GENETIC AND HOUSEHOLD RISK FACTORS FOR SCHISTOSOMA JAPONICUM INFECTION IN THE PRESENCE OF LARGER SCALE ENVIRONMENTAL DIFFERENCES IN THE MOUNTAINOUS TRANSMISSION AREAS OF CHINA

EDMUND Y. W. SETO,* BO ZHONG, JOHN KOUCH, ALAN HUBBARD, AND ROBERT C. SPEAR
School of Public Health, University of California, Berkeley, California; Institute of Parasitic Disease, Sichuan Center for Disease Control and Prevention, Chengdu, Sichuan, People's Republic of China; University of Connecticut Medical School, Farmington, Connecticut

Abstract. Schistosoma japonicum egg excretion and kinship relationship data from 13 endemic villages in the mountainous transmission area near Xichang, in Sichuan province, China, were analyzed via a variance components methodology to assess the relative contribution of kinship, shared household, and shared village to the risk of infection. Large intervillage differences in egg counts exist in this region due to differences in transmission potential related to environmental differences in snail density and agricultural practices. After accounting for these intervillage differences, there was no kinship or household effect on egg excretion. This reinforces earlier findings that suggest environmental factors dominate risk in this region.

INTRODUCTION

Schistosomiasis remains a serious problem in many mountainous villages of Sichuan province. Recent studies have found considerable variability in Schistosoma japonicum infection among villages near Xichang city. In some villages, the prevalence of infection was over 70%, while in villages only a few kilometers away, the prevalence was less than 10%. These differences in prevalence were attributed to differences in environmental risk factors, including snail density and agricultural practices. However, even within a village, where residents share environmental risk, there is great variability in infection intensities among individuals. Based on schistosome egg excretion data, a few individuals harbor the majority of worms. A cost-effective control strategy depends upon the identification of these wormy individuals, however this can be difficult. In this paper, we consider the role of heritable susceptibility as one potential risk factor that may explain why certain individuals are wormier than others.

There is evidence in support of genetic susceptibility from schistosomiasis studies from around the world. Segregation and gene mapping studies conducted in Brazil and Sudan have found two distinct genetic loci related to human susceptibility to Schistosoma mansoni infection and disease progression. The first major gene related to infection levels, SM1, was found from a study in Caatinga, Brazil, and was mapped to a region in the genome, 5q31-q33. In a Sudanese study, a second gene SM2, distinct from SM1, was found to be related to disease progression (fibrosis and portal hypertension). In China, two major genes have been found to be related to the severity of liver fibrosis. In a study of 230 schistosomiasis japonica patients, the frequency of an HLA class II gene and polymorphism in the interleukin (IL)-13 gene were elevated in fibrotic patients (OR = 24.5). The odds-ratio for the presence of both genes is higher than that for the presence of either gene alone, suggesting that the genes act synergistically to enhance the development of fibrosis.

Despite the identification of genes related to infection intensity and disease, recent studies from Brazil and Kenya have reported different estimates of the effect of heritable susceptibility (kinship or additive genetic factors) on schistosomiasis infection. A study from Brazil looking at detailed genealogic data from 597 individuals belonging to 6 pedigrees living in 145 households found that 21–43% of variability in S. mansoni egg counts were attributed to additive genetic effects and 12–28% were attributed to the effects of a shared household. However, using similar methods, a study from Kenya looking at 4,408 individuals belonging to 912 households and 241 pedigrees found that only 9–20% and 7–10% of infection intensity of Schistosoma haematobium was attributed to kinship and shared household, respectively. A literature search of recent English and Chinese language journals found a Chinese review of the methodology for kinship analysis and a discussion of the findings from Brazil, but no articles that describe the impact of kinship on S. japonicum infection. This paper represents the first estimates of the kinship effect for S. japonicum.

