### Title

Exploring the Contribution of Host Susceptibility to Epidemiological Patterns of Schistosoma japonicum Infection Using an Individual-Based Model




### Journal

American Journal of Tropical Medicine and Hygiene, 92(6)

### ISSN

0002-9637



### Authors

Wang, Shuo
Spear, Robert C



### Publication Date

2015-06-03

### DOI

10.4269/ajtmh.14-0691

Peer reviewed

*Am. J. Trop. Med. Hyg.*, 92(6), 2015, pp. 1245–1252
doi:10.4269/ajtmh.14-0691


# Exploring the Contribution of Host Susceptibility to Epidemiological Patterns of *Schistosoma japonicum* Infection Using an Individual-Based Model

Shuo Wang and Robert C. Spear*
*Center for Occupational and Environmental Health, School of Public Health, University of California, Berkeley, California*

*Abstract.* We recently reported the analysis of epidemiological data suggesting variability in individual susceptibility to infection by *Schistosoma japonicum* among rural villagers who reside in Sichuan Province of southwestern China. By supplementing the data used in the earlier analysis from other studies we have reported from this region, we presented improved estimates of cercarial exposure, which in turn, result in stronger evidence of susceptibility. This analysis was conducted using an individual-based mathematical model (IBM) whose use was motivated by the nature and extent of field data from the low-transmission environments exemplified by one of our datasets and typical of the current situation in most endemic areas of China. In addition to individual susceptibility and water contact, the model includes stochastic aspects of cercarial exposure as well as of diagnostic procedures, the latter being particularly relevant to the low-transmission environment. The simulation studies show that, to produce key aspects of the epidemiological findings, the distribution of susceptibility ranges over several orders of magnitude and is highly right skewed. We found no compelling evidence that the distribution of susceptibility differed between the two populations that underlie both the epidemiological and simulation results.

## INTRODUCTION

Almost 20 years ago, Woolhouse and others[1] published an influential paper exploring the validity and implications of the "20/80 rule" in the context of infectious disease control programs. The rule implies that 20% of the exposed population is responsible for 80% of the disease transmission potential for a variety of diseases including vector-borne parasites. These authors concluded that "heterogeneities in contact rates lead to consistent and substantial increases" in the basic reproductive number and that this heterogeneity is "likely to be an important determinant of the epidemiology of vector-borne diseases ...." Clearly, a quantitative understanding of the determinants of "contact rate" in particular situations is also essential in identifying and targeting the 20% for surveillance or intervention. Here, we explore the determinants of contact rate using epidemiological data on *Schistosoma japonicum* infections among rural villagers in Sichuan Province of southwestern China.

Woolhouse and others[1] addressed the heterogeneity issue using mathematical models of disease transmission. Recently, Civitello and Rohr[2] further pursued aspects of heterogeneity in transmission and noted that classic models of parasitic disease transmission generally represent transmission (or contact) rates by single parameters that lump the effects of both exposure and susceptibility, the latter being the risk or intensity of infection for a given exposure. They use laboratory data on snail infections from *Schistosoma mansoni* miricidia to explore the interaction of these two factors via modeling studies and show that models that account for exposure and susceptibility separately best predicted infection prevalence across their dataset. Following the exposure/susceptibility distinction, we previously reported a statistical analysis of two longitudinal datasets on human infection and reinfection by the waterborne helminth *S. japonicum* among rural villagers who reside in hilly and mountainous agricultural settings.[3] That analysis suggested substantial variability in susceptibility to infection among individuals comprising these populations; evidence for which has been summarized by Quinnell.[4] If this finding can be confirmed, and if susceptibility were to be a stable or slowly varying property of an individual, the practical implications may be particularly important in moving from low-risk environments to the termination of transmission.

A weakness in our earlier analysis arose from the indirect methods used to estimate exposure. Here we use the same datasets used in the earlier analysis, but supplemented with other data that allow a more direct quantification of exposure and the estimation of infection intensity to better estimate the role of susceptibility. This analysis was conducted using an individual-based mathematical model (IBM) whose structure and mechanistic assumptions are consistent with the earlier work at the community level by our group and others. The use of an IBM for this purpose was motivated by the nature and extent of field data from the low-transmission environments exemplified by one of our datasets and typical of the current situation in most endemic areas of China.

The IBM evolved from our use of differential equation models of the form initially developed by MacDonald and later revised and popularized by Anderson and May and others.[5–15] These community-based models (CBMs) typically define mean worm burden in humans and infection prevalence in snails as state variables and yield outcomes at the population level, although some variants have introduced more specific subpopulations stratified by, for instance, occupation, age, or infection intensity. In contrast, IBMs track the actions of autonomous individuals and their interactions with the environment.[16] This modeling paradigm has two distinguishing features.[17] First, because the IBM tracks every individual in the population of interest, individual data are used and its heterogeneity represented to the greatest possible extent. Second, an IBM allows for simulation of each individual with their specific temporal and spatial information embedded; the latter being particularly crucial when interactions between humans and the environment are associated with the study question. The linkage back to results obtained with the CBM or to epidemiological summary statistics is that the collection of individual simulations allows the description of emergent behavior at the community

*Address correspondence to Robert C. Spear, 50 University Hall, Berkeley, CA 94720-7360. E-mail: spear@berkeley.edu

level. Hence, the IBM provides a means of investigating the origins of the patterns of behavior of the transmission process assumed and/or observed at the community level.[18]

## MATERIALS AND METHODS

Like the CBM, the IBM mimics each step of the life cycle of schistosome between the definitive and intermediate hosts.[19,20] Our current version of the CBM includes three state variables: the mean worm burden of the population, the mean infected snail density in the environment, and the mean level of acquired immunity in the human population. In the IBM, however, we have not incorporated acquired immunity because the levels of worm burden involved are below the threshold where these effects are likely to affect transmission rates as determined in earlier work. Details can be found elsewhere.[19–23]

As noted above and consistent with a great deal of epidemiological evidence, implicit in most CBMs is the assumption that the distribution of worm burden in the human population is highly right skewed with a small fraction of the population carrying the majority of the total burden in the community. Most commonly, this distribution is assumed to be well-represented by a negative binomial distribution. This is an assumption that has proven quite useful if infections are near endemic equilibrium levels but questionable under nonequilibrium conditions.[24] This assumption is not embedded in the IBM and we will subsequently be interested in observing if, and to what degree, this pattern emerges from the IBM simulations.

