RESEARCH ARTICLE

Occupancy Modeling for Improved Accuracy and Understanding of Pathogen Prevalence and Dynamics

Michael E. Colvin[1]*[¤], James T. Peterson[2], Michael L. Kent[3], Carl B. Schreck[2]

1 Oregon Cooperative Fish and Wildlife Research Unit, Department of Fisheries and Wildlife, Oregon State University, 104 Nash Hall, Corvallis, Oregon, 97331, United States of America, 2 Oregon Cooperative Fish and Wildlife Research Unit, U.S. Geological Survey-Department of Fisheries and Wildlife, Oregon State University, 104 Nash Hall, Corvallis, Oregon, 97331, United States of America, 3 Department of Microbiology, Oregon State University, 220 Nash Hall, Corvallis, Oregon, 97331, United States of America

¤ Current address: Department of Wildlife Fisheries and Aquaculture, Mississippi State University, PO Box 9690, Mississippi State, MS, 39762, United States of America
* Michael.colvin@msstate.edu

Abstract

Most pathogen detection tests are imperfect, with a sensitivity < 100%, thereby resulting in the potential for a false negative, where a pathogen is present but not detected. False negatives in a sample inflate the number of non-detections, negatively biasing estimates of pathogen prevalence. Histological examination of tissues as a diagnostic test can be advantageous as multiple pathogens can be examined and providing important information on associated pathological changes to the host. However, it is usually less sensitive than molecular or microbiological tests for specific pathogens. Our study objectives were to 1) develop a hierarchical occupancy model to examine pathogen prevalence in spring Chinook salmon *Oncorhynchus tshawytscha* and their distribution among host tissues 2) use the model to estimate pathogen-specific test sensitivities and infection rates, and 3) illustrate the effect of using replicate within host sampling on sample sizes required to detect a pathogen. We examined histological sections of replicate tissue samples from spring Chinook salmon *O. tshawytscha* collected after spawning for common pathogens seen in this population: *Apophallus*/echinostome metacercariae, *Parvicapsula minibicornis*, *Nanophyetus salmincola*/ metacercariae, and *Renibacterium salmoninarum*. A hierarchical occupancy model was developed to estimate pathogen and tissue-specific test sensitivities and unbiased estimation of host- and organ-level infection rates. Model estimated sensitivities and host- and organ-level infections rates varied among pathogens and model estimated infection rate was higher than prevalence unadjusted for test sensitivity, confirming that prevalence unadjusted for test sensitivity was negatively biased. The modeling approach provided an analytical approach for using hierarchically structured pathogen detection data from lower sensitivity diagnostic tests, such as histology, to obtain unbiased pathogen prevalence estimates with associated uncertainties. Accounting for test sensitivity using within host replicate samples also required fewer individual fish to be sampled. This approach is

useful for evaluating pathogen or microbe community dynamics when test sensitivity is <100%.

Introduction

Parasites and pathogens (hereafter collectively referred to as pathogens) are routinely surveyed and monitored in domestic (e.g., fish culture facilities) [1] and free-ranging (hereafter, natural) wildlife populations [2]. Accurate assessment of pathogen presence and prevalence is also relevant for populations destined for geographic translocation and introduction [2–4]. This is especially critical when those introductions may lead to pathogen introductions and potential population declines of endangered species. For example, with salmonid fishes, introduction of the causative agent of whirling disease, *Myxobolus cerebralis*, in the United States intermountain West through introductions of infected fish resulted in population declines and negative biological and economic impacts [5–7]. Similarly, translocations of raccoon *Procyon lotor* captured in Texas and Florida resulted in the introduction of parvoviral enteritis and rabies to recipient populations [8,9]. In an extreme case, the loss of approximately half of Hawaiian avian fauna was attributed, at least in part, to the introductions of avian pathogens [10].

