Sandwich mapping of schistosomiasis risk in Anhui Province, China

Yi Hu,[1-3] Robert Bergquist,[4] Henry Lynn,[1-3] Fenghua Gao,[5] Qizhi Wang,[5] Shiqing Zhang,[5] Rui Li,[1-3] Liqian Sun,[1-3] Congcong Xia,[1-3] Chenglong Xiong,[1] Zhijie Zhang,[1-3] Qingwu Jiang[1,2]

[1]Department of Epidemiology and Biostatistics, School of Public Health, Fudan University, Shanghai; [2]Key Laboratory of Public Health Safety, Ministry of Education, Shanghai; [3]Laboratory for Spatial Analysis and Modeling, School of Public Health, Fudan University, Shanghai, China; [4]Ingerod, Brastad, Sweden; [5]Anhui Institute of Parasitic Diseases, Wuhu, China

Abstract

Schistosomiasis mapping using data obtained from parasitological surveys is frequently used in planning and evaluation of disease control strategies. The available geostatistical approaches are, however, subject to the assumption of stationarity, a stochastic process whose joint probability distribution does not change when shifted in time. As this is impractical for large areas, we introduce here the sandwich method, the basic idea of which is to divide the study area (with its attributes) into homogeneous subareas and estimate the values for the reporting units using spatial stratified sampling. The sandwich method was applied to map the county-level prevalence of schistosomiasis japonica in Anhui Province, China based on parasitological data collected from sample villages and land use data. We first mapped the county-level prevalence using the sandwich method, then compared our findings with block Kriging. The sandwich estimates ranged from 0.17 to 0.21% with a lower level of uncertainty, while the Kriging estimates varied from 0 to 0.97% with a higher level of uncertainty, indicating that the former is more smoothed and stable compared to latter. Aside from various forms of reporting units, the sandwich method has the particular merit of simple model assumption coupled with full utilization of sample data. It performs well when a disease presents stratified heterogeneity over space.

Introduction

Schistosomiasis japonica, caused by the trematode worms of *Schistosoma japonicum*, is prevalent in some Asian countries and remains a major public health problem in China (Yang *et al*., 2014) and the Philippines (Soares Magalhães *et al*., 2014). Although great progress has been achieved in China through the implementation of a multitude of control measures for more than six decades, approximately 65 million people are still at risk of schistosomiasis infection with 0.36 million individuals infected in 2009 (Hao *et al*., 2009). The burden of this disease fluctuates currently (Zhang *et al*., 2012b; Hu *et al*., 2014a) reflecting that control activities have entered into an unstable situation, where the trend is no longer unidirectional. The national surveillance system therefore needs to be scrutinized and strengthened. Availability of reliable and up-to-date risk maps at the basic administrative levels (*e.g.*, census tract or county) would facilitate both planning and evaluation of the control strategy.

In China, schistosomiasis surveillance in controlled regions has been operating since the 1990s (Wu *et al*., 2005) with reports on crude prevalence of the disease at the various administrative units produced annually. For example, county-level prevalence data are either actively collected through village-based field surveys, then reported to towns and finally summed at the county level (Zhang *et al*., 2012a; Hu *et al*., 2014b) or, alternatively, passively obtained via case reports from local anti-schistosomiasis stations with the data summed at the county level (Chen *et al*., 2002, 2003; Hao *et al*., 2009), the latter method currently being predominant. It can be argued that the validity and accuracy of this kind of prevalence mapping is questionable, since inaccurate or missing reports are inevitable during the process of reporting to higher administrative levels. Ideally, the prevalence should be estimated directly from village-based field surveys, but survey costs would soar in

this way, as more sample villages in each administrative unit would be needed to achieve a reliable estimate of the true prevalence.

