. And the parameter P_i is modeled through a logit transformation $\log it(P_i)$, expressed as:

$$\log it(P_i) = \log\left[\frac{P_i}{(1.0 - P_i)}\right] = \mu + v_i + \varepsilon_i$$

where μ (beta0 in WinBUGS) is the intercept term (mean), which we used for the neural-tube birth defect ratio, v_i is spatially structured auto-regression, and spatially unstructured ε_i is the random effects. The Bayesian



a



b

Figure 2
a) Villages in study region, b) Voronoi polygons

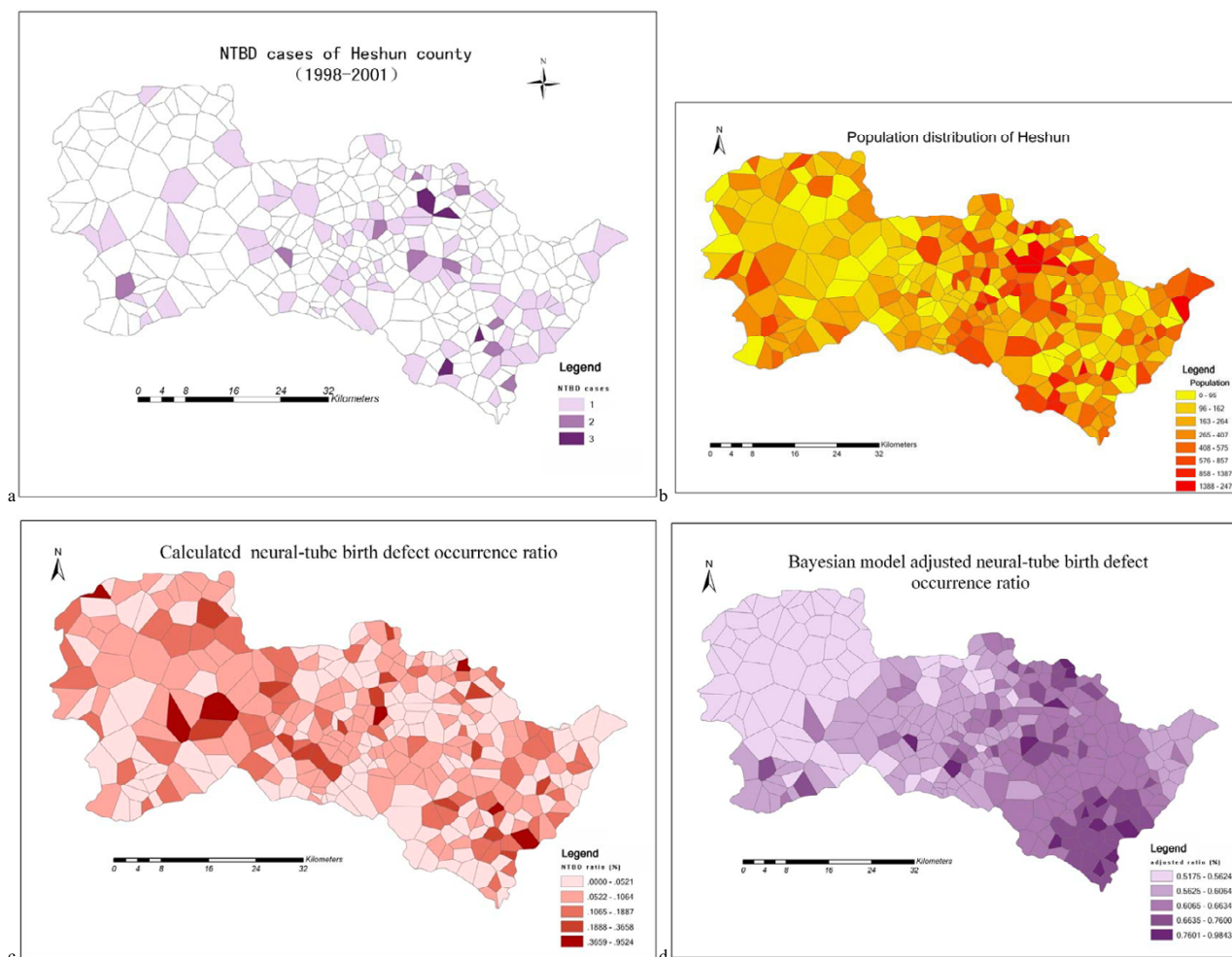


Figure 3
 a) Neural-tube birth defect cases, b) Population of Heshun, c) Calculated NTBD occurrence ratio, d) Bayesian adjusted NTBD occurrence ratio