The ability to understand the pathogen-host dynamics and develop effective pathogen mitigation strategies and risk analyses depends upon the quality of the data. Many pathogen detection tests are imperfect (i.e., may not detect a pathogen even if it is present) and therefore pathogen detection probabilities are generally less than one [11]. Imperfect detection increases the likelihood that a pathogen present in a population is missed and induces negative biases in the estimation of pathogen prevalence. Molecular and other similar screening methods that use nonspecific tests (e.g., PCR tests that use universal DNA primers) are particularly vulnerable to false negatives regarding detection of pathogens in low abundance, especially when sympatric with other species that are in much greater abundance, regardless of virulence of any particular pathogen [12–14]. Histopathology has been used in recent surveys for diseases in wild animal populations, particularly for screening of pathogens or non-infectious lesions without *a priori* restrictions [15–20], or when toxicopathic changes were a target endpoint [21,22]. However, a recent evaluation on Chinook salmon *Oncorhynchus tshawytscha* concluded that histological methods were considerably less sensitive than microbiological tests for certain pathogens, resulting in numerous false negatives [17]. The lack of sensitivity of histology compared to microbiological tests that involve either amplification of pathogens in culture or DNA by PCR would certainly apply to other pathogen surveys [13,23]. Imperfect detection of pathogens can substantially influence sample design. Greater sample sizes are required to detect a pathogen in a population, to compensate for false negatives, which may pose a significant constraint if sampling is lethal, something of great importance when lethal take involves threatened and endangered wildlife species. Researchers also have an ethical obligation to minimize lethal sampling of vertebrate species required by Institutional Animal Care and Use Committees [24]. Therefore, analytical approaches that provide accurate estimates of pathogen presence and prevalence while reducing the number of individuals sampled and handled are desirable. Confidence regarding parasite prevalence estimates is also vital for studies concerned with pathogen community ecology.

Test sensitivity is traditionally calculated by comparing the number of detections from one test with those of a high sensitivity "gold standard" diagnostic, such as highly sensitive PCR test, thereby providing an estimate of the probability of detecting the pathogen when truly

present [25,26]. The gold standard test is assumed to have perfect pathogen detection (i.e., sensitivity = 100%); however it is possible that these tests are imperfect as well. For example, Kent et al. [17] detected spring Chinook salmon pathogens in histological samples that were not detected in corresponding pathogen-specific gold standard tests recommended by the American Fisheries Society-Fish Health Section [27]. Similarly, false negatives occurred between bacteriological and histopathological tests, the confirmatory tests for bovine tuberculosis [28]. There still remains need for a flexible, cost-effective approach to accurately estimate pathogen prevalence and detection probability (i.e., detecting a pathogen given it is present) that does not require restrictive assumptions. Approaches have been developed to estimate pathogen prevalence in pooled samples using imperfect test [29,30]. Recently occupancy-detection estimation has been proposed as a method to account for imperfect detection of *Plasmodium* species and provide estimates of test sensitivity [31]. Occupancy-detection estimation includes the collection of replicate tissue samples from individual hosts that are used to estimate pathogen detection probabilities and provide better estimates of test sensitivity compared to ad hoc comparisons between diagnostic tests (e.g., a test compared to a gold standard test). Application of occupancy-detection models to the study of pathogens in natural populations has been limited to the individual host-level [32–34]. Additional insight into within host-pathogen dynamics can be gained by evaluating multiple organ tissues within the host in a hierarchical structure. Our study objectives were to 1) develop a hierarchical occupancy model to estimate pathogen prevalence (hereafter infection rate) in spring Chinook salmon and their distribution among host organ tissues 2) use the model to estimate pathogen- and tissue-specific test sensitivities and host- and organ-level infection rates, and 3) illustrate the effect of using replicate within host sampling on number of hosts sampled required to detect a pathogen in a population given it is present (i.e., power). This manuscript reports on the successful achievement of the study objectives.

Methods

Ethics statement

This study was performed under the auspices of animal use protocol AUP # 4438. The protocol was approved by the Institutional Animal Care and Use Committee of Oregon State University. This particular study utilized fish that were already dead as a result of routine spring Chinook salmon spawning operations conducted by Oregon Department of Fish and Wildlife. Since this study utilized existing dead fish no field permit was required from state regulatory agencies.

