Methods

Overview of PIHM approaches

We propose and two alternative approaches for estimating LD50 and PD from parasite intensity data. Our two approaches, as well as the Adjei Method, follow the same general steps. First, all approaches assume that the distribution of parasites across hosts in the pre-mortality population follows a negative binomial distribution and, using the Crofton Method, each approach starts by estimating the total number hosts (N_p), the mean number of parasites per host (μ_p), and the aggregation of the parasites across hosts before PIHM (k_p) (see Appendix X for a description of the Crofton Method). We discuss in the following sections that our alternative approaches do not have to start with the Crofton Method, but is a useful starting point when comparing our methods to the Adjei Method. Second, each method assumes that the probability of host survival with x parasites is given by the logistic host-survival function

$$h(x|a,b) = h_x = \frac{e^{a+b\log(x)}}{1+e^{a+b\log(x)}} \tag{1}$$

Using equation 1, each method uses either a likelihood or chi-squared approach to estimate a and b. Finally, from the estimated parameters a and b, LD_{50} and PD can be calculated by the equations (Adjei *et al.* 1986)

$$LD_{50} = \exp(\frac{a}{b}) \tag{2}$$

and

$$PD = 1 - \sum_{x=0}^{\infty} h_x * f_x \tag{3}$$

where f_x is the probability mass function for the pre-mortality distribution of

parasites per hosts which follows a negative binomial distribution with mean μ_p and aggregation parameter k_p. For each method, any differences in predictions of LD_{50} and PD are a result of how a given method each calculates a and b.

Estimating a and b with the Adjei Method

To estimate a and b, the Adjei Method proceeds as follows (Adjei *et al.* 1986)

1. Estimate N_p, μ_p, and k_p using the Crofton Method
2. Given N_p, μ_p, and k_p, estimate the expected number of hosts in category i via equation $N_p f_i$. i specifies either a given parasite intensity or some range of parasites intensities (i.e. number of hosts with 100 parasites or number of hosts with 100-150 parasites) and there are m categories.
3. Calculate the observed number of hosts in each category i. Assume that the observed number of hosts in each category i is binomial distributed with the total number of "trials" equal to the expected number of hosts in i.
4. If observed number of hosts in category i is greater than the expected number of hosts in i, let observed in i equal to expected in i.
5. Run a generalized linear model in with a binomial random component and a logistic link to estimate a and b.

See Adjei *et al.* (1986) and Appendix X for a more thorough description of this approach and some supporting examples.

Estimating a and b with chi-squared approach

The first alternative approach that we propose relies on estimating a and b by minimizing the χ^2 statistic. The approach proceeds as follows

1. Estimate N_p, μ_p, and k_p using the Crofton Method.
2. Specify m categories of parasite intensities where each category i either contains a single parasite intensity or a range of parasite intensities. Calculate the expected number of hosts in each category i by N_p times the probability that a host is category i and is alive. This can be written as $N_p * f_i * h_i$ where the appropriate summation is taken if i is a range of parasite intensities.

3. Calculate the observed number of hosts in each category i.
4. Calculate the χ^2 statistic for the observed and expected n categories using the standard notation $\chi^2 = \sum_{i=1}^{m} \frac{(observed_i - expected_i)^2}{expected_i}$
5. Find a and b that minimize this χ^2 statistic.

The advantage of this approach over the Adjei approach is that is does not require any alteration of the observed data. Moreover, this approach can theoretically be extended such that N_p, k_p and μ_p can be estimated directly from minimization of the χ^2 statistic rather than via the Crofton Method. In practice, this minimization is difficult because host-parasite datasets are often not large enough and/or PIHM is not strong enough to uniquely estimate 5 parameters. Therefore, the Crofton Method provides an easy way to estimate N_p, k_p, and μ_p before minimizing the χ^2 statistic to find a and b.

Estimating a and b with the likelihood approach

The second alternative method relies on maximizing the likelihood of the observed data under a PIHM model. This method proceeds as follows.

1. Estimate N_p, μ_p, and k_p using the Crofton Method.
2. Specify the probability distribution that a host has x parasites and is alive as

$$p(x; a, b, \mu_p, k_p) = \phi h_x f_x \tag{4}$$

   where ϕ is a normalizing constant. equal to $(\sum_{x=0}^{\infty} h(x; a, b) * f(x; \mu_p, k_p))^{-1}$.
3. Assuming hosts are independent, the likelihood of a datasets with n hosts can be written as $L(a, b|x, \mu_p, k_p) = \Pi_{i=0}^{n} p(x_i; a, b, \mu_p, k_p)$.
4. Use standard optimization techniques to estimate a and b

The likelihood approach is advantageous over both of the aforementioned approaches because 1) it requires only estimating 4 parameters (a, b, μ_p, and k_p) rather than 5 parameters 2) all 4 parameters can either be estimated jointly or μ_p

and k_p can first be estimated using the Crofton method and 3) standard statistical techniques, such as likelihood ratio tests or AIC, can be used to assess whether a model with PIHM is better than a model without PIHM.