MATERIALS AND METHODS

Location. The study was conducted in 13 endemic villages surrounding Ongohai Lake near Xichang city in Sichuan province, China. The villages belong to four townships (Chuanxing, Daxing, Gaoqian, and Hainan) spanning a 12 × 12 km area in an elevated valley of the Daliang Mountains (elevation 1,500 m). The ethnic composition of our study villages is primarily Han Chinese, with only 2% belonging to the Yi ethnic group, who have their own language and generally live in separate villages from the Han. The agriculture of the region consists primarily of rice, tobacco, corn, and various cash crops during the spring-summer season, and wheat, bean, and garlic, during the fall-winter season. The intermediate host for schistosomiasis transmission is the Oncomelania hupensis robertsoni snail that lives in the irrigation ditches and along the terraced walls lining the fields surrounding each village’s cluster of residences. The daily agricultural activities of adult villagers (and during the summer months, for children, swimming) places them at risk for exposure to water contaminated with infective cercariae. The transmission process is exacerbated by the inadequate treatment of infective stool, and the use of this stool as fertilizer in the fields. There are few buffalo or cattle in this region, and in many villages,
transmission is perpetuated by human activity and behavior alone. The area has been intensively studied as part of two U.S.-funded projects to identify spatial risk factors responsible for transmission and to model and simulate the transmission process and the efficacy of different control strategies, as well as by Chinese researchers, trying to identify effective control strategies for hyperendemic areas.

**Study population.** The 13 villages are a subset of 20 villages, for which we have previously documented environmental and epidemiologic data. The study targeted all residents aged 4–60, but was open to villagers of any age. In each village, we collected data on infection, demography, water contact, and used global positioning systems (GPS) to map the spatial distributions of snail density and cercarial risk as measured by mouse bioassays. Mouse bioassays are routinely used in China and are preferred over filter-based cercariometry methods due to the stickiness of *S. japonicum* cercariae and the bioassay’s ability to measure infective cercariae. Each mouse bioassay consists of five mice in a cage, exposed to surface water 2 hours each day for 5 days, for a cumulative exposure time of 10 hours. Six to 16 cages per village were surveyed in late July for the 13 villages. After a parasite maturation period of 35 days the number of worms per mouse was determined from dissection. Human data were collected with informed consent, using protocols approved by the Institutional Review Boards of the University of California, Berkeley, the National Institutes of Health, the Sichuan Institute of Parasitic Diseases, and the University of Connecticut Medical School.

**Kinship.** In the demographic survey every individual in the study was assigned a unique personal identifier and a household ID that uniquely identifies the house in which they lived. To establish kinship relationships within each village, we asked the village leader to construct the pedigrees for each individual in the village, by naming the father and mother for each person in the village, and the parents of those parents as far back as could be recalled. In total, the pedigree dataset consisted of 3,389 individuals, 1,742 male and 1,647 female, belonging to 411 pedigrees and 773 households within the 13 villages, with the largest pedigree consisting of 202 individuals.

**Infection intensity.** *S. japonicum* infection was determined via stool samples and examination at the end of the transmission season in the year 2000. Prior to this exam, little to no organized control had been done in these villages other than routine treatment of those who visited the control station on their own due to symptoms of schistosomiasis. The exam followed Ministry of Health protocols, which call for a miracidial hatch test which involves the collection of at least 30 g from each person on three separate days, and incubation to visually inspect for hatched miracidia. In addition participants submitted an entire day’s stool for Katz-Katz egg counts, as recommended by the WHO Steering Committee for Schistosomiasis. Three 41.5 mg slides were examined from homogenized stool samples. Egg per gram of stool (EPG) data were collected for 2,287 individuals in the 13 villages. All EPGs were transformed to square root of EPG, to normalize the skewed data.

**Water contact.** A 25% subsample of the study participants was targeted for a water contact questionnaire that ascertained the frequency, duration, and location of nine common water contact activities by month (from April to October). Activities include washing clothes or vegetables, washing agricultural tools, washing hands and feet, playing or swimming in water, ditch operation or maintenance, ploughing, rice cutting, and fishing. Reported frequencies and durations were subsequently scaled to units of surface area minutes of contact, which adjust for the portion of the body typically covered for each activity. These scaled measures were then summed across all activities and all months to achieve a single water contact metric per individual. Because respondents were surveyed at the end of the transmission season, the data are subject to recall errors. In the 2,287 individuals with egg count data, 710 (31%) had responded to the water questionnaire, which was greater than our original goal of a 25% subsample. Of the subsequent kinship analyses described below that include water contact as a covariate, water contacts for individuals who did not participate in the survey were treated as missing data.

