SOIL-TRANSMITTED HELMINTHIASES IN CHINA: A SPATIAL STATISTICAL ANALYSIS

Dejian Lai and Bartholomew P Hsi

School of Public Health, University of Texas Health Science Center, Houston, TX 77030, USA

Abstract. The prevalence of the soil-transmitted helminthiases is extremely high in some rural areas in China. The endemic foci of these human parasites constitute one of the most neglected public health problems. A large scale survey on human parasites (1988-1992) was conducted in China. The prevalences of the soil-transmitted helminthiases in the 30 provinces, autonomous regions and municipalities were reported and correlated with ecological factors (Yu *et al*, 1994; Xu *et al*, 1995). In this paper, we re-examined and analyzed the reported results with a spatial pattern test statistic, the *D* statistic. The values of the *D* statistic indicated that the spatial correlation of the prevalence of the soil-transmitted helminthiases in China was significant. We also found that the spatial correlation of the prevalence could largely be explained by the paddy field area *per capita*. The correlation of the prevalence of the soil-transmitted helminthiases and the paddy field area *per capita* was higher than the correlation of the prevalence with the factors found previously by Xu *et al* (1995).

INTRODUCTION

The high prevalence of the soil-transmitted intestinal nematodes, round-worms *(Ascaris lumbricoides)*, hookworms and whipworms *(Trichuris trichiura)*, in many tropical and subtropical developing countries poses a substantial public health problem to human beings. The prevalence of these human parasites are higher in children and teenagers than in other age groups (Yu *et al*, 1994; Xu *et al* 1995; Evans and Stephenson, 1995). Infections of these parasites retard children's growth rates, decrease physical and mental energy levels, and increase morbidity and mortality. The clinical course of these infections varies from relatively mild symptoms and signs to very severe ones.

China is the largest developing country with a high proportion (80%) of agricultural and rural populations in its 1.2 billion total population. It is located on the Eurasian plate with a large area of territory. The varied and complicated landforms create a diverse climate with many different temperature zones. The population densities and life styles in China vary from region to region.

Soil-transmitted helminthiases were highly endemic in some Chinese rural areas. The results of the 1988-1992 national survey of human parasites in China were reported by Yu *et al* (1994) and Xu *et al* (1995). The prevalence of the soil-transmitted helminthiases was found to be closely associated with some climatic and geographic factors (Xu *et al*, 1995). In this paper, we re-analyzed the reported results using a statistical spatial pattern test, the *D* statistic. A simple linear regression model which correlates the prevalence rates with a new factor, the paddy field area *per capita* was established. The new factor has a higher correlation with the prevalence of the soil-transmitted helminthiases than the factors examined previously by Xu *et al* (1995).

Correspondence: Dejian Lai, RAS E510, School of Public Health, University of Texas at Houston, Houston, TX 77030, USA.

MATERIALS AND METHODS

Spatial analyses have been widely used to quantify the visual perception of spatial patterns in mapped data (Cliff and Ord, 1975; Grimson and Rose, 1991; Cuzick and Edwards, 1990, Whittemore *et al*, 1987). More recent work and references can be found in Oden (1995) and Sun (1995). Suppose some health related aggregate statistics, such as life expectancies or mortality rates, are observed from each of several regions. Spatial correlation and its statistical significance can be determined by using a class of permutation statistics (Walter, 1994). The permutation test statistic can be described as follows. Let Y_i be the aggregate statistics observed from region i. We assume that Y_i s are independent and identically distributed random variables. Let X_i be a function of the observed values of $Y_1, Y_2, \ldots, Y_n$ and $h_n(X_i, X_j)$ be a real-value symmetric function. That is, $h_n(X_i, X_j) = h_n(X_j, X_i)$. The statistic

$$W_n = \sum\sum_{1 \le i \ne j \le n} w_{ijn} h_n(X_i, X_j)$$
$$= {}^2\sum\sum_{1 \le i < j \le n} w_{ijn} h_n(X_i, X_j) \qquad (1)$$

is called a permutation statistic with weight W_{ijn}. In this paper, we assume that $W_{ijn} = W_{jin}$. This statistic can be used to detect spatial pattern and quantify associations among the aggregate statistics. If W_{ijn} is 1 and $h_n(X_i, X_j) = h(X_i, X_j) / [n(n-1)]$, the statistic W_n is known as a *U*-statistic. Its asymptotic distribution has been investigated (Lee, 1990; Serfling, 1980). If W_{ijn} is a constant and X_i's are independent and identically distributed, the limit distribution of W_n was reported by Shapiro and Hubert (1979) and O'Neil and Redner (1993). The statistic W_n of this weighted form is widely used in biometry (Mantel 1967) and in other fields (Blom, 1976; Brown and Kildea, 1978; Jammalamadaka and Janson, 1986). If we let $W_{ijn} = d_{ij} / (\sum\sum d_{ij})$, $X_i = \text{rank}(Y_i)$, and $h_n(X_i, X_j) = (X_i - X_j)$ where

$$d_{ij} = \begin{cases} 1 & \text{if region } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

this form of W_n is called the *D* statistic which is a special case of the permutation statistic and has been used in testing spatial patterns in regional health data (Walter, 1994).