Fish sampling

Fish were sampled during an ongoing study of adult spring Chinook salmon in the upper Willamette River (UWR), Oregon, in September 2012. The threatened status of UWR spring Chinook under the Endangered Species Act limits lethal sampling of live fish in this system [35]. However, conservation hatchery operations capture spring Chinook at Dexter Dam (Dexter, Oregon; 43.919937N, 122.815739W) and move them to Willamette Hatchery to mature over summer. Oregon Department of Fish and Wildlife (ODFW) hatchery spawning operations allowed us to opportunistically collect a sufficient sample of fresh carcasses and tissues therein for pathogen detection. Estimates of pathogen prevalence from previous monitoring at this location varied from 15 to 95% [17,36] and were used as guides to estimate required sample sizes. An *a priori* simulation evaluation of the sample size requirements indicated that at least 23 fish were needed to obtain a 90% probability of collecting at least one infected fish assuming a parasite infection rate of 10% assuming 100% sensitivity (M. Colvin unpublished simulation data). The number of fish sampled was 26 to ensure adequate sampling and because sufficient
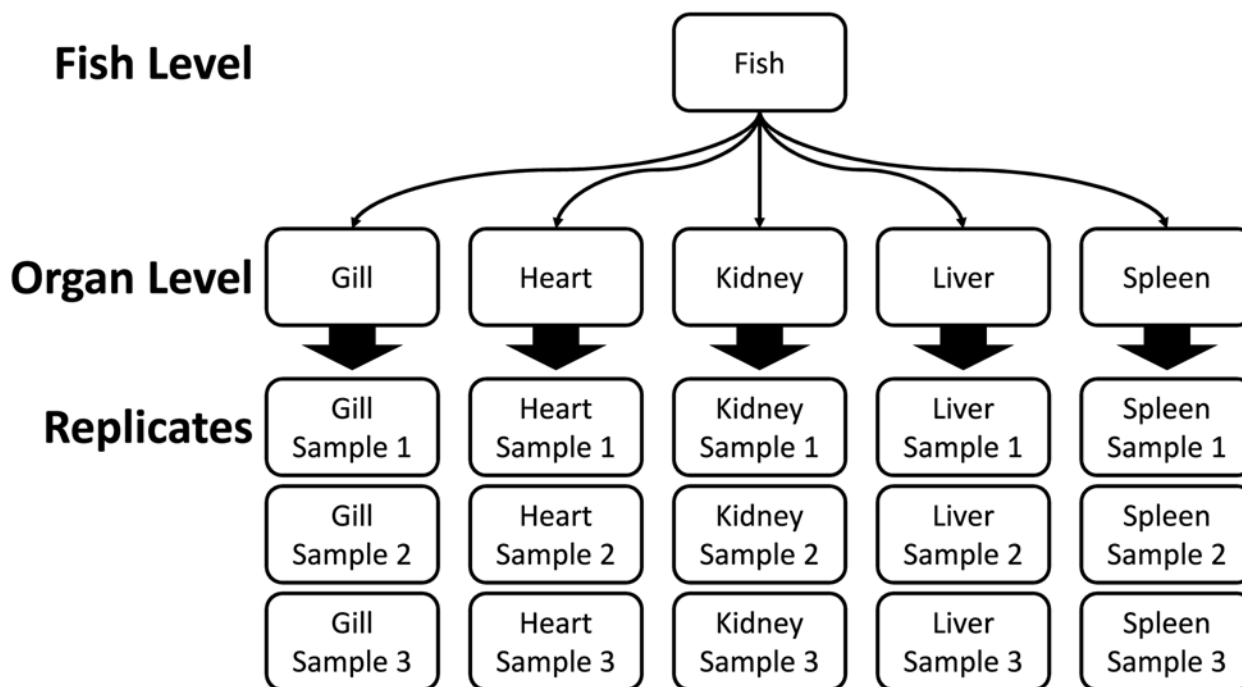
Fig 1. Data structure illustration for pathogens detected by histological analysis of tissues sampled from Willamette Hatchery (Oakridge, OR) spring Chinook salmon. Histological samples were processed to minimize the number of slides needed and therefore multiple tissues were processed on a single slide when feasible.

numbers of fish were available. Sampled fish were euthanized and spawned by ODFW employees and the carcass necropsied within 30 minutes of death.