**Statistical methods.** The underlying statistical methods used here follow principles of genetic relatedness. If there exists a heritable component to schistosomiasis susceptibility, then those who are closely related should have similar egg counts. Moreover, these similarities should be evident even after adjusting for the confounding effects of shared household and village, and other potential covariates such as age, sex, education, and water contact. The degree to which individuals are closely related can be measured by the kinship coefficient. The kinship coefficient describes the probability that two individuals share an identical allele by common descent. For any two individuals, a kinship coefficient that is greater than zero implies that the two individuals share a common ancestor. The coefficient can be computed based on each person’s position in the pedigree. The construction of pedigrees and computation of kinship coefficients were performed using the PEDSYS program. PEDSYS uses the Stevens-Boyce algorithm for computing the Kinship coefficients. The algorithm is based on the probability that an individual’s allele came from a particular parent is 0.5. And the probability that that parent got the allele from a specific parent of his/hers is 0.5. Hence the probabilities multiply with each generation. The algorithm consists of searching the pedigree and finding the common ancestor linking two individuals, counting the number of individuals in that linkage to compute the kinship probability for that particular ancestor, and summing the probabilities across all common ancestors (if there are more than one):

\[
f = \sum_{i=1}^{l} \left[ \frac{1}{2} \right] n_i \left( 1 + f_a \right)
\]

where \( f \) is the inbreeding coefficient of an individual, \( i \) is the number of common ancestors found, \( n_i \) is the number of individuals linking the two persons through the common ancestor, and the \( (1 + f_a) \) term adjusts for the potential that the common ancestor is inbred.

A variance component analysis was conducted using the Sequential Oligogenic Linkage Analysis Routines (SOLAR) program to estimate the variance in egg counts due to kinship. Specifically, SOLAR produces maximum likelihood estimates for a random effects regression model of square root of EPG in which the random effects are characterized by a covariance matrix:

\[
\text{Cov}(X_i, X_j) = 2 \sigma^2 \phi_{ij} + \sigma^2 \Delta_{ij} + \sigma^2 \delta_{ij}
\]
where $\sigma_k^2$ is the variance due to the kinship effect as measured by a matrix of kinship coefficients, $\psi_{ij} \sigma_k^2$ is the variance due to shared household, where the elements of the $\Delta_k$ matrix are 1 if $i$ and $j$ share a residence, and 0 otherwise; and $\sigma_e^2$ is the residual variance of the model ($\delta_0$ is the identity matrix).\(^{15,16}\) A $\chi^2$ test of log-likelihoods was used to evaluate various models that included different combinations of kinship, shared household, and covariates (fixed effects) for village, age, sex, education, and water contact.

In the original presentation of the above method, Lange and others\(^{16}\) point out the above variance components method solved by maximum likelihood is sensitive to departures from normality of the trait. They suggest the use of transformations of the trait to provide more normally distributed data, which was done in our study, as well as those by Bethony and others\(^8\) and King and others.\(^9\) However, despite square root transformations, egg count data remained strongly overdispersed. Hence, to verify the results of the analysis of variance method, we used a second method based on negative binomial regression, which more directly addresses the natural aggregation of egg counts within the population. The negative binomial regression was implemented in STATA.\(^{17}\) and modeled individual EPG as a function of the average EPG of their relatives at varying magnitudes of kinship coefficients, accounting for clustering on shared household, and fixed effects for village membership.

**RESULTS**

Figure 1 shows the location of the 13 villages in our study, shaded according to prevalence, and labeled with average village EPG, snail density, average worms per mouse from the cercarial bioassay, and proportion of fields devoted to growing rice. The figure illustrates the regional differences in risk within this endemic area. Figure 2 shows the correlations that exist between environmental factors that influence infection risk. Spearman rank correlation between village prevalence and average EPG was 0.88 ($P = 0.0001$); between prevalence and snail density was 0.48 ($P = 0.1$); between prevalence and average worms per mouse was 0.91 ($P < 0.0001$); and between prevalence and proportion of fields devoted to rice was −0.80 ($P = 0.001$). The correlation between prevalence and snail density was 0.68 ($P = 0.01$) with one outlying village removed.

