Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Bayesian mapping of neural tube defects prevalence in Heshun County, Shanxi Province, China during 1998~2001*

CHI Wen-xue†1, WANG Jin-feng†‡2, LI Xin-hu[2], ZHENG Xiao-ying[3], LIAO Yi-lan[2]

([1]China University of Geosciences, Beijing 100083, China)

([2]Institute of Geographical Sciences and Nature Resources Research, Chinese Academy of Sciences, Beijing 100101, China)

([3]Institute of Population Research, Peking University, Beijing 100871, China)

†E-mail: chiwx@126.com; wangjf@lreis.ac.cn

Received Aug. 10, 2006; revision accepted Nov. 14, 2006

Abstract: Objective: To estimate the prevalence rates of neural tube defects (NTDs) in Heshun County, Shanxi Province, China by Bayesian smoothing technique. Methods: A total of 80 infants in the study area who were diagnosed with NTDs were analyzed. Two mapping techniques were then used. Firstly, the GIS software ArcGIS was used to map the crude prevalence rates. Secondly, the data were smoothed by the method of empirical Bayes estimation. Results: The classical statistical approach produced an extremely dishomogeneous map, while the Bayesian map was much smoother and more interpretable. The maps produced by the Bayesian technique indicate the tendency of villages in the southeastern region to produce higher prevalence or risk values. Conclusions: The Bayesian smoothing technique addresses the issue of heterogeneity in the population at risk and it is therefore recommended for use in explorative mapping of birth defects. This approach provides procedures to identify spatial health risk levels and assists in generating hypothesis that will be investigated in further detail.

Key words: Birth defects, Neural tube defects (NTDs), Disease map, Spatial analysis, Bayesian smoothing, China
doi:10.1631/jzus.2007.A0921 **Document code:** A **CLC number:** TP317; R18

INTRODUCTION

As with the analysis of any set of data, it is always good practice to begin by producing and inspecting graphs. A feel for the data can then be obtained and any outstanding features be identified. In spatial epidemiology this is called disease mapping (Berke, 2004). Disease mapping is a method used by epidemiologists, medical demographers and biostatisticians to understand the geographical distribution of a disease (Meza, 2003). Public health managers often want to map the locations of the occurrences of a disease and to study its spatial distribution in the area. Proper visualization of the spatial pattern and distribution of the disease can guide such managers in finding the priorities for allocating budget, personnel, equipment, etc. Mapping of prevalence rates over the different geographical regions helps to have an idea of environmental determinants of a specific disease (Maiti, 1998).

An abnormal neural tube closure occurs between the third and fourth weeks of gestational age that results in structural defects which occur anywhere along the neuroaxis from the developing brain to the sacrum and often results in the exposure of neural tissue (Frey and Hauser, 2003). Neutral tube defects (NTDs) are a leading cause of death for very young children and long-lasting impacts on many children. The specific causes of most NTDs are not known. Shanxi Province, a northern region in China, has the highest ratio of NTDs in the world (Wu *et al*., 2004). The understanding of spatial distribution of NTDs in the rural area of China plays a key role in successful

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 40471111 and 70571076), and the Ministry of Science and Technology of China (No. 2001CB5103)

application of prevention and medical treatment strategies. But the geographical analysis of a disease risk is particularly difficult when the disease is not frequent and the area units are small. The usual procedure is to map the prevalence rate across different geographical regions. However, the mapping of raw prevalence rates has been found to be inappropriate since it does not account for the spatial heterogeneity of the population at risk. Bayesian mapping could be the better key to model the geographical distribution of a disease. Previous studies have already used a Bayesian approach for analysis of childhood mortality in rural Burkina Faso (Sankoh *et al*., 2002). Further studies on Bayesian applications can be found in (Jarup *et al*., 2002; Berke, 2004; Bergamaschi *et al*., 2006). Nowadays, geographic information system (GIS) is often used to visualize spatial patterns in the geographical distribution of a disease, usually for explorative and descriptive purposes, as well as for providing information for further studies. It can also be used to gain important clues on the aetiology of a disease. Therefore, the purpose of our study is to estimate the prevalence rates of NTDs by Bayesian analysis based on data obtained from Heshun County, Shanxi Province, China during 1998~2001.