The use of geostatistics has contributed to great progress in disease mapping in recent decades (Berke, 2004; Biggeri *et al*., 2006; Goovaerts, 2006). For example, block Kriging (Isaaks and Srivastava, 1989) can be employed to estimate schistosomiasis risk through parasitological data from sample villages, but the assumption of *stationarity* makes this method impractical when applied to large areas (Vounatsou *et al*., 2009). A stationary process is a stochastic process, whose joint probability distribution does not change when shifted in time, *i.e.* parameters do not follow any trends since statistical properties such as mean, variance, autocorrelation, etc. remain constant in a stationary time series. The precision of this model-based approach would, however, be poor if the spatial stratification were strong and spatial autocorrelation weak (Wang *et al*., 2009). Traditionally, some design-based approaches, such as random sampling, require as many samples as possible to give an unbiased and stable estimate, which is usually impractical.

In the present study, we employed the recently proposed sandwich method (Wang *et al*., 2013), to estimate the county-level prevalence of schistosomiasis in Anhui Province, China. The aim was to compare this method with block Kriging and discuss the results in the context of merits and limitations of this new method. Our overall objective was to identify a superior way of estimating the mean prevalence and its uncertainty for each county.

Materials and Methods

Study area

Anhui Province spans approximately 139,600 km^2 in the East of China and has a population of 59.9 million (2012). The northern and north-central parts of the province are flat, while mountains and a series of hills occupy south-western and south-eastern Anhui. Major rivers include the Huaihe River in the North and the Yangtze River in the South. The province enjoys a subtropical humid monsoon climate with much rain in June and July, which is conducive to growth of *Oncomelania hupensis*, the intermediate snail host of schistosomiasis.

Parasitological data

Schistosomiasis prevalence data were obtained from a cross-sectional survey carried out in 2001 by health professionals of the Anhui Institute of Parasitic Diseases (AIPD) in 1549 randomly selected villages spread over the province. The data were collected through several field surveys using a two-pronged diagnostic approach: screening by a serological test of all residents 5 to 65 years old and confirmed by Kato-Katz stool examination (Yu *et al*., 2007). All egg-positive individuals were unequivocally deemed to be infected with *S. japonicum*. Figure 1 shows the locations of the sample villages in the endemic area.

Land use

A dataset concerning the provincial land use in 2000, provided by the Data Center for Resources and Environmental Sciences, the Chinese Academy of Sciences (RESDC) (http://www.resdc.cn), was used as auxiliary information. The types of land cover included paddy field, dry field, forested land, grass land, water body, land used for industry and buildings, and unused land.

Statistical analysis

Spatial stratification of schistosomiasis risk was first checked using a geographical power of determinant (PD) indicator proposed by Wang *et al.* (2010). The PD value, ranging from 0 to 1, quantifies how similar the spatial distribution of a disease is in comparison to a particular risk factor. If the PD value is close to 1, the disease has a relatively similar spatial stratification as the risk factor in question; if closer to 0, the spatial stratifications of the two variables are quite different.

Figure 2 presents the main points (or levels) of the sandwich method. This approach is composed of three geographical layers: a sampling layer, a zoning layer and a reporting layer.

The sampling layer can be an attribute surface/study area, from which observations are collected. In our study, this layer was composed of the sampled villages.
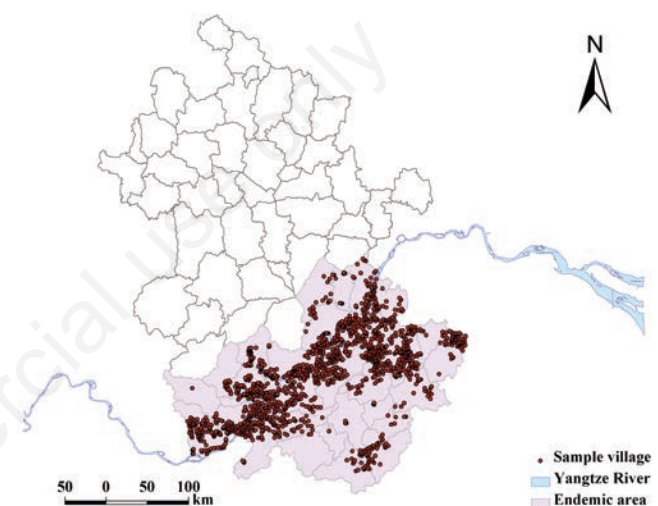

