**Title:**

Aggregation of *Schistosoma haematobium* increases under intense control efforts, potentially complicating schistosomiasis elimination

**Short title:**

*Schistosoma haematobium* aggregation and schistosomiasis elimination

**Authors:**

Christopher M. Hoover1, Stefanie Knopp2,3, Joseph A. Lewnard4, Alan E. Hubbard5, Jay P. Graham1, Giulio A. de Leo6,7, Susanne H. Sokolow6,7, David Rollinson8, Fatma Kabole9, Justin V. Remais1

1 Division of Environmental Health Sciences, University of California, Berkeley School of Public Health, Berkeley, CA 94720 USA

2 Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland

3 University of Basel, Petersplatz 1, 4001 Basel, Switzerland

4 Division of Epidemiology, University of California, Berkeley School of Public Health, Berkeley, CA 94720 USA

5 Division of Biostatistics, University of California, Berkeley School of Public Health, Berkeley, CA 94720 USA

6 Department of Biology, Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950 USA

7 Woods Institute for the Environment and Center for Innovation in Global Health, Stanford University, Stanford, CA 94305 USA

8 Global Schistosomiasis Alliance, Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom

9 Neglected Diseases Program, Zanzibar Ministry of Health, P.O. Box 236, Unguja, United Republic of Tanzania

## Abstract:

**Background:** The aggregation of helminthic parasites across individuals with varying worm burdens has important implications for the transmission of neglected tropical diseases including schistosomiasis. Only mated worm pairs contribute to transmission, and the mating frequency of worms is heavily influenced by their aggregation amongst human hosts. Here we use data from the Zanzibar Elimination of Schistosomiasis Transmission (ZEST) project to determine how aggregation changes in response to control efforts such as mass drug administration (MDA).

**Methods:** We first quantify aggregation via maximum likelihood estimation of the dispersion parameter, , by fitting negative binomial distributions to egg burden data collected from children enrolled in the study across 90 small administrative units called shehias from 2012-2017. We then use generalized estimating equations to estimate changes in as a function of mean egg burdens. In addition, we use approximate Bayesian computation to estimate mean community parasite burdens and mating probabilities for every shehia-year under different assumptions regarding the distribution of male and female parasites.

**Results:** We estimate values of ranging from 0.001 to 0.067 (median 0.0074) across all shehia-years and find a 0.012 (95% CI 0.011 – 0.013) increase in , implying decreasing aggregation, associated with an interquartile range increase in the mean shehia egg burden. This negative association between aggregation and burden was also captured in estimated worm burdens, though was sensitive to the assumed distribution of male and female parasites. Mating probability was found to remain high in some shehias even as parasite burden decreased.

**Conclusions:** Increasing aggregation at lower transmission intensities may complicate surveillance and elimination efforts in low-transmission settings. Ensuring treatment of high burden individuals and focusing on environmental interventions may aid in achieving elimination even in the context of increasing aggregation.

## Short Summary:

Parasite aggregation is an important factor in the elimination of helminthic diseases, but has not been quantified in elimination settings. Using data from a cluster randomized trial conducted on the Zanzibar archipelago, where urinary schistosomiasis is targeted for transmission interruption and elimination, we estimated the dispersion parameter of *Schistosoma haematobium* egg burdens across human hosts to quantify aggregation. We found that aggregation increases, implying more parasites are harbored by fewer human hosts, as mean community egg burden decreases. This phenomenon could complicate elimination efforts, particularly if individuals that retain high burdens are systematically untreated or have inherently higher susceptibility. Achieving elimination in the context of increasing aggregation may be more likely if mass drug administration strategies are supplemented by environmental interventions such as snail control and behavioral change interventions that ensure high-burden individuals are able to access and comply with treatment.

# Background

Decades of mass drug administration (MDA) have drastically reduced the prevalence and burden of neglected tropical diseases, particularly helminthiases including schistosomiasis, onchocerciasis, lymphatic filariasis (LF), and soil transmitted helminthiases (STHs). National control programs enacted by ministries of health in endemic countries—often supported by organizations such as the World Health Organization and drug companies which donate necessary drugs for treatment—have made substantial progress reducing the global burden of helminthiases. [1] More recently, the global conversation has shifted from morbidity control towards elimination, especially as some areas have made substantial progress towards elimination with integrated control approaches that utilize MDA in addition to environmental and behavioral interventions. [2–6] Globally, progress towards elimination is mixed and some areas of sub-Saharan Africa, where the vast majority of schistosomiasis and other helminthiases occur, have struggled to achieve transmission control and progress towards elimination. [7]

With sufficient coverage and frequency of treatment, elimination using MDA alone is believed to be possible in low and moderate transmission settings. [8] However, achieving necessary treatment conditions is challenging from a practical standpoint due to the programmatic limitations of many control campaigns. School-based interventions, for instance, are often the primary method of drug delivery in communities since they are straightforward to implement and reach school-aged children, who are widely considered the most vulnerable and most heavily infected segment of the population. However, school-based strategies, and community-wide treatments that seek to reach all individuals, often miss infected individuals who continue to contribute to transmission. [9,10]

Because of this, control efforts are likely to have profound implications for parasite aggregation (i.e. the distribution of parasites among the human host population). While the dynamics of parasite prevalence and mean community infection following MDA and other interventions have been studied extensively, the corresponding dynamics of aggregation are largely unexplored. Aggregation is measured by the aggregation parameter, (also frequently referred to as “dispersion parameter”) of a negative binomial distribution. While is often considered constant in models and other forward-looking analyses of control efforts, there is reason to believe that it is itself a dynamic variable that is positively correlated with transmission intensity. More transmission leads to less-skewed distributions of infection, while reductions in transmission from interventions such as MDA lead to more skewed distributions in which fewer individuals host larger proportions of the total parasite population. Individuals that remain infected in these scenarios may be those who miss treatment, are heavily exposed, are particularly susceptible to infection, or have parasite strains that are resistant to treatment. [11–15]