Worm development in an individual is tracked in the IBM and includes the following components: water contact, cercarial density in water to which the individual is exposed, individual susceptibility to worm development, and the fraction of parasites surviving to maturity in vivo. To complete the loop from worm burden in a village population through fecal egg excretion, miracidial hatching and transport to cercarial density produced by infected snail populations poses a significant challenge in the low-transmission environment. First, the detection limits of snail surveys and cercarial bioassays, central to model calibration in our earlier work, are no longer capable of identifying the spatial and temporal distributions of infected snails and the free-swimming forms of the schistosome within villages.[25] A further difficulty is that virtually all of our work with the CBM was focused on internal transmission among humans and snails within a village in contrast with the very ill-defined external sources of the free-swimming forms of the parasite that are likely to be a much more important contributors to village level infection risk in the low-transmission environment.[26] These factors caused us to treat cercarial density as an input variable to the IBM to use exposure-related field data available to us from the low-prevalence end of our earlier studies and focus on the effects of individual exposure modifiers, together with susceptibility and diagnostic sensitivity, as they condition host response to environments so defined.

**Epidemiological data and supplemental exposure data.** Our earlier statistical analysis, which resulted in evidence of variable host susceptibility to schistosome infection, was based on longitudinal data from two cohorts, Cohort 1 composed of 424 individuals from 10 villages monitored from 2000 to 2006 and Cohort 2 composed of 400 individuals from 27 villages monitored from 2007 to 2010.[3] Infection risk was generally higher in the population from which Cohort 1 was drawn than that of Cohort 2, the latter being typical of the low-transmission environment now common in most endemic regions of China. In both cohorts, infection was determined by a positive result using either the Kato-Katz thick smear procedure or a miracidial hatching test commonly used in China.

In this analysis, our first objective is to better describe the two components of individual cercarial exposure, frequency and intensity of water contact and cercarial density in that water. Water contact was assessed in both populations in earlier studies by sample surveys that allowed a common estimate of cumulative annual water contact for each individual in both cohorts based on the frequency and duration of various tasks as well as the task-specific amount of skin exposure. The survey design details are presented elsewhere.[27,28] Because we are not here interested in seasonality but in year-end infection intensity, each individual was assigned a constant time-weighted average water contact value over the annual infection season as detailed further in Supplemental Material Item 1 (SM1).

Cercarial density presents a greater challenge since the only directly relevant data available pertains to Cohort 1 and consists of mouse bioassay data from 5 of the 10 villages collected mainly over the 2001 infection season plus the consistent absence of any mouse infections in the studies of Cohort 2. The 2001 studies indicated substantial differences between villages in cercarial density, only modest within village changes over the infection season, and insensitivity to short-term weather fluctuations.[29] Hence, we chose to model cercarial density throughout the infection season as a stationary random process with a village-specific mean. The variance of the process was based on that of the mouse bioassay data adjusted for the duration of an exposure episode that we define as being composed of one or more 30-minute episodes.[30] The absence of any data on the possible autocorrelation of sequential measurements led us to assume that each 30-minute exposure episode was to an independently drawn cercarial density value. For example, if an individual was exposed while tending irrigation ditches for an hour and a half, three independent 30-minute cercarial density values would be drawn from the distribution pertaining to that village.

The total number of cercariae to which individual $i$ is exposed in 1 year, $C_i$, is estimated as:

$$C_i = \sum_{j=1}^{n_i} r_s s_i c_{jk} \Delta t = r_s s_i \Delta t \sum_{j=1}^{n_i} c_{jk} \quad (1)$$

where $n_j$ is the annual number of exposure intervals as defined in Supplemental Material Item 1 (SM1), $c_{jk}$ is the cercarial density in the $k$th village in number of cercariae per $m^2$ of surface water associated with the $i$th individual's $j$th water contact episode; $s_i$ is the skin exposure rate in $m^2$/min, and $\Delta t$ is the number of minutes per exposure episode. The parameter $r_s$ is the acquisition rate of cercariae, that is, the fraction of available cercariae that attach to and penetrate the skin in fraction/min. A value of one is used below in the absence of field data although we speculate its value to depend in part on water velocity. Here, in each simulation run, $s_i$ is treated as a constant for each individual since it is based on an estimate of their cumulative annual water contact. Hence, in this implementation, $c_{jk}$ is the only stochastic factor associated with each water contact episode. (Because cercariae are surface seeking, units of water surface area are used instead of water volume.)

In earlier simulations of transmission in the Cohort 1 villages using the CBM, mean worm burdens measured in 2000 and

2002 allowed the estimation of village annual average cercarial densities for those years.[26] These estimates were used in the IBM simulations for these villages and given in Supplemental Material Item 2 (SM2). In the Cohort 2 villages, the absence of infected snails and positive mouse bioassays made it necessary to rely on the Cohort 1 data as the basis for estimation of the cercarial densities in Cohort 2 villages. Specifically, it was found that mean worm burdens in humans less than 10 are generally associated with infected snail densities less than $0.15/m^2$. Hence, in villages where the mean worm burden is below 5 and no infected snails are observed, an upper bound on infected snail density is about $0.06/m^2$ that produces cercarial densities of about $8/m^2$.

An additional complication in the Cohort 2 villages was the presence of significant numbers of bovines, some of which were found to be infected. Hence, we chose to assign mean annual cercarial density estimates for those villages based on human prevalence and the relative number of infected bovines as detailed in Supplemental Material Item 3 (SM3).

