



ELSEVIER



<http://intl.elsevierhealth.com/journals/ijid>

A multi-level analysis of risk factors for *Schistosoma japonicum* infection in China

Jianzhou Yang^{a,e,g,h}, Zhengyuan Zhao^{b,c}, Yuesheng Li^{b,d},
Daniel Krewski^{e,f}, Shi Wu Wen^{c,f,g,h,*}

^a Department of Preventive Medicine, Changzhi Medical College, Changzhi, Shanxi, China

^b Hunan Institute of Parasitic Diseases, Yueyang, Hunan, China

^c School of Public Health, Central South University, Changsha, Hunan, China

^d Molecular Parasitology Laboratory, Australian Centre for International and Tropical Health and Nutrition, the Queensland Institute of Medical Research, Brisbane, Qld, Australia

^e McLaughlin Centre for Population Health Risk Assessment, Institute of Population Health, University of Ottawa, Ottawa, Ontario, Canada

^f Department of Epidemiology and Community Medicine, University of Ottawa Faculty of Medicine, Ottawa, Ontario, Canada

^g OMNI Research Group, Department of Obstetrics and Gynecology, University of Ottawa Faculty of Medicine, 501 Smyth Rd, Box 241, Ottawa, Ontario, Canada, K1H 8L6

^h Ottawa Health Research Institute, Clinical Epidemiology Program, Ottawa, Ontario, Canada

Received 21 May 2008; received in revised form 16 December 2008; accepted 4 February 2009

Corresponding Editor: William Cameron, Ottawa, Canada

KEYWORDS

Schistosoma japonicum;
China;
Risk factors;
Multi-level model;
Logistic regression model

Summary

Objective: The aim of this study was to explore the risk factors of schistosomiasis japonica in China, using a hierarchical multi-level model with individuals nested within villages.

Methods: A cross-sectional survey of schistosomiasis japonica was conducted in 16 villages in the Chinese province of Hunan. A multi-level modeling technique (HLM version 6.04) was used to assess risk factors of schistosomiasis. The results from this multi-level model were compared with those from a conventional single-level logistic regression model.

Results: A total of 10 245 individuals were enrolled in this study, of whom about 4.1% were infected with *Schistosoma japonicum*. In the multi-level model analysis, individual level variables such as gender, age, and occupation, and village level variables such as type of *S. japonicum* endemic area, drinking water source, sewage treatment, June temperature, and April rainfall were associated with schistosomiasis japonica infection. Conventional single-level logistic regression analysis selected more independent variables, and had narrower confidence intervals around the corresponding regression coefficients. In particular, per capita income, precipitation in October, and density of infected snails were statistically significant in the conventional single-level logistic regression analysis but not in the multi-level model.

* Corresponding author. Tel.: +1 613 737 8899x73912; fax: +1 613 739 6266.
E-mail address: swwen@ohri.ca (S.W. Wen).

Conclusions: Multi-level modeling is a useful tool in the analysis of risk factors of schistosomiasis japonica. Because the multi-level model captures the hierarchical structure of the data, it may be considered a more appropriate analytical tool for data of this type. This technique may also be useful in the analysis of other infectious diseases with a similar hierarchical structure.

© 2009 International Society for Infectious Diseases. Published by Elsevier Ltd. All rights reserved.

Introduction

Schistosomiasis is a major public health problem, affecting 200 million people in 74 countries worldwide.¹ The Dongting Lake region is one of the endemic regions of schistosomiasis in China, with a total of one million people and more than 300 000 domestic animals infected with *Schistosoma japonicum*. It is estimated that snail habitats in the lake region cover 3795 km².²

Although the immediate route of schistosomiasis japonica infection is through contact with contagious water, the mode of transmission is complex. A variety of social, biological, and environmental factors can affect the likelihood of infection with schistosomiasis japonica. These include socio-demographic factors such as sex, age, occupation, income, education, agricultural activities, and water sources;^{3,4} biological factors such as genetics and infectious status of snails and livestock;⁵ and natural environmental factors such as temperature and precipitation.⁶ These risk factors can be divided into two broad groups: factors that are unique to each individual, such as sex, age, occupation, income, education, and genetics, and ecological factors that are shared by individuals living in the same community (e.g., village), such as water sources, infectious status of snails and livestock, temperature, and precipitation.