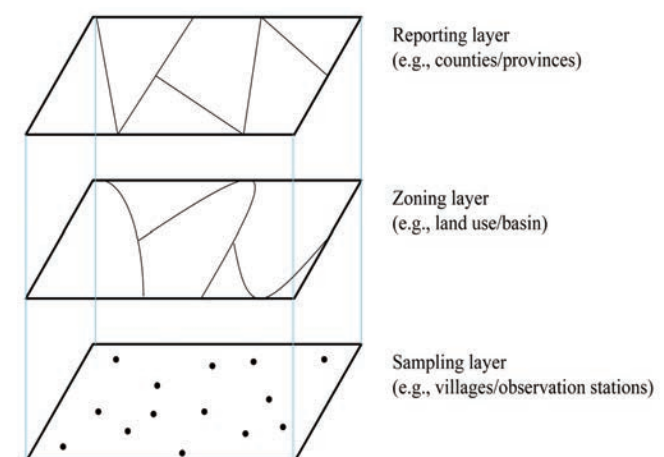Figure 1. Locations of sample villages in Anhui Province, China.

Figure 2. Structure diagram of the sandwich method. Based on Wang *et al.*, 2013.

The zoning layer can take any form of (physical or man-made) zoning that partitions the study area into subareas. The perfect selection of this layer is the one, in which the subareas are spatially homogeneous (constant mean) in terms of the attribute surface. In our study, the zoning layer was composed of land cover, since it is well recognized that the schistosomiasis risk is highly dependent on the prevailing type of land cover.

The reporting layer can consist of any spatial set of administrative units, *e.g.* the county used in our study.

We used a bottom-up process, in which the sample villages were first drawn from within each type of land use with the prevalence data from this zoning surface subsequently transferred to each county (Figure 2). The specific data processing was done as follows:

Zone value estimation

For each subarea in the zoning layer, we estimated the mean and its variance based on the sample dataset. The sample mean ($\bar{y}_i$) of a subarea i is:

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \qquad \text{eq. 1}$$

the variance $v(\bar{y}_i)$ of which is

$$v(\bar{y}_i) = E[\bar{y}_i - E(\bar{y}_i)] = \left(\frac{n_i - 1}{N_i}\right)\left(\frac{1}{n_i}\right)s_i^2 \qquad \text{eq. 2}$$

$$s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \qquad \text{eq. 3}$$

where n_i is the number of sample villages within subarea i; y_{ij} the schistosomiasis prevalence for the sample village j within subarea i and N_i the number of villages (including sample villages and unsampled villages) within subarea i.

Estimating the values of the reporting units

The values for the reporting units (*i.e.* the counties in our study) were estimated from the zoning unit estimates. A reporting unit may overlap one or several subareas from the zoning layer. According to the stratified sampling method (Cochran, 1977), the mean prevalence ($\bar{r}_k$) for a reporting unit k is

$$\bar{r}_k = \frac{1}{N_k} \sum_{i=1}^{L} N_{ki}\bar{y}_i = \sum_{i=1}^{L} W_{ki}\bar{y}_i \qquad \text{eq. 4}$$

and its variance $v(\bar{r}_k)$:

$$v(\bar{r}_k) = \sum W_{ki}^2 v(\bar{y}_i) \qquad \text{eq. 5}$$

where N_k is the number villages within the reporting unit k, N_{ki} the number of villages within the area intersected by the subarea i and the reporting unit k, W_{ki} the weight and L the number of parts in the zoning layer (*i.e.* the subareas) within the reporting unit k.

The PD value of land use was calculated by the specific software (http://www.sssampling.org/Excel-GeoDetector/) and the sandwich software (http://www.sssampling.org/sandwich/) was used to map the

schistosomiasis risk. Spatial stratification and spatial autocorrelation of the risk were checked by PD value and empirical variogram, respectively. For comparison, we also estimated the county-level prevalence by block Kriging.