**IBM implementation.** With the estimates of annual average cercarial density assigned, we now turn to the linkage of these data to year-end infection intensity in individuals as measured by the Kato-Katz procedure or the miracidial hatch test. This process has several elements that are inherently stochastic. We begin by noting that, given the number of cercarial hits experienced by an individual $C_i$, the annual number entering the systemic circulation is $J_i = \alpha_i C_i$. Note that the parameter describing the fraction of cercarial hits surviving to adult worms in vivo, $J_i$ is defined on an annual basis rather than per exposure episode and also that only an integer number of acquisitions is meaningful. We also import from the CBM a survival function to account for those parasites that die before maturity. This function is treated as the probability of "succeeding" in a Bernoulli trial, each trial corresponding to the successful development of one schistosome. The rationale is related to conducting simulations on the individual level in low-exposure environments: if multiplied by this fraction and having the decimal part rounded toward zero again, infection with one worm will then have no infection by definition, and infection with a few worms will have none or much lower chance of pairing and producing eggs. Thus, each of the $J_i$ schistosomes is simulated separately with stochasticity in its survival incorporated. Finally, the number of worms that individual $i$ has acquired by the end of each year is calculated by:

$$w_i = \sum_{J_i} B(1, e^{-\mu_w \tau_w}) \tag{2}$$

The last element of the IBM relates to infection testing. As both the miracidial hatch test and the Kato-Katz test depend on the presence of schistosome eggs, the first step is to determine the number of single and paired worms. To do so, two assumptions are made: 1) each worm is either female or male with equal probability, and 2) any two unpaired worms of the opposite sex will pair. In the simulation, accordingly, the sex of each worm is randomly assigned upon development and the number of worm pairs calculated. In Supplemental Material Item 4 (SM4), we have detailed the estimation of schistosome eggs per gram of stool sample (EPG) given the number of worm pairs and the probability of detecting eggs in a stool sample using either the Kato-Katz procedure or the miracidial hatch test.

There are two types of infection status outcome in the simulation for each individual: 1) the "true" status defined by the existence/absence of worms, which can be determined directly by counting the number of worms, and 2) the "observed" status as the result of infection testing, which includes a series of stochastic factors including the number of worm pairs, the EPG level on the day(s) when infection testing is conducted (simulated), and the detection probabilities of both methods.

**Individual susceptibility.** The CBM includes a population mean susceptibility parameter, denoted by $\alpha$, which represents the mean number of cercarial hits to the skin that survive to begin development in vivo to an adult worm. This parameter is poorly defined in the literature and was generally given a wide range of possible values in our earlier Monte Carlo simulation studies.[31] While we do not regard the available data as allowing definitive estimation of individual susceptibility, $\alpha_i$, it does appear feasible to gain insight into its variability across the population and determine at least its relative magnitude using the data summarized above.

Recall that the number of annual cercarial hits to an individual that begin development is $J_i = \alpha_i C_i$ and that, on average, one pair of worms can produce about 1.4 EPG per day. On the basis of Equation 1, the resulting EPG expected in a fecal sample, $E_i$, is approximately:

$$E_i \approx \alpha_i r_s s_i \Delta t \sum_{j=1}^{n_i} c_j \tag{3}$$

and, $\alpha_i$ is approximately:

$$\alpha_i \approx \frac{E_i}{r_s s_i \Delta t} \cdot \frac{1}{\sum_{j=1}^{n_i} c_j} \tag{4}$$

For subsets of individuals with complete data from both cohorts, it is possible to estimate $\alpha_i$ and gain some insight into its variability for this subset. For those individuals, the principle issue is to address the stochastic variabililty assigned to the cercarial density associated with each water contact interval. Monte Carlo simulation was used to take this variability into account based on individual water contact data and village of residence as described earlier.

Individual data collected from the two cohorts were used separately for the estimation of $\alpha_i$ and only individuals with both $E_i$ and $n_i$ greater than zero were included for obvious reasons. Uncertainty in the number of exposure episodes and in cercarial density was included in the algorithm used to estimate each individual's susceptibility parameter $\alpha_i$.

1. For each individual, estimation error in total water contact was assumed to vary between 0.8 and 1.2 times his/her original annual $n_i$ as derived from the survey data to yield the number of water contact intervals.
2. A random sample was drawn from the cercarial density associated with each interval from the negative binomial distribution estimated for the village of residence. To be consistent with the epidemiological data, each simulation was conducted on a 2-year basis for Cohort 1, with the first and the second years' $\bar{c}_v$ of each village selected from our previous analysis.[26] For Cohort 2, the simulation duration was 1 year, and $\bar{c}_v$ of each village is predetermined based on the values given in Supplemental Material Item 3 (SM3).

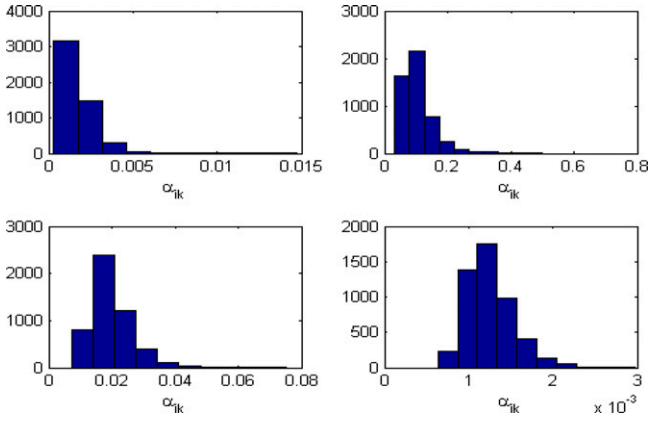


FIGURE 1. Distribution of 5,000 simulated $\alpha_{i,k}$ values for four individuals from Cohort 1.

The model is coded in MATLAB© (The MathWorks Inc., Natick, MA), and a total of 5,000 values of $\alpha_{i,k}$ were calculated for each individual in each cohort. Because these results are needed for the main simulations studies, preliminary results are presented at this point.