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 average 1300 m altitude. The maximum distances from east to west and from north to south are about 75 km and 35 km respectively in this county. The area has a total population of 134522 and comprises 2250 square kilometers (Fig.1). Records of birth defect cases for four years (1998~2001) were acquired based on hospital registers and investigation in villages. Infants are routinely examined by a health care provider at birth and hospital discharge, during the first week of life, at the end of the second week of life and again at ≈42 days of age. Infants born at home (<10% of total births) are examined once or twice during the first week of life, and then followed using the same schedule as infants delivered in hospitals. In addtion, this investigation was carried out in a systematic manner so that surveillance methods were similar throughout the villages. These cases were divided into NTDs and other birth defects by organ system. There are 322 villages and one town in the study area. The total of NTDs in the 322-villige area was 80 cases during 1998~2001 (Table 1). Because the main objective of this study was to map prevalence rate of NTDs in these rural villages, the town was not included. So birth defects registers in the town were removed from the study as well. 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 ArcGIS9.0 GIS software. As there were no boundaries defined for the villages, we drew them for each village using a Voronoi chart.

Table 1 Cases of birth defects in Heshun County, Shanxi Province, China during 1998~2001

	1998	1999	2000	2001	Total
NTDs	19	14	14	33	80
Other types of birth defects	18	12	17	8	55
Total	37	26	31	41	135

Classical statistical method

In this study, the crude prevalence rates were calculated for each village by simply dividing the number of NTDs by the resident population. Within a map of N spatial units, let n_i denote the population and y_i the number of corresponding diagnosed NTD results in the ith village. The observed prevalence for each spatial unit is given by the ratio $p_i=y_i/n_i$ (Staubach *et al*., 2002). A major advantage to mapping crude rates is that it assists the reader in identifying where relatively large numbers of negative health events are occurring. However, note that the prevalence rate is often zero due to the no-NTD cases observed in the respective village (y_i=0). Fig.2 shows the observed raw prevalence rates p_i per 10000 resident populations.

Empirical Bayesian smoothing

Mapping the raw prevalence rates of NTDs can lead to spurious spatial features. Villages of small populations can appear highly variable and these apparently variable regions contain a disproportionate number of high (or low) parameter estimates. These problems can be overcome by building the geographical distribution map with a Bayesian method

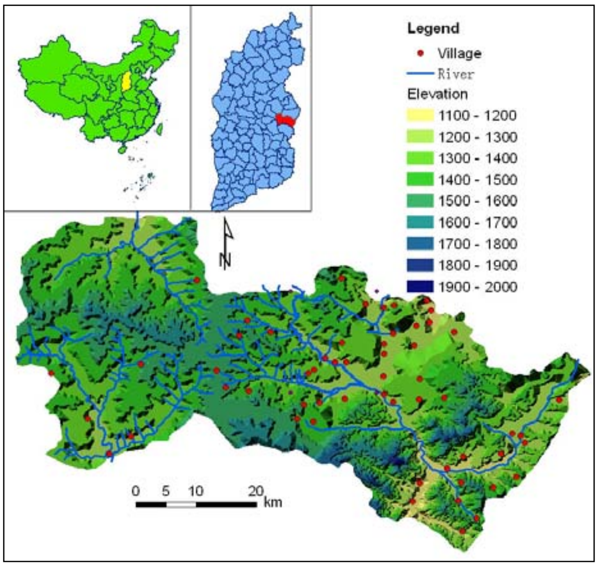

Fig.1 Location of NTD cases in Heshun County, Shanxi Province, China during 1998~2001

