The application of spatial empirical Bayesian smoothing method in spatial analysis of bacillary dysentery: A case study in Yudu County, Jiangxi Province

Yuwei Wang 1A,∗, Wang Gao1,b

1School of Mathematics and Computer Science, Jianghan University, Wuhan 430056, China

aweberwang@jhun.edu.cn, bgao@jhun.edu.cn

*Corresponding author

Abstract. Bacillary dysentery (BD) has become one of the major public health threats to the sustainable development of human beings. The primary objective of this paper is to validate the effectiveness of applying spatial empirical Bayesian smoothing (SEBS) method in spatial analysis of BD in Yudu County, Jiangxi Province. By using SEBS method in calculating BD incidence, we compared the raw incidence data and smoothed incidence data. Spatial distribution map and global autocorrelation analysis were adopted to explore the effect of SEBS method in BD analysis. The result showed that SEBS method provided a stable incidence estimate in epidemic research. The proposed method could be used as an effective tool for studying the spatial distribution of BD, especially in town level study.

1. Introduction

Bacillary dysentery (BD) is a kind of intestinal infectious diseases caused by the bacteria of the genus Shigella, which is listed as class-B notifiable infectious disease in China[1,2]. The incidence of BD is listed in the top 5 place of class-A and class-B notifiable infectious diseases in China, affecting the health and life quality of people seriously[3]. At the same time, millions of people are infected with BD every year worldwide[4]. From a global perspective, BD is a crucial public health problem in the world, hindering the sustainable development of human in 21st century[5,6].

With the assistance of geographic information system and spatial statistic technology, the spatial distribution characteristics of disease can be mapped, so as to monitor and analyse its spatial distribution pattern of incidence[7,8]. Nevertheless, when the population size or the infected cases of the study area is small, especially in town scale study, the raw incidence of the disease is calculated directly by infected cases and population, the results tend to be fluctuant, making it difficult to identify real regional differences. Spatial empirical Bayesian smoothing (SEBS) is a method for adjusting or smoothing spatial variables, which can eliminate the influence of unreliable information such as small sample or extreme value, preferably obtaining a stable incidence estimate in epidemic research[9–11].

Based on SEBS method, this paper analysed the spatial distribution pattern of BD in Yudu County of Jiangxi Province in 2017. We calculated the raw incidence and SEBS-based incidence of BD, compared the spatial distribution and autocorrelation of the two situations, intended to prove that SEBS method was more effective for spatial analysis of BD incidence, especially in small spatial scale.
2. Study area, data and methodology

2.1. Study Area
Yudu County is located in the south of Jiangxi Province, extending between 25°35′8″ to 26°20′53″ North and 115°11′ to 115°49′ East. The whole county covers an area of 2893 km², including 23 towns and villages (Figure 1). Yudu County belongs to subtropical humid monsoon climate zone, with warm and humid climate, four distinct seasons, sufficient light and rainfall[12].

![Administrative division map of Yudu at town level.](image)

2.2. Data
The data of BD cases in 2017 used in this paper were retrieved from the Center for Disease Control and Prevention of Yudu County. The time information of the original data were collected daily, the location information were accurate to the township street, and other information include gender, age, and occupation. Demographic data originated from the 2017 China county statistical yearbook. The data were analysed at town level, the incidence of BD was calculated as the number of new cases per 100,000 population in one year.

2.3. Methodology
Spatial empirical Bayesian smoothing method defines a neighborhood for each spatial area, the spatial smoothing is conducted based on neighborhood scale, neighborhood incidence, the mean and variance of neighborhood area. As a result, the incidence of the disease can be recalculated by SEBS method. The equation of SEBS approach is defined as follows[13]:

\[
\hat{r}(u_i) = \lambda(u_i)z(u_i) + [1 - \lambda(u_i)]m^*(u_i)
\]  

(1)

where \(u_i\) represents study area \(i\), \(\lambda(u_i)\) is the Bayesian shrinkage factor, \(z(u_i)\) is the raw incidence, \(m^*(u_i)\) is population-weighted average incidence in the neighborhood. The smoothed incidence data can...
reduce the spatial differences of incidence caused by small population base, and ensure that the results are more in line with the first law of geography. The raw and smoothed incidence data were displayed through two incidence maps at town level firstly, in order to visualize the spatial distribution of BD incidence for comparison.