In Cohort 1, $\alpha_i$ was calculated for 108 out of 470 people; in Cohort 2, only 24 out of 608 people had EPG values greater than zero and were therefore available to be used for the simulation. Shown in Figure 1 are the histograms of the 5,000 runs for four individuals in Cohort 1. It can be seen that, mainly depending on the number $n_i$, the distribution of $\alpha_{i,k}$ for each individual has quite diverse patterns, varying from highly right skewed (top left) to nearly normal (bottom right). Furthermore, when $n_i$ is relatively small, the total number of cercarial hits will be zero in many simulations, thereby preventing estimation of $\alpha_{i,k}$. To account for all simulation results, for each individual, the median of the 5,000 $\alpha_{i,k}$ values was adopted as a point estimate of the individual $\alpha_i$.

The sample distribution function (SDF) for the 132 estimated median values of $\alpha_i$ are shown in Figure 2 for the combined cohorts (note the logarithmic scale of the $x$-axis). Compared with the rest of the people in Cohort 1, the mean water exposure minutes of the 108 people with measureable EPG was only 6% greater, and the difference was insignificant ($P = 0.54$). However, the 24 infected people in Cohort 2, on average, had only less than half of the exposure minutes of the other 584 people. Therefore, the 132 people included in the $\alpha_i$ calculation are convincingly the more susceptible group, and, considering the majority who had considerable exposure but no detectable EPG, the overall distribution of individual susceptibility appears to be highly right skewed. Hence, the SDF shown in Figure 2 assumes these values to constitute the upper 12% of the combined populations for which $\alpha_i$ values could be estimated as discussed in the following paragraphs.

In both cohorts the values of $\alpha_{i,k}$ range between about $10^{-3}$ and $10^{-1}$. Many fewer points were available from Cohort 2 and they tended to indicate somewhat higher values of $\alpha_i$ as would be expected due to the generally lower exposures experienced by infected individuals in that cohort. However, the uncertainties in cercarial density estimation between the two cohorts lead us to regard this difference as insufficient to suggest that the susceptibility estimates characteristic of the two cohorts are derived from different underlying distributions.

The final step is to specify the overall distributions of individual susceptibility that will be assumed to apply to the whole population for subsequent simulation purposes. Combining the two cohorts, the smallest estimated $\alpha_i$ of $2.98 \times 10^{-4}$ of those with measureable EPG corresponds to the 88th percentile of the population distribution. That, plus the evidence for skewness of the population distribution, results in hypothesizing a series of log-normal curves with different combinations of geometric mean (GM) and geometric standard deviation (GSD) to be fit to the data, each of which has the same 88th percentile. As the GSD is a parameter reflecting the variance of the distribution, three levels of GSD, respectively 1.5, 3.0, and 4.5, are used to represent low, medium, and high variance possibilities. The resulting cumulative density curves of the three log-normal distributions of $\alpha_i$ used in subsequent simulations are shown in Figure 3.

*Simulation studies.* As discussed in the previous paragraphs, the two cohorts showed considerable variability in annual water contact as well as in village-specific cercarial density.

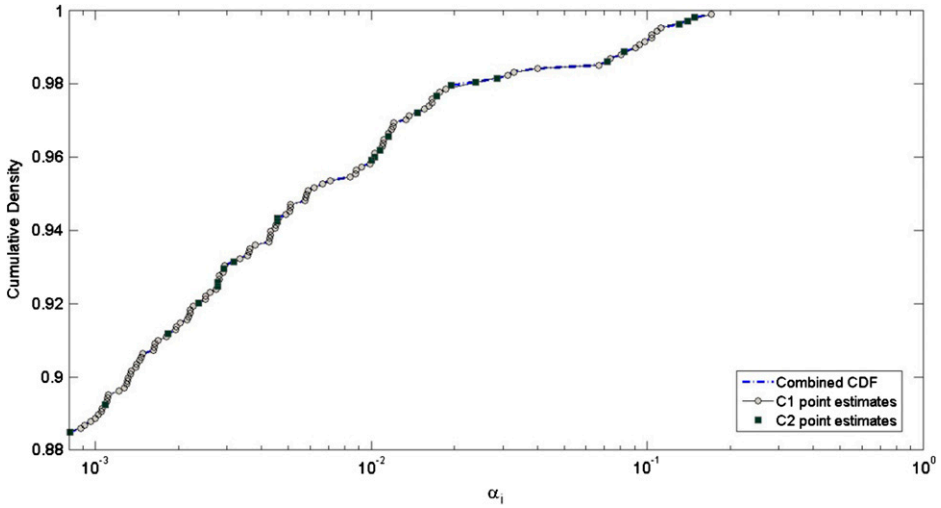


FIGURE 2. Estimated median values of $\alpha_i$ for 108 people in Cohort 1 (circles) and 24 people in Cohort 2 (squares).

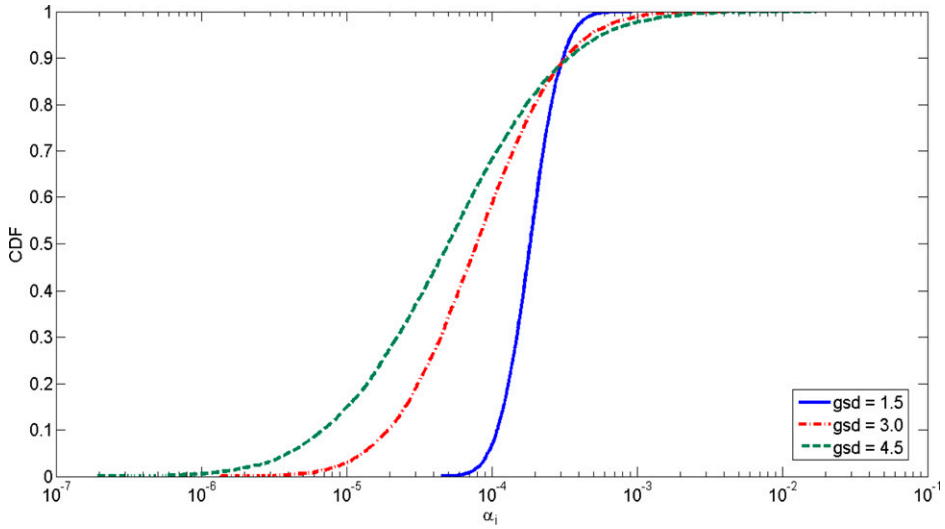


FIGURE 3. Distribution of $\alpha_i$ corresponding to GSD of 1.5, 3.0, and 4.5. GSD = geometric standard deviation.