adjusted rate was generated by WinBUGS under the binomial-logistic modeling assumption. (See WinBUGS code and sample results in Appendixes I & II) [5]. The μ (beta0) has a uniform prior and the v_i and ε_i with a gamma prior. The car.normal in WinBUGS was chosen to specify the intrinsic Gaussian CAR prior distribution of v_i . Figure 3c shows the population distribution and the ratios of neural-tube birth defects cases, which is the maximum likelihood estimate of P_i and figure 3d shows the adjusted ratios.

Spatial statistics methods

Getis's G_i^* statistics, developed by Getis and Ord (1992), is a multiplicative measurement of the overall spatial association of values which fall within a critical distance

of each other. It can be used as a method of detecting hotspots, and can be expressed as follows [6,7]:

$$G_i^* = \frac{\sum_j^n w_{ij}(d) \cdot \gamma_i - W_i^* \cdot \bar{\gamma}}{S \cdot \{[nS_{1i}^* - W_i^{*2}]/(n-1)\}^{1/2}}$$

Here, S is the standard variance of the birth defects occurrence ratio, When the distance from village j to i is within distance d , then $w_{ij}(d) = 1$; otherwise $w_{ij}(d) = 0$, and S_{1i}^* , W_i^* . The higher the value of G_i^* is, the greater the influence of village i is at a given distance d , indicating that village i is a hotspot of the region.

Table 1: Typical distance scales and their meanings

| Statistical items | Distance scale | Meanings |
|--|----------------|----------------------------------|
| Nearest distance among remote villages | 6.2 – 9.3 km | Socio-economic activities scopes |
| Differentiated distance of soil types | 19.5 – 30 km | Geological variance distance |

