

1 Supplementary Data

2 1 Implementation of the Crofton Method

3 The algorithm for fitting the Crofton Method (Crofton 1971) proceeds as follows.
4 First, obtain a dataset with n hosts where each host has some parasite intensity 0
5 to p_{max} . Starting with the full dataset, guess a vector of pre-mortality parameters
6 $(N_{p1}, \mu_{p1}, k_{p1})$ where N_{p1} is the total number of hosts before mortality, μ_{p1}
7 is the parasite intensity before mortality, and k_{p1} is the parasite aggregation
8 before mortality. Given these parameters, use a negative binomial distribution
9 to calculate the predicted number of hosts with $0, 1, 2, \dots, p_{max}$ parasites. Compare
10 the expected number of hosts with $0, 1, 2, \dots, p_{max}$ parasites to the observed
11 number hosts with $0, 1, 2, \dots, p_{max}$ parasites and calculate the χ^2 -squared statistic
12 associated with the observed and predicted vectors. In reality, one often has to bin
13 the parasite intensity data because all parasite intensities are not represented in
14 the dataset. Continue to guess $(N_{p1}, \mu_{p1}, k_{p1})$ vectors until a set of parameters is
15 found that minimizes the χ^2 -squared statistic.

16 Second, choose a truncation value (t_2) such that $t_2 < p_{max}$. Truncate the
17 data such that $\text{data}_{\text{truncated}} \leq t_2$ and repeat the above iterative procedure to
18 calculate another set of parameters $(N_{p2}, \mu_{p2}, k_{p2})$ that minimizes the χ^2 -squared
19 statistic on the truncated data. Choose a new truncated value $t_3 < t_2$ and repeat
20 the first two steps. Continue to truncate the dataset until it only contains hosts
21 with 0, 1, and 2 parasites (or 3 bins). Because the method attempts to estimate
22 three parameters, at least 3 classes are needed for all 3 parameters to be identifiable
23 (Royce and Rossignol 1990).

24 Once the iterative procedure has been completed, parasite-induced host
25 mortality is traditionally identified by plotting the different truncation values t_i
26 against the different values of N_{pi} and looking for a distinct “kink” in the resulting

plot. Once the “kink” as occurred, the values of N_{pi} will typically remain close to constant as t_i is decreased further. The “true” pre-mortality parameters N_{pt} , μ_{pt} , and k_{pt} are taken to be at the point where the “kink” occurs.

We provide an implementation and unit tests of the Crofton Method in Supplementary Information 4. Figure S?? visually shows that our implementation of the Crofton Method agrees with results previously published by Crofton (1971).

2 Implementation of the Adjei Method

The Adjei Method for estimating PIHM has two steps (Adjei et al. 1986). The first step is to estimate the parameters of the pre-mortality host-parasite distribution using the Crofton Method (see *SI* 1). The three parameters estimated are the total number of hosts before mortality N_p , the mean number of parasites per host before mortality μ_p , and the aggregation of parasites before mortality given by the parameter k_p from a negative binomial distribution. When k_p is small, parasites are highly aggregated among hosts and when k_p is large parasites are more evenly distributed across hosts (Wilson et al. 2002). The implementation of the Crofton Method has been discussed at length elsewhere (e.g. Royce and Rossignol 1990; Lester 1984, and in *SI* 1) and we provide a tested implementation of the method in *SI* 4.

The second step of the Adjei Method is to make the assumption that infection, host mortality, and sampling occur in that order and are temporally separate (Adjei et al. 1986). Next, Adjei et al. assume that the host survival function follows the logistic form

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

where x is the parasite intensity in a given host and a and b are the two parameters

of the logistic function. Generally, a larger a allows for hosts to tolerate larger parasite intensities before experiencing parasite-induced mortality and a larger b leads to a more rapid decline in the probability of host survival as parasite intensity increases. The value $\exp(a/b)$ is referred to as the LD_{50} . Individuals with loads higher than this will have a greater than 50% chance of death.

By taking the first and second derivatives of equation 1, one can easily find that the maximum rate of decline in host survival probability with increasing parasite intensity occurs at the LD_{50} and has a value of $b/4$. This is in many ways analogous to the parasite pathogenicity parameter α given in classic macroparasite models, which specifies the slope of the linear relationship between host death rate and parasite intensity (Anderson and May 1978; Isham 1995). The parameter a is easily interpreted by holding b constant and looking at how a one unit change in a affects the log parasite intensity at which some percentage p of hosts experience mortality. Letting a_1 and a_2 be two different values of a and x_1 and x_2 be two different parasite intensities, a bit of rearranging of equation 1 gives.

$$\begin{aligned}\log \frac{p}{1-p} &= a_1 - b \log x_1 \\ -\log \frac{p}{1-p} &= a_2 - b \log x_2 \\ 0 &= a_1 - a_2 - b \log x_1 + b \log x_2 \\ a_2 - a_1 &= b(\log x_2 - \log x_1)\end{aligned}$$

If $a_2 - a_1 = 1$, then the change in log parasite intensity at which p percentage of hosts survive is $1/b$.