In this paper, the *D* statistic is used in quantifying the spatial patterns of the soil-transmitted helminthiasis prevalence rates in China. The standardized *D* statistic is obtained by subtracting the mean from its original observed value and then dividing the remainder by its standard deviation. The expectation and the variance of the *D* statistic are $(n+1)/3$ and $[n(n=1)-A]/(18A)$, respectively, where $A = \sum\sum_{i \ne j} d_{ij}$ (Walter, 1994). We computed the standardized *D* statistic for the soil-transmitted parasite prevalence rates in the 30 provinces, autonomous regions, and municipalities (P/A/M) in China.

In the analysis of the correlation of the prevalence of the soil-transmitted helminthiases with the paddy field area *per capita* we used the usual Pearson correlation coefficient and the simple least square linear regression as in Xu *et al* (1995), in which they studied the correlations of the prevalences of *Ascaris lumbricoides*, hookworms and *Trichuris trichiura* with several factors.

The prevalence of these soil-transmitted helminthiases were extracted from Xu *et al* (1995).

The paddy field area *per capita* was calculated by dividing the total paddy field area in 1990 of each P/A/M (Zhao, 1994) by the total population in 1990 of each P/A/M (Li, 1992).

The adjacent matrix of the P/A/M of China is presented in Table 1 and the locations of the P/A/M in China are shown in Fig 1. The prevalence of the soil-transmitted helminth infections and the paddy field area *per capita* are summarized in Table 2 for easy reference.

RESULTS

We first analyzed the spatial patterns of the prevalence rates of the soil-transmitted helminthiases (*Ascaris lumbricoides*, hookworms and *Trichuris trichiura)* in each province, autonomous region, and municipality (P/A/M) in China using the *D* statistic discussed in the previous section. The values for the standardized *D* statistic with simple symmetric binary weights for each of these prevalence rates are -4.4962, -5.9291, and -4.8210, respectively. The standardized *D* statistic can be reasonably assumed to be asymptotically distributed as standard normal under the null hypothesis that the aggregate statistics are independently and identically distributed (Walter, 1994). The values of the observed *D* statistic suggest that the spatial autocorrelations of soil-transmitted helminth infections in the 30 P/A/M in China are statistically significant as expected.

The correlation coefficients of the prevalence of *Ascaris lumbricoides*, hookworms and *Trichuris trichiura* in China with the paddy field area *per capita* were 0.6416, 0.6739 and 0.6667, respectively. These correlation coefficients were higher than the coefficients of the prevalence rates with other factors reported in the original paper by Xu *et al* (1995).

Using the simple least square method, we constructed linear relationships between the prevalence rates of soil-transmitted helminth infections and the paddy field area *per capita*. Let x be the paddy field area *per capita* and y be the prevalence of soil-transmitted helminth infection. For the prevalence of *Ascaris lumbricoides*, hookworm and

Table 1

Adjacency matrix for the 30 provinces/autonomous regions/municipalities (P/A/M) of the People's Republic of China.

P/A/M						
Beijing	00110	00000	00000	00000	00000	00000
Shanghai	00000	00000	00000	11000	00000	00000
Tianjin	10010	00000	00000	00000	00000	00000
Hebei	10101	11000	00001	00000	10000	00000
Shanxi	00010	10001	00000	00000	10000	00000
Nei Mongol	00011	01111	11000	00000	00000	00000
Liaoning	00010	10100	00000	00000	00000	00000
Jilin	00000	11010	00000	00000	00000	00000
Heilongjiang	00000	10100	00000	00000	00000	00000
Shaanxi	00001	10000	11000	00000	11000	01000
Gansu	00000	10001	01110	00000	00000	01000
Ningxia	00000	10001	10000	00000	00000	00000
Qinghai	00000	00000	10010	00000	00000	01001
Xingjiang	00000	00000	10100	00000	00000	00001
Shandong	00010	00000	00000	10100	10000	00000
Jiangsu	01000	00000	00001	01100	00000	00000
Zhejiang	01000	00000	00000	10111	00000	00000
Anhui	00000	00000	00001	11010	11000	00000
Jiangxi	00000	00000	00000	01101	01110	00000
Fujian	00000	00000	00000	01010	00010	00000
Henan	00011	00001	00001	00100	01000	00000
Hubei	00000	00001	00000	00110	10100	01000
Hunan	00000	00000	00000	00010	01011	01100
Guangdong	00000	00000	00000	00011	00101	10000
Guangxi	00000	00000	00000	00000	00110	10110
Hainan	00000	00000	00000	00000	00011	00000
Sichuan	00000	00001	10100	00000	01100	00111
Guizhou	00000	00000	00000	00000	00101	01010
Yunnan	00000	00000	00000	00000	00001	01101
Xizang	00000	00000	00110	00000	00000	01010

Trichuris trichiura, we found the least square regression lines to be $y = 26.4427 + 43.7970X$, $y = 2.1190 + 47.2712X$, and $y = 0.0586 + 43.1457X$, respectively. These regression equations are all statistically significants (p-values were all < 0.0001).