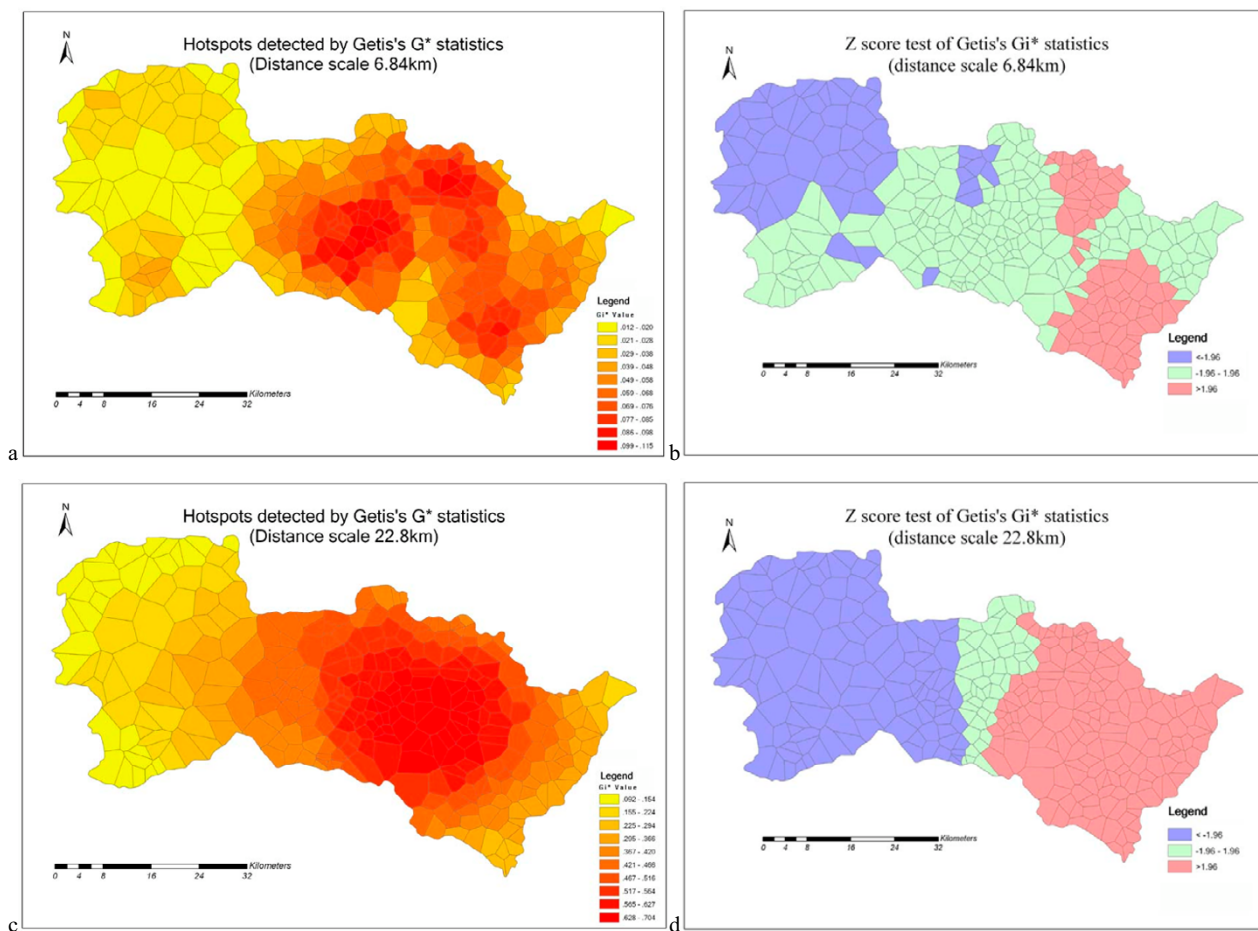


Figure 4
 Getis's G_i^* statistics a) Hotspots detected by Getis's G_i^* statistics at 6.84 km (Grouped distribution), b) Z scale test of G_i^* (at 6.84 km), c) Hotspots detected by Getis's G_i^* statistics at 22.8 km (Grouped distribution), d) Z scale test of G_i^* (at 22.8 km)

Results

Distance scale

As Getis's G_i^* statistics require a critical distance value as a threshold, typical distances in this region were calculated and described as shown in table 1.

Hotspots detect

Figure 4 shows two typical hotspot distributions of NTBD occurrence ratios calculated by Getis's G_i^* statistics.

Discussion

Birth defects are becoming a major cause of rising infant mortality. It has been found that birth defects account for a gradual increase in infant mortality from one fourth to

one third of all births in the 1990's. In order to prevent birth defects or enable early intervention, risk factors have been assessed in laboratories using many analytical methods [8]. However, as people live in different kinds of environments and have different socioeconomic statuses, no laboratory environment can fully simulate the conditions associated with risk factors. So the analytical results determined in laboratories can only explain a small fraction of the risk factors for birth defects.

Socioeconomic status is the most obvious potential obstacle in any spatial analysis of health outcomes. There has been little research on the strength of the relation between socioeconomic status and the risk of congenital anomaly (birth defects) (H. Dolk, 1998) [9]. Our work suggests that there is a typical grouped distribution of hotspots when distance scopes based on residents' common socioeconomic activities are taken as a critical distance value. We therefore think that socioeconomic status may affect the scope of risk factors. For example, the scope of inter-

marriage usually falls within the definition of social activities distance, and the male and female usually have similar socioeconomic status when they get married. This may indicate that they have been exposed to some common risk factors, which would accelerate the occurrence of birth defects.

Another possible obstacle to assessing risk factors for birth defects is the presence of toxins in the environment. Birth defects are heterogeneous in pathogenesis and aetiology. We selected neural-tube birth defects as the main object, and the study area has a zonal distribution of soil types [10] (Figure 5). According to general knowledge of local residents, areas with high occurrences of birth defects are usually diggings areas. So we took the soil variance distance scale as the critical distance value in Getis's G_i^* statistics, and found a similar zonal distribution of birth defects hotspots.

