Title: Detecting and quantifying parasite-induced host mortality from intensity data: method comparisons and limitations

Authors: Mark Q. Wilber, Sara B. Weinstein and Cheryl J. Briggs

Affiliations: 1) Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, Santa Barbara California, United States of America

Corresponding author:
Mark Q. Wilber
University of California, Santa Barbara
Department of Ecology, Evolution, and Marine Biology
2111 Noble Hall
Santa Barbara, CA 93106
mark.wilber@lifesci.ucsb.edu

Note: Supplementary data associated with this article

Abstract

Parasites can significantly impact animal populations by changing host behavior, reproduction and survival. Detecting and quantifying these impacts is critical for understanding disease dynamics and managing wild animal populations. However, for wild hosts infected with macroparasites, it is notoriously difficult to quantify the fatal parasite load and number of animals that have died due to disease. When ethical or logistical constraints prohibit experimental determination of these values, examination of parasite intensity and distribution data may offer an alternative solution. In this study we introduce a novel method for using intensity data to detect and quantify parasite-induced mortality in wildlife populations. We use simulations to show that this method is more reliable than previously proposed methods while providing quantitative estimates of parasite-induced mortality from empirical data that are consistent with previously published qualitative estimates. However, this method, and all techniques that estimate parasite-induced mortality from intensity data alone, have several important assumptions that must be scrutinized before applying them to real-world data. Given that these assumptions are met, our method is a new exploratory tool that can help inform more rigorous studies of parasite-induced host mortality.

Keywords: parasite aggregation, negative binomial distribution, Crofton Method, host survival function, lethal dose

1 Introduction

Infectious agents can impact animal populations by changing population dynamics and stability (**??**), altering predator-prey interactions (**?**), and even causing species' decline and extinction (**??**). Accurately estimating the impact of these infectious agents in wildlife is critical to understanding what regulates host and parasite populations, making predictions about disease transmission, and managing disease outbreaks (**?**). The impact of pathogens, such as rabies (**?**), bovine tuberculosis (**?**), and rinderpest (**?**), are typically modeled based on the presence or absence of disease, such that host survival is not generally considered to be a function of the number of infectious agents present

within the host. In contrast, models of macroparasites generally assume that pathology increases with parasite burden and host survival probability must be treated as a function of infection intensity (**?**). Helminths exhibiting this intensity-dependent pathology have significant impacts on human health (**?**), domestic livestock economics (**?**), and wildlife survival (**??**). While it is generally assumed that some fraction of wild host populations succumb to parasitic infection, it is notoriously difficult to actually quantify parasite-induced host mortality (PIHM) in wild animal populations because it is difficult to observe the dead or dying hosts most impacted by parasitism (**?**).

Ideally, parasite-induced host mortality is quantified by experimentally infecting and tracking individual hosts in the wild population; however, for logistical and ethical reasons this method is rarely feasible (**?**). Snapshot data of parasite intensities across multiple hosts is much easier to collect and has often been used to identify the presence of PIHM (**??????**) and to quantify the relationship between infection intensity and host mortality (**?**).

? first proposed that PIHM could be identified from parasite intensity data by comparing the observed parasite distribution in sampled hosts to the distribution predicted in the absence of parasite-induced mortality. This method assumes that, prior to host mortality, infection intensity in the host population follows a negative binomial distribution and the tail of the distribution is truncated as intensity dependent pathology removes the most heavily infected hosts. Assuming mortality occurs only in heavily infected hosts, evidence of this parasite-induced mortality should then be detectable by iteratively fitting a negative binomial distribution to hosts with lower and lower parasite intensities, and comparing these truncated predicted distributions to the corresponding truncated observed parasite data (Fig. 1, see Supplementary Data S1 for additional detail).

While the Crofton Method detects the presence of PIHM, it makes no attempt to quantify the relationship between infection intensity and host survival probability; information that is necessary for estimating parasite impacts on host populations (**??**).