To estimate the parameters in equation 1, the Adjei Method first calculates the expected number of hosts with a given parasite load x by using the equation $g(x; \mu_p, k_p) * N_p$, where $g(x; \mu_p, k_p)$ is the negative binomial pre-mortality

70 distribution. Second, the observed and predicted number of hosts with x parasites
71 are paired as a single data point and the method then assumes that this data
72 point follows a binomial distribution with the total number of “trials” equal to
73 the predicted number of hosts and the total number of “successes” equal to the
74 observed number of hosts. In some cases, the observed number of hosts is greater
75 than the expected number of hosts and the Adjei Method alters the data so that
76 the observed is equal to the predicted (Adjei et al. 1986). After this questionable
77 manipulation, the (observed, predicted) pairs are fit to a standard Generalized
78 Linear Model (McCullagh and Nelder 1989) with a binomial response variable
79 and a logistic link function given by equation 1. This model provides estimates for
80 parameters a , b and LD_{50} .

81 While not included in the original implementation of the Adjei Method,
82 a χ^2 test with a degrees of freedom of 1 can be used to assess whether a GLM
83 model that includes parasite intensity as a predictor of host survival probability is
84 a “better” model than a GLM without this predictor. This allows the Adjei Method
85 to determine whether PIHM is a significant factor in a host-parasite system.

86 The Adjei Method’s most glaring deficiency is the need to alter the observed
87 data in order to fit the model into the binomial GLM framework. A second more
88 subtle problem with the Adjei Method is the potential need to bin data in order
89 to predict greater than one host in a given parasite intensity class. For example,
90 if the total number of hosts pre-mortality was 50, the mean number of parasites
91 per host pre-mortality was 100 and the aggregation parameter was 1, applying
92 the equation $g(x; \mu_p = 100, k_p = 1) * 50$ would result in less than 1 individual in
93 all parasite intensities x . In other words, the Adjei Method cannot be applied to
94 samples with either very high mean parasite loads, small sample sizes, or both
95 without some sort of binning of the data. While this is not a flaw *per se*, it does
96 add a certain level of subjectivity (i.e. which bins should you use?) to a method
97 that already has serious potential issues. In this analysis, we always assume the

98 Adjei Method is not binning the data, though we provide code for applying the
99 binning method in Supplementary Data S3.

100 **3 Code and unit tests for estimating parasite-induced host** 101 **mortality**

102 Python code, unit tests, and a help file for the Crofton Method, the Adjei Method
103 and the Likelihood Method can be found at [https://github.com/mqwilber/](https://github.com/mqwilber/parasite_mortality)
104 `parasite_mortality`

105 **References**

- 106 1.
107 Adjei, E.L., Barnes, A., Lester, R.J.G., 1986. A method for estimating possible
108 parasite-related host mortality, illustrated using data from *Callitetrarhynchus*
109 *gracilis* (Cestoda: Trypanorhyncha) in lizardfish (*Saurida* spp.). *Parasitology*. 92,
110 227–243.
- 111 2.
112 Anderson, R.M., May, R.M., 1978. Regulation and stability of host-parasite
113 interactions: I. Regulatory processes. *J. Anim. Ecol.* 47, 219–247.
- 114 3.
115 Crofton, H.D. 1971. A quantitative approach to parasitism. *Parasitology*. 62,
116 179–193.
- 117 4.
118 Isham, V. 1995. Stochastic models of host-macroparasite interaction. *Ann. Appl.*
119 *Probab.* 5, 720–740.
- 120 5.
121 Lester, R.J.G. 1984. A review of methods for estimating mortality due to parasites
122 in wild fish populations. *Helgolander Meeresun.* 37, 53–64.
- 123 6.
124 McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models* 2nd edn. Chapman
125 & Hall, New York.
- 126 7.
127 Royce, L.A., Rossignol, P., 1990. Epidemiology of honey bee parasites. *Parasitol.*
128 *Today*. 6, 348–353.

129 8.
130 Wilson, K., Bjoernstad, O.N., Dobson, A.P., Merler, S., Poglayen, G., Read, A.F.,
131 Skorpning, A., 2002. Heterogeneities in macroparasite infections: patterns and
132 processes. In: Hudson, P.J., Rizzoli, A., Grenfell, B., Heesterbeek, H., Dobson,
133 A., (Eds.), *The Ecology of Wildlife Diseases*. Oxford University Press, Oxford, pp.
134 6–44.