The distribution of log transformed EPG amongst individuals for one of the 13 villages (Daxing Xinming, Production Group 3) is shown in Figure 3 by age and sex. Peak infection occurs in mid-ages, 20–29 for males, and later 40–49 for females. Although there are slight differences in the distributions for the remaining villages, the general trend of higher EPGs for middle-aged villagers is a consistent finding.

The average amount of self-reported water contact across all 13 villages was 255 m-s-minutes over the infection season. Figure 4 shows how water contact differs with gender and changes with age. Generally, young children, particularly boys have the highest amount of water contact, which is due to playing in water during the summer months. After young childhood, water contact generally increases with age, with typically slightly higher water contact for men. Although there are slight differences in the distributions for one of the 13 villages (Daxing Xinming, Production Group 3) is shown in Figure 3 by age and sex. Peak infection occurs in mid-ages, 20–29 for males, and later 40–49 for females. Although there are slight differences in the distributions for the remaining villages, the general trend of higher EPGs for middle-aged villagers is a consistent finding.

Figure 2. Relationships between human infection prevalence and intensity (a), snail density (b), average worms per mouse from the mouse bioassay (c), and proportion of agriculture that is devoted to rice (d).
genic model with only a kinship effect, a village model with only a shared village effect, and saturated models that included kinship with either household or village effects. Kinship, household, and village effect models all explained significant proportions of the variance in egg counts (kinship 29%, household 16%, and village 19% of variance, and \( P < 0.0001 \), \( P = 0.007 \), and \( P < 0.00001 \), respectively). The saturated model with a household effect suggested that 19% of egg variability was attributed to kinship after accounting for shared household. However, the saturated model with village effect suggested that there was no effect of kinship after accounting for shared village.

Table 1 also shows the results of four models adjusted for covariates (age, sex, education, and water contact). Due to the importance of village in the unadjusted model, we incorporated the village effect by using indicator variables for village membership as fixed covariates to the adjusted models. The small household effect present in the saturated model that included all covariates, village, household, and kinship effects was not significant (\( P > 0.5 \)). Hence, after including indicators for village there was little to no attributable variance in egg counts due to kinship or household.

An issue of concern in the above variance components analyses lies in the inherent assumption of normally distributed trait data. Previous studies of egg count data have dismissed this concern after log transformations of the egg counts. However, with our study data, strong over-dispersion even after square root or log transformation of the egg counts prompted a second negative binomial regression analysis. The results of the negative binomial model with only average EPG of close kin as a predictor of individual EPG confirmed the presence of a statistically significant (\( P < 0.05 \)) correlation between individual EPG and average EPG of closely related kin (\( \Phi_{ij} \geq 0.125 \)). A comparison of the negative binomial model to a Poisson model, also confirmed that the negative binomial was more appropriate for the overdispersed data. As before with the variance components analyses, after adjusting for water contact, clustering on household, and village fixed effects, the EPG of closely related kin was no longer a significant predictor of individual EPG. Interestingly though, though not significant, the coefficients for the effect of the EPG of closely related kin on individual EPG was negative, suggesting the possibility of a protective kinship effect. This protective effect is discussed below.

### DISCUSSION

Our methods and findings differ from those used by Bethony and others and King and others in *S. mansoni* in Brazil and *S. hematobium* in Kenya. This is due in large part to the uniqueness of *S. japonicum* and our study area in China. Similar to their studies, we used variance components models on egg count data. This analysis resulted in estimates of the kinship effect (19–29%) that lie in the middle of those reported for Brazil (21–43%) and Kenya (9–20%). However, after accounting for intervillage differences in environmental risk, which we knew *a priori* were large for our study area, the effect of kinship disappeared. We verified this finding by fitting variance component models to villages individually, as well as via an alternative negative binomial regression approach which also resulted in no kinship effect.