After establishing the linear regression lines, we calculated the D statistic for the residuals from each model. The standardized D statistic from the residuals of these linear regressions were -1.0954, -2.7767 and -1.5921, respectively.

DISCUSSION

In this paper, we applied the D statistic and the simple linear regression model to study the spatial correlation of the prevalences of soil-transmitted helminthiases in China. The D statistic indicated that all the three major types of the soil-transmitted helminthiases were highly spatially correlated.

The simple linear regression of the prevalence of the soil-transmitted helminthiases (*Ascaris lumbricoides*, hookworms and *Trichuris trichiura*)

Fig 1–Map of the People's Republic of China

and the *per capita* paddy field area is highly statistically significant. This explanatory variable used in the linear regression model has a higher correlation than any other factors previously identified in Xu *et al* (1995).

The standardized values of the *D* statistic for the residuals from the above regression are much smaller, in a absolute value sense, than the standardized values of the *D* statistic for the original prevalence rates of the soil-transmitted helminthiases. This fact indicates that the spatial patterns of the soil-transmitted helminthiases in China can be explained mostly by the paddy field area *per capita* in each P/A/M. In fact, the spatial pattern of the *per capita* paddy field area itself is statistically significant. The standardized value of the *D* statistic for the *per capita* paddy field area is -3.9994.

We associated the prevalence rates of the soil-transmitted helminthiases in China with the *per capita* paddy field area and found that this area field has a higher correlation coefficients with the prevalence rates than those reported in Xu *et al* (1995). We know also that the infectious mechanism of the soil-transmitted helminthiases is simple, but the prevalences of the these infections are influenced by the joint effects of many social, biological and economic factors. The *per capita* paddy field area is only a proxy of these factors. Hence, we should not conclude from this paper that the prevalence of the soil-transmitted helminthiases in China are determined solely by the paddy field area *per capita.*

The ultimate goal of any analysis or control programs for an infectious disease is to eradicate the infection. Theoretically, it can be achieved by either completely interrupting the transmitting mechanism, or by curing all currently infected people, or by vaccinating against the infection. Unfortunately, there are no vaccines against the soil-transmitted helminthiases. In most of the developed countries, the goal of eradicating the soil-transmitted helminthiases is achieved mainly by preventing transmission with good sanitation systems and food hygiene. This is costly and takes time to complete and may not be a suitable approach for Chinese rural areas at the present.

In order to achieve a better control of the infection of *Ascaris lumbricoides,* hookworms and *Trichuris trichiura,* some other important factors should also be included. Because human nightsoil is the main fertilizer used in China's agriculture, in the foreseeable future, agriculture production would be depressed if human nightsoil were not used. Hence, for breaking the life cycle of these parasites, it is very important to find proper ways to disinfest and maintain human nightsoil as fertilizer.

It has been widely accepted that the most attention should be given to school children who have the highest prevalence rates of the soil-transmitted helminthiases among all age groups (Xu *et al,* 1995). Due to the high population density, poor sanitation systems and inadequate supply of clean water in rural China, it is a difficult task to reduce the prevalence of infection at the present stage of economic and scientific development in these rural areas. However, China has a relative better and broader population coverage by the elementary education system compared to many other developing countries. If knowledge of the infectious mechanism and periodic chemotherapy are given in the schools, the prevalence of the soil-transmitted helminthiases could be reduced.

Our analysis provides a different angle for studying the survey results of the soil-transmitted helminthiases in China. The significant spatial patterns of the prevalence rates and their relation with the paddy field area *per capita* may suggest some effective ways such as spatial (regional) cooperation among agencies in different areas in educating people susceptible to infection, inducing voluntary effort for controlling the soil-transmitted helminthiases and improving the health of the Chinese people, especially those in rural areas in China.

Table 2
The prevalence rates of the soil-transmitted helminthiases, the total population (in 10 thousand) and the area of paddy field (in 10 thousand *mu**) of the 30 provinces, autonomous regions and municipalities (P/A/M) of the People's Republic of China.