Aggregation dynamics near the elimination threshold are important from a number of practical and theoretical vantages. [16] As mean community infection is reduced, so too is the number of individuals contributing to and sustaining transmission—a realization of the “20/80 rule” in which 20% of individuals contribute 80% of the transmission potential. [11] Identifying and treating these individuals is complicated by systematic non-access to treatment, treatment failures, and insensitive diagnostics, but will be essential to achieving elimination. [11,17–19] A better understanding of aggregation dynamics could aid the planning and implementation of surveillance and control efforts in low transmission and elimination settings. [17]

Here, we examine evidence for dynamic aggregation from a cluster randomized trial (CRT) that was part of the Zanzibar Elimination of Schistosomiasis Transmission (ZEST) project. [20,21] We revisit classic assumptions of how adult schistosomes are distributed among human hosts and introduce a novel, mechanistic data-generating process for worm acquisition and resulting mating dynamics. We then explore how these assumptions and underlying data-generating processes affect estimation of the mating probability and discuss implications for ongoing control and elimination efforts for schistosomiasis and other helminthiases.

# Methods

## Data

The CRT of the ZEST project enrolled 45 small administrative regions called shehias on the island of Pemba with the goal of eliminating urogenital schistosomiasis as a public health problem (defined as reducing the prevalence of heavy infections below ) and 45 shehias on the island of Unguja with the goal of interrupting *S. haematobium* transmission (defined as reducing the number of incident cases to zero). [20,21] On each island, shehias were randomly assigned to receive biannual MDA, biannual MDA plus snail control, or biannual MDA plus a behavioral intervention. [21–23] Cross-sectional parasitological surveys were conducted annually from 2012-2017 in each shehia. Data used here from both islands were acquired via a formal data request to the Schistosomiasis Consortium for Operational Research and Evaluation (SCORE) that supported the research of the ZEST project and are comprised of 27,626 observations of individual egg burdens among adults and 75,039 observations of individual egg burden among school-aged children, each measured as *Schistosoma haematobium* eggs per 10mL urine.

## Aggregation and its effects on transmission

From a theoretical standpoint, the conditions under which MDA alone can achieve elimination is dependent on reducing the parasite population below its breakpoint. Breakpoints in helminth transmission are expected to arise from mate limitation in low transmission settings whereby the probability of individuals being infected with fecund worms decreases as prevalence and transmission decrease. Mate limitation in helminth populations is generally quantified in terms of the probability that a given female parasite in a population is successfully mated, , which is estimated as a function of the mean community worm burden, , and aggregation parameter, . [24,25] In endemic settings, parasite distributions among a population of definitive human hosts are well-represented by a negative binomial distribution with below 1. This implies a heavily right-skewed distribution in which most parasites are aggregated in a few human hosts.

For constant aggregation parameter, , in the range as has been widely observed previously, as , giving rise to the hypothesized breakpoint below which worm mating becomes insufficient to sustain transmission. [26] As transmission decreases, the response of the aggregation parameter could have profound implications for the breakpoint. For instance, if as , implying increasingly skewed distributions that can be realistically interpreted as all parasites being harbored by one individual, remains large and the breakpoint is vanishingly small. [25] However, derivations of the mating probability based on the negative binomial distribution of adult parasites are derived from endemic settings, and stochasticity in worm acquisition at low transmission intensities is likely to play an increasingly important role on the distribution of adult parasites as elimination is approached. [24,25] This raises doubt as to whether analytic estimates of the mating probability and breakpoint, which rely on an assumption of negative binomially distributed parasites, are reliable in elimination settings.

## Aggregation dynamics from observed egg burdens

The likelihood of the inverse negative binomial aggregation parameter, , and mean community egg burden, , can be estimated for each shehia, , and year, , from individual, , egg counts, , as:

While aggregation is most frequently reported and discussed in terms of , its inverse, has more desirable properties for statistical estimation and inference (see [27]). Therefore results are presented in terms of following trivial transformation from estimates derived using .

Mean community egg burden, , is simply estimated as the first moment of individual egg counts, and maximum likelihood estimates of were estimated using the Brent method within the optim function in R. [28] Uncertainty in estimates of were derived from the Hessian matrix. Both and were estimated among the adult (A), child (C), and total (T) populations in each shehia-year.

To estimate the relationship between aggregation and mean community egg burden, weighted generalized estimating equations (GEE) with unstructured correlation matrices were used of the form:

Weights were assigned as the inverse of the standard error of . Models stratifying by population (child, adult, or total), by intervention type (MDA only, behavioral, or snail control), and by island (Pemba or Unguja) were also estimated. Results are reported for each stratum as the change in associated with an interquartile range increase in mean community egg burden, . Clustered nonparametric bootstrapping was used to estimate uncertainty in estimates of with bootstrapped samples at the shehia level.

## Estimating mean community worm burden and aggregation

Because egg counts are an indirect estimate of worm burden, are prone to undercounts at low prevalence, and may change non-linearly with mean community burden [29], we next sought to determine if changes in aggregation as measured by egg counts are indicative of changes in worm aggregation. We used approximate Bayesian computation (ABC) to estimate mean community worm burden and aggregation, denoted and , respectively, under four proposed data-generating mechanisms, (described below). Briefly, ABC proceeds by: 1) sampling parameter sets from a prior distribution, ; 2) simulating datasets, , from the priors under the given data generating mechanism; 3) deriving summary statistics, , from the generated data to compare to the observed data, ; 4) generating distance metrics, , to assess the fit of the generated data to the observed data; and 5) accepting parameter sets that fall within a provided distance tolerance. [29] The accepted parameter sets thus represent an approximation of the posterior distribution of , but the estimation procedure does not require exact calculation of the likelihood of every proposed prior. Here, ABC was preferred to Bayesian MCMC methods due to the number of estimates needed as well as the complexity of the likelihood function of the parasite population estimated from observed egg counts (see e.g. [30,31]).