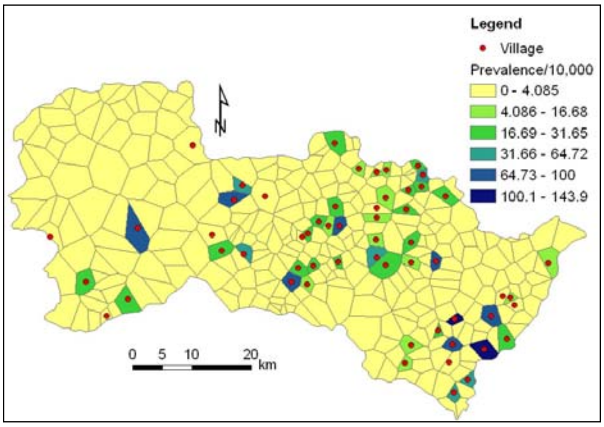

Fig.2 Map of the crude neural tube defects prevalence rates

that can model the random and the true variations separately (Berke, 2005). When the random noise from the final map is filtered out, the map is finally interpretable as a reliable estimate of reality. The resulting smoothed regional estimates have a variance-stabilizing side effect by borrowing strength from (local or global) neighborhood information. The outline of the empirical Bayes approach for smoothing regional rates of rare diseases is based on the Poisson model (Berke, 2004). Assume the cases in every region i are independently Poisson distributed with the unknown parameter θ_i which has an unknown prior distribution associated with expectation $E(\theta_i)=\pi$ and variance $Var(\theta_i)=\varphi^2$. Then the totals of cases from the ith region, e.g. m_i (i=1,…,N) are distributed as follows: $m_i|\theta_i,n_i$~Possion(θ_i,n_i), θ_i~Normal(π, φ^2), where n_i denotes the population in the ith village. The method of moment estimator (MME) of the unknown hyperparameters are $\hat{\pi}=\sum m_i/\sum n_i$ for the prior mean and $\hat{\varphi}^2=\sum n_i(p_i-\hat{\pi})^2/\sum n_i-\hat{\pi}/\overline{n}$ for the prior variance, where $\overline{n}$ denotes the mean regional 'at-risk' population and the summation is over the range of i. The empirical Bayesian estimates then becomes $\hat{\theta}_i=\rho_i p_i+(1-\rho_i)\hat{\pi}$ with shrinkage weights $\rho_i=\hat{\varphi}^2/(\hat{\varphi}^2+\hat{\pi}/n_i)$. It should be noted that Bayes estimates, i.e. $\hat{\theta}_i$, generally have a smaller associated variance than the corresponding frequentist estimates, i.e. p_i. The practical use of the Bayesian modeling, instead of the classical statistical one, is applied to study the geographical variation of NTDs across the county of Heshun. Bayesian modeling methods were used through WinBUGS software, and the results were input into the ArcGIS9.0 for mapping.

RESULTS

We built the maps of the geographical variations of NTD prevalence rates across the 322 villages of Heshun County both with a classical statistical approach (Fig.2) and a Bayesian approach (Fig.3).

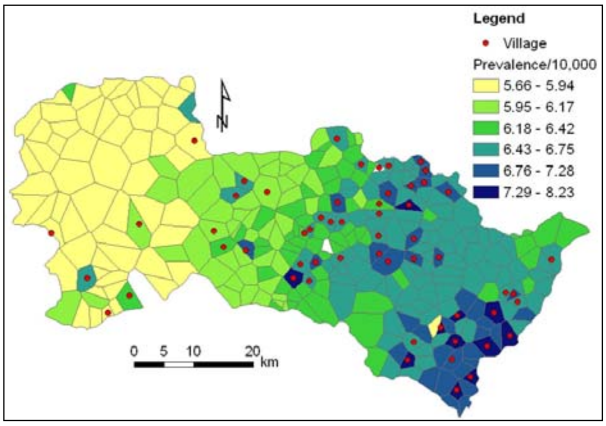

Fig.3 Map of the Bayesian estimates of neural tube defects prevalence rates

The maps had to be interpreted by considering that different shades of color were proportional to the prevalence value. So the darker the area was, the higher was the prevalence of disease. The classical statistical approach produced an extremely dishomogeneous map, strongly influenced by random variability. This map was dominated by extreme prevalence values found in the areas with very small population. On the contrary, the Bayesian map (Fig.3) was much smoother and more interpretable. The smoothed prevalence rates under the Poisson model for rare phenomena are more appropriate for disease mapping than the raw rates in Fig.2. The Bayesian area-specific prevalence ranged from 5.66 to 8.23 per 10000 inhabitants. Visual map perception reveals some potential high-risk regions in the southeastern part of the study area.

DISCUSSION

Maps of disease occurrence have been used for at least a century to identify potential spatial patterns in health outcomes and generate hypotheses on etiologic risk factors (Wall and Devine, 2000). Advances in GIS and mapping technologies now provide opportunities for public health professionals to rapidly analyze spatial relationships and disease risk factors to facilitate policy planning and implementation. When we practice analysis and surveillance of diseases, many and various maps of disease prevalence are used (Lawson *et al*., 1999).