These factors, together with the central question regarding the variability in individual susceptibility (here assumed site invariant and described by the same distribution for both cohorts) yield two transmission scenarios that we expect to result in medium- and low-mean infection intensities. But, is that a reasonable expectation? That is, given the residual uncertainty and variability in the individual parameters, is the low- or medium-transmission level a relatively stable emergent property at the community level or might different or mixed patterns of community response be observed? Moreover, do the distributions of individual infection intensity and the reinfection characteristics in these communities mimic the qualitative patterns seen in the epidemiological data? To investigate these issues, two hypothetical populations, denoted as HP1 and HP2, corresponding to the two original cohorts are created such that:

1. The number of people in HP1 and HP2 follow the original data. There were 529 people from 10 villages in Cohort 1 and 727 people from 28 villages in Cohort 2, for whom the respective water contact distributions were estimated.
2. For each hypothetical individual, residential village is randomly assigned with its corresponding annual mean cercarial density $\bar{c}_v$. Note again that for HP1, the annual cercarial density was estimated separately for two consecutive years. Accordingly, each simulation was conducted on a two-year basis. For HP2, on the other hand, the simulation interval is 1 year.
3. The values of three other parameters must be specified for each individual: individual susceptibility ($\alpha_i$), the water contact minutes (or equivalently, $n_i$), and the mean body surface area ($s_i$). The $\alpha_i$ of each individual is randomly selected from one of the three distributions shown in Figure 3. Derived from the survey data for each cohort is a negative binomial distribution that is parameterized to describe the total number of water contact minutes for a randomly selected individual. Finally, we chose to sample the values of $s_i$ for each individual from empirical cumulative density distributions based on the originally calculated body surface area of each cohort from the survey data. The specific data corresponding to each of the above components of the simulations are given in Supplemental Material Item 5 (SM5).

Two series of simulations were conducted to account for stochasticity from these various sources. The first series mimics transmission among the same group of individuals over multiple years. So the simulation results reflect the fluctuations of water contact and random variability in cercarial exposure from year to year. Hence, each individual's $\alpha_i$, $s_i$, and $\bar{c}_v$ (or equivalently, the village in which this individual resides) are fixed in all simulations once randomly assigned. For water exposure, it is assumed that the total number of 30-minute intervals randomly varies between 0.8 and 1.2 of the originally selected $n_i$ (the same assumption as that earlier used in simulations to estimate $\alpha_i$). For the second series, on the other hand, each individual's exposure profile is randomly reassigned in every simulation, so a new hypothetical population is created each time. The main goal of conducting Series II simulations is to confirm that the patterns observed in the Series I simulations are not due to chance, that is, they do not apply to only one specific population at one particular time, but represent relatively stable patterns of infection intensity in the population. Both series of simulations were conducted 2,000 times for each cohort.

The second epidemiological finding that was addressed in the simulations relates to the role of individual susceptibility to reinfection after treatment with praziquantel. Recall that it was the analysis of epidemiological data on these two cohorts that initially suggested the existence of differential susceptibility in the populations and that also served as the basis for our estimation above the values of the susceptibility parameter $\alpha_i$ for a subset of individuals. The epidemiological finding was that repeated infections were found to be nonrandomly distributed in the population even when controlled for cercarial exposure. That is, people who were infected in previous surveys were more likely to be reinfected after treatment than those who were not previously infected.[3] However, the earlier statistical analysis used proxies for cercarial exposure such as

village infection prevalence, country of residence, and individual demographic variables as substitutes. The IBM offers an alternative, mechanism-oriented model structure, incorporates the hypothetical populations, and uses explicit estimates of cercarial density. The question then, is if the same pattern of reinfection occurs in the hypothetical populations.

In each simulation, the model is run for two time units (TUs) two 2-year intervals for HP1 and one 2-year interval for HP2. At the end of each TU, a mass chemotherapy treatment is simulated assuming perfect compliance and efficacy. That is, every individual's worm burden is reset to zero. Any infection occurring at TU1 will be regarded as baseline infection and that of TU2 is the reinfection. If reinfection is random among the population, the proportion of people who are reinfected in each simulation should be approximately $P_R \approx IR_1 \times IR_2$, where $IR_1$ and $IR_2$ are fraction of the population infected at TU1 and TU2. On the other hand, the actual proportion of reinfections in each simulation is $P_A = N_R/N_{HP}$, where $N_R$ is the number of people who have infections (EPG $> 0$) at both time points. $N_R$ is directly counted from each simulation. $N_{HP}$ is the number of people in the corresponding hypothetical population. The ratio of the two proportions is an index for comparison between the two HPs that represent two different levels of cercarial exposure and transmission intensity where $R_{AR} = P_A/P_R$. In this context, the epidemiological finding was that $R_{AR}$ was significantly greater than unity for both cohorts and that the ratio increased as risk of infection decreased, that is, the ratio was larger for Cohort 2 than Cohort 1.

## RESULTS

The first pattern of interest regards the distribution of infection intensity among the hypothetical populations. Our epidemiological data from previous infection surveys, as well as that of many other investigators, have shown that EPG in population groups is approximately negative binomially distributed for most transmission levels and exposure durations.[15,32] To verify this pattern is also characteristic of the two HPs, each simulation was run for three TUs—that is, three 2-year intervals for HP1 and three 1-year intervals for HP2.