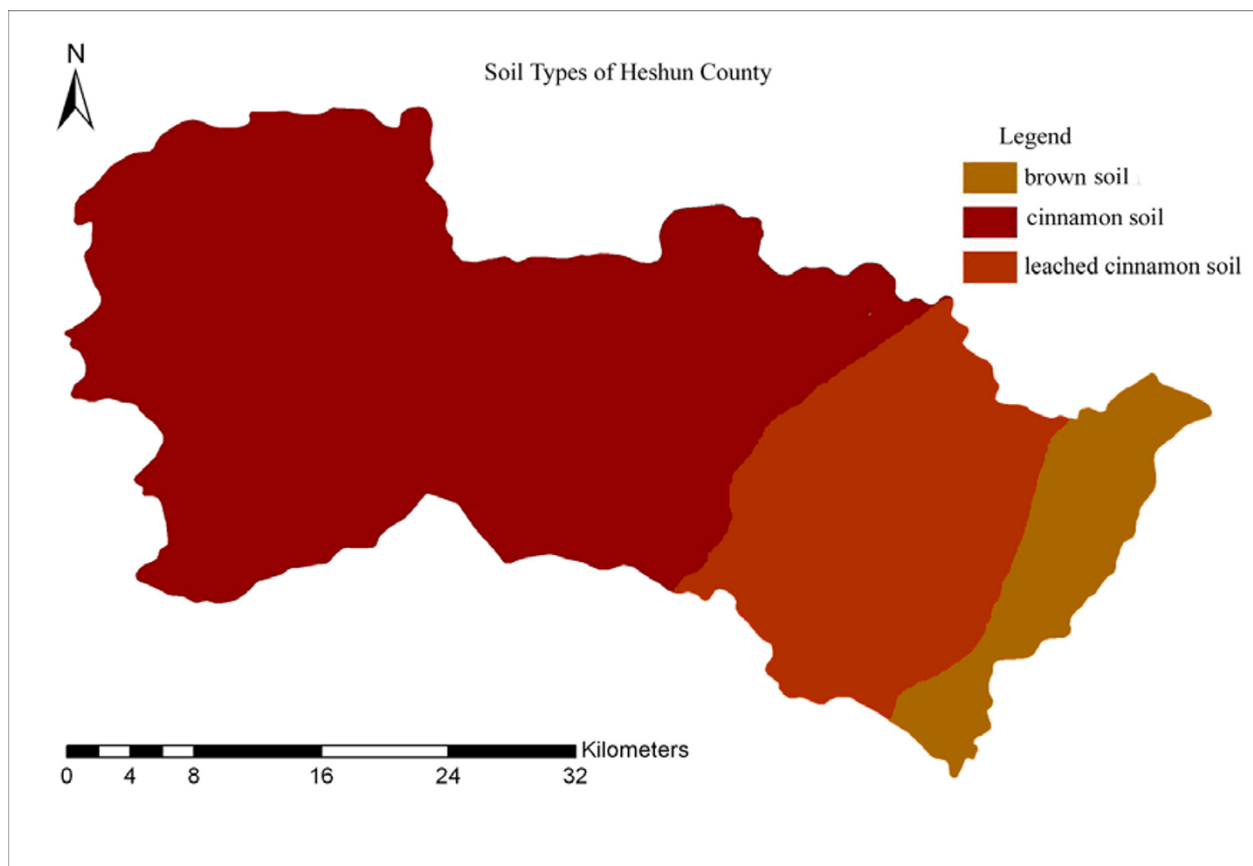


Figure 5
Soil types distribution (zonal type) Data source: Chinese Map Press. 1978