The first two data-generating mechanisms considered correspond to the two classical assumptions in models of schistosome mating dynamics [24] that male and female schistosomes are “distributed together” from the same negative binomial distribution (Case1) or “distributed separately” from independent negative binomial distributions assumptions (Case 2). Case 1 is the canonical assumption used in essentially all analyses of schistosome aggregation. A third data-generating mechanism corresponds to a mechanistic model (Case 3) that considers susceptibility and exposure as independent parameters, and a fourth data-generating mechanism (Case 4) assumes that a proportion of parasites are distributed together while the remaining parasites are distributed separately, essentially a hybrid of Case 1 and 2.

Prior parameter distributions and data generating processes that simulate individual worm burdens and egg counts are delineated in Table 1. Briefly, Case 1 assumes worms are distributed among human hosts according to a single negative binomial distribution with 1:1 sex ratio. Case 2 assumes male, , and female, , worms are distributed according to separate negative binomial distributions, each with mean . Case 4 relies on a partitioning parameter, , drawn from a standard uniform prior which determines the proportion of worms that follows a Case 1 “together” distribution with the remaining () portion of worms following a Case 2 “separate” distribution.

Case 3 is mechanistic in that it considers susceptibility and exposure as distinct processes with independent distributions. It assumes susceptibility, , interpreted as the probability a given cercarial exposure will result in an adult worm, follows a gamma distribution, and cercarial exposures follow a negative binomial distribution independent of susceptibility. Worm pairing in individual hosts is then determined from a hypergeometric distribution that incorporates individual susceptibility, cercarial exposure, and 1:1 sex ratio among cercarial exposures (Table 1). The susceptibility distribution parameters are fixed, with parameters derived from previous work on *S. japonicum* [14,15], such that Case 3 estimation is based on the same number of parameters as Case 1 and 2, and estimates of the mean community worm burden and aggregation parameter can be obtained from the expectation and combined variances of the Case 3 susceptibility and exposure parameters.

The number of eggs shed per mated worm pair per day is assumed to follow a negative binomial distribution with mean and dispersion for all four cases (Table 1). Additionally, egg output from mated pairs is reduced by density dependent fecundity (DDF). [32–34] The DDF parameter , and the egg shedding parameters and are fixed across all data-generating mechanisms (Table 1).

Summary statistics used in ABC to compare the observed and simulated datasets consist of the mean community egg burden, ; standard error of individual egg counts, ; and an adjusted egg-prevalence measure—the number of egg positive individuals squared over the number of individuals—for every shehia-year in which at least one individual had an egg count . The distribution of this adjusted prevalence measure—as opposed to the prevalence (bounded between 0 and 1) or raw number of individuals infected (an integer)—leads to more stable ABC estimates. We used the default standardized distance metric in the R package ABC [35] to compare the observed summary statistics, , to simulation summary statistics,, derived from simulated datasets for each data-generating Case in each eligible shehia-year. The parameter sets with the smallest estimated distance for each simulation (e.g. tolerance of ) were accepted. These accepted parameter sets were weighted based on their distance from the observed data to derive estimates of the posterior distributions of .

## Parasite aggregation dynamics and implications for mating probability

Posterior predictive checks were performed to assess the ability of posterior parameter sets of each data-generating Case to emulate observed data from ZEST. For each data-generating Case and each eligible shehia-year, weighted draws from the posterior distribution, , were used to simulate community-wide egg burdens as delineated in Table 1. These egg burdens were then summarized using the same summary statistics as used in ABC estimation and compared to the observed summary statistics values visually and via the mean squared error between the distribution median and the observed summary statistics.

Posterior distributions of and were extracted as part of the posterior data generation and input into GEE models of the same form used for mean community egg burdens to estimate the change in worm aggregation per unit change in worm burden. Weights for shehia-year observations were derived as the inverse of the sum of the mean squared errors between the observed and median generated summary statistics.

Finally, analytic and simulated estimates of the mating probability were compared to evaluate the influence of dynamic aggregation and assumptions on the underlying data-generating mechanisms. Analytic formulas derived previously [24] for Case 1 together and Case 2 separate assumptions with constant aggregation were used to estimate the mating probability associated with mean community worm burdens in the range . Next, analytic estimates of the mating probability allowing for dynamic aggregation were derived by incorporating as a function of from the marginal estimates of the egg burden GEE, and from the marginal estimates of the worm burden GEEs fitted for each data-generating Case. Simulated estimates of the mating probability for every shehia-year were recorded from posterior samples as the number of mated worms per worm () in the simulated data.

**Table 1**: Priors and data-generating processes for approximate Bayesian computation estimation of mean community worm burden and aggregation. Data generation proceeds by first drawing from the multivariate prior, , containing shehia-year level summaries of mean worm burden and aggregation (for Cases 1, 2, and 4) or of individual susceptibility and exposure (for Case 3), then generating individual worm burdens, worm pairs, and finally egg burdens. Summaries of individual egg burdens for each shehia-year are then compared to observed data summaries from the Zanzibar Elimination of Schistosomiasis Transmission project to derive a posterior estimate of model parameters.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Case 1 (Males and females distributed together) | Case 2 (Males and females distributed separately) | | Case 3 (Explicit susceptibility and exposure) | Case 4 (Hybrid together and separate) |
| Priors |  |  | |  |  |
| Fixed Parameters |  |  | | ; |  |
| Worms |  |  | |  |  |
| Pairs |  | | | | |
| Eggs |  | | | | |
| : multivariate prior for data-generating case,  : mean community worm burden in shehia, , and year,  : individual, , worm burden in shehia, , and year, , with sex (*m*ale or *f*emale) and distribution assumption (*t*ogether or *s*eparate),  : aggregation parameter in shehia, , and year,  : mean cercarial exposure in shehia, , and year, | | | : shape parameter of gamma-distributed susceptibility,  : rate parameter of gamma-distributed susceptibility,  : mean eggs produced per mated worm pair  : aggregation parameter of negative binomially distributed daily egg release  : density dependent fecundity parameter | | |

All analyses were performed in R version 4.0.2 utilizing the geepack [36], tidyverse [37], and ABC [35] packages. The data that support the findings of this study are openly available in ClinEpiDB as the dataset “Study: SCORE Zanzibar *S. haematobium* Cluster Randomized Trial” at <https://clinepidb.org/ce/app/record/dataset/DS_eddb4757ba>. All code necessary to reproduce the analysis can be found at <https://github.com/cmhoove14/SchistoDynAgg>.