Results

The PD value of land use was 0.71. Figure 3 displays the empirical variogram of schistosomiasis prevalence, in which the rising trend is not very evident. The partial sill, *i.e.* the difference between the sill (defined as the total variance where the empirical variogram appears to level off) and the nugget, which refers to a discontinuity at the origin in the variogram, and the range (defined as the distance where the empirical variogram starts to flatten) were 0.17e-04 and 16.07 km, respectively.

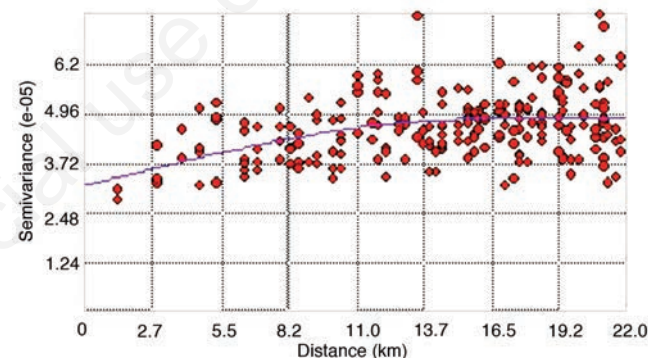Figure 4 shows the schistosomiasis prevalence in the form of maps

Figure 3. Empirical variogram of the village-level prevalence of schistosomiasis in Anhui Province, China. The curve was produced using the spherical function.
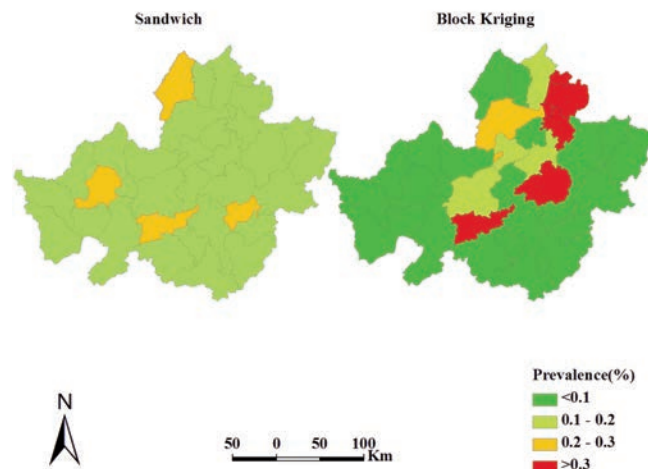

Figure 4. Predicted prevalence of schistosomiasis at the county level in Anhui Province, China: comparison between the sandwich method and block Kriging.

at the county level in the endemic area using both the sandwich method and block Kriging. As shown in Table 1, the sandwich estimates range from 0.17 to 0.21% with a standard deviation (SD) of 1.06e-03, while the Kriging estimates varied from 0 to 0.97% with a SD of 2.48e-03. The former is clearly more smoothed than the latter.

Figure 5 represents the estimation uncertainties with respect to the two methods. As seen in the maps, a lower level of uncertainty (ranging from 0.09e-04 to 0.32e-04) is apparent with the sandwich method, while a higher level of uncertainty (ranging from 0.61e-02 to 0.73e-02) stands out when Kriging modelling is applied.

Discussion

When spatially arranged units are sampled and geocoded, Kriging is useful for the production of a smoothed map. However, the larger the area, the more impractical the assumption of stationarity becomes. To overcome this problem, a new approach, sandwich mapping was used for the estimation of the county-level schistosomiasis prevalence based on village-level parasitological data. The basic idea was to divide the area of certain attributes into homogeneous or quasi-homogeneous subareas and estimate their values for the reporting units using spatial stratified sampling. We believe this method to be novel in that it avoids the heavy dependence upon the intra-area sampling that characterizes conventional design-based methods (*e.g.*, random sampling) (Rao, 2003). The sandwich method can not only be applied without the parametric assumptions required in the case of probabilistic, model-based methods (*e.g.*, Kriging) (Christakos, 2005), but provides also a new way to estimate prevalence at various administrative units using actively collected data instead of passively collected data (*e.g.*, reported data), which are often used in the annual schistosomiasis reports in China (Chen *et al*., 2002, 2003; Hao *et al*., 2009).