The histograms of prevalence and population mean EPG for the 2,000 runs at the end of 2 years for HP1 and 1 year for HP2 are shown in Supplemental Material Item 6 (SM6) and show normal-like distributions for both, as the central limit theorem predicts. Moreover, the variability in both is modest, which indicates that the pattern of response in these variables is consistent across the 2,000 runs. Figure 4 shows the intensity distributions corresponding to the infection intensity of each individual averaged over of the 2,000 runs at the end of 2, 4, and 6 years and for the three susceptibility GSDs for HP1. Qualitatively, all of the plots present clear right skewed distributions that are phenomenologically consistent with the survey data of Cohort 1 (Supplemental Material Item 7 [SM7]). Thus, the selection of the GSD of the distribution of individual susceptibility relies on quantitative comparisons between the simulated and survey results for the same duration of transmission—one TU, or 2 years. Table 1 presents this

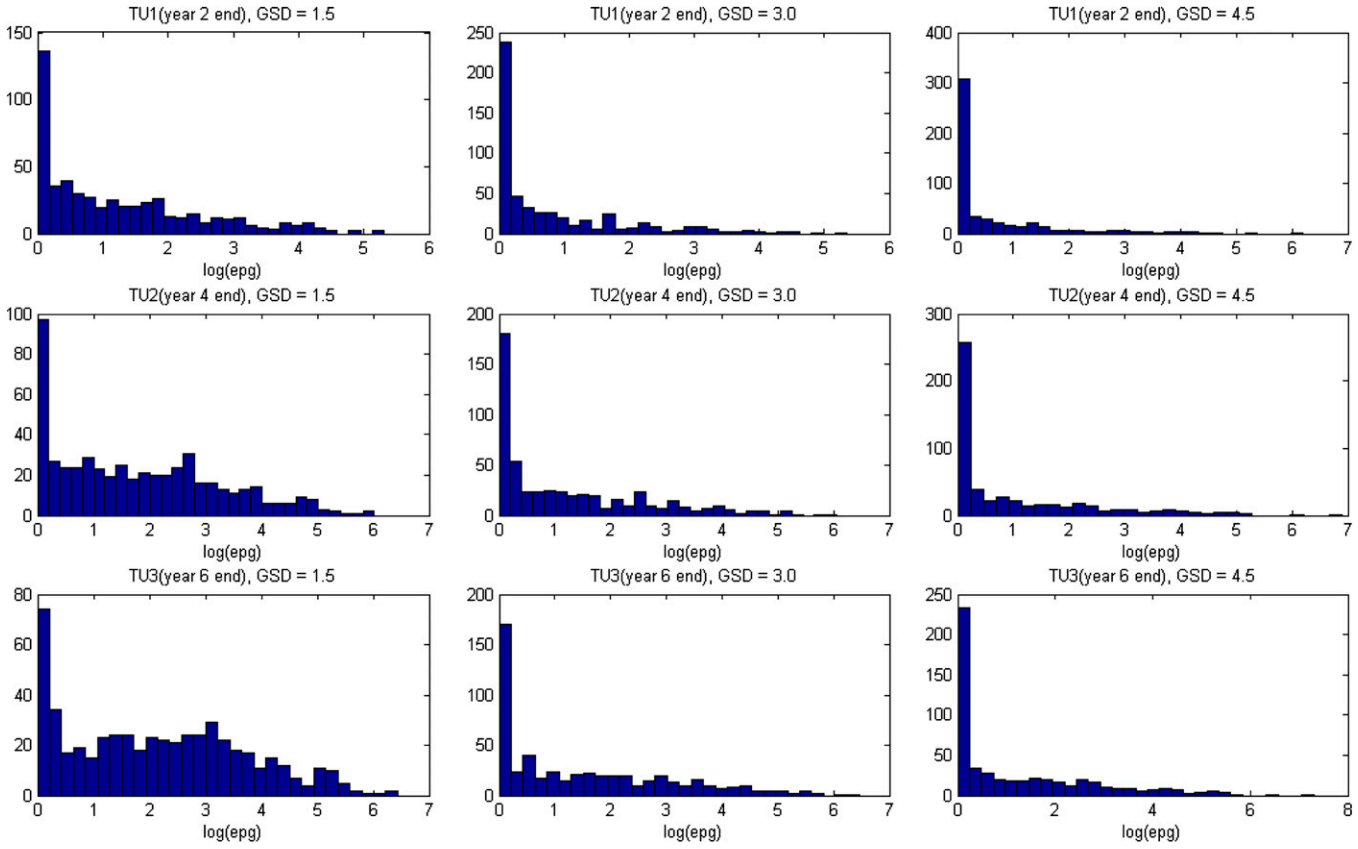


FIGURE 4. EPG histograms corresponding to three GSDs of the individual susceptibility distribution for HP1 based on 2,000 Series I simulations. GSD = geometric standard deviation.

TABLE 1

Mean, variance, and aggregation parameter of EPG distributions for Cohort 1 (survey results) and the corresponding HP1 (simulation)

| | Prevalence (%) | Mean (EPG) | Var (EPG) | $k_{EPG}$ |
|---|---|---|---|---|
| Cohort 1 survey data | 23.0 | 8.5 | 824.0 | 0.08 |
| GSD = 1.5 | 51.8 | 9.0 | 278.7 | 0.30 |
| GSD = 3 | 34.9 | 6.3 | 426.9 | 0.09 |
| GSD = 4.5 | 27.8 | 6.9 | 692.1 | 0.07 |

EPG = eggs per gram of stool sample; GSD = geometric standard deviation.

comparison including the aggregation parameter $k_{EPG}$ that assumes the data to be from a negative binomial distribution.

The susceptibility distribution with GSD of 1.5 yielded more than twice as high a prevalence in HP1 than in Cohort 1 (51.8% versus 23.0%) despite having a similar mean EPG (9.0 versus 8.5). In addition, the variance and $k_{EPG}$ of the simulated data were not close to the survey results. As can be seen from Table 1, the results corresponding to GSD of 4.5 were, in general, more consistent with the survey results without sacrificing the accuracy of $k_{EPG}$ estimation.