# Results

There were 93 shehia-years in which no individuals had positive egg counts and 16 shehia-years in which the maximum individual egg count was 1 and therefore variance estimates for could not be derived. This left 447 observations of and , shown in Figure 1A along with the marginal estimate of from the fitted GEE. Estimates of ranged from 0.001-0.067 (median 0.0074). We estimate a median 0.012 (Bootstrapped IQR: 0.011-0.013) increase in associated with an IQR increase in , implying a negative relationship between mean community egg burden and aggregation. This relationship appears to be driven by dynamics in children, as there was weak evidence for a relationship between and among adults (Fig 1B). Mean egg burden among adults ( 1.36) was substantially lower than in children ( 4.65), however. Shehias receiving the MDA+snail control and MDA+behavior change interventions appeared to have a stronger burden-aggregation relationship (larger , Fig 1B). An additional GEE model including main effects of intervention arm and island showed that reductions in were significantly larger in shehias receiving the MDA+snail control and MDA+behavioral change interventions ( 0.01 and 0.016, respectively) compared to the MDA-only intervention.

Chart

Description automatically generated

Figure 1: Relationship between mean community egg burden and aggregation. Panel A shows the scatterplot of all estimates of and where each point is sized according to its weight, derived as the inveerse of the standard error of . Points are also symbolized according to their assigned intervention group—MDA only (▲), MDA+snail control (♦), or MDA+behavioral intervention (■)—and the island of the shehia. The solid black line in panel A represents the marginal estimate of from the fitted GEE. Panel B shows estimates of , the change in aggregation for an interquartile range increase in mean community egg burden stratified by population, treatment group, and island. The black point and error bars correspond to the unstratified marginal estimate of as in panel A. Error bars correspond to the interquartile range of estimates derived from B=10,000 bootstrapped samples.

Chart

Description automatically generated

Figure 2: Comparison of observed summary statistics to the median summary statistics generated from 1,000 simulated datasets derived from the ABC posterior distributions for every shehia-year. Colors and right-hand facet labels indicate the data generating Case and the 1:1 line implying perfect agreement between observed and generated data is shown. Posterior predictions from the mechanistic Case 3 data-generating mechanism (row 3, purple) were found to better reproduce the observed data, particularly the adjusted prevalence measure (see annotated means of mean-squared errors across all shehia-years). MSE: mean squared error

Examination of the Bayes factors comparing the posterior probabilities of each Case shows that the mechanistic Case 3 performs better than the canonical Case 1 (mean Bayes factor = 1.97, range 0-12.78), while Case 2 (mean Bayes factor = 0.89, range 0.07-5.98), and the hybrid Case 4 (mean Bayes factor = 0.97, range 0-6.9) generally performed worse than Case 1. Superior performance of Case 3 was further confirmed by examination of the mean squared errors between the generated and observed data, with Case 3 generated data resulting in the lowest MSE across all three summary statistics (Fig 2).

Case 1 and 3 estimates of and exhibited a similar pattern to and , namely increasing aggregation (indicated by decreasing ) with decreasing mean community worm burden (Fig 1 and Fig 3a and 3c). However, Case 2 and Case 4 estimates showed the opposite effect of decreasing aggregation with decreasing mean community worm burden (Fig 3b and 3d), though it should be noted that these two data-generating cases performed poorly both in terms of the MSE of the summary statistics and the Bayes factor.

This poor performance of Case 2 and 4 prompted further comparisons of Cases 1 and 3. Figure 4 shows that Case 3 estimates of the aggregation parameter were nearly identical to those from Case 1 estimation at higher mean community egg burdens, but diverged at lower mean community egg burdens. In particular, Case 3 estimates of the aggregation parameter were much lower than those of Case 1.

Chart, scatter chart

Description automatically generated

Figure 3: Summary of approximate Bayesian computation estimates of mean community worm burden () and aggregation () for Case 1 (Together, A), Case 2 (Separate, B), Case 3 (Mechanistic, C), and Case 4 (Hybrid, D). Dashed lines depict the marginal estimates of the fitted GEE for each Case: . Axes are log transformed to aid visualization.

Chart, scatter chart

Description automatically generated

Figure 4: Difference between Case 3 (mechanistic) and Case 1 (together) aggregation estimates across mean community egg burdens. At low infection intensities, canonical Case 1 estimates tend to predict higher values of , implying lower aggregation and higher mating probability, than mechanistic Case 3 estimates.

Chart

Description automatically generated

Figure 5: Comparison of mating probabilities derived analytically and from approximate Bayesian computation posterior distributions. Points depict the median mating probability across the median community worm burden from data generated for every shehia-year and Case from ABC posterior distributions. Solid lines depict analytical estimates of the mating probability assuming constant aggregation parameter, , equal to the median across all shehia-year observations. Dotted lines depict analytical estimates of the mating probability assuming aggregation changes as a function of mean community egg burden according to the fitted egg burden GEE, and dashed lines depict analytical estimates of the mating probability assuming aggregation changes as a function of mean community worm burden according to the fitted worm burden GEE for each case (shown in Figure 3). Note that Case 2 analytical mating probability estimates are derived according to the Case 2 (distributed separately) mating probability function from [32], while all others are derived for the canonical Case (distributed together) mating probability function.

Figure 5 shows the mating probability produced from the posterior distributions of all four data-generating cases. These estimates are also compared to analytic predictions of the mating probability for both static aggregation (Fig 5 solid lines) and dynamic aggregation (, , Fig 5 dashed and dotted lines, respectively). These analytic predictions generally align well with Case 1 and Case 3 estimates, but do frequently underestimate the mating probability at low mean community worm burdens.