70 **?** suggested that this relationship could be calculated by first using the Crofton Method
71 to estimate the pre-mortality parasite distribution and then using this distribution to
72 calculate the probability of host survival with increasing parasite intensity. To do this,
73 **?** modeled host survival as a logistic function and then used a generalized linear model
74 (GLM) to estimate the parameters of the host survival function (see Supplementary Data
75 S2 for a technical description of the Adjei Method). Although this method can predict the
76 host survival function, it has several technical drawbacks. When mean infection intensity
77 is high or sample sizes are small the observed intensity data must be subjectively binned
78 into intensity ranges in order to fit the GLM framework. Furthermore, for the Adjei
79 Method to work, any observed intensity values greater than predicted values must be
80 modified and set equal to the predict values (see Supplementary Data S2 for details); a
81 questionable act of data manipulation. These manipulations may introduce bias, reduce
82 the precision and limit the power of this method to detect and quantify parasite-induced
83 host mortality.

84 After 30 years, and despite clear limitations (**?**), these methods (particularly the
85 Crofton Method) are still discussed among parasitologists and are the primary techniques
86 for examining population-level impacts of parasitism using parasite intensity data. In
87 these methods, PIHM can only be identified by visually examining plots of the pre-
88 mortality parameters predicted by the Crofton Method and determining whether they
89 show a "kink" over a range of truncation values (Fig. 1B; **??**). These qualitative criteria
90 makes it difficult to compare PIHM between studies and a more rigorous and quantitative
91 method is needed to both detect and quantify host mortality. The survival function given
92 by the Adjei Method may be used to do this; however, it requires manipulating the
93 original data and its accuracy remains untested.

94 In this study, we propose a novel method for detecting and quantifying PIHM that
95 ameliorates many of the aforementioned deficiencies of the previous methods. Our method
96 does not require data alteration, is highly generalizable, and uses standard statistical
97 techniques to quantitatively determine whether PIHM is occurring in a system. We use

simulations to compare our method with the Adjei Method to test the ability of both to (1) detect occurrence of PIHM and (2) estimate the host survival function. We then apply both methods to real datasets previously used in PIHM analyses and compare the results. Finally, we discuss the limitations of inferring PIHM from intensity data and how these methods fit in modern quantitative parasitology.

2 Materials and methods

2.1 A novel, likelihood-based method for estimating PIHM

Our method (henceforth the Likelihood Method) begins with the same assumptions as the Adjei Method: namely that infection has occurred and hosts with fatal parasite loads have died prior to the population sampling. As discussed by **?**, this is not necessarily unrealistic as some parasite infections occur primarily in younger hosts with parasite-induced mortality occurring soon after infection (e.g. **??**).

The Likelihood Method then assumes that prior to mortality the parasite distribution can be described by the distribution $g(x;\phi)$, which specifies the probability of a host having x parasites before mortality occurs. ϕ is a vector of parameters that describes the shape of this distribution. The probability of a host surviving with x parasites from infection until sampling is given by the host survival function $h(\text{survival};x,\theta)$ where θ specifies any additional parameters needed to define the host survival function.

With these two assumptions, we can define a distribution that gives the probability of having a parasite load of x parasites conditional on host survival, $P(x|\text{survival})$. Using standard rules of conditional probability this distribution can be written as

$$P(x|\text{survival}) = \frac{P(\text{survival}|x) * P(x)}{P(\text{survival})} \tag{1}$$

$P(\text{survival}|x)$ is the survival function $h(\text{survival};x,\theta)$, $P(x)$ is the pre-mortality parasite distribution $g(x;\phi)$ and $P(\text{survival}) = \sum_{x=0}^{\infty} P(\text{survival}|x) * P(x) = \sum_{x=0}^{\infty} h(\text{survival};x,\theta) *$

$g(x;\phi)$. Therefore, equation 1 can be written as

$$P(x|\text{survival}) = \frac{h(\text{survival};x,\theta) * g(x;\phi)}{\sum_{x=0}^{\infty} h(\text{survival};x,\theta) * g(x;\phi)} \quad (2)$$

Using this probability distribution, one can then find the parameters θ and ϕ that maximize the likelihood of an observed host-parasite dataset. To estimate the significance of PIHM in a host-parasite system, a likelihood ratio test can be used in which the full model is given by equation 2 and the reduced model is given by the pre-mortality distribution $g(x;\phi)$. If PIHM is not significant in the system, the resulting likelihood ratio statistic should approximately follow a χ^2 distribution with degrees of freedom equal to the number of parameters in the full model with parasite-induced mortality minus the number of parameters in the reduced model without parasite-induced mortality (**?**).