Then spatial autocorrelation analysis was integrated to explore the spatial association of the BD incidence. In this study, the global Moran’s I index was adopted to make a comparison between raw incidence data and smoothed incidence data, in order to explore the difference of the two datasets and to test the effect of SEBS method. The range of Global Moran’s I index is between -1 and 1. A positive value of Moran’s I index indicates a positive spatial association, that is, nearby areas had similar incidence. Conversely, the negative value represents a concentration of different incidence[14,15]. Particularly, approaching 0 means a random distribution of the BD incidence.

3. Result

3.1. Incidence Maps

178 cases of BD were reported in Yudu County during 2017. Among the 23 towns of Yudu County, besides Xiaoxi Township, Tieshanlong, Duanwu Township, Qiaotou Township and Ge’ao, all other towns reported infected cases in 2017.

Combined with the population base of the 23 townships in Yudu, the annual incidence of each township in Yudu County was calculated. We conducted a comparison between the raw incidence and SEBS-based incidence of the spatial distribution of BD, the result was shown in Figure 2.

Figure 2. Spatial distribution map of raw incidence (a) and SEBS-based incidence (b).

Figure 2(a) and (b) demonstrated the spatial distribution of raw incidence and SEBS-based incidence in 2017 respectively. The high incidences were concentrated in Pangushan, Jingshi Township and Licun Township in two maps. It is worthwhile to pay attention to these high incidence rate townships. Compared to the raw incidence map, the incidence of 23 townships in Figure 2(b) was smoother than Figure 2(a). In detail, the range of annual incidence in (a) was 0.00 to 138.63, compared with 0.74 to 133.81 for (b). This result suggested that the SEBS method eliminated the impact of
missing cases on the incidence rate. Meanwhile, for the towns with small population size or infected cases, SEBS method smoothed the incidence to make it closer to local average value.

3.2. Spatial autocorrelation analysis

Global Moran’s I index was employed to estimate the spatial autocorrelation of BD incidence in Yudu. The spatial autocorrelation analysis result for BD incidence—raw incidence and SEBS-based incidence—is listed in Table 1.

| Incidence                  | Moran’s I | Z score | p-value |
|----------------------------|-----------|---------|---------|
| Raw incidence              | 0.25      | 1.75    | 0.07*   |
| SEBS-based incidence       | 0.39      | 2.72    | 0.02**  |

* Statistical significance at 10% level; ** Statistical significance at 5% level.

In general, the global Moran’s I for BD incidence in 2017 was positive, suggesting a clustering distribution of similar incidence rates at the town level in Yudu. The Moran’s I value of SEBS method was bigger, implying that the spatial correlation of smoothed incidence across Yudu County was stronger.

The major difference between two methods was the statistical significance level, the global Moran’s I value under SEBS method was significant at level 0.05, while the global Moran’s I value from raw data was significant at level 0.1. This result further confirmed the effectiveness of SEBS method from a statistical view.

4. Discussion

Analyzing the spatial characteristics of BD is of great theoretical and practical significance for monitoring, preventing and controlling the epidemic. In this research, spatial empirical Bayesian smoothing approach was used to analyse the spatial distribution pattern of BD in Yudu County, Jiangxi Province, China. According to the demographic data and BD cases, both the population and infected cases were unevenly distributed across the region. The small number of cases and population may result in unstable incidence. Furthermore, the number reported in each township may have missing cases, which increased the instability of incidence as well. Based on the above issues, SEBS was applied to recalculate the incidence of BD.

Here, we compared the spatial distribution map and global Moran’s I index of BD raw incidence and SEBS-based incidence. It was found that SEBS-based incidence map could reduce the effect of small population and cases, eliminating the impact of spatial outliers, making the incidence map smoother. This discovery was consistent with some previous studies[16–18].

Global Moran’s I index was chosen to estimate spatial autocorrelation because it measured the spatial dependence at global scale[19,20]. By defining the similarity between spatial weights and spatial attributes, the distribution pattern of BD incidence can be reflected. Based on the contrast of Moran’s I index, it can be concluded that the clustering trend of BD after smoothing is more significant, and the spatial dependence of SEBS-based incidence is more convincing under statistical significance test. Overall, the SEBS method should be taken into consideration in spatial epidemiology study, especially in region scale study.

Due to the limited data availability, this study targets the effect of SEBS method in Yudu County for one year. Nevertheless, more related data should be input to test the effectiveness of SEBS method. We call for future research on SEBS method for different spatial and temporal scales.

5. Conclusion
This paper aims to incorporate SEBS method in spatial analysis of BD at town level. Through the comparison of distribution map and spatial autocorrelation analysis, the SEBS method taken in the incidence calculation was more rational than raw incidence in town level study. The findings of this study suggest that the SEBS method is recommended when the infected cases or the population of the study area is small, especially in town scale research. In addition, the result can provide valuable reference for further epidemiological investigation and disease prevention.