# Discussion

Here we use data from the ZEST project to explore how parasite aggregation dynamics change as urogenital schistosomiasis elimination is approached due to intense control efforts. We find that parasite aggregation as measured directly from community egg distributions and via estimation of community worm burden distributions likely increases as elimination is approached. In addition, the classic (Case 1) assumption that male and female parasites are distributed together appears to hold at large mean community worm burdens. However, as proposed previously, this assumption appears to break down at smaller community worm burdens as elimination is approached. [24]

The distribution of individual egg counts from the ZEST study participants was highly aggregated, with in the range 0.001-0.067. Community egg aggregation was found to significantly increase as the mean community worm burden decreased, and this effect was driven by children. As children frequently have the highest exposure and may be most susceptible to infection due to lack of acquired immunity, [35] this is not surprising. We also found that the relationship between the aggregation parameter and mean community egg burden was greater in shehias from the ZEST study that received snail control or behavioral interventions in addition to biannual MDA, particularly among children and on the island of Unguja. This could be driven by higher starting prevalence and burden among these intervention arms. [21]

We propose a novel, mechanistic data-generating process (Case 3) that explicitly models the interaction of host susceptibility and exposure to allow the data to determine the most likely distribution of male and female parasites. Using approximate Bayesian computation to estimate mean worm burdens and worm aggregation, we find that this Case 3 data-generating process generally provides a better fit to the observed ZEST data than classical assumptions. Due to the flexibility and superior fit of the proposed Case 3 mechanism, we believe it represents a compelling theoretical basis for analyzing and modelling the dynamics of parasite aggregation and mating.

Case 3 estimation suggest that male and female parasites are distributed together (Case 1) at high transmission intensities that result in high worm burdens. Case 3 estimates of worm burden and aggregation were more similar to estimates recovered from Case 1 estimation at higher worm burdens. However, as infection intensities decrease, Case 3 estimates diverge from Case 1, with estimates from Case 3 generally smaller than those from Case 1. This, together with the superior posterior fits of Case 3, suggest that Case 1 dynamics may fail to capture the aggregation dynamics and resulting distribution of mated pairs of schistosomes in very low transmission intensity environments such as those found in elimination settings.

One explanation for this is that the distribution of male and female parasites is not strictly “together” (Case 1) or “separate” (Case 2), but is rather a blend of both that is dependent on the mean community worm burden itself. [24] We tested this hypothesis explicitly with the Case 4 data-generating distribution, but found that Case 4 generally provided a poor fit to the observed data. The fact that Case 3 provides the best fit suggests that capturing variability in both exposure and susceptibility is important to capture aggregation dynamics at very low transmission intensities.

This suggests that identifying individuals who are particularly susceptible to infection, or who are heavily exposed due to their occupation or behaviors, could be extremely valuable for control efforts, since these individuals are both most susceptible to pathology associated with infection and are most likely contributing to sustaining transmission. These individuals could be identified through a sufficiently accurate biomarker or simply by taking note of individuals who are most heavily infected. This information could then be used to inform targeted drug administration strategies that prioritize treatment of such high-risk individuals, while maintaining broad coverage of the population. Such strategies could be explored in subsequent modeling analyses that build on the results presented here.

Targeted strategies should also be pursued with caution if relying on diagnostics such as hematuria or egg-assays that are known to have low sensitivity for light infections. Indeed, reliance on these diagnostics represents a limitation of this analysis as sensitivity in egg counts is not explicitly considered in analysis of the community egg burdens or in the data-generating Cases considered. However, high variability in daily egg counts produced by even heavily infected individuals [30] is a part of the data-generating processes of the ABC estimation, meaning that generated egg distributions would be expected to match observed ones if diagnostic sensitivity in the real world were perfect. Future studies using more sensitive diagnostics such as circulating cathodic antigen (CCA) or circulating anodic antigen (CAA) assays, molecular methods, or incorporating additional diagnostics into the data generation procedures could be pursued to confirm these findings.

Heterogeneity in susceptibility and exposure (and their joint distribution) are not the only processes that could explain the dynamics of egg and worm burden and aggregation identified here, though they most likely play a significant role. Recent analyses have found that mated pairs may survive praziquantel treatment and resume egg-laying shortly after treatment. [11] This would lead individuals harboring such resistant pairs to maintain their high worm burden following treatment, while others are cleared of adult worms, thus increasing aggregation as is seen in these analyses. Similarly, individuals who miss treatment due to absence, pregnancy, or non-compliance would also maintain their high worm burdens as others are treated and cleared of adult worms, leading to increased aggregation as the mean community burden decreases.

Regardless of the exact mechanism or combination of mechanisms that give rise to increased aggregation, the practical implications are noteworthy. [38] Because some individuals maintain high worm burdens and egg output, they are more likely to maintain transmission among the snail population, which could contribute to infection rebound in the entire community. From a theoretical perspective, this can be expressed in terms of the mating probability. Our results show that increased aggregation may lead to mating probabilities higher than are commonly predicted assuming Case 1 dynamics with constant aggregation.

Deriving an analytic solution for the mating probability as a function of the mean worm burden—or better still the mean egg burden—from the proposed Case 3 data distribution would be valuable to further explore aggregation dynamics arising from heterogeneities in susceptibility and exposure. However, we also believe it is likely that Case 3 simply provides a more flexible framework to capture the high amount of stochasticity inherent in systems approaching elimination, meaning any analytic estimate may prove unreliable.