P/A/M	*Ascaris lumbricoides (%)*	*Trichuris trichiura (%)*	Hookworm (%)	Pop	Paddy field
Beijing	29.5	0.8	0.0	1,082	49.0
Shanghai	29.8	23.2	3.8	1,334	428.0
Tianjin	28.3	6.9	0.4	879	66.6
Hebei	31.8	0.6	0.3	6,108	210.6
Shanxi	25.7	1.1	0.02	2,876	16.1
Nei Mongol	16.6	0.2	0.087	2,146	114.3
Liaoning	55.5	6.6	0.03	3,946	805.5
Jilin	28.4	0.4	0.0	2,466	627.8
Heilongjiang	10.3	0.6	0.0	3,521	1,022.1
Shaanxi	41.4	5.3	0.01	3,288	257.2
Gansu	37.5	2.1	0.01	2,237	11.4
Ningxia	24.5	0.4	0.4	466	262.6
Qinghai	33.9	0.6	0.0	446	0.0
Xingjiang	9.1	1.2	0.3	1,516	122.0
Shandong	38.3	13.6	6.1	8,439	240.5
Jiangsu	39.5	27.3	21.8	6,706	4,206.2
Zhejiang	60.0	40.3	28.2	4,145	2,144.1
Anhui	46.4	17.4	33.4	5,618	2,772.3
Jiangxi	71.1	17.1	17.6	3,771	2,989.2
Fujian	57.1	41.0	21.6	3,005	1,498.7
Henan	41.4	8.2	20.7	8,551	596.5
Hubei	39.5	18.3	8.8	5,397	2,807.7
Hunan	67.7	20.2	22.9	6,066	3,948.0
Guangdong	46.4	33.2	22.3	6,283	2,849.8
Guangxi	66.0	47.7	37.9	4,225	2,377.3
Hainan	61.8	66.7	60.9	656	383.0
Sichuan	68.4	30.4	40.9	10,722	4,833.8
Guizhou	71.1	29.1	22.9	3,239	1,167.9
Yunnan	59.6	27.3	19.3	3,697	1,469.2
Xizang	6.0	2.3	0.4	220	0.9

*One *mu* equals 666.7 m^2, or 1/6 acre, or 1/15 hectare.

REFERENCES

Besag J, Newell J. The detection of clusters in rare disease. *J R Stat Soc Ser A* 1991; 154 : 143-55.

Blom G. Some properties of incomplete *U*-statistics. *Biometrika* 1976; 63 : 573-80.

Brown BM, Kildea DG. Reduced *U*-statistics and the Hodges-Lehmann estimator. *Ann Stat* 1978; 6 : 828-35.

Cliff AD, Ord JK. Model building and the analysis of spatial pattern in human geography. *J R Stat Soc Ser B*, 1975; 37 : 297-348.

Cuzick J, Edwards R. Spatial clustering for inhomogeneous population. *J R Stat Soc Ser B*, 1990; 52 : 73-104.

Evans AC, Stephenson LS. Not by drug alone: the fight against parasitic helminths. *World Health Forum* 1995; 16 : 258-61.

Grimson RC, Rose RD. A versatile test for clustering and proximity analysis of neurons. *Meth Info Med* 1991; 30 : 421-41.

Jammalamadaka SR, Janson S. Limit theorems for a triangular scheme of *U*-statistics with applications to inter-point distances. *Ann Probability* 1986; 14 : 1347-58.

Lee AJ. *U*-statistics : Theory and Practice. New York : Marcel Dekker, 1990.

Li CR. A Study of China's Population. Beijing: Foreign Languages Press, 1992.

Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res* 1967; 27 : 209-20.

Oden N. Adjusting Moran's *I* for population density. *Stat Med* 1995; 14 : 17-26.

O'Neil KA, Redner RA. Asymptotic distributions of weighted *U*-statistics of degree 2. *Ann Probability,* 1993; 21 : 1159-69.

Serfling RJ. Approximation Theorems of Mathematical Statistics. New York: John Wiley, 1980.

Shapiro CP, Hubert L. Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions. *Ann Stat* 1979; 7 : 788-94.

Sun P. Further study on the statistical test to detect spatial pattern. *Biometric J* 1995; 37 : 199-203.

Walter SD. A simple test for spatial pattern in regional health data. *Stat Med* 1994; 13 : 1037-44.

Whittemore AS, Friend N, Brown BW, Holly EA. A test to detect clusters of disease. *Biometrika* 1987; 74 : 631-5.

Xu LQ, Yu SH, Jiang ZX, *et al.* Soil-transmitted helminthiases: nationwide survey in China. *Bull WHO* 1995; 73 : 507-513.

Yu SH, Xu LQ, Jiang ZX, *et al.* Special report: nationwide survey of human parasites in China. *Southeast Asian J Trop Med Public Health* 1994; 25 : 4-10.

Zhao SQ. Geography of China: environment, resources, population, and development. New York : John Wiley, 1994.