The precision of the Kriging interpolation is poor when the spatial stratification is strong and the spatial autocorrelation weak. We used a PD indicator to check the spatial stratification of schistosomiasis risk. The PD value of 0.71 for land use shows that the spatial distribution of land use can explain 71% of the risk variation, which indicates that the spatial distribution of risk is strongly similar to that of specific land use; on the other hand, the barely visible rise of the trend in the variogram in Figure 3 suggests a weak spatial correlation. Therefore, the accuracy of county-level mapping based on Kriging must be deemed to be insufficiently precise.

The sandwich approach depends on a different model and different assumptions than that of Kriging. In the sandwich framework, the subareas divided by the zoning layer are assumed to be homogeneous or quasi-homogeneous with the spatial stratified sampling used to estimate the attribute value for the reporting units. This framework provides an easy and relatively simple way (*e.g.*, there is no assumption on probability in the model) to tackle prediction problems at the reporting unit level. Another advantage is that the method is easily implemented,

and that the sample data are fully used when estimating the attribute value for a reporting unit. As shown in Figure 2, the prevalence of schistosomiasis in a county was estimated from data for each type of land cover and calculated using all the sample villages within each type of land use in the study area. Hence, both sample villages within and around that county were used when estimating the prevalence, which is the basic idea of the Kriging method. Indeed, attributes in neighbouring units are not necessarily the most appropriate units to borrow information from when spatial heterogeneity exits (Haining, 2003). The difference is apparent in Figures 4 and 5, in which the sandwich estimate is more smoothed and stable as more sample data were included when estimating the prevalence in each county. Another advantage is that the reporting units can take any form of spatial unit (*e.g.*, census tract, economic region or basin) depending on the research need.

Sandwich estimation may not bring significant benefits if an arbitrary zoning layer is constructed (Wang *et al*., 2013) since this step is critical for the accuracy of the sandwich estimate requiring prior knowledge about the surface. In our study, it was well recognized that land use is identified as an important determinant that controls the spatial distribution of schistosomiasis (Steinmann *et al*., 2006; de Souza *et al*., 2012; Zhou *et al*., 2013). As explained above, the PD value of 0.71 for land use indicates that schistosomiasis presents some degree of spatial homogeneity within each type of land use. Hence, land use can be seen as an appropriate zoning layer.

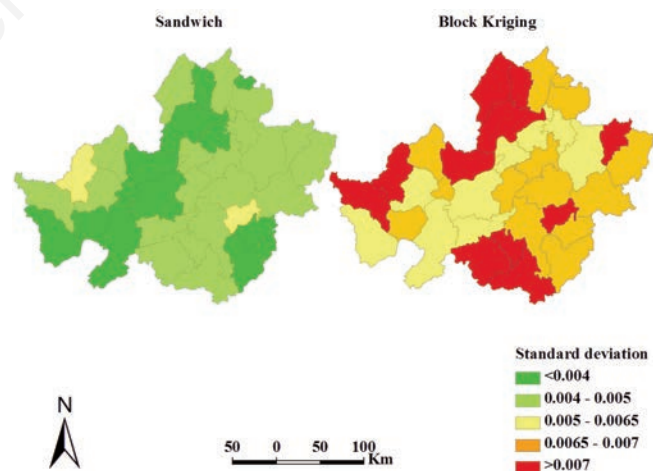Although our analysis indicates that land use is an appropriate zon-

Figure 5. Uncertainties of the model predictions of the prevalence of schistosomiasis at county level in Anhui Province, China: comparison between the sandwich method and block Kriging.