These results imply that the hypothesized breakpoint in schistosomiasis transmission could be exceedingly small or non-existent for practical purposes. Widespread resilience to elimination [39–43] and the recent outbreak of schistosomiasis in Corsica that may have been introduced by a single egg-shedding individual [44] also provide evidence in support of this finding. Breaking schistosomiasis transmission to achieve elimination may therefore be more dependent on reducing environmental transmission through snail control, sanitation improvements, and behavioral interventions and on identifying and successfully treating the few individuals who remain infected and maintain transmission as elimination is approached. [16,45–47]

A renewed focus on snail control as an important component of schistosomiasis elimination is encouraging. Heterogeneities in susceptibility, exposure, and sex differentiation—and resulting aggregation dynamics—are just as complex in intermediate host snail populations as in human hosts. [48] Additional focus on even more integrated strategies that incorporate snail control, MDA, environmental remediation, sanitation, and behavioral changes could prove even more successful, as they were in bringing about widespread elimination of schistosomiasis transmission in China. [2,49] Emerging evidence that zoonotic components of transmission such as hybrid *S. bovis* and *S. haematobium* [50] strains and infected rodents [51] are capable of sustaining transmission even in the presence of widespread MDA make such integrated strategies even more important. Identifying such integrated intervention strategies that result in the highest probability of successful elimination, particularly in the context of resistance to elimination due to dynamic aggregation as described here, is the subject of an ongoing modeling analyses that will be presented separately.

In conclusion, we have identified dynamic aggregation as an important aspect of schistosomiasis transmission in elimination settings. We propose a novel, generalizable quantitative framework for modeling parasite acquisition and subsequent mating dynamics among the human host population that suggests the canonical Case 1 assumption that male and female parasites are distributed together may not be applicable approaching elimination. This implies that the mating probability of adult worms remains high as elimination is approached, and the feasibility of achieving elimination using MDA alone may be overestimated under current modeling paradigms. We suggest more targeted treatment strategies and a focus on integrated intervention strategies that target other components of the schistosome lifecycle to combat this source of resilience and maintain global progress towards the elimination of schistosomiasis.

# Acknowledgments

CMH and JVR acknowledge funding from the National Institutes of Health (1R01TW010286 and R01AI125842), the National Science Foundation Water, Sustainability and Climate program (1360330 and 1646708), and the University of California Multicampus Research Programs and Initiatives (MRP-17-446315). CMH additionally acknowledges support from the NSF Data Science for the 21st Century Training program (1450053). SK and DR are thankful for financial support from the University of Georgia Research Foundation Inc., which is funded by the Bill & Melinda Gates Foundation for the Schistosomiasis Consortium for Operational Research and Evaluation (SCORE) projects (prime award no. 50816, sub-award no. RR374-053/4893206 to DR and sub-award no. RR374-053/4893196 to SK). Additionally, SK is financially supported by a direct grant from the Bill & Melinda Gates Foundation (Investment ID: OPP1198086). GADL and SHS have been supported by the National Science Foundation through grants # DEB – 2011179 and Belmont CEH/NSF ICER-2024383

# Author Contributions

|  |  |
| --- | --- |
| **Conceptualization** | Christopher M. Hoover, Justin V. Remais |
| **Data Curation** | Stefanie Knopp, David Rollinson, Fatma Kabole |
| **Formal Analysis** | Christopher M. Hoover |
| **Funding Acquisition** | Justin V. Remais |
| **Investigation** | Christopher M. Hoover, Stefanie Knopp |
| **Methodology** | Christopher M. Hoover |
| **Project Administration** | All authors |
| **Resources** | All authors |
| **Software** | Christopher M. Hoover |
| **Supervision** | Justin V. Remais, Joseph A. Lewnard, Jay P. Graham, Alan E. Hubbard |
| **Validation** | All authors |
| **Visualization** | Christopher M. Hoover |
| **Writing – Original Draft Preparation** | Christopher M..Hoover |
| **Writing – Review & Editing** | All authors |

# References

1. World Health Organization. Ending the neglect to attain the Sustainable Development Goals: A road map for neglected tropical diseases 2021–2030. Geneva; 2020. Available: https://www.who.int/neglected\_diseases/Revised-Draft-NTD-Roadmap-23Apr2020.pdf?ua=1

2. Yang Y, Zhou YB, Song XX, Li SZ, Zhong B, Wang TP, et al. Integrated control strategy of schistosomiasis in The People’s Republic of China: Projects Involving agriculture, water conservancy, forestry, sanitation and environmental modification. Adv Parasitol. 2016;92: 237–268. doi:10.1016/bs.apar.2016.02.004

3. Spear R, Zhong B, Liang S. Low Transmission to Elimination: Rural Development as a Key Determinant of the End-Game Dynamics of Schistosoma japonicum in China. Trop Med Infect Dis. 2017;2: 35. doi:10.3390/tropicalmed2030035

4. Liang S, Abe EM, Zhou X-N. Integrating ecological approaches to interrupt schistosomiasis transmission: opportunities and challenges. Infect Dis Poverty. 2018;7: 124. doi:10.1186/s40249-018-0506-4

5. Ending the neglect to attain the sustainable development goals. A Global Strategy on Water, Sanitation and Hygiene to Combat Neglected Tropical Diseases 2021–2030. Geneva, Switzerland; 2021. Available: https://www.who.int/publications/i/item/9789240022782

6. Secor W, Colley D. When Should the Emphasis on Schistosomiasis Control Move to Elimination? Trop Med Infect Dis. 2018;3: 85. doi:10.3390/tropicalmed3030085

7. Trippler L, Ame SM, Hattendorf J, Juma S, Abubakar S, Ali SM, et al. Impact of seven years of mass drug administration and recrudescence of Schistosoma haematobium infections after one year of treatment gap in Zanzibar: Repeated cross-sectional studies. Montresor A, editor. PLoS Negl Trop Dis. 2021;15: e0009127. doi:10.1371/journal.pntd.0009127

8. Anderson RM, Turner HC, Farrell SH, Yang J, Truscott JE. What is required in terms of mass drug administration to interrupt the transmission of schistosome parasites in regions of endemic infection? Parasites and Vectors. 2015;8: 553. doi:10.1186/s13071-015-1157-y

9. Faust CL, Osakunor DNM, Downs JA, Kayuni S, Stothard JR, Lamberton PHL, et al. Schistosomiasis Control: Leave No Age Group Behind. Trends in Parasitology. Elsevier Ltd; 2020. pp. 582–591. doi:10.1016/j.pt.2020.04.012