Necropsy, tissue sampling, and histological processing

Tissue samples from organs were collected for histological examination during necropsy. Three tissue samples of gill, heart, posterior kidney, spleen, and liver were removed and immediately fixed in 10% buffered neutral formalin for at least 7 d (Fig. 1). Fixed tissues were trimmed and placed in tissue cassettes, then dehydrated, sectioned, and affixed to slides using standard histological techniques by the Oregon State University Veterinary Diagnostic Laboratory. Processing resulted in 6 slides for each individual fish; 3 slides for visceral organs (heart, kidney, spleen, liver contained on a slide) and 3 slides for gill tissues. Slides were prepared using standard histological protocols and stained with hematoxylin and eosin [37]. All slides were evaluated for pathogen presence by one of the co-authors (M.K.), who is an experienced fish pathologist. Organ-specific pathogen detections were represented as a series of 0's and 1's. The following pathogens were encountered in tissue sections in many fish, and thus were included in our analysis: 1) metacercariae in the gill (either *Apophallus* sp. or echinostome type), 2) metacercariae of *Nanophyetus salmnicola* in various organs, 3) *Renibacterium salmoninarum*, diagnosed by the presences of granulomatous lesions with no other associated pathogens in various organs, or 4) *Parvicapsula minibicornis* in kidney glomeruli or tubule tissue.

Sample and organ unadjusted prevalence

Prevalence unadjusted for test sensitivity, hereafter unadjusted prevalence (α), was calculated for each pathogen encountered. Typical sampling would include a single tissue sample for each

organ and those values aggregated to the host-level [17]; however this dataset included 3 organ-level tissue samples (hereafter, replicates). To approximate the typical approach, organ-level and host-level prevalence was calculated for each replicate. Organ-level unadjusted prevalence was calculated for each tissue sample replicate by dividing the number of detections by the sample size for each pathogen detected. For example, if *R. salmoninarum* was detected 10, 12, and 9 times (i.e., 3 replicate tissue samples) out of 26 hosts then the prevalence would be 38.4, 46.1, and 34.6% respectively. Host-level unadjusted prevalence was calculated for each replicate by aggregating among organ detections to the host-level and dividing the number of infected hosts by the number of hosts. For example, if a pathogen was detected in kidney and heart tissues these detections were aggregated to 1 at the host- level. These values were then graphically compared to occupancy model estimated prevalence (described below) for each pathogen to examine discrepancies between approaches.

Pathogen occupancy model overview, fitting, and inference

The hierarchical occupancy model developed in this study estimated 3 quantities: 1) the probability an individual host is infected by a pathogen (infection rate; Ψ); 2) the probability an organ is infected given the host is infected (conditional organ infection rate; φ); and 3) the probability a pathogen is detected in an organ given the organ is infected (sensitivity; *s*) (Table 1). Hierarchical occupancy models incorporate additional information from replicate tissue samples of secondary sampling units, organs within fish in this case (Fig. 1). The model also assumed pathogens were correctly identified when detected and that organ tissue-specific detection probabilities did not depend on overall pathogen burden (i.e., sensitivity was equal for low and high burdens). These assumptions can be relaxed as discussed later. We believe these assumptions were reasonable given the extensive experience of the slide reader, and there was no possibility for morphological changes in organ-specific pathogen occupancy because tissues were preserved in a formalin-based fixative.

Given the complex nature of the hierarchical modeling framework, we used a Bayesian approach and state space formulation [38] to estimate host and organ-level pathogen infection rates (Ψ and φ) and pathogen- and tissue-specific sensitivities (*s*). Construction of the likelihood of the data given the occupancy model is difficult due to hierarchical dependencies (i.e., organs nested within fish) arising from the sampling process. However, the likelihood of the model given the data can be constructed using hierarchical state space models [38]. These models can be fit using a Bayesian approach to estimate Ψ and φ and *s*.

Host-level pathogen infection state was modeled as a Bernoulli distributed variable that estimated the unobserved infection state for each fish (i.e., uninfected = 0, infected = 1) with probability Ψ_i as:

$$Z_i \sim Bern(\Psi_i) \tag{1}$$

Table 1. Parameters, symbols, and descriptions of terms used in this study.