Previous studies correctly argue for the need to compensate for shared household, which can confound the relationship between kinship and infection intensity. Individuals who live together are typically closely related and share risk (common domestic water contact behaviors: frequency, duration, and locations of contact activities). However, the notion of common environmental risk extends beyond the household, and operates at many different social and geographic scales. Accounting for shared household insufficiently accounts for correlation at the village level. In China when a son marries, he
is given land within the village and a house of his own. Hence, within a village individuals can be closely related but still live in different households. This social behavior facilitates the collection of multi-household pedigree data. As individuals marry and move into separate households, they will often still share environmental risk factors that operate at the village scale, such as the surrounding snail density and the risky behaviors of other neighboring households (what they grow, how they treat their waste, and fertilize their crops).

The strength of village-level risk factors (Figure 2) and absence of more focal ways of describing risk such as via kinship and shared household effects creates to some extent more of a problem than it solves. From a control perspective, there is great need for a clear set of risk factors by which to screen individuals for potential infection that are less expensive and time-consuming than parasitological procedures. However, when considering individual risk factors we find that there is no clear difference in egg counts between males and females. Also, the risk of infection spans a wide range of ages. There is a tendency to infer the presence of age-acquired immunity to schistosomiasis based on lower infections with age, which other researchers have documented in other regions of China. However, based on our data, to the extent that it exists in the mountains in Sichuan, the effect of age-acquired immunity is quite mild. There is not as dramatic a drop in infection with age as that seen in other studies. Despite high water contact, egg counts are low for children younger than 10 years of age. Increasing water contact with age corresponds to gradual increases egg counts, with only a fall off of egg counts in older, 50+ aged individuals.

Based on the transmission cycle, water contact should be an important risk factor for infection. However, other studies have had difficulty in accurately measuring water contact. Methods involving self-reporting are subject to recall errors. Direct observation can affect the behavior of those being observed. Bethony and others included water contact as a covariate in their model, however it was not significant. The plot of water contact versus EPG for one of our villages (Figure 5) suggests that without any adjustment, water contact is a poor indicator of infection risk. However, in the final adjusted model, water contact is a small, but significant predictor of egg counts ($\beta_{\text{water}} = 3.64 \times 10^{-4}$, $P = 0.03$). It was also a significant covariate in the negative binomial regression models we fit. One reason for this may be the importance of better characterizing cercarial exposure as opposed to water contact alone. Not all water contact leads to infection, however infection is more likely to occur where there are higher concentrations of cercariae. Hence, spatial heterogeneities in cercarial concentrations are important. In our model, we account for differences in cercarial concentrations via adjustment at the household and village levels. Currently, the lack of good cercarial detection methods limits our ability to characterize the spatial distribution of risk at a more local scale, and is a future research need.

While the results of negative binomial regression generally matched the findings of the analysis of variance method, an important difference is that the regression technique provided coefficients that assessed the direction and magnitude of the kinship and water contact effects. The negative coefficient for the average EPG of close relatives suggests that having a closely related family member who is heavily infected may lower an individual’s risk of infection. In fact, in villages with higher infection, as an individual’s EPG increases the water contact scores of their most closely related kin generally decreases. It perhaps suggests complex social behaviors where either most of the burden of water contact activities are handled by particular family members or an avoidance of water contact once a relative becomes infected with schistosomiasis.

Although no organized control programs existed at the time of our survey in 2000, testing and treatment is generally available in this region of China for those individuals who seek it. Presumably this would lead to lower infection intensities in those individuals most recently treated. However, treatment history data, which were collected for our study participants suggest otherwise. On average there were considerably higher EPGs among those most recently treated, suggesting that reinfection rates are high. Additional analyses on those individuals with no treatment over the past two years found similar results for kinship effect after adjustment for shared village.