The parameterization of equation 2 depends on the parasite system of interest. Here, we assume that the pre-mortality parasite distribution $g(x;\phi)$ follows a negative binomial distribution with two parameters mean parasite intensity (μ_p) and aggregation (k_p, where smaller k_p indicates a more aggregated parasite population) before mortality (**???**). A variety of different biological and statistical assumptions can result in an equilibrium parasite distribution that follows a negative binomial distribution (**???**). Furthermore, the negative binomial distribution is an incredibly flexible distribution that fits many host-parasite systems even when the underlying mechanisms determining the empirical distribution are unknown (**?**).

The function for $h(\text{survival};x,\theta)$ is also system specific. Many theoretical models of parasite-induced host mortality assume that the parasite-induced death rate of hosts is a linear function of parasite intensity (**???**). In systems where there is truly a linear relationship between infection intensity and survival probability it will be nearly impossible to use intensity data to detect parasite-induced host mortality (**?**). However, some systems do exhibit non-linear host survival functions (**?**), in which case these methods would be applicable.

To compare the Likelihood Method and the previously proposed Adjei Method, we adopt the non-linear, logistic host-survival function used in the earlier study given by

$$h(\text{survival}; x, a, b) = \frac{\exp\left(a - b\log(x)\right)}{1 + \exp\left(a - b\log(x)\right)} \tag{3}$$

Generally, a larger b leads to a more rapid decline in the probability of host survival as parasite intensity increases, with the maximum rate of decline having a value of $b/4$ (Supplementary Data S2). b is in many ways analogous to the pathogenicity parameter (α) in classic macroparasite models that gives the parasite intensity dependent host death rate (**??**). When b is held constant, a larger a allows for hosts to tolerate larger parasite intensities before experiencing parasite-induced mortality. More specifically, for every one unit increase in a the log parasite intensity at which any percent of hosts survive (e.g. 99% of hosts survive) increases by $1/b$ (Supplementary Data S2).

The equation $\exp(a/b)$ can also be used to calculate the parasite LD_{50}, here defined as the infection intensity above which a host has greater than 50% probability of dying. Equation 3 is commonly used in toxicology and has the useful properties of being bounded between 0 and 1 and being differentiable for all x (**?**). That being said, it is phenomenological and is used simply because it tends to fit survival data. However, given that a goal of these analyses is to compare the Likelihood Method's results to the Adjei Method, it is natural to adopt the same host-survival function to facilitate comparison. When applying the Likelihood Method to other systems more mechanistic host-survival functions can be used in place of equation 3.

2.2 Evaluating the Adjei and Likelihood Methods

Question 1: Can we detect PIHM?

We used statistical power and Type I error to test the ability of the Adjei Method and the Likelihood Method to correctly identify the presence of PIHM on simulated data with known pre-mortality parameters. The power of a method is the probability of correctly detecting PIHM given that it is occurring and the Type I error is the probability

of incorrectly identifying PIHM given that it is not occurring. If a method has low Type I error we can be confident that when we detect PIHM it is actually occurring. If one method has higher power for detecting PIHM than another, we will need to sample fewer hosts to detect PIHM.

Consistent with the model assumption that parasite infection, host mortality, and population sampling are temporally separate events, we first created a pre-mortality host population by drawing N_p randomly infected hosts from a negative binomial distribution with parameters μ_p and k_p. This represents a host population that has become infected but not yet experienced parasite-induced mortality (**?**). In the Adjei Method and Crofton Method, N_p is a necessary parameter defined as the number of hosts in the population before parasite-induced mortality. More accurately, N_p is the number of hosts that would have been sampled had parasite-induced host mortality not occurred. This parameter is not necessary when using the Likelihood Method because, unlike the Adjei Method and Crofton Method which estimate parasite-induced mortality using absolute numbers of hosts, the Likelihood Method estimates parasite-induced mortality using probabilities. However, to compare the results of the Likelihood Method with the Adjei Method, we specified a value for N_p for all simulations.