Parameter	Occupancy-detection analog	Symbol	Description
Infection rate	Occupancy	Ψ	Probability of a host being infected by a pathogen.
		φ	Probability of an organ being infected by a pathogen given the host is infected.
Prevalence	Proportion occupied units	P	Proportion of hosts or organs predicted to be infected calculated as the number of infected divided by the total.
Sensitivity	Detection probability	s	Probability of detecting a pathogen in organ tissue samples given the pathogen is present.
Unadjusted prevalence	Naïve prevalence	α	The proportion of hosts in a sample infected by a pathogen calculated as the observed number of infected hosts/number of hosts sampled.

doi:10.1371/journal.pone.0116605.t001

where, Z_i was the unobserved infection state for fish i and Ψ_i is the probability that fish i is infected by the pathogen. Infection rate (Ψ_i) was predicted using a linear model with a logit link as:

$$\ln\left(\frac{\Psi_i}{1 - \Psi_i}\right) = \omega_{0i}, \tag{2}$$

where, ω_{0i} is the intercept of the linear model. Linear models were used in equation 2 and in following equations to facilitate model fitting. The infection rate of organ j within fish i conditional on fish i being infected (i.e., $Z_i = 1$) was estimated as:

$$z_{ij} \sim Bern(Z_i \cdot \varphi_{ij}), \tag{3}$$

where, z_{ij} was the unobserved infection state for organ j within fish i, Z_i was the host-level infection state of the ith fish and φ_{ij} was the probability that organ j was infected given the fish was infected (i.e., $Z_i = 1$) otherwise $\varphi_{ij} = 0$. A linear model with a logit link was used to estimate φ_{ij} as:

$$\ln\left(\frac{\varphi_{ij}}{1 - \varphi_{ij}}\right) = \beta_{0ij}, \tag{4}$$

where β_{0ij} is the intercept of a linear model predicting infection rate for organ j within fish i.

A detection model was used to link observed tissue level pathogen detections to the process model described above (equations 1–4) as:

$$y_{ijk} \sim Bern(z_{ij} \cdot s_{ij}), \tag{5}$$

where y_{ijk} was the observed pathogen detections (0 or 1) for the replicate k from organ j within fish i, z_{ij} was the unobserved organ-specific infection state for fish i, and s_{ij} was the conditional organ tissue-level sensitivity given the organ was infected (i.e., $z_{ij} = 1$), otherwise $s_{ij} = 0$. Organ tissue-specific sensitivity (s_{ij}) was predicted using a linear model as:

$$\ln\left(\frac{s_{ij}}{1 - s_{ij}}\right) = \theta_{0ij}, \tag{6}$$

where θ_{0ij} was the intercept term for organ j within fish i

Separate hierarchical occupancy models for each pathogen encountered in the study were fit using Markov Chain Monte Carlo (MCMC) implemented in WinBUGS 2.14 [39]. Each model was fit using diffuse priors and three chains with 130k iterations and 30k burn-in samples (i.e., the first 30k iterations were discarded). Chains were thinned by 10 for *Apophallus*/echinostome metacercariae 20 for *P. minibicornis* and *N. salmnicola*, and 50 for *R. salmoninarum* to minimize autocorrelation in the MCMC. The Gelman-Rubin convergence statistic and visual inspection of the 3 MCMC chains were used to ensure adequate chain mixing was achieved [40]. Posterior predictive check was used to calculate a Bayesian goodness-of-fit (GOF) value for fitted models to assess whether the model adequately fit the data [38,41]. Goodness-of-fit values vary from 0 to 1 with extreme values indicating lack of fit and values close to 0.5 indicating adequate model fit. The R programming environment, WinBUGS 2.14, and the R2WinBUGS package were used for all analyses [39,42,43]. Separate host- and organ-level infection rates could not be estimated for *Apophallus*/echinostome metacercariae because this parasite was only detected in gill tissue. Therefore, the organ-level infection rate was set to one during model fitting to estimate a host-level infection rate and prevalence. Further information

regarding the occupancy model and example code and data for replicating the analysis are available in S1 Information.

Effects of replicates and hierarchical sampling on pathogen detection

The approach developed in this study provides a framework to analyze pathogen detection data from multiple tissue types within a host and replicate tissue samples surveyed, which represents a tradeoff between sampling more primary sampling units (i.e., hosts) or more intensive tissue sampling within hosts (i.e., > 1 replicate). Because the primary objective of any pathogen survey is to detect a pathogen in sample given the pathogen is present in the population, we used simulation to calculate the probability of detecting a hypothetical pathogen (i.e., power) giving varying levels of: (1) host-level infection rate, 0.1 to 0.9 by increments of 0.1; (2) conditional organ tissue level infection rate, 0.1 and 0.3 (i.e., organ infection rate of infected hosts); and (3) test sensitivity, 10% and 50%; (4) number of organ surveyed, 1, 3, and 5; (5) number of replicates surveyed, 1–4; and (6) host sample sizes, 5–60. Two thousand replicate datasets then were simulated for all combinations of these parameters and the proportion of the 2000 replicate datasets where the hypothetical pathogen was detected at least once was calculated. The minimum number of hosts needed to detect the pathogen with a power of 80% was determined and graphically summarized. See S1 Information for R code used to replicate analysis.