Previous studies of risk factors for schistosomiasis japonica have been conducted using traditional analytical methods at either the individual level or the ecological level, thereby ignoring the hierarchical structure of the data. Failure to take into account the hierarchical structure of data can underestimate the standard errors of the model parameters, and lead to inflated type I error rates. These problems can be avoided using multi-level modeling, which incorporates the hierarchical structure of the data. Multi-level models were first developed by statisticians in the UK and the USA in the 1980s based on variance component analysis.^{7–9} More recently, it has become a popular statistical tool in the social sciences.¹⁰ The objective of the present study was to explore the uses of multi-level models in the assessment of risk factors of schistosomiasis japonica.

Materials and methods

We conducted a cross-sectional survey in 16 villages in the Chinese province of Hunan in 2005, representing the national schistosomiasis japonica surveillance sites in the province. These villages represent various geographic areas with different snail habitat types and severity of schistosomiasis japonica. These areas are divided into four loosely defined types: lake-embankment type, lake-beach type, inside embankment type, and hill type. Lake-embankment type represents villages that are located close to the lake but inside the embankment. Residents in these villages get infec-

tion of schistosomiasis japonica mainly through contacts with water in the lake or on the beach. Lake-beach type represents villages that are located close to the lake with no embankment. Residents in these villages get infection of schistosomiasis japonica largely through contacts with water in the lake or on the beach. Inside embankment type represents villages that are located inside the embankment but far away from the lake. Residents in these villages get infection of schistosomiasis japonica mainly through farming or recreation activities in snail-ridden areas within the embankment. Hill type represents villages that are located in mountain and hill areas. Residents in these villages get infection of schistosomiasis japonica through farming or recreation activities in streams, drains, culverts, ravines, and other fields with snail habitats. All residents aged 6 years or older were invited to participate in the study.

The outcome variable was the status of schistosomiasis japonica infection of the individual study subjects. The indirect hemagglutination test (IHA) was first used to screen for the presence of antibodies to *S. japonicum*. IHA-positive cases were then tested by stool examination. One stool specimen was obtained from each patient and three Kato–Katz thick smears were prepared. Schistosomiasis japonica infection was diagnosed if any of the three Kato–Katz smears was positive. Schistosomiasis japonica infection in a sample of livestock (including 818 cattle, 201 sheep, and two horses) was measured by miracidium hatching methods.¹¹ A snail survey using a combined method of environmental sampling and systematic sampling techniques was conducted in April of the survey years, and snail infection density (per 0.11 m²) was estimated accordingly.

Individual level independent variables considered in this study representing potential risk factors for schistosomiasis included sex, age, and occupation. The village was chosen as the geographical unit for ecological level variables, with a single value of each ecological variable considered calculated or determined for each village. Thus, all study subjects living in the same village had the same value of each village-level risk factor for schistosomiasis. Ecological variables considered in this study included geographic characteristics of the study villages (lake-embankment, lake-beach, inside embankment, and hill type), socioeconomic status (per capita income), drinking water source (proportion of subjects using safe drinking water), sewage treatment (proportion of subjects using hygienic lavatories), prevalence of schistosomiasis japonica infection among livestock, density of snail infection, temperature, and precipitation. The continuous variables were categorized into subgroups. The cut-off points for continuous variables were based on the literature and the preliminary analysis of the variables collected in the current study (Table 1).

Because our data were hierarchical in nature, with individuals nested within the villages, we used a two-level random-intercept multi-level model for statistical analysis.