We next chose values of a and b for the host survival function and calculated the probability of survival for all N_p hosts using equation 3. Then, to simulate the period in which hosts died due to infection, for each host we drew a random number from a uniform distribution between 0 and 1 and if the calculated host survival probability was less than this random number, the host experienced parasite-induced mortality. The surviving individuals represent the post-mortality hosts that would be sampled in the field.

We then used these simulated pre-mortality and post-mortality datasets to test the ability of both methods to correctly determine whether or not PIHM was occurring when the parameters N_p, μ_p and k_p were known. Although the parameters N_p, μ_p, and k_p are always unknown in real systems, a method that fails under these ideal simulation

conditions with known parameters will certainly also fail when these values must be estimated from empirical data. In practice, for the Adjei Method, N_p, μ_p, and k_p are estimated using the Crofton Method (**?**), while μ_p and k_p in the Likelihood Method can be estimated jointly with a and b or via the Crofton Method.

We compared the two methods using three different mean parasite intensity values (μ_p = 10, 50, 100) and three different host survival functions (gradual, moderate, and steep decreases in the host survival with increasing parasite intensity, Fig. 2A). For a given μ_p, each survival function had the same LD_{50} ([μ_p = 10, LD_{50} = 7.39], [μ_p = 50, LD_{50} = 35.57], [μ_p = 100, LD_{50} = 77.3]), but different values of a and b. We examined each μ_p-survival function pair at three levels of parasite aggregation, k_p = 0.1, 0.5, and 1 — realistic values of parasite aggregation in natural populations (**?**). For each of these 27 parameter combinations we simulated 150 datasets and tested the probability of each method correctly identifying PIHM in the post-mortality dataset (power) and incorrectly identifying PIHM in the pre-mortality dataset (Type I error). For each method, we used a likelihood ratio test to determine whether the full model with PIHM provided a significantly better fit than the reduced model without PIHM at significance level of 0.05. We also examined the impact of sample size by simulating each parameter for pre-mortality sample sizes of N_p = [50, 100, 200, 300, 400, 500]. Wild host populations were assumed to be sampled after PIHM has occurred, thus we calculated the sample size in the power simulations as the average number of surviving hosts over all 150 simulations for each parameter combination. The distribution of surviving hosts over the 150 simulations was generally symmetrical and the standard deviation was small compared to the mean (maximum coefficient of variation was approximately 0.06 across all parameter combinations), suggesting that the mean number of surviving hosts was an adequate summary statistic of the number of hosts sampled post-mortality.

We then tested the ability of the Likelihood Method to correctly identify PIHM under the more realistic condition of unknown pre-mortality parameters. Based on the first set of simulations, we excluded the Adjei Method and only examined the power

of the Likelihood Method under "best-case" scenario parameter values, setting $\mu_p = 10$ and $k = 1$ because PIHM is most detectable when parasites are less clumped and mean intensity is low. We examined the impact of survival function shape and sample size on the Likelihood Method's ability to identify PIHM when the pre-mortality parameters μ_p and k_p and the survival function parameters a and b needed to be estimated. We performed 500 simulations over a range of different samples sizes for gradual, moderate, and steep survival functions, following the simulation procedure described above.

Question 2: Can we estimate properties of the host survival function?

In the previous section we compared the ability of the Adjei Method and the Likelihood Method to provide a "yes" or "no" answer for whether or not PIHM was occurring in a system. In this section we compared the ability of the Adjei Method and the Likelihood Method to estimate properties of the survival function such as the parameters a, b and LD_{50}. Using the same simulation procedure and parameter combinations described above, we simulated 150 datasets, estimated a, b, and LD_{50} and calculated the standardized bias and precision for these estimates (**?**). Because estimating properties of the host survival function requires more information than simply detecting PIHM, we used larger values of N_p for this simulation (N_p = [300, 500, 1000, 2000, 5000, 7500, 10000]). We used the average number of surviving hosts for each set of 150 simulated datasets as our measure of sample size. Although both a and b are necessary to estimate LD_{50}, the two parameters showed similar patterns of bias and precision so we only show the results for a.

2.3 Application to real data

We tested the ability of the Adjei Method and the Likelihood Method to identify PIHM in six host-parasite datasets given in **?** and four datasets given in **?** (Table 1). **?** analyzed infection patterns in the snail *Gammarus pulex* infected with the acanthocephalan *Polmorphus minutus*. **?** analyzed males and females of two species of lizard fish *Saurida tumbil* and *S. undosquamis* that were infected by the cestode

Callitetrarhynchus gracilis.