Results

Overview of the data and model fits

Pathogens detected by histological examination of organ tissues included: gill metacercariae, *R. salmoninarum*, *N. salmincola*, and *P. minibicornis*. *Apophallus*/echinostome metacercariae and *N. salmincola* metacercariae were detected in gill tissues. Kidney tissues contained detectable levels of *N. salmincola*, *R. salmoninarum*, and *P. minibicornis*. Within the kidney, *P. minibicornis* was detected in glomeruli and tubule tissues. *N. salmincola* was detected in heart tissues. Only *R. salmoninarum* was detected in splenic tissues. Pathogen detections were not perfect (i.e., not detected in every replicate) for all pathogens encountered. In particular, false negatives, identified as detection histories including a combination of 0s and 1s accounted for 4% to 88% of the observed detections (Table 2). Goodness-of-fit checks indicated adequate fit for pathogen-specific occupancy models fit to detection data with p-values that varied from 0.48 to 0.58.

Fish and organ-level unadjusted pathogen prevalence

Unadjusted pathogen prevalence (α) varied at the host- and organ-levels, and organ-level prevalence was always lower than host-level prevalence (Table 3). Host-level unadjusted prevalence was highest for *N. salmincola* (0.96–1.00) and unadjusted prevalence among organs was lower, varying from 0.73 to 0.88. *P. minibicornis* host-level unadjusted prevalence varied among replicates from 0.84 to 0.96 and organ-level unadjusted prevalence was lower varying from 0.64 to 0.81. *R. salmoninarum* prevalence was lowest among pathogens detected, and among replicate host-level unadjusted prevalence varied from 0.08 to 0.15. *Apophallus*/echinostome metacercariae were only detected in gill tissue and unadjusted prevalence varied among replicates from 0.08 to 0.27.

Model estimates

Sensitivity estimates.—Model estimated organ tissue-specific sensitivities (*s*) were all less than one and varied among pathogens and organs (Table 4). Estimated sensitivity was lowest for

Table 2. Frequency of detection combinations for the 26 fish sampled.

		Detection type		
		Non-detection	Imperfect	Perfect
Apophallus/echinnostome metacercariae	Gill	17	6	3
Renibacter salmoninarum	Kidney	23	23	23
	Liver	25	1	0
	Spleen	25	25	25
Nanophyetus salmincola	Gill	0	9	17
	Heart	3	12	11
	Kidney	0	13	13
Parvicapsula minibicornis	Glomerulus	4	9	13
	Tubules	5	4	17

Non-detections represent three replicate pathogen non-detections (i.e., 000). Imperfect detections represent possible combinations of replicate pathogen detections and non-detections (i.e., 100, 010, 001, 110, 101, 011). Perfect detections represent three replicate pathogen detections (i.e., 111).

doi:10.1371/journal.pone.0116605.t002

Apophallus/echinostome metacercariae in gill tissue (0.54) and greatest for *P. minibicornis* in kidney glomeruli (0.91). Sensitivity also varied more among pathogens than for individual pathogens among organs. For instance, sensitivity for *R. salmoninarum* in three different organs averaged 0.66 (range 0.63–0.69), whereas *N. salmincola* sensitivity averaged 0.85 (range 0.81–0.88). Most sensitivity estimates were relatively precise with the exception of *R. salmoninarum* estimates for liver and spleen, which had very large credible intervals (Table 4) likely due to a very low number of detections.

Host- and organ-level infection rates

Host- and organ-level infection rates (Ψ, φ) also varied among pathogens detected and ranged from a low of 0.41 for *Apophallus*/echinostome metacercariae to a high of to 0.99 for *N. salmincola* (Table 5). In some cases where pathogens infected multiple organs, organ-level

Table 3. Fish and organ-level prevalence (unadjusted for test sensitivity; *a*) for replicated tissue-level detection/non-detection data.