Table 1. Statistics for the predicted prevalence of schistosomiasis at the county level in Anhui Province, China, using sandwich and block Kriging.

Modeling	Minimum	Median	Mean	Maximum	Standard deviation
Sandwich	0.166	0.175	0.188	0.206	1.060e-03
Block Kriging	0	0.038	0.156	0.971	2.482e-03

ing layer, it may not be the optimal one. Any other risk factor, such as land surface temperature, rainfall, distance to water bodies (*e.g.*, lake, river, *etc.*) are definitely potential candidates. We compared different zoning layers determined by the factors mentioned using the PD value, but none was better than the land cover parameter. A possible reason is that these factors are continuous over the study area, and subareas divided through arbitrary discretization (*e.g.*, equal intervals, quintiles, *etc.*) may not be homogeneous. Even so, more knowledge on how to zone the surface can be explored by further study geared at obtaining more accurate sandwich estimates. Another limitation of the current study is that the prevalence of schistosomiasis in each sample village may not be perfectly accurate, as we could not rule out diagnostic errors completely because the sensitivity and specificity of the tests were not as high as they would need to be for a fail-safe result (Wang *et al.*, 2008). Further studies are needed to take the diagnostic error into account when modelling schistosomiasis risk. Using uncertainties of estimates might make it difficult to fully clarify which approach is better, but it could confirm which one is the more stable and reliable. This is a limitation inherently coupled to difficulties in carrying out validation. The focus here is to estimate the county-level prevalence using village-level one; hence, there are no observations at the county level needed to conduct a cross-validation.

Conclusions

This study presents an application of the sandwich method in mapping the county-level prevalence of schistosomiasis in Anhui Province. The merits of this method are simple model assumption, full utilization of sample data, the possibility of various forms of reporting units and the fact that it performs well in the case of stratified, heterogeneous areas. Many environment-related diseases (*e.g.*, malaria and schistosomiasis) depend strongly on environmental determinants, which present heterogeneity over space and time. In this case, the sandwich method would be more reliable compared to other stationarity-based methods (*e.g.*, Kriging). Nevertheless, we should still use the sandwich method with caution when mapping a disease as the accuracy is so strongly dependent on the zoning layer, which in turn is highly related to prior knowledge about disease prevalence.

References

Berke O, 2004. Exploratory disease mapping: kriging the spatial risk function from regional count data. Int J Health Geogr 3:18.

Biggeri A, Dreassi E, Catelan D, Rinaldi L, Lagazio C, Cringoli G, 2006. Disease mapping in veterinary epidemiology: a Bayesian geostatistical approach. Stat Methods Med Res 15:337-52.

Chen X, Jiang Q, Wang L, Zhao G, Zhao Q, Gu Y, 2002. Schistosomiasis situation in the People's Republic of China in 2001. Chin J Schisto Control 14:241-3.

Chen X, Wu X, Wang L, Dang H, Wang Q, Zheng J, Guo J, 2003. Schistosomiasis situation in People's Republic of China in 2002. Chin J Schisto Control 15:241-4.

Christakos G, 2005. Random field models in earth sciences (Dover Books on Science). Dover Publ., San Diego, CA, USA.

Cochran WG, 1977. Sampling techniques. John Wiley, New York, NY, USA.

de Souza GE, Leal-Neto OB, Albuquerque J, Pereira DSH, Barbosa CS, 2012. Schistosomiasis transmission and environmental change: a spatio-temporal analysis in Porto de Galinhas, Pernambuco, Brazil. Int J Health Geogr 11:51.

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.

Haining R, 2003. Spatial data analysis: theory and practice. Cambridge University Press, Cambridge, UK.

Hao Y, Zheng H, Zhu R, Guo J, Wang L, 2009. Schistosmiasis situation in People's Republic of China in 2009. Chin J Schisto Control 22:521-7.