In both earlier studies, the authors reported PIHM in some of the datasets and we tested whether the Adjei Method and/or the Likelihood Method also predicted PIHM. For the six datasets from **?**, we used the general conclusions of the author and truncated the data at four parasites, applied the Crofton Method to estimate the pre-mortality distribution, and then ran the Likelihood Method and Adjei Method using these pre-mortality parameters. For the **?** datasets, we followed the same procedure as the authors and first truncated the data at two parasites and then fit the Crofton Method for the female fish of both species. Then, following the **?**'s methods, we parameterized the male pre-mortality distributions for each species with the results from the females. Finally, we applied the Adjei Method and the Likelihood Method to determine whether or not PIHM was significant for these species and compared our results to those given by the authors. All code for the analyses is provided in Supplementary Data S3.

3 Results

3.1 Question 1: Detecting presence of PIHM

The power of the Adjei Method to detect PIHM in a system was close to unity for larger sample sizes and tended to decrease as sample size decreased for all survival functions (Fig. 2C; Supplementary Fig. S1-S3). The Likelihood Method had a power close to unity for all parameter combinations and sample sizes considered. With gradual survival functions, the power of the Likelihood Method decreased slightly for small samples sizes (Fig. 2C, Supplementary Fig. S1-S3).

The Adjei Method had highly inflated Type I error rates (i.e. falsely detected PIHM) for all parameter combinations that we considered (Fig. 2B; Supplementary Fig. S1-S3). This method also showed the unintuitive pattern of decreasing Type I error rate with decreasing sample size. This occurred because, at small samples sizes, intensity data must be binned before the Adjei Method can be used (Supplementary Data S2). In contrast, the Likelihood Method showed a Type I error rate at or near the pre-set level of 0.05 for all parameter combinations and sample sizes considered (Fig. 2B; Supplementary

Fig. S1-S3).

When all parameters were jointly estimated, the Likelihood Method showed highly context-dependent results even when detecting PIHM under the best-case scenario of $\mu_p = 10$ and $k_p = 1$. For steep survival curves, PIHM could be detected with a power of greater than 0.8 from a sample of less than 100 hosts (Fig. 3). However, for moderate survival functions over 400 hosts had to be sampled to achieve the same power and for gradual survival functions, no tested sample size ever achieved a power greater than 0.8 (Fig. 3).

3.2 Question 2: Estimating the LD_{50} and survival function

The Likelihood Method gave asymptotically unbiased estimates of the LD_{50} for all combinations of parameters examined in this study (Fig. 4, Supplementary Fig. S4-S6). Even for the smallest sample sizes we considered, the Likelihood Method's estimate of LD_{50} was largely unbiased, with small biases occurring for gradual host survival functions. The precision of the Likelihood Method's LD_{50} estimates decreased (increasing coefficient of variation) as sample size decreased for all parameter combinations we examined (Fig. 4, Supplementary Fig. S4-S6).

The Adjei Method produced biased estimates of the LD_{50} across nearly all parameter combinations, tending to underestimate the true value of the parameter (Fig. 4, Supplementary Fig. S4-S6). For $\mu_p = 10$, the LD_{50} estimates from the Adjei Method were largely unbiased for large samples sizes, but as μ_p increased, the Adjei Method produced biased estimates of LD_{50} across all sample sizes, with bias increasing as sample size decreased (Fig. 4, Supplementary Fig. S4-S6). The LD_{50} estimates from the Adjei Method also showed large decreases in precision with the steepest survival function across all values of μ_p (Fig. 4, Supplementary Fig. S4-S6).

In terms of the host survival function, the Likelihood Method gave unbiased estimates of survival function parameter a when sample sizes were large, however as sample size decreased these estimates became severely biased (Fig. 4, Supplementary Fig. S7-S9) The Adjei Method produced biased estimates of the host survival function

across all sample sizes, with consistently greater bias for steeper survival functions and higher mean parasite loads. (Fig. 4, Supplementary Fig. S7-S9).