Using a susceptibility GSD of 4.5, the results of Series I simulations for HP2 yield a mean incidence quite close to the survey result for Cohort 2 (11.4% versus 11.7%), but the mean EPG is considerably lower than the survey result (1.14 versus 4.78). Indeed, as can be seen from the distribution of mean EPG for the 2,000 runs for HP2 in SM6, 4.78 would be in the extreme right tail of that distribution. In tracing back, we found that 75% of the total EPG observed in Cohort 2 was attributed to two individuals and, upon exclusion of these two outliers, the survey would have yielded a mean EPG of 1.19. Thus, while the variability in susceptibility and exposure in HP2 does not account for these outliers, the simulation results in the two hypothetical populations have proved reasonably capable of capturing the incidence rate and infection intensity for the two cohorts.

Regarding reinfection, and continuing to use a susceptibility GSD of 4.5, the simulation results shown in Table 2 reveal two features that are qualitatively consistent with the epidemiological results. First, reinfection is not random among the population—individuals who are infected in the first TU are more likely to be reinfected in the second TU. Second, the ratio of $R_{AR}$ is greater in the low-transmission than in high-transmission environment. Clearly, the values of $R_{AR}$ are higher in the simulations than seen in the epidemiological data and the implications of that discrepancy are discussed below.

## DISCUSSION

Recall that our overall objective was to explore via simulation the role of individual susceptibility and its interaction with cercarial exposure in search of further evidence for the existence of differential susceptibility to infection by *S. japonicum*. The results of the simulations show that regardless of the degree of variability of susceptibility in the population, the distribution of infection intensity is highly right skewed. Moreover, the variance of the distribution of individual susceptibility in the population is important in mimicking the quantitative epidemiological findings with respect to the distribution of EPG in both the medium- and low-risk transmission environments. Of the three log-normal distributions used to describe susceptibility, the largest GSD most closely matched the simulation outcomes with the epidemiological results. These findings suggest that significant differential susceptibility exists and, further, that there is a high degree of variability within the human population.

TABLE 2

Epidemiological estimates and simulated values of reinfection ratio ($R_{AR}$) for several GSDs

| | Cohort 1/HP1 | Cohort 2/HP2 |
|---|---|---|
| Epidemiology | 1.63 | 4.04 |
| Simulation | 2.56 | 5.72 |

The other principal point of comparison between the simulations and the epidemiological data relates to the characteristics of reinfection. In the simulation we find, as common sense suggests, that the more susceptible individuals are more prone to reinfection. Further, this clustering of reinfection among the most susceptible increases as the overall prevalence of infection decreases. This finding, too, mimics the epidemiological data. However, from a quantitative perspective, the degree of clustering as measured by the reinfection ratio $R_{AR}$ was greater in the simulations than that found epidemiologically. This observation is of some interest in that it relates to the interaction of exposure and susceptibility.

In exploring the exposure/susceptibility interaction, note that Equation 3 illustrates that, in the model, schistosome egg excretion by the infected human host is the product of the susceptibility parameter, water contact, and the cercarial density in that water, summed over the number of exposure events in an infection season. Both in our earlier epidemiological studies and those of others, infection intensity was found to be relatively insensitive to degree of water contact.[27,33] This was also the case in these simulation studies (see Supplemental Material Item 8 [SM8]). Hence, in the simulations, and very likely in the field, cercarial density and individual susceptibility are the principal determinants of the population distribution of EPG and of prevalence.

While we are tempted to speculate that some of the discrepancies between the simulation and epidemiological results suggest a somewhat higher mean and/or greater variability in cercarial density or a different distribution type for $\alpha_i$, we do not believe that the underlying data are sufficient to support any further fine tuning. In addition, it is also possible that there is some degree of residual acquired immunity in Cohort 1 in particular, which could contribute to the overestimate of reinfection shown in Table 2. Whatever the explanation, it would be of considerable value to have a method for directly measuring cercarial density in water on one hand and a biological marker of susceptibility of individuals to infection on the other. While the former seems an obvious conclusion, the benefits of developing a biomarker of susceptibility need further exploration. We are currently using the IBM described here to investigate the usefulness of such a biomarker to disease surveillance and that will be the subject of a future report.

Received November 1, 2014. Accepted for publication March 11, 2015.

Published online April 13, 2015.

Note: Supplemental material items, tables, and figures appear at www.ajtmh.org.

Acknowledgments: The research on which this article is based was supported by grant number 5RO1AI68854 from the National Institute Allergy and Infectious Diseases.

Authors' addresses: Shuo Wang and Robert C. Spear, Center for Occupational and Environmental Health, School of Public Health, University of California, Berkeley CA, E-mails: wangshuo00@gmail.com and spear@berkeley.edu.