Table 1 Results of the univariate analysis of risk factors for schistosomiasis infection

Variable	No. of study subjects	No. of infected subjects	Infection rate (%)	<i>p</i> -Value ^a
Individual level				
Sex				
Male	5172	287	5.55	0.000
Female	4936	107	2.17	
Age (years)				
6–19	2106	32	1.52	0.000
20–39	2624	110	4.19	
40–59	3697	198	5.36	
≥60	1681	54	3.21	
Occupation				
Farmer	7448	326	4.38	0.000
Fisherman	103	23	22.33	
Other	2557	45	1.76	
Village level				
Endemic type				
Hill	6337	205	3.23	0.000
Lake-embankment	1989	121	6.08	
Lake-beach	859	65	7.57	
Inside embankment	923	3	0.33	
Per capita income				
<3000 Yuan	6424	302	4.70	0.000
≥3000 Yuan	3684	92	2.50	
Drinking water sources				
Safe water sources	8040	227	2.82	0.000
Infested water sources	2068	167	8.08	
Proportion using hygienic lavatory				
<50%	6574	277	4.21	0.025
≥50%	3534	117	3.31	
Infected rate of livestock				
<0.1%	3383	103	3.04	0.000
0.1–1%	4073	150	3.68	
≥1%	2652	141	5.32	
Density of infected snail				
<0.002/0.11 m ²	4553	116	2.55	0.000
≥0.002/0.11 m ²	5555	278	5.00	
Average temperature in June				
<27 °C	3140	69	2.20	0.000
≥27 °C	6968	325	4.66	
Average precipitation in April				
<100 mm	3496	110	3.15	0.005
≥100 mm	6612	284	4.30	
Average precipitation in September				
<50 mm	4963	167	3.36	0.007
≥50 mm	5145	227	4.41	
Average precipitation in October				
<100 mm	6240	176	2.82	0.000
≥100 mm	3868	218	5.64	

^a Chi-square test.

This method allows for variation in the baseline schistosomiasis rate across villages, while assuming that the effect of each variable in the model is the same within each village. Since the dependent variable was binary, we used the logistic form of the multi-level model in HLM version 6.04.¹²

We first conducted a series of univariate analyses to assess the association between each of the individual and ecological risk factors and schistosomiasis infection. We then built a null model, which included no individual level or village level risk factor. This null model served as a baseline for comparison with other models, from which we estimated the intra-class

correlation (ICC) in the data. The ICC represents the proportion of the total variance explained by the presence of clusters (villages) in the data. The correlation coefficients estimated from the null model were used to compute the expected schistosomiasis infection rate for a cluster (village). All variables that were significant ($p < 0.05$ in univariate analyses) were included as covariates in the full model, and a backward stepwise regression method ($p < 0.05$ for entry and $p > 0.1$ for removal) was used to identify the significant independent variables from the initial multi-level models. Finally, we used conventional logistic regression models, with the same dependent and independent variables used in the multi-level models, to estimate the adjusted odds ratios (and 95% confidence intervals) for schistosomiasis, and compared the results from the single-level logistic regression models with those from multi-level models. This study was approved by the Research Ethical Board of Hunan Institute of Parasitic Diseases.

Results

A total of 10 245 individuals were tested by IHA, of whom 2457 were positive. Among the 2457 IHA-positive subjects, 2320 were examined by the Kato–Katz thick test, from which 394 positives were diagnosed. Accounting for the loss of follow-up from IHA test to stool examination, the estimated overall prevalence was 4.1% ($(394/2320) \times (2457/10\ 245)$).

In the univariate analysis, both individual-level variables (specifically, sex, age, and occupation) and village-level variables (endemic type, per capita income, drinking water source, proportion using hygienic lavatory, proportion of infection of livestock, density of infected snail, average temperature in June, and average precipitation in April, September and October) were strongly associated with schistosomiasis infection (Table 1).

The null model demonstrated significant ($p < 0.01$) variation in the schistosomiasis japonica infection rate among villages; in particular, the ICC indicated that 52% of the total variance could be attributed to village level variables.

The results of multi-level model analysis showed that both individual-level variables (specifically, gender, age, and occupation) and village-level variables (type of schistosomiasis endemic area, drinking water source, sewage treatment, June temperature, and April rainfall) were associated with schistosomiasis infection (Table 2). Compared with the multi-level hierarchical model, conventional single-level regression analysis selected more independent variables and yielded narrower confidence intervals on the regression coefficients in the model, especially for village level variables.