3.3 Application to real data

The previous authors qualitatively detected PIHM in 7 of the 10 datasets considered (Table 1). The Likelihood Method parameterized from the pre-mortality parameters of the Crofton Method detected significant PIHM in 6 of these 7 datasets at a significance level of 0.05. The only dataset in which the Likelihood Method did not detect a significant effect of PIHM was the Adjei dataset for female *S. tumbil*. For this dataset there was a marginally significant effect of PIHM ($\chi^2_{df=2} = 5.34; p = 0.069$). The Adjei Method detected PIHM in 9 of the 10 datasets (Table 1), consistent with our simulation results showing that the Adjei Method has a high Type I error rate. Moreover, the Adjei Method estimates of the LD_{50} were quite variable for the Crofton data, consistent with our simulation results that the Adjei Method LD_{50} estimates could be imprecise for sample sizes of less than 1000 hosts (Supplementary Fig. S4-S6).

4 Discussion

Our likelihood-based method to estimate parasite-induced host mortality from observed parasite intensity data is a significant improvement over the previous methods. In simulations, it had greater power for detecting PIHM over a wider range of parameter values and also exhibited fewer false detection events (Type I errors) in both simulations and when applied to published datasets previously used in PIHM analyses. The Likelihood Method was also generally less biased and more precise when quantifying parasite-induced mortality via the host survival function for the parameters we considered. The superior performance of the Likelihood Method over the Adjei Method can be attributed to its fewer parameters, its lack of unnecessary data alteration, and its applicability across a variety of different parameter combinations. In short, the Likelihood Method is a better method for detecting and quantifying PIHM than the previously proposed Adjei Method.

Although superior to the Adjei Method, the Likelihood Method still cannot be applied to all real datasets. For host-parasite systems where host mortality occurs as a

steep, non-linear function of parasite intensity only 75 hosts must be sampled to have an 80% power in detecting PIHM. However, as the maximum slope of the survival function decreases and the function becomes somewhat linear, hundreds, or possibly thousands of hosts would have to be sampled to achieve the same result. This is consistent with previous studies which illustrate the difficulty of detecting PIHM from linear host survival functions (**?**). While it may be feasible to sample several hundred invertebrates or small fish, even the smallest sample sizes are completely unfeasible for many vertebrates, particularly the species of conservation concern where addressing the impact of parasitism would be most important. An even larger sample size would be required to identify PIHM when parasites are highly aggregated, mean infection intensity is high, or parasite prevalence is low, all of which are common in many parasitic helminths. Moreover, while linear functions make PIHM undetectable, at the other extreme, steep, non-linear survival curves produce severely biased estimates of the survival function. Given the interaction between all of these different factors, the Likelihood Method is probably limited to detecting PIHM in systems where greater than 100 hosts can be collected, parasites are common and only moderately aggregated, and substantial host mortality occurs at relatively low parasite intensity.

While we have improved on the existing methods for quantifying PIHM from parasite intensity data, all such methods require several fundamental assumptions. Nearly all current methods derive from **?** (but see **?**) and assume that, prior to any PIHM, parasites are distributed in the host population following a negative binomial distribution. But, it is fundamentally impossible to know what the pre-mortality parasite distribution was in a wild host population and it is widely recognized that different processes can lead to a variety of parasite distributions in hosts (**??**). However, the negative binomial is extremely flexible and there is substantial empirical and theoretical evidence to support the assumption that, prior to any PIHM, parasite distributions can be fit by a negative binomial distribution (**???**).

It is important to note that the flexibility of the negative binomial distribution may

also reduce our ability to detect PIHM. If a negative binomial can be fit to the observed post-mortality parasite distribution then, regardless of how lethal the parasite was, it will be impossible to detect PIHM because there is no need for a more complex model. Many observed parasite distributions are well-fit by the negative binomial distribution (**?**), suggesting that systems where these methods are applicable without any *a priori* knowledge may be uncommon. However, if one has *a priori* knowledge about some aspect of the pre-mortality distribution (e.g. assumes/knows the value of k_p, **?**), then the Likelihood Method could be applicable even if the the post-mortality distribution was well-fit by a negative binomial.