## REFERENCES


1. Woolhouse MEJ, Dye C, Etard JF, Smith T, Charlwood JD, Garnett GP, Hagan P, Hii JLK, Ndhlovu PD, Quinnell RJ, Watts CH, Chandawana SK, Anderson RM, 1997. Heterogeneties in the transmission of infectious agents: implications for the design of control programs. *Proc Natl Acad Sci USA 94:* 338–342.
2. Civitello DJ, Rohr JR, 2014. Disentangling the effects of exposure and susceptibility on transmission of the zoonotic parasite *Schistosoma mansoni*. *J Anim Ecol 83:* 1379–1386.
3. Carlton EJ, Hubbard A, Wang S, Spear RC, 2013. Repeated *Schistosoma japonicum* infection following treatment in two cohorts: evidence for host susceptibility to helminthiasis? *PLoS Negl Trop Dis 7:* e2098.
4. Quinnell R, 2003. Genetics of susceptibility to human helminth infection. *Int J Parasitol 33:* 1219–1231.
5. Anderson RM, May RM, 1985. Helminth infections of humans: mathematical models, population dynamics, and control. *Adv Parasitol 24:* 1–101.
6. Gambhir M, Michael E, 2008. Complex ecological dynamics and eradicability of the vector borne macroparasitic disease, lymphatic filariasis. *PLoS ONE 3:* e2874.
7. Michael E, Simonsen PE, Malecela M, Jaoko WG, Pedersen EM, Mukoko D, Rwegoshora RT, Meyrowitsch DW, 2001. Transmission intensity and the immunoepidemiology of bancroftian filariasis in East Africa. *Parasite Immunol 23:* 373–388.
8. Castillo-Chavez C, Feng Z, Xu D, 2008. A schistosomiasis model with mating structure and time delay. *Math Biosci 211:* 333–341.
9. Chan MS, Bundy DA, 1997. Modelling the dynamic effects of community chemotherapy on patterns of morbidity due to *Schistosoma mansoni*. *T Roy SocTrop Med Hyg (Geneve) 91:* 216–220.
10. Feng Z, Li CC, Milner FA, 2002. Schistosomiasis models with density dependence and age of infection in snail dynamics. *Math Biosci 177–178:* 271–286.
11. Finkelstein JL, Schleinitz MD, Carabin H, McGarvey ST, 2008. Decision-model estimation of the age-specific disability weight for schistosomiasis japonica: a systematic review of the literature. *PLoS Negl Trop Dis 2:* e158.
12. Ishikawa H, Ohmae H, 2009. Modeling the dynamics and control of transmission of *Schistosoma japonicum* and *S. mekongi* in southeast Asia. *Korean J Parasitol 47:* 1–5.
13. Medley GF, Bundy DA, 1996. Dynamic modeling of epidemiologic patterns of schistosomiasis morbidity. *Am J Trop Med Hyg 55:* 149–158.
14. Milner FA, Zhao R, 2008. A deterministic model of schistosomiasis with spatial structure. *Math Biosci Eng 5:* 505–522.
15. Anderson RM, May RM, 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford, New York: Oxford University Press.
16. Bonabeau E, 2002. Agent-based modeling: methods and techniques for simulating human systems. *Proc Natl Acad Sci USA 99 (Suppl 3):* 7280–7287.
17. Burke DS, Epstein JM, Cummings DA, Parker JI, Cline KC, Singa RM, Chakravarty S, 2006. Individual-based computational modeling of smallpox epidemic control strategies. *Acad Emerg Med 13:* 1142–1149.
18. Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke H-H, Weiner J, Weigand T, DeAngelis DL, 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science 310:* 987–991.
19. Liang S, Seto EY, Remais JV, Zhong B, Yang C, Hubbard A, Davis GM, Gu X, Qiu D, Spear RC, 2007. Environmental effects on parasitic disease transmission exemplified by schistosomiasis in western China. *Proc Natl Acad Sci USA 104:* 7110–7115.
20. Liang S, Spear RC, Seto E, Hubbard A, Qiu D, 2005. A multigroup model of *Schistosoma japonicum* transmission dynamics and control: model calibration and control prediction. *Trop Med Int Health 10:* 263–278.
21. Liang S, Spear RC, 2008. Model-based insights into multi-host transmission and control of schistosomiasis. *PLoS Med 5:* e23.
22. Maszle DR, 1998. *Dynamic modeling for the control of schistosomiasis in China in light of parametric uncertainty. Graduate Group in Bioengineering*. Berkeley, CA: University of California.
23. Seto EY, Carlton EJ, 2010. Disease transmission models for public health decision-making: designing intervention strategies for *Schistosoma japonicum*. *Adv Exp Med Biol 673:* 172–183.
24. Yakob L, Williams GM, Gray DJ, Halton K, Solon JA, Clements ACA, 2013. Slaving and release in co-infection control. *Parasit Vectors 6:* 157.
25. Carlton EJ, Bates MN, Zhong B, Seto EY, Spear RC, 2011. Evaluation of mammalian and intermediate host surveillance methods for detecting schistosomiasis reemergence in southwest China. *PLoS Negl Trop Dis 5:* e987.
26. Spear RC, 2012. Internal versus external determinants of *Schistosoma japonicum* transmission in irrigated agricultural villages. *J R Soc Interface 9:* 272–282.
27. Seto E, Lee Y, Liang S, Zhong B, 2007. Individual and village level study of water contact patterns and *Schistosoma japonicum* infection in mountainous rural China. *Trop Med Int Health 12:* 1199–1209.
28. Spear RC, Seto E, Liang S, Birkner M, Hubbard A, Qiu D, Yang C, Zhong B, Xu F, Gu X, Davis GM, 2004. Factors influencing the transmission of *Schistosoma japonicum* in the mountains of Sichuan Province of China. *Am J Trop Med Hyg 70:* 48–56.
29. Spear RC, Zhong B, Mao Y, Hubbard A, Birkner M, Remais J, Qiu D, 2004. Spatial and temporal variability in schistosome cercarial density detected by mouse bioassyas in village irrigation ditches in Sichuan, China. *Am J Trop Med Hyg 71:* 554–557.
30. Wang S, 2013. *Explortion of Surveillance and Control Strategies for Re-emerging Schistosomiasis Environments in Sichuan Province, China: The Development and Application of an Individually-based Model. Graduate Group in Environmental Health Sciences*. Berkeley, CA: University of California, 105.
31. Spear RC, Hubbard A, 2010. Parameter estimation and site-specific calibration of disease transmission models. *Adv Exp Med Biol 673:* 99–111.
32. Liang S, Maszle D, Spear RC, 2002. A quantitative framework for a multi-group model of *Schistosoma japonicum* transmission dynamics and control in Sichuan, China. *Acta Trop 82:* 263–277.
33. Scott JT, Diakhaté M, Vereecken K, Fall A, Diop M, Ly A, De Clercq D, De Vlas SJ, Berkvens D, Kestens L, Gryseels B, 2003. Human water contacts patterns in *Schistosoma mansoni* epidemic foci in northern Senegal change according to age, sex and place of residence, but are not related to intensity of infection. *Trop Med Int Health 8:* 100–108.
34. Hubbard A, Liang S, Maszle D, Qiu D, Gu X, Spear RC, 2002. Estimating the distribution of worm burden and egg excretion of *Schistosoma japonicum* by risk group in Sichuan Province, China. *Parasitology 125:* 221–231.