It is important to note that estimates of heritable effects are largely dependent upon the population and overall environment under study. This is the first study of kinship effects for S. japonicum, however, the findings may not extrapolate to all of China. There are clear regional differences in intermediate hosts, their ecologies, and modes of transmissions in China. In fact, snails from one province are refractory to infection of schistosomes of other provinces. Thus, different heritable effects may be found throughout China, and for schistosomes in other countries.

The implications of our study on transmission control relate

### Table 1

| Kinship | Household | Village | Residual | Likelihood
|---------|-----------|---------|----------|-------------
| Unadjusted | | | | |
| Sporadic | | | | |
| Household | 0.29 ± 0.04 | 0.16 ± 0.02 | 1.00 | −2446.35
| Polygenic | | | 0.84 ± 0.02 | −2408.10
| Village | 0.19 ± 0.06 | 0.19 ± 0.06 | 0.81 ± 0.06 | −2243.14
| Saturated (H) | 0.00 | 0.19 ± 0.05 | 0.73 ± 0.04 | −2401.00
| Saturated (V) | 0.00 | 0.07 ± 0.03 | 0.81 ± 0.06 | −2243.14
| Adjusted (covariates for age, sex, water, education, and village) | | | |
| Sporadic | | | | |
| Household | 0.00 | 0.00 | 1.00 | −730.76
| Polygenic | 0.00 | | 1.00 | −730.76
| Saturated | 0.00 | 0.01 ± 0.05 | 0.99 ± 0.05 | −730.93
to the importance of village-level environmental risk factors. Although finding suitable individual-level risk factors remains a challenge, the village level is still a good scale in which to organize control. Based on our prior work, existing geographic information systems and remote sensing methods are available for creating village-level maps of the potential for transmission. These maps identify environmental risk factors, such as areas of snail habitat and high risk agriculture. In China there is a tendency to fund short-term focused and intense control strategies, such as chemotherapy and molluscidic programs, which initially have very good results in certain places. However disease often reemerges in controlled locations after only a few years. Because GIS/RS maps have the ability to document the inherent potential for transmission irrespective of historical control, they can be a useful tool in informing ongoing surveillance and control.

Received March 3, 2005. Accepted for publication July 7, 2005.

Acknowledgments: The authors thank Dr. Jeff Bethony for his initial guidance and encouragement at the outset of the study. The PEDSYS program used in this study was supported by the National Institutes of Health. Both PEDSYS and SOLAR may be downloaded from the Southwest Foundation for Biomedical Research Web site: http://www.sfbr.org/pedsys/pedsys.html and http://www.sfbr.org/solar/index.html. This study was funded in part by grants from the Medical Dean’s Award of the University of Connecticut Medical School and National Institutes of Health (ROI AI-50612).

Authors’ addresses: Edmund Y. W. Seto, School of Public Health, 140 Warren Hall, University of California, Berkeley, CA 94720, Telephone: 510-642-9219, Fax: 510-642-5815, E-mail: seto@berkeley.edu. Bohu Zhong, Institute of Parasitic Disease, Sichuan Center for Disease Control and Prevention, 10 University Road, Chengdu, Sichuan, 610041, People’s Republic of China, E-mail: zhongbo1968@163.com. John Kouch, University of Connecticut Medical School, Medical Student Affairs Office, AG062, 263 Farmington Avenue, Farmington, CT 06030, E-mail: JKouch@mph.uche.edu. Alan Hubbard, School of Public Health, 140 Warren Hall, University of California, Berkeley, CA 94720, Telephone: 510-642-8365, Fax: 510-642-5815, E-mail: hubbard@stat.berkeley.edu. Robert C. Spear, School of Public Health, 140 Warren Hall, University of California, Berkeley, CA 94720, Telephone: 510-642-0761, Fax: 510-642-5815, E-mail: spear@berkeley.edu.

Reprint requests: Edmund Y. W. Seto, School of Public Health, 140 Warren Hall, University of California, Berkeley, CA 94720, Telephone: 510-642-9219, Fax: 510-642-5815, E-mail: seto@berkeley.edu.

REFERENCES


14. Southwest Foundation for Biomedical Research, 1999. PEDSYS v2.03. San Antonio, TX: Southwest Foundation for Biomedical Research.