Hu Y, Xiong C, Zhang Z, Luo C, Cohen T, Gao J, Zhang L, Jiang Q, 2014a. Changing patterns of spatial clustering of schistosomiasis in Southwest China between 1999-2001 and 2007-2008: assessing progress toward eradication after the world bank loan project. Int J Environ Res Public Health 11:701-12.

Hu Y, Xiong C, Zhang Z, Luo C, Ward M, Gao J, Zhang L, Jiang Q, 2014b. Dynamics of spatial clustering of schistosomiasis in the Yangtze River Valley at the end of and following the World Bank Loan Project. Parasitol Int 63:500-5.

Isaaks E, Srivastava R, 1989. Applied geostatistics. Oxford University Press, Oxford, UK.

Rao JNK, 2003. Small area estimation. John Wiley, New York, NY, USA.

Soares Magalhães RJ, Salamat MS, Leonardo L, Gray DJ, Carabin H, Halton K, McManus DP, Williams GM, Rivera P, Saniel O, Hernandez L, Yakob L, McGarvey S, Clements A, 2014. Geographical distribution of human *Schistosoma japonicum* infection in the Philippines: tools to support disease control and further elimination. Int J Parasitol 44:977-84.

Steinmann P, Keiser J, Bos R, Tanner M, Utzinger J, 2006. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. Lancet Infect Dis 6:411-25.

Vounatsou P, Raso G, Tanner M, N'Goran EK, Utzinger J, 2009. Bayesian geostatistical modelling for mapping schistosomiasis transmission. Parasitology 136:1695-705.

Wang JF, Christakos G, Hu MG, 2009. Modeling spatial means of surfaces with stratified nonhomogeneity. IEEE T Geosci Remote 47:4167-74.

Wang JF, Haining R, Liu TJ, Jiang CS, 2013. Sandwich estimation for multi-unit reporting on a stratified heterogeneous surface. Environ Plann A 45:2515-34.

Wang JF, Li XH, Christakos G, Liao YL, Zhang T, Gu X, Zheng XY, 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. Int J Geogr Inf Sci 24:107-27.

Wang XH, Zhou XN, Vounatsou P, Chen Z, Utzinger J, Yang K, Steinmann P, Wu XH, 2008. Bayesian spatio-temporal modeling of *Schistosoma japonicum* prevalence data in the absence of a diagnostic 'gold' standard. PLoS Neglect Trop D 2:e250.

Wu XH, Chen MG, Zheng J, 2005. Surveillance of schistosomiasis in five provinces of China which have reached the national criteria for elimination of the disease. Acta Trop 96:276-81.

Yang GJ, Liu L, Zhu HR, Griffiths SM, Tanner M, Bergquist R, Utzinger J, Zhou XN, 2014. China's sustained drive to eliminate neglected tropical diseases. Lancet Infect Dis 14:881-92.

Yu JM, de Vlas SJ, Jiang QW, Gryseels B, 2007. Comparison of the Kato-Katz technique, hatching test and indirect hemagglutination assay (IHA) for the diagnosis of *Schistosoma japonicum* infection in

China. Parasitol Int 56:45-9.

Zhang Z, Zhu R, Bergquist R, Chen D, Chen Y, Zhang L, Guo J, Fei Z, Jiang Q, 2012a. Spatial comparison of areas at risk for schistosomiasis in the hilly and mountainous regions in the People's Republic of China: evaluation of the long-term effect of the 10-year World Bank Loan Project. Geospatial Health 6:205-14.

Zhang Z, Zhu R, Ward MP, Xu W, Zhang L, Guo J, Zhao F, Jiang Q, 2012b.

Long-term impact of the World Bank Loan Project for schistosomiasis control: a comparison of the spatial distribution of schistosomiasis risk in China. PLoS Neglect Trop D 6:e1620.

Zhou YB, Liang S, Chen GX, Rea C, Han SM, He ZG, Li YP, Wei JG, Zhao GM, Jiang QW, 2013. Spatial-temporal variations of *Schistosoma japonicum* distribution after an integrated national control strategy: a cohort in a marshland area of China. BMC Public Health