If one has evidence that the pre-mortality is not negative binomial, the generality of our method easily allows another distribution to be specified for $g(x, \phi)$. For example, one could use the resulting stationary host-parasite distribution from a stochastic host-parasite model without parasite-induced host mortality (**?**) to specify the form of $g(x, \phi)$ and then apply the techniques discussed in this paper to detect PIHM. The general requirement for the Likelihood Method to detect PIHM in a stochastic host-parasite process is that the stationary distribution of the process with mortality is significantly different than the stationary distribution without mortality. It is widely recognized that parasite-induced host mortality decreases the aggregation of host-parasite distributions relative to those without mortality (**?**), suggesting that the Likelihood Method could be generally applicable to host-parasite systems that follow the assumptions of many stochastic host-parasite models. This is an intriguing area for further research.

If the Likelihood Method is applicable and the truncation of the negative binomial distribution is detected, one must be aware that the truncation pattern may be caused by other processes such as within host density dependence, age dependent variation in host resistance and/or heterogeneous infection rates (**???**). This means that in the event that PIHM is detected, it may actually not be the result of PIHM. Moreover, if host mortality depends on parasite intensity and additional variables (e.g. host sex, host size), failure to identify these important confounding variables could significantly affect the ability

of these methods to correctly identify PIHM. However, both of these issues – inferring process from pattern and confounding variables – are well-recognized limitations of most statistical inference and are addressed via judicious model specification and selection (**?**).

As suggested by **?** these methods for estimating PIHM can provide preliminary insight into whether or not PIHM is worth further exploration. However, we stress that these methods are an exploratory tool for assessing the role of PIHM in a system, and potential users should critically evaluate whether they think they have a large enough sample size and an appropriate host survival function/post-mortality distribution for the methods developed in this paper to be applicable. Even if they are applicable, inferring PIHM from distributional data is no substitute for field experiments and an in depth understanding of the natural history of the host-parasite system under consideration.

Acknowledgments

We thank the Briggs Lab group, Kuris/Lafferty Lab group, Theoretical Ecology group at University of California, Santa Barbara (United States), and two anonymous reviewers for helpful feedback. MW was supported by National Science Foundation Graduate Research Fellowship (Grant No. DGE 1144085) and the University of California Regents (United States). CB was supported by National Institute of Health (United States) grant 1R01GM109499 from the Ecology of Infectious Disease program.

Figure 1: A schematic representation of the iterative approach of the Crofton Method. (A) The light gray shows the pre-mortality distribution that the Crofton Method is trying to estimate from the dark grey post-mortality distribution. The Crofton Method proceeds by truncating the post-mortality data at different levels (t_i, e.g. $i = 0, \ldots, 5$) and finding the pre-mortality host population size (N_p), pre-mortality mean parasite intensity (μ_p), and pre-mortality parasite aggregation (k_p) that best fit the truncated data. (B) The parameter N_p is then plotted against the truncation level t_i to determine if a "kink" occurs in the parameter values (**?**). This "kink" indicates that PIHM is occurring in the system. In the above example, PIHM is occurring in the system as visualized by the distinct "kink" at t_4.

Figure 2: The simulation results comparing the power and the Type I error of the Adjei Method and the Likelihood Method across a range of different sample sizes. (A) Five potential shapes for a host-survival functions. In the simulations we used a gradual survival function (dotted line), and moderate survival function (dashed line), and a steep survival function (solid line). The linear and immediate survival functions represent two potential extremes that we do not include in the simulations. For each of these survival functions and the parameter combinations described in the main text, we tested the Type I error and power of the Likelihood (Like.) Method and Adjei Method. (B) Gives the Type I error of each method over a range of pre-mortality sample sizes with a pre-mortality mean parasite intensity (μ_p) of 50 and pre-mortality parasite aggregation (k_p) at 0.5. The red line shows the pre-set significance level of 0.05. (C) Gives the power of each method for detecting PIHM over a range of post-mortality sample sizes for $\mu_p = 50$ and $k_p = 0.5$. In general, the Likelihood Method has higher power and lower Type I error than the Adjei Method. See the Supplementary Fig. S1-S3 for Type I error and power results for all parameter combinations.