Discussion

Multi-level modeling has been widely used in research in education, demography, psychology, sociology, and agriculture.^{13–15} To our knowledge, application of this analytical tool in studies of infectious disease has been limited. A few studies have been carried out on schistosomiasis infection in Africa,^{16–18} mostly used Bayesian analysis or Bayesian spatial analysis of schistosomiasis infection in school children, including one study on the relationships between anemia and parasitic infections (including schistosomiasis) in Kenyan school children.¹⁹ However, no study has compared the

results obtained from multi-level models with conventional logistic regression analysis, and no application of this analytical tool for *S. japonicum* research has been reported in China. Hierarchical data is commonly encountered in epidemiologic studies, such as in cross-sectional studies based on cluster sampling. Multi-level models provide a powerful tool for the analysis and evaluation of data with a hierarchical structure, because it allows for the simultaneous examination of the effects of cluster-level and individual-level covariates.²⁰ Inter-individual and inter-group variation can be examined, and the relative contributions of individual-level and cluster-level variables can be assessed.

Based on the null model used in the present study, prevalence of schistosomiasis japonica varied significantly among villages (with about 50% of the total variation arising from village-level variables in the null model), suggesting that characteristics of the village play an important role in schistosomiasis japonica infection. The fact that the risk factors for schistosomiasis japonica identified in our study were mainly at the village level underscores the need for application of multi-level models in the analysis of hierarchical data of the type considered here.

In the final multi-level model, individual level variables accounted for only 0.64% of the total variation. This was due mostly to unavailability of important individual level variables. A particularly important missing individual level variable was exposure to potentially contaminated water, which is the immediate cause of schistosomiasis infection. Another important missing individual level variable was treatment history, which can change the infection status of the individual. Of the variables available to us, village-level variables accounted for 89.1% of the total variation in schistosomiasis infection rates in the final model.

Conventional single-level logistic regression analysis selected more independent variables and had resulted in narrower confidence intervals than did the multi-level analysis. Per capita income, precipitation in October, and density of infected snails were statistically significant in the conventional logistic regression analysis, but were not significant in the multi-level analysis. For independent variables that were selected both by conventional logistic regression and multi-level modeling, the confidence intervals on the corresponding regression coefficients were all narrower using logistic regression. These discrepancies occurred because conventional logistic regression analysis ignores the hierarchical structure of the data, leading to overstatement of precision and inflated false-positive rates.²⁰

It is worth discussing the practical implications of our findings. The social, biological, and environmental risk factors of schistosomiasis infection identified in the present study are largely consistent with those identified in previous studies.^{21–23}

For example, the present and previous studies all showed that the risk of schistosomiasis japonica infection was much higher in lake and embankment areas as compared with hilly areas. The frequency of contacts with infected water by humans and livestock in the lake and embankment areas was much higher, so prevalence of schistosomiasis japonica was much higher in these areas as well. Other findings, such as being male, being 20–60 years old, and being a fisherman showed an increased risk of schistosomiasis japonica infection, and were also consistently observed in the present and previous studies.

Table 2 Comparison of results obtained from multi-level modeling and conventional single-level logistic regression of risk factors of schistosomiasis