Good maps of birth defects risk have long been recognized as an important tool for birth defects control. In particular, the analysis of the geographical distribution of NTDs is appealing, as it can give clues to an etiological explanation and can test specific hypotheses on suspected ecological risk factors. Meanwhile, more hypotheses could be generated through mapping process. Unfortunately, this kind of research is particularly difficult to carry out employing a classical statistical approach, especially when the aim is to study an uncommon disease, such as NTDs, and when the disease is distributed in small areas with relatively small population (Bergamaschi *et al*., 2006). Here, the term 'small-area' is used to describe an area with small 'at-risk' population, but not necessarily small in geographical size or scale. Crude rates for areas with small 'at-risk' population are subject to high chance variation and do not take into account underlying population or exposure characteristics. A drawback to mapping crude rates is that because the village populations differ by age or other factors, the rates may not be very useful for comparing underlying health risks among areas. Under these conditions, the real effects in which we are interested can be masked by random noise and conversely random noise can be mistaken for real effects. Therefore, different geographical distributions that are possibly due to chance (mainly in those areas where the values are extreme and the population is small) can be incorrectly interpreted as true variations of epidemiological interest (Bergamaschi *et al*., 2006). As a consequence, the map was not interpretable from an epidemiological point of view.

On the contrary, the Bayesian approach overcame these problems with spatial smoothing of the rates, both allowing us to estimate area-specific prevalence rates and filtering out the random variations from the estimated prevalence due to the small number of cases in each village (the NTDs are low probability events, as many as four years' NTD cases were added in estimating the prevalence of the disease for each village in the study area), thus showing the true underlying variations of NTD prevalence. Bayesian spatial analysis and smoothing approach enables data sharing (i.e. risk smoothing) over space, which often results in more reliable risk prediction (MacNab, 2004). In addition, the Bayesian method was chosen because it easily allowed for the inclusion of a spatially structured variation (Mather *et al*., 2006). Therefore, the Bayesian map we obtained was more homogeneous than the previous one and more likely to represent the true distribution of the disease, even a Bayesian approach reserves potential drawbacks. Bayesian disease mapping models are essentially conservative, with high specificity even when data are very sparse, but conversely they have a low sensitivity, especially if the raised risk-areas have only a moderate excess or are not based on substantial expected counts (Bergamaschi *et al*., 2006).

Anyway, map smoothing provides reliable NTD prevalence rates by aggregating data over contiguous geographic units. In this process, true differences (clusters and outliers) in small areas may be removed (Goujard, 1999; Antó and Cullinan, 2001). In this study, it is seen that the Bayesian method is helpful for visual identification of clustering in the southeastern part of the study region.

Contributions to the area over the last four years have been instrumental in helping to pinpoint potential causes of NTD and to provide a strategy for effective allocation of health resource. The Bayesian mapping could be useful in research studies of the relation of prevalence to explanatory variables. These findings are consistent with several hypotheses for the etiology of NTDs in the study area, including near coal mine pathogens or toxins either directly inducing disease or indirectly leading to disease by drinking the polluted water at local wells (Zheng *et al*., 1999; Shen, *et al*., 2001; Verkasalo *et al*., 2004; Finkelman, 2004; Dolk, 2004; Sun, 2004). Ongoing investigations into the management and cause of NTDs can use this information to develop management strategies and more efficiently target future investigation.

CONCLUSION

We have explored how a Bayesian approach can be used to obtain reliable maps of NTDs in Heshun County, Shanxi Province, China. Bayesian mapping approaches might well be applicable in small areas with small populations of geographical variation in epidemiological measurements where conventional statistical methods cannot adequately allow for the mapping of the data. In conclusion, our study underlines the usefulness of Bayesian methods to obtain reliable maps of NTD prevalence and to identify possible clusters of NTDs for further carrying out of epidemiological investigations.

References

Antó, J.M., Cullinan, P., 2001. Clusters, classification and epidemiology of interstitial lung diseases concepts, methods and critical reflections. *Eur. Resp. J.*, **18**(Suppl. 32):101-106.

Bergamaschi, R., Montomoli, C., Candeloro, E., Monti, M., Cioccale, R., Bernardinelli, L., Fratino, P., Cosi, V., 2006. Bayesian mapping of multiple sclerosis prevalence in the province of Pavia, northern Italy. *J. Neurol. Sci.*, **244**: 127-131. [doi:10.1016/j.jns.2006.01.013]

Berke, O., 2004. Exploratory disease mapping: kriging the spatial risk function from regional count data. *Int. J. Health Geogr.*, **3**:18. [doi:10.1186/1476-072X-3-18]

Berke, O., 2005. Exploratory spatial relative risk mapping. *Prevent. Veter. Med.*, **71**:173-182. [doi:10.1016/j.prevetmed.2005.07.003]

Dolk, H., 2004. Epidemiologic approaches to identifying environmental causes of birth defects. *Am. J. Med. Genet.*, **125C**:4-11. [doi:10.1002/ajmg.c.30000]