Figure 3: The power of the Likelihood Method (Like.) to detect PIHM for gradual, moderate, and steep survival functions when all four parameters μ_p, k_p, a, and b were jointly estimated. The curves were generated from 500 simulations for 10 pre-mortality

sample sizes, N_p. The vertical, dotted-dashed lines indicate the sample size at which the power for the Likelihood Method with steep and moderate survival functions is 0.8 (75 hosts for steep functions and 408 for moderate functions). The Likelihood Method with a gradual survival function never has a power above 0.8.

Figure 4: Bias and precision (coefficient of variation) for the Likelihood Method (Like.) and Adjei Method estimates of the a parameter and the LD_{50} of the host survival function based on simulated PIHM data over a range of post-mortality sample sizes. As the coefficient of variation increases, precision decreases. The pre-mortality parameters for this simulation were $\mu_p = 50$ and $k_p = 0.5$. The figure shows the simulations for three different host survival functions (gradual, moderate, and steep), each with the same LD_{50}. Bias and precision results of LD_{50} and a for all other parameter combinations can be found in Supplementary Fig. S4-S9.

Figure S1: The type I error rate and the power of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 10$ for various shapes of the host survival function and levels of aggregation k_p. The first column gives the type I error rate of each method for falsely detecting PIHM when none is present. The red line gives the the pre-set type I error rate of $\alpha = 0.05$. The second column gives the power of a given method to detect PIHM when it is actually occurring.

Figure S2: The type I error rate and the power of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 50$ for various shapes of the host survival function and levels of aggregation k_p. The first column gives the type I error rate of each method for falsely detecting PIHM when none is present. The red line gives the the pre-set type I error rate of $\alpha = 0.05$. The second column gives the power of a given method to detect PIHM when it is actually occurring.

Figure S3: The type I error rate and the power of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 100$ for various shapes of the host survival function and levels of aggregation k_p. The first column gives the type I error rate of each method for falsely detecting PIHM when none is present. The red line gives the

the pre-set type I error rate of $\alpha = 0.05$. The second column gives the power of a given method to detect PIHM when it is actually occurring.

Figure S4: The bias and the precision of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 10$ for various shapes of the host survival function and levels of aggregation k_p when estimating LD_{50}. The first column gives the bias of each method's LD_{50} estimate over 150 simulations. The second column gives the precision of each method's LD_{50} estimate over 150 simulations.

Figure S5: The bias and the precision of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 50$ for various shapes of the host survival function and levels of aggregation k_p when estimating LD_{50}. The first column gives the bias of each method's LD_{50} estimate over 150 simulations. The second column gives the precision of each method's LD_{50} estimate over 150 simulations.

Figure S6: The bias and the precision of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 100$ for various shapes of the host survival function and levels of aggregation k_p when estimating LD_{50}. The first column gives the bias of each method's LD_{50} estimate over 150 simulations. The second column gives the precision of each method's LD_{50} estimate over 150 simulations.

Figure S7: The bias and the precision of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 10$ for various shapes of the host survival function and levels of aggregation k_p when estimating the a parameter of the host survival function. The first column gives the bias of each method's a estimate over 150 simulations. The second column gives the precision of each method's a estimate over 150 simulations.

Figure S8: The bias and the precision of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 50$ for various shapes of the host survival function and levels of aggregation k_p when estimating the a parameter of the host survival function. The first column gives the bias of each method's a estimate over 150 simulations. The second column gives the precision of each method's a estimate over 150 simulations.

Figure S9: The bias and the precision of the Likelihood Method (black lines) and

the Adjei Method (green lines) when $\mu_p = 100$ for various shapes of the host survival function and levels of aggregation k_p when estimating the a parameter of the host survival function. The first column gives the bias of each method's a estimate over 150 simulations. The second column gives the precision of each method's a estimate over 150 simulations.

Figure S10: A comparison of this paper's implementation (solid lines, circles) of the Crofton Method with the results given in **?** (dashed lines, diamonds). (A) compares the predicted number of hosts in a population pre-mortality (N_p). (B) compares the predicted parasite aggregation pre-mortality (k_p). (C) compares the χ^2 statistic for each implementation. Three of the 6 stations fit by **?** are shown here and all show that our implementation gives very similar results to those given by **?**.

Crofton, H.D. 1971. A quantitative approach to parasitism. Parasitology. 62, 179–193.

References