10. Adriko M, Faust CL, Carruthers L V., Moses A, Tukahebwa EM, Lamberton PHL. Low Praziquantel Treatment Coverage for Schistosoma mansoni in Mayuge District, Uganda, Due to the Absence of Treatment Opportunities, Rather Than Systematic Non-Compliance. Trop Med Infect Dis. 2018;3. doi:10.3390/tropicalmed3040111

11. Faust CL, Crotti M, Moses A, Oguttu D, Wamboko A, Adriko M, et al. Two-year longitudinal survey reveals high genetic diversity of Schistosoma mansoni with adult worms surviving praziquantel treatment at the start of mass drug administration in Uganda. Parasites and Vectors. 2019;12: 607. doi:10.1186/s13071-019-3860-6

12. Carlton EJ, Hubbard A, Wang S, Spear RC. Repeated Schistosoma japonicum Infection Following Treatment in Two Cohorts: Evidence for Host Susceptibility to Helminthiasis? Webster JP, editor. PLoS Negl Trop Dis. 2013;7: e2098. doi:10.1371/journal.pntd.0002098

13. Irvine MA, Reimer LJ, Njenga SM, Gunawardena S, Kelly-Hope L, Bockarie M, et al. Modelling strategies to break transmission of lymphatic filariasis - aggregation, adherence and vector competence greatly alter elimination. Parasit Vectors. 2015;8: 547. doi:10.1186/s13071-015-1152-3

14. Wang S, Spear RC. Exploring the Contribution of Host Susceptibility to Epidemiological Patterns of Schistosoma japonicum Infection Using an Individual-Based Model. Am J Trop Med Hyg. 2015;92: 1245–1252. doi:10.4269/ajtmh.14-0691

15. Wang S, Spear RC. Exposure versus Susceptibility as Alternative Bases for New Approaches to Surveillance for Schistosoma japonicum in Low Transmission Environments. Mutapi F, editor. PLoS Negl Trop Dis. 2016;10: e0004425. doi:10.1371/journal.pntd.0004425

16. Collyer BS, Anderson RM. Probability distributions of helminth parasite burdens within the human host population following repeated rounds of mass drug administration and their impact on the transmission breakpoint. 2021. doi:10.1098/rsif.2021.0200

17. Toor J, Truscott JE, Werkman M, Turner HC, Phillips AE, King CH, et al. Determining post-treatment surveillance criteria for predicting the elimination of Schistosoma mansoni transmission. Parasit Vectors. 2019;12: 437. doi:10.1186/s13071-019-3611-8

18. Kittur N, Castleman JD, Campbell CH, King CH, Colley DG. Comparison of Schistosoma mansoni Prevalence and Intensity of Infection, as Determined by the Circulating Cathodic Antigen Urine Assay or by the Kato-Katz Fecal Assay: A Systematic Review. Am J Trop Med Hyg. 2016;94: 605–610. doi:10.4269/ajtmh.15-0725

19. Knopp S, Ame SM, Hattendorf J, Ali SM, Khamis IS, Bakar F, et al. Urogenital schistosomiasis elimination in Zanzibar: Accuracy of urine filtration and haematuria reagent strips for diagnosing light intensity Schistosoma haematobium infections. Parasites and Vectors. 2018;11: 552. doi:10.1186/s13071-018-3136-6

20. Knopp S, Ame SM, Person B, Hattendorf J, Rabone M, Juma S, et al. A 5-Year intervention study on elimination of urogenital schistosomiasis in Zanzibar: Parasitological results of annual cross-sectional surveys. French M, editor. PLoS Negl Trop Dis. 2019;13: e0007268. doi:10.1371/journal.pntd.0007268

21. Knopp S, Person B, Ame SM, Ali SM, Hattendorf J, Juma S, et al. Evaluation of integrated interventions layered on mass drug administration for urogenital schistosomiasis elimination: a cluster-randomised trial. Lancet Glob Heal. 2019;7: e1118–e1129. doi:10.1016/S2214-109X(19)30189-5

22. Person B, Knopp S, Ali SM, A’Kadir FM, Khamis AN, Ali JN, et al. Community co-designed schistosomiasis control interventions for school-aged children in Zanzibar. J Biosoc Sci. 2016;48: S56–S73. doi:10.1017/S0021932016000067

23. Celone M, Person B, Ali SM, Lyimo JH, Mohammed UA, Khamis AN, et al. Increasing the reach: Involving local Muslim religious teachers in a behavioral intervention to eliminate urogenital schistosomiasis in Zanzibar. Acta Trop. 2016;163: 142–148. doi:10.1016/j.actatropica.2016.08.004

24. May RM. Togetherness among Schistosomes: its effects on the dynamics of the infection. Math Biosci. 1977;35: 301–343. doi:10.1016/0025-5564(77)90030-X

25. Bradley DJ, May RM. Consequences of helminth aggregation for the dynamics of schistosomiasis. Trans R Soc Trop Med Hyg. 1978;72: 262–273. doi:10.1016/0035-9203(78)90205-5

26. Anderson RMRM, May RMRM. Infectious Diseases of Humans. New York, NY, NY: Oxford University Press; 1991.

27. Lloyd-Smith JO. Maximum Likelihood Estimation of the Negative Binomial Dispersion Parameter for Highly Overdispersed Data, with Applications to Infectious Diseases. Rees M, editor. PLoS One. 2007;2: e180. doi:10.1371/journal.pone.0000180

28. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.