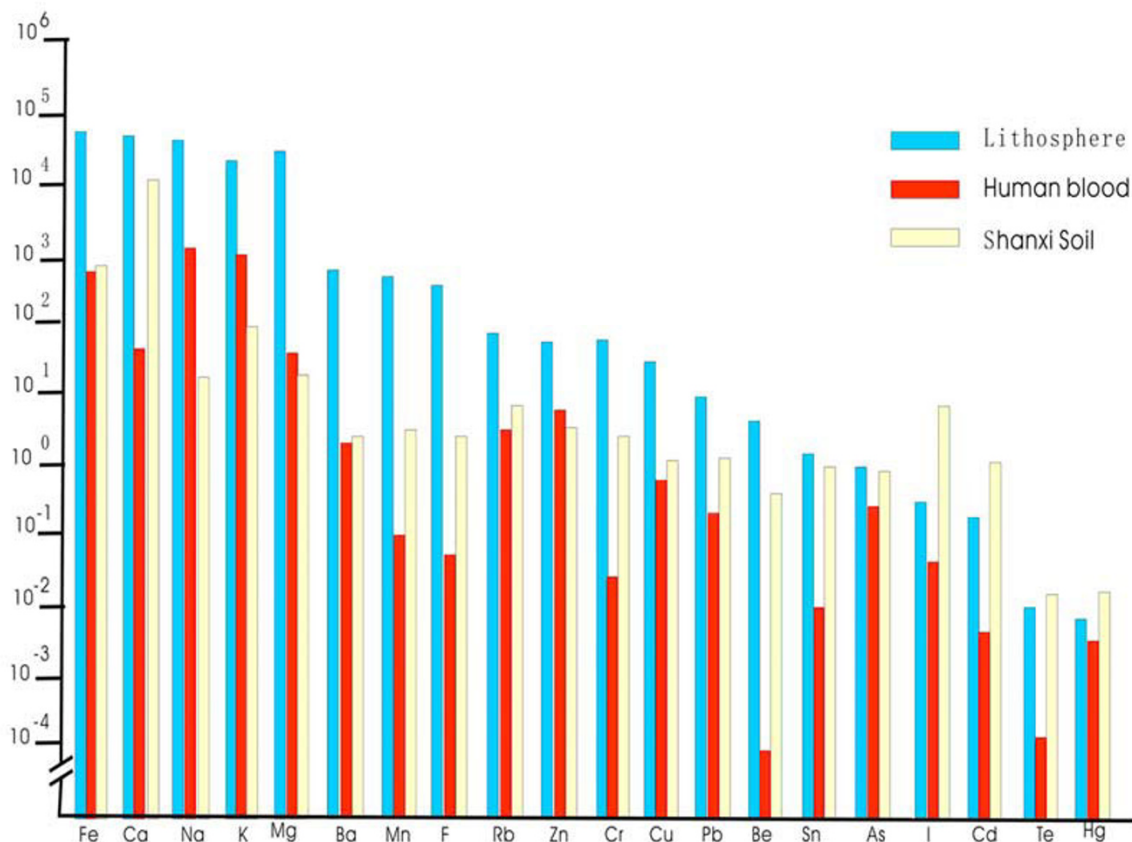


Figure 6
 Contents of twenty chemical elements in the worldwide lithosphere, human blood and soil samples from Shanxi province * Na, K, Ca, Mg, Fe were calculated for the mathematical average value for different contents of all soil types (%), and others were calculated for geometrical the average value for different soil type distribution areas (mg/kg). (Data source: Xiaolan AN, 1995; Chongwen SHI et al. 1996)

Table 2: Difference in chemical element contents of three main soil types in Heshun

| Soil types | High content | Average content | low content |
|-----------------------|----------------------|-------------------|-------------|
| leached cinnamon soil | Co Pb Cu Ni Cr Zn Mn | As F Hg | Cd |
| cinnamon soil | Pb Ni Cr Zn Mn Hg Cd | As Co F Cu | |
| brown soil | As F Hg Cd | Co Pb Ni Cr Zn Mn | Cu |

* High content elements refers to the contents 10% above average and low contents refers to contest 10% below the average in Shanxi province. (Data source: Chongwen SHI et al. 1996)

Table 3: Nutrient of the three soil types

| Soil types | pH | Organic (%) | Available Nitrogen(mg/kg) |
|-----------------------|------|-------------|---------------------------|
| leached cinnamon soil | 6.9 | 3.73 | 98.63 |
| cinnamon soil | 7.28 | 5.47 | 106.50 |
| brown soil | 6.62 | 3.00 | 75.00 |

(Data source: Lingrao MENG, 2002)

The phenomena of hotspot distribution at this distance scale could be used as a basis to inform further study, as soil variance can reflect the changes of lithology, and the minerals or rocks may have some specific chemicals or chemical mixtures. According to the dualism of the balance between the environment and human beings, the contents of a human being's blood are consistent with chemical elements in the lithosphere. If this balance is broken, diseases will break out. [8,11] The contents of the chemical elements of the lithosphere, human blood and soil samples from Shanxi province are shown below (Figure 6), in which we can see differences in the distribution of chemical elements. The elements in the three main soil types of Heshun County were tested and compared to the base values for the whole province (Table 2). The main difference in the three soil types is the metal content. Leached cinnamon soil and cinnamon soil have high contents of metals which may have an influence on human health, and brown soil has average contents of those metals. Analysis of the nutrient contents of these three soil types, [12] indicates that with high/low contents of organic matter and high/low pH values, these health-related metals have different statuses. In soil with high organic matter content and pH value, metals usually form complexes and thus cannot easily be absorbed by crops. This finally affects the diet of local residents. In soil with low organic matter content and pH value, metals are easily lost with soil erosion, Only small amounts of these metals are absorbed by plants and do harm and so there is less danger to local residents. Soil with average organic matter and pH value, such as leached cinnamon soil, may contain a higher risk factor for neural-tube birth defects (Table 3). However, these conclusions would need further analysis to determine environmental potential hazard. And direct measurement of soil chemicals and residents' exposure to them would help to assess the risk factors.