		Replicate		
Pathogen	Level	1	2	3
Apophallus/echinostome metacercariae	Gill[a]	0.08	0.27	0.27
Renibacter salmoninarum	Fish	0.15	0.08	0.08
	Kidney	0.08	0.08	0.08
	Liver	0.04	0.00	0.00
	Spleen	0.04	0.00	0.00
Nanophyetus salmincola	Fish	1.00	1.00	0.96
	Gill	0.84	0.85	0.85
	Heart	0.80	0.81	0.81
	Kidney	0.88	0.73	0.80
Parvicapsula minibicornis	Kidney[a]	0.84	0.96	0.96
	Glomerulus	0.64	0.81	0.80
	Tubules	0.68	0.73	0.80

Fish level prevalence represents the aggregation of among organ detections.

[a] Gill and kidney represent the highest level of detection for *Apophallus*/echinostome metacercariae and *P. minibicornis* and therefore fish level prevalence is equal to these values.

doi:10.1371/journal.pone.0116605.t003

Table 4. Estimated organ-specific sensitivity (*s*) and 95% credible intervals for the occupancy model fit to spring Chinook collected from Willamette Hatchery, Oregon.

Pathogen	Organ	Estimate	95% credible interval	
			Lower	Upper
Apophallus/echinostome metacercariae	Gills	0.544	0.307	0.754
Renibacter salmoninarum	Kidney	0.694	0.304	0.947
	Liver	0.626	0.018	0.999
	Spleen	0.660	0.020	0.999
Nanophyetus salmincola	Gills	0.841	0.760	0.995
	Heart	0.883	0.796	0.909
	Kidney	0.813	0.719	0.884
Parvicapsula minibicornis	Tubules	0.887	0.799	0.948
	Glomerulus	0.912	0.829	0.967

Estimates reported for tissues where pathogens were detected.

doi:10.1371/journal.pone.0116605.t004

infection rates exceeded host-level infection rates, which can occur due to the conditional specification of organ-level infection rates given the host was infected. The unconditional organ-level infection rate are the product of host and organ-level infection rates and are equal to or lower than the host-level infection rate. For example, the probability that heart tissue from a randomly selected fish was infected with *N. salmincola* is 0.95•0.99 = 0.94. Conditional organ-level *R. salmoninarum* infection rate was highest for the kidney (0.22) and lowest for the liver and spleen, with very imprecise estimates for the latter two organs (Table 5).

Prevalence comparison

Host-level prevalence unadjusted for test sensitivity (α) was lower than model estimated infection rates (Ψ; Fig. 2). The negative discrepancy decreased as unadjusted prevalence approached

Table 5. Fish and organ-level infection rates (Ψ, φ) and prevalence estimates (*P*) and credible intervals for the occupancy model fit to spring Chinook collected from Willamette Hatchery, Oakridge, Oregon.

| Pathogen	Level	Infection rate			Prevalence		
		Estimate	95% credible interval		Estimate	95% credible interval	
			Lower	Upper		Lower	Upper
Apophallus/echinostome metacercariae	Gills[a]	0.414	0.228	0.665	0.401	0.346	0.577
Renibacter salmoninarum	Fish	0.564	0.181	0.962	0.567	0.192	0.962
	Kidney	0.219	0.043	0.703	0.123	0.115	0.192
	Liver	0.161	0.010	0.876	0.076	0.038	0.423
	Spleen	0.159	0.011	0.914	0.073	0.038	0.423
Nanophyetus salmincola	Fish	0.953	0.862	0.995	1.000	1.000	1.000
	Gills	0.990	0.939	1.000	1.000	1.000	1.000
	Heart	0.901	0.753	0.990	0.906	0.846	0.962
	Kidney	0.990	0.936	1.000	1.000	1.000	1.000
Parvicapsula minibicornis	Fish	0.931	0.815	0.989	0.967	0.961	1.000
	Tubules	0.860	0.707	0.964	0.849	0.846	0.885
	Glomerulus	0.823	0.653	0.945	0.809	0.808	0.846