Finkelman, R.B., 2004. Potential health impacts of burning coal beds and waste banks. *Int. J. Coal Geol.*, **59**:19-24. [doi:10.1016/j.coal.2003.11.002]

Frey, L., Hauser, W.A., 2003. Epidemiology of neural tube defects. *Epilepsia*, **44**(Suppl. 3):4-13. [doi:10.1046/j.1528-1157.44.s3.2.x]

Goujard, J., 1999. Clusters of birth defects: emergency and management. A review of some publications. *Eur. J. Epidemiol.*, **15**:853-862. [doi:10.1023/A:1007530216916]

Jarup, L., Best, N., Toledano, M.B., Wakefield, J., Elliott, P., 2002. Geographical epidemiology of prostate cancer in Great Britain. *Int. J. Cancer*, **97**:695-699. [doi:10.1002/ijc.10113]

Lawson, A.B., Böhning, D., Biggeri, A., Lesaffre, E., Viel, J.F., John, W., Sons, L., 1999. Disease Mapping and its Uses. *In*: Lawson, A.B. (Ed.), Disease Mapping and Risk Assessment for Public Health. John Wiley & Sons, New York, p.3-13.

MacNab, Y.C., 2004. Bayesian spatial and ecological models for small-area accident and injury analysis. *Accident Anal. & Prevent.*, **36**:1019-1028. [doi:10.1016/j.aap.2002.05.001]

Maiti, T., 1998. Hierarchical Bayes estimation of mortality rates for disease mapping. *J. Statist. Planning and Inference*, **69**:339-348. [doi:10.1016/S0378-3758(97)00165-1]

Mather, F.J., Chen, V.W., Morgan, L.H., Correa, C.N., Shaffer, J.G., Srivastav, S.K., Rice, J.C., George, B., Swalm, C.M., Wu, X.C., *et al*., 2006. Hierarchical modeling and other spatial analyses in prostate cancer incidence data. *Am. J. Prevent. Med.*, **30**(2S):88-100. [doi:10.1016/j.amepre.2005.09.012]

Meza, J.L., 2003. Empirical bayes estimation smoothing of relative risks in disease mapping. *J. Statist. Planning and Inference*, **112**:43-62. [doi:10.1016/S0378-3758(02)00322-1]

Sankoh, O.A., Berke, O., Simboro, S., Becher, H., 2002. Bayesian and GIS Mapping of Childhood Mortality in Rural Burkina Faso. Control of Tropical Infectious Diseases. Uni-Heidelberg Discussion Paper.

Shen, X.M., Wu, S.H., Yan, C.H., 2001. Impacts of low-level lead exposure on development of children recent studies in China. *Clin. Chim. Acta*, **313**:217-220. [doi:10.1016/S0009-8981(01)00675-1]

Staubach, C., Schmid, V., Knorr-Held, L., Ziller, M., 2002. A Bayesian model for spatial wildlife disease prevalence data. *Prevent. Veter. Med.*, **56**:75-87. [doi:10.1016/S0167-5877(02)00125-3]

Sun, G.F., 2004. Arsenic contamination and arsenicosis in China. *Toxicol. Appl. Pharm.*, **198**:268-271. [doi:10.1016/j.taap.2003.10.017]

Verkasalo, P.K., Kokki, E., Pukkala, E., Vartiainen, T., Kiviranta, H., Penttinen, A., Pekkanen, J., 2004. Cancer risk near a polluted river in Finland. *Environ. Health Perspect.*, **112**(9):1026-1031.

Wall, P.A., Devine, O.J., 2000. Interactive analysis of the spatial distribution of disease using a geographic information system. *J. Geogr. Syst.*, **2**:243-256. [doi:10.1007/PL00011457]

Wu, J.L., Wang, J.F., Meng, B., Chen, G., Pang, L.H., Song, X.M., Zhang, K.L., Zhang, T., Zheng, X.Y., 2004. Exploratory spatial data analysis for the identification of risk factors to birth defects. *BMC Public Health*, **4**:23. [doi:10.1186/1471-2458-4-23]

Zheng, B.S., Ding, Z.H., Huang, R.G., Zhu, J.M., Yu, X.Y., Wang, A.M., Zhou, D.X., Mao, D.J., Su, H.C., 1999. Issues of health and disease relating to coal use in southwestern China. *Int. J. Coal Geol.*, **40**:119-132. [doi:10.1016/S0166-5162(98)00064-0]