29. Van Lieshout L, Polderman AM, De Vlas SJ, De Caluwé P, Krijger FW, Gryseels B, et al. Analysis of Worm Burden Variation in Human Schistosoma mansoni Infections by Determination of Serum Levels of Circulating Anodic Antigen and Circulating Cathodic Antigen. J Infect Dis. 1995;172: 1336–1342. doi:10.1093/infdis/172.5.1336

30. Hubbard A, Liang S, Maszle D, Qiu D, Gu X, Spear RC. Estimating the distribution of worm burden and egg excretion of Schistosoma japonicum by risk group in Sichuan Province, China. Parasitology. 2002;125: 221–31. Available: http://www.ncbi.nlm.nih.gov/pubmed/12358419

31. De Vlas SJ, Van Oortmarssen GJ, Habbema JD, Gryseels B, Polderman AM. A model for variations in single and repeated egg counts in schistosoma mansoni infections. Parasitology. 1992;104: 451–460. doi:10.1017/S003118200006371X

32. Medley G, Anderson RM. Density-dependent fecundity in schistosoma mansoni infections in man. Trans R Soc Trop Med Hyg. 1985;79: 532–534. doi:10.1016/0035-9203(85)90087-2

33. Polman K, De Vlas SJ, Van Lieshout L, Deelder AM, Gryseels B. Evaluation of density-dependent fecundity in human Schistosoma mansoni infections by relating egg counts to circulating antigens through Deming regression. Parasitology. 2001;122: 161–167. doi:10.1017/S0031182001007193

34. Neves MI, Gower CM, Webster JP, Walker M. Revisiting density-dependent fecundity in schistosomes using sibship reconstruction. Beechler BR, editor. PLoS Negl Trop Dis. 2021;15: e0009396. doi:10.1371/journal.pntd.0009396

35. Csilléry K, François O, Blum MGB. Abc: An R package for approximate Bayesian computation (ABC). Methods Ecol Evol. 2012;3: 475–479. doi:10.1111/j.2041-210X.2011.00179.x

36. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. J Stat Softw. 2006;15: 1–11. doi:10.18637/jss.v015.i02

37. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4: 1686. doi:10.21105/joss.01686

38. Klepac P, Metcalf CJE, McLean AR, Hampson K. Towards the endgame and beyond: complexities and challenges for the elimination of infectious diseases. Philos Trans R Soc B Biol Sci. 2013;368: 20120137. doi:10.1098/rstb.2012.0137

39. Landouré A, Dembélé R, Goita S, Kané M, Tuinsma M, Sacko M, et al. Significantly Reduced Intensity of Infection but Persistent Prevalence of Schistosomiasis in a Highly Endemic Region in Mali after Repeated Treatment. Brooker S, editor. PLoS Negl Trop Dis. 2012;6: e1774. doi:10.1371/journal.pntd.0001774

40. Toor J, Alsallaq R, Truscott JE, Turner HC, Werkman M, Gurarie D, et al. Are We on Our Way to Achieving the 2020 Goals for Schistosomiasis Morbidity Control Using Current World Health Organization Guidelines? Clin Infect Dis Achiev WHO Schistosomiasis Goals • CID. 2018: 66. doi:10.1093/cid/ciy001

41. Colley DG, Andros TS, Campbell CH. Schistosomiasis is more prevalent than previously thought: what does it mean for public health goals, policies, strategies, guidelines and intervention programs? Infect Dis Poverty. 2017;6: 63. doi:10.1186/s40249-017-0275-5

42. Spear RC, Seto EYW, Carlton EJ, Liang S, Remais J V, Zhong B, et al. The challenge of effective surveillance in moving from low transmission to elimination of schistosomiasis in China. Int J Parasitol. 2011;41: 1243–7. doi:10.1016/j.ijpara.2011.08.002

43. Arakala A, Hoover CM, Marshall JM, Sokolow SH, De Leo GA, Rohr JR, et al. Estimating the elimination feasibility in the “end game” of control efforts for parasites subjected to regular mass drug administration: Methods and their application to schistosomiasis. Dobson AP, editor. PLoS Negl Trop Dis. 2018;12: e0006794. doi:10.1371/journal.pntd.0006794

44. Boissier J, Grech-Angelini S, Webster BL, Allienne J-F, Huyse T, Mas-Coma S, et al. Outbreak of urogenital schistosomiasis in Corsica (France): an epidemiological case study. Lancet Infect Dis. 2016;16: 971–979. doi:10.1016/S1473-3099(16)00175-4

45. Sokolow SH, Wood CL, Jones IJ, Swartz SJ, Lopez M, Hsieh MH, et al. Global assessment of schistosomiasis control over the past century shows targeting the snail intermediate host works best. Caffrey CR, editor. PLoS Negl Trop Dis. 2016;10: e0004794. doi:10.1371/journal.pntd.0004794

46. Lo NC, Gurarie D, Yoon N, Coulibaly JT, Bendavid E, Andrews JR, et al. Impact and cost-effectiveness of snail control to achieve disease control targets for schistosomiasis. Proc Natl Acad Sci U S A. 2018;115: E584–E591. doi:10.1073/pnas.1708729114

47. Gurarie D, Lo NC, Ndeffo-Mbah ML, Durham DP, King CH. The human-snail transmission environment shapes long term schistosomiasis control outcomes: Implications for improving the accuracy of predictive modeling. Basáñez M-G, editor. PLoS Negl Trop Dis. 2018;12: e0006514. doi:10.1371/journal.pntd.0006514

48. Civitello DJ, Rohr JR. Disentangling the effects of exposure and susceptibility on transmission of the zoonotic parasite *Schistosoma mansoni*. Fenton A, editor. J Anim Ecol. 2014;83: 1379–1386. doi:10.1111/1365-2656.12222

49. Liu Y, Zhong B, Wu Z-S, Liang S, Qiu D-C, Ma X. Interruption of schistosomiasis transmission in mountainous and hilly regions with an integrated strategy: a longitudinal case study in Sichuan, China. Infect Dis Poverty. 2017;6: 79. doi:10.1186/s40249-017-0290-6

50. Webster BL, Diaw OT, Seye MM, Webster JP, Rollinson D. Introgressive Hybridization of Schistosoma haematobium Group Species in Senegal: Species Barrier Break Down between Ruminant and Human Schistosomes. Secor WE, editor. PLoS Negl Trop Dis. 2013;7: e2110. doi:10.1371/journal.pntd.0002110

51. Rudge JW, Webster JP, Lu D-B, Wang T-P, Fang G-R, Basáñez M-G. Identifying host species driving transmission of schistosomiasis japonica, a multihost parasite system, in China. Proc Natl Acad Sci U S A. 2013;110: 11457–62. doi:10.1073/pnas.1221509110