Independent variable	Multi-level model		Conventional logistic regression	
	OR	95% CI	OR	95% CI
Individual level				
Sex				
Male	2.73	2.18, 3.41	2.72	2.16, 3.43
Female (reference)	1.0	-	1.0	-
Age				
6–19 (reference)	1.0	-	1.0	-
20–39	2.96	1.75, 5.03	2.87	1.91, 4.31
40–59	4.00	2.37, 6.75	3.85	2.63, 5.64
≥60	2.45	1.40, 4.27	2.39	1.53, 3.74
Occupation				
Farmer	0.94	0.61, 1.47	-	-
Fisherman	4.27	2.14, 8.50	4.82	2.77, 8.40
Other (reference)	1.0	-	1.0	-
Village level				
Endemic type				
Hill (reference)	1.0	-	1.0	-
Lake-embankment	8.64	1.83, 40.70	4.37	1.99, 9.56
Lake-beach	18.77	3.80, 92.64	16.55	5.96, 45.96
Inside embankment	6.32	1.07, 37.30	-	-
Per capita income				
<3000 Yuan (reference)			1.0	-
≥3000 Yuan			0.70	0.56, 0.88
Drinking water sources				
Safe water sources (reference)	1.0	-	1.0	-
Infested water sources	4.88	2.65, 8.99	1.77	1.16, 2.70
Proportion of using hygienic lavatory				
<50%	3.82	1.62, 9.02	4.06	2.42, 6.79
≥50% (reference)	1.0	-	1.0	-
Average precipitation in April				
<100 mm (reference)	1.0	-	1.0	-
≥100 mm	3.26	1.77, 6.01	2.16	1.74, 2.67
Average precipitation in October				
<100 mm (reference)			1.0	-
≥100 mm			2.49	1.88, 3.30
Average temperature in June				
<27 °C (reference)	1.0	-	1.0	-
≥27 °C	3.94	1.85, 8.40	6.03	4.05, 8.97
Density of infected snail				
<0.002/0.11 m ² (reference)			1.0	-
≥0.002/0.11 m ²			2.22	1.41, 3.49

OR, odds ratio; CI, confidence interval.

However, our multi-level analysis did not find a significant, independent effect of income and density of infected snails on schistosomiasis japonica infection. On the other hand, low income and high density of infected snails have been consistently found to be associated with schistosomiasis infection in previous studies that have ignored the hierarchical structure of the data, and have treated the data at a single level (by either lumping individuals into a single cluster or allocating cluster average values to individuals).^{24,25} These findings may have a bearing on the development of schistosomiasis japonica infection prevention programs, such as

whether massive application of molluscicides to control snails in endemic areas is an effective preventive measure.^{26,27} It is therefore important both practically and theoretically to explore the reasons for the differences in results obtained from multi-level and single-level analyses.

Acknowledgements

This study was funded by a grant from the surveillance sites project of schistosomiasis, Ministry of Public Health, P.R. China. Dr Wen is a recipient of the Ontario Women's Health

Council-Institute of Gender and Health of CIHR Mid Career Award and a recipient of an R&D Research Allowance from CIHR. Dr Krewski is the NSERC/SSHRC/McLaughlin Chair in Population Health Risk Assessment at the University of Ottawa. The authors thank the staff of the Hunan Institute of Parasitic Diseases who participated in the fieldwork and the residents of the 16 participating villages in Hunan, China for their support of the study.

Ethical approval: This study was approved by the Research Ethical Board of Hunan Institute of Parasitic Diseases.

Conflict of interest: No conflict of interest to declare.