References
[1] Hao Y, Liao W, Ma W, Zhang J, Zhang N, Zhong S, Wang Z, Yang L and Huang C 2019 Science of the Total Environment Effects of ambient temperature on bacillary dysentery : A multi-city analysis in Anhui Province, China Sci. Total Environ. 671 1206–13
[2] Yang T, Cao C, Guo J, Xu M and Tian H 2019 Analysis of factors evident in the relation between railways and the incidence of dysentery using linear regression Geomatics, Nat. Hazards Risk 10 1459–74
[3] Zhang H, Si Y, Wang X and Gong P 2016 Patterns of Bacillary Dysentery in China, 2005-2010 Int. J. Environ. Res. Public Health 13 164
[4] Liu Z, Tong M X, Xiang J, Dear K, Wang C, Ma W, Lu L and Liu Q 2020 Daily Temperature and Bacillary Dysentery : Estimated Effects, Attributable Risks, and Future Disease Burden in 316 Chinese Cities Health Perspect. 128 057008
[5] Lee H S, Hoang T T H, Pham-duc P, Lee M, Grace D, Phung D C and Thuc V M 2017 Seasonal and geographical distribution of bacillary dysentery (shigellosis) and associated climate risk factors in Kon Tum Province in Vietnam from 1999 to 2013 Infect. Dis. Poverty 6 113
[6] Xu C, Li Y, Wang J and Xiao G 2017 Spatial-temporal detection of risk factors for bacillary dysentery in Beijing, Tianjin BMC Public Health 17 743
[7] López-quílez A 2019 Spatio-Temporal Analysis of Infectious Diseases Int. J. Environ. Res. Public Health 16 669
[8] Yi L et al 2019 The impact of climate variability on infectious disease transmission in China: Current knowledge and further directions Environ. Res. 173 255–61
[9] Ayubi E, Mansournia M A, Motlagh A G, Mosavi-jarrahi A, Hosseini A and Yazdani K 2017 Exploring neighborhood inequality in female breast cancer incidence in Tehran using Bayesian spatial models and a spatial scan statistic Epidemiol. Health 39 e2017021
[10] Jr K O and Owens C J 2009 Monitoring county-level chlamydia incidence in Texas, 2004 - 2005 : application of empirical Bayesian smoothing and Exploratory Spatial Data Analysis (ESDA) methods Int. J. Health Geogr. 8 12
[11] Saita S, Silawan T, Parker D M, Sriwichai P, Phuanukoonnont S, Sudathip P, Maude R J, White L J and Pan-ngum W 2019 Spatial Heterogeneity and Temporal Trends in Malaria on the Thai-Myanmar Border (2012-2017): A Retrospective Observational Study Trop. Med. Infect. Dis. 4 62
[12] Liu W, Ouyang G, Luo X, Luo J, Hu L and Fu M 2020 Moisture content, pore-water pressure and wetting front in granite residual soil during collapsing erosion with varying slope angle Geomorphology 362 107210
[13] Goovaerts P 2006 Geostatistical analysis of disease data : accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging Int. J. Health Geogr. 5 52
[14] Docheshmeh A and Ozgur G 2017 Groundwater quality ranking for drinking purposes, using the entropy method and the spatial autocorrelation index Environ. Earth Sci. 76 269
[15] Yan C, Chen Y, Miao Z, Qin S, Gu H and Cai J 2018 Spatiotemporal Characteristics of Bacillary Dysentery from 2005 to 2017 in Zhejiang Province, China Int. J. Environ. Res. Public Health 15 1826
[16] Tsai P-J 2012 Application of Moran ’ s Test with an Empirical Bayesian Rate to Leading Health Care Problems in Taiwan in a 7-Year Period Glob. J. Health Sci. 4 63
[17] Berke O 2004 Exploratory disease mapping: kriging the spatial risk function from regional count data *Int. J. Health Geogr.* **3** 18

[18] Cardoso D T, de Souza D C, de Castro V N, Geiger S M and Barbosa D S 2019 Identification of priority areas for surveillance of cutaneous leishmaniasis using spatial analysis approaches in Southeastern Brazil *BMC Infect. Dis.* **19** 318

[19] Koo H, Wong D W S and Chun Y 2019 Measuring Global Spatial Autocorrelation with Data Reliability Information *Prof. Geogr.* **71** 551–65

[20] Zhang H, Si Y, Wang X and Gong P 2017 Environmental Drivers and Predicted Risk of Bacillary Dysentery in Southwest China *Int. J. Environ. Res. Public Health* **14** 782