For the data used in this study, birth defects may not have been fully reported in this area as some pregnant women chose home births rather than hospital births. So we used data from hospital records and investigations in villages. Some women may have relocated their place of residence during their pregnancy, so there could be a migration bias in risk factor identification. However, as chemicals accu-

multate in the body over time and there have been no large scale movements of people in this region, we think there is little migration bias for the selected study area. [13] As birth defects are low-probability events and the family planning policy was carried out strictly, we added four years of birth defects cases together to calculate the occurrence ratio. Though this may magnify the occurrence ratio, little bias would be introduced for long-existing environmental risk factors for birth defects in the spatial dimension.

Conclusions

This study analyzed the spatial distribution of neural-tube birth defects by exploratory spatial data analysis methods. The results showed that two typical clustering phenomena were present for two scales. These two critical distances in

Getis's G_i^* statistics have some relationship to socio-economic and geological environments. These give effective clues to detect the environmental risk factors for neural-tube birth defects. This exploratory spatial data analysis has proven the efficiency of using the proposed method to seek clues to environmental factors that increase the risk of birth defects. The next step of this study should be to analyze geological samples by chemical testing in the laboratory to identify environmental risk factors. Following this, animal models for exposure risk should be established and re-tested. Meanwhile, socio-economic activities such as the scope of inter-marriage should be investigated as well.

Authors' contributions

This study was conceived and completed by Jilei Wu. Jinfeng Wang and Xiaoying Zheng supervised this study, and Bin Meng assisted with the WinBUGS program to calculate the neural-tube birth defects occurrence ratio. Gong Chen, Lihua Pang and Keli Zhang assisted with the study and analyses of socio-economic factors and soil analysis and Xinming Song and Ting Zhang assisted with the medical analysis of birth defects.

Acknowledgements

This study was supported by grants 2001CB5103 from the National "973" Program, 30025042, JJ03000101 and 49871064 from the National Nature Science Foundation of China, 2002AA135230 from the National High Tech-

nology Research and Development Program and the Science Innovation Project of the Chinese Academy of Sciences and the "211" program of Peking University.

References

1. International Birth Defects Information Systems: [<http://ibis-birthdefects.org/>].
2. Rothman KJ: **A sobering start for the cluster busters' conference.** *Am J Epidemiol* 1990, **132**:S6-S13.
3. **MMWR of CDC, 1990/39(RR-11)** :1-16 [<http://www.cdc.gov/mmwr/preview/mmwrhtml/00001797.htm>].
4. Cressie N, Read TRC: **Spatial Data Analysis of Regional Counts.** *Biom J* 1989, **31(6)**:699-719.
5. Haining R: *Spatial Data Analysis, theory and practice* Cambridge University Press; 2003.
6. Getis A, Ord JK: **The Analysis of Spatial Association by Use of Distance Statistics.** *Geographical Analysis* 1992, **24**:189-206.
7. Aldstadt J, Getis A: **Point Pattern Analysis in an ArcGIS Environment.** [<http://www.dpi.inpe.br/gilberto/csiss/papers/aldstadt.pdf>].
8. An XL, Fu SHL: *Environmental Eugenics* Beijing Medical University and Chinese Union Medical University Union Press; 1995.
9. Dolk H, Vrijheid M, Armstrong B, Abramsky L, Bianchi F, Garne E, Nelen V, Robert E, Scott JES, Stone D, Tenconi R: **Risk of congenital anomalies near hazardous-waste landfill sites in Europe: the EUROHAZCON study.** *Lancet* 1998, **352**:423-427.
10. Institute of Soil Research, CAS: *Soil types distribution of China* Chinese Map Press; 1978.
11. Sh C, Zhao L, Guo X, Gao S, Yang J, Li J: **The distribution of chemical elements base values and influenced factors in Shanxi province of China.** *Agro-Environmental Protection* 1996, **15(1)**:24-28.
12. Meng L: **To the comparisons of the loss for natural soil's nitrogen and phosphorus.** *Geography and Territorial Research* 2002, **18(4)**:98-103.
13. Armstrong B: **Study design for exposure assessment in epidemiological studies.** *the science of the total Environment* 1995, **168**:187-194.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2458/4/23/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