References

- World Health Organization Expert Committee. Prevention and control of schistosomiasis and soil-transmitted helminthiasis: report of the WHO expert committee. WHO Technical Report Series No. 912. Geneva, Switzerland: World Health Organization; 2002, p. 1–57.
- Li YS, Sleight AC, Ross AG, Williams GM, Tanner M, McManus DP, et al. Epidemiology of *Schistosoma japonicum* in China: morbidity and strategies for control in the Dongting Lake region. *Int J Parasitol* 2000;30:273–81.
- Zheng J, Qian K, Zhu QH, Chen SH, Zhang R, Wang YA, et al. Effect of man's activity on the family aggregation of schistosomiasis patients. *Chin J Schisto Control* 1989;1:62–3.
- He N, Yuan HC, Zhang SJ, Jiang QW. A quantitative study on human water contact in a hilly rural endemic area of *Schistosoma japonicum*. *Chin J Schisto Control* 1996;8:1–5.
- Zhang W, Wong CM. Evaluation of the 1992–1999 World Bank schistosomiasis control project in China. *Acta Trop* 2003;85:303–13.
- Zhao ZQ, Li L, Zhao ZC. Preliminary study on the effects of atmospheric temperature and rainfall in July on detectable rate of acute schistosomiasis and prediction in Four-lake area of Hubei Province. *J Environment Health* 2000;17:160–2.
- Goldstein H. Multilevel mixed model analysis using iterative generalised least squares. *Biometrika* 1986;73:43–56.
- Aitkin M, Longford N. Statistical modelling in school effectiveness studies (with discussion). *J R Stat Soc Series A* 1986;149:1–43.
- Raudenbush SW, Bryk AS. A hierarchical model for studying school effects. *Sociol Educ* 1986;59:1–17.
- Smolkowski K, Biglan A, Dent C, Seeley J. The multilevel structure of four adolescent problems. *Prev Sci* 2006;7:239–56.
- Goff WL, Ronald NC. Miracidia hatching technique for diagnosis of canine schistosomiasis. *J Am Vet Med Assoc* 1980;177:699–700.
- Raudenbush SW, Bryk AS, Cheong Y, Congdon RT. *HLM6: Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International; 2004.
- Afshartous D, de Leeuw J. An application of multilevel model prediction to NELS:88. *Behaviormetrika* 2004;31:43–66.
- Poelmans S, Sahibzada K. A multi-level model for studying the context and impact of work–family policies and culture in organizations. *Human Resource Manage Rev* 2004;14:409–31.
- Hoshino S. Multilevel modeling on farmland distribution in Japan. *Land Use Policy* 2001;18:75–90.
- Beck-Worner C, Raso G, Vounatsou P, N'Goran EK, Rigo G, Parlow E, et al. Bayesian spatial risk prediction of *Schistosoma mansoni* infection in western Cote d'Ivoire with a remotely-sensed digital elevation model. *Am J Trop Med Hyg* 2007;76:956–63.
- Rudge JW, Stothard JR, Basanez MG, Mgeni AF, Khamis IS, Khamis AN, et al. Micro-epidemiology of urinary schistosomiasis in Zanzibar: local risk factors associated with distribution of infections among schoolchildren and relevance for control. *Acta Trop* 2008;105:45–54.
- Clements AC, Brooker S, Nyandindi U, Fenwick A, Blair L. Bayesian spatial analysis of a national urinary schistosomiasis questionnaire to assist geographic targeting of schistosomiasis control in Tanzania. *East Africa Int J Parasitol* 2008;38:401–15.
- Koukounari A, Estambale BB, Njagi JK, Cundill B, Ajanga A, Crudder C, et al. Relationships between anaemia and parasitic infections in Kenyan schoolchildren: a Bayesian hierarchical modeling approach. *Int J Parasitol* 2008;38:1663–71.
- Diez Roux AV. A glossary for multilevel analysis. *J Epidemiol Community Health* 2002;56:588–94.
- Huang YX, Manderson L. Socioeconomic factors and prevalence of schistosomiasis japonica in a rural area of China. *Chin J Schisto Control* 1999;11:137–42.
- Cao Q, Tan XD, Kong QL, Xu XJ. Reforming latrine for schistosomiasis control in Hubei. *J Public Health Prev Med* 2006;6:28–32.
- Li BG, Guo XG, Su GC, Li WS, Liu DQ, Wu X. Relationship between number and quality of private lavatory and social economic situation in schistosomiasis endemic areas. *J Pract Parasit Dis* 1998;6:185–6.
- Zheng J, Wang XF, Wang YA, Hua ZH, Zhu HG, Wu X, et al. The relationship between socioeconomic factors and schistosomiasis in mountainous region in Yunnan. *Chin J Pract Parasit Dis* 1996;14:106–9.
- Li BG, Guo XG, Li WS. Impact of social-economic level to schistosomiasis endemicity and control program in Heqing County. *J Pract Parasit Dis* 1998;6:85–6.
- Jiang QW, Wang LY, Guo JG, Chen MG, Zhou XN, Dirk E. Morbidity control of schistosomiasis in China. *Acta Trop* 2002;82:115–25.
- Lin DD, Hu GH, Zhang SJ. Optimal combined approaches of field intervention for schistosomiasis control in China. *Acta Trop* 2005;96:242–7.