Dataset simulation and comparison of PIHM approaches

To compare the ability of these methods to recover LD_{50} and PD, we randomly generated data using the following procedure. First, we drew N_p randomly infected hosts from a negative binomial distribution with parameters μ_p and k_p. Second, we chose values of a and b and calculated the probability of survival for all N_p hosts using equation 1. Third, we drew N_p random numbers from a uniform distribution between 0 and 1 and if host survival probability was less than this random number, the host experienced parasite-induced mortality. The surviving hosts comprised the dataset that would be obtained in the field.

Using the simulated datasets, we devised two scenarios to test the above approaches. In the first scenario, we assumed that the values of N_p, μ_p, and k_p were known and tested the ability of both methods to recover the true values of a and b over increasing values of N_p. While this first scenario is unrealistic because the parameters N_p, μ_p, and k_p are always unknown, we implemented this scenario as a baseline to establish the efficacy of the methods independent of the Crofton Method used to estimate N_p, μ_p and k_p. If a method could not return unbiased [word choice] estimates of LD_{50} and/or PD under these idealized conditions, we took this as evidence for unreliability of this method. The second scenario proceeded exactly as the first, except that we estimated N_p, mu_p, and k_p using the Crofton Method.

For each scenario, we used two different LD_{50} values: $\exp(2) = 7.39$ parasites with $\mu_p = 10$ and $\exp(3) = 20.08$ parasites with $\mu_p = 30$. For each LD_{50} value we choose three combinations of a and b that represented three different

patterns of how host survival decreased with parasite intensity: gradual, moderate and a sharp decrease in host survival probability (Figure X). Finally we examined three levels of parasite aggregation, $k_p = 0.1, 0.5, 1$, which are realistic values of parasite aggregation in natural populations (Shaw *et al.* 1998). For each of these parameter combinations we simulated 150 datasets, estimated a, b, LD_{50}, and PD and calculated the standardized bias and precision (Walther & Moore 2005) for these estimates over varying pre-mortality host population sizes N_p = [300, 500, 1000, 2000, 5000, 7500, 10000]. N_p is not technically the sample size on which the methods are being tested because parasite-induced mortality reduces N_p for each simulated dataset. We therefore used the average number of surviving hosts over all 150 simulations for a given parameter combination as our measure of sample size.

Results

Scenario I: Known pre-mortality parameters

Across all parameter combinations we examined, the two alternative approaches always provided less or equal bias when estimating LD_{50} and PD than the Adjei Method. For large sample sizes, all three methods were approximately equally effective at predicting LD_{50} and PD (Figure X). However, as sample size decreased the Adjei Method began to show significant bias in its predictions of LD_{50} and PD. As sample sized decreased, the Adjei Method tended to first show bias in host-parasite systems when the host-survival decreased sharply with increasing parasite intensity (i.e. $a = 30, b = -15$). Approaching the smallest sample size, the Adjei Method generally showed the bias for all types of host-survival functions.

For all three methods, the aggregation in the pre-mortality host population had strong effects on the bias of the LD_{50} and PD estimates. As k_p decreased,

the bias increased for all methods. This trend was most pronounced in the Adjei Method followed by the Chi-squared Method and was least pronounced in the Likelihood Method. These general trend held for different LD_{50}s (Appendix X). [Run same analyses with a different mean]

Scenario II: Unknown pre-mortality parameters

When the Crofton Method was first used to estimate

Scenario III: Efficacy of alternative methods under-realistic conditions

As both alternative methods were superior to the Adjei Method for estimating the host survival function, we focused whether either of these methods would be practical under realistic field conditions.

Discussion

References

1.
Adjei, E. L., Barnes, A. & Lester, R. J. G. (1986). A method for estimating possible parasite-related host mortality, illustrated using data from Caliitetrarhynchus gracilis (Cestoda: Trypanorhyncha) in lizardfish (Saurida spp.). *Parasitology*, 92, 227–243.

2.
Shaw, D. J., Grenfell, B. T. & Dobson, a. P. (1998). Patterns of macroparasite aggregation in wildlife host populations. *Parasitology*, 117 (Pt 6, 597–610.

3.
Walther, B. a. & Moore, J. L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, 28, 815–829.