doi:10.1371/journal.pone.0116605.t005
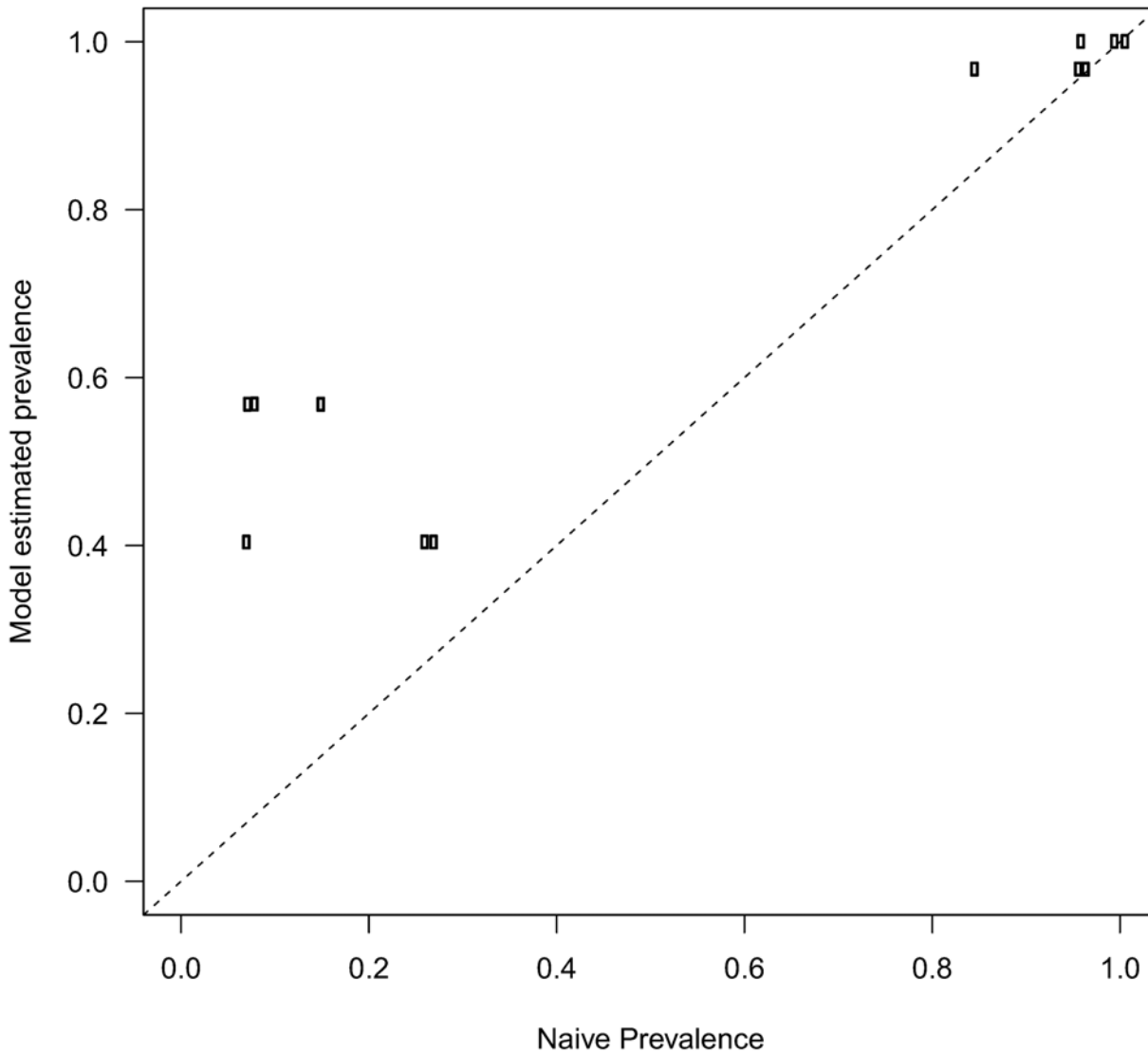
Fig 2. Comparison of host-level model estimated (Ψ) and unadjusted prevalence (α) for pathogens detected. Unadjusted prevalence (α) represents estimates for each replicate (3 per pathogen, 12 estimates total). The dotted line denotes a 1:1 relationship. A small amount of random noise was added due to overplotting. Vertical lines denote 95% credible intervals.

one and therefore biases were relatively small for widespread pathogens such as *N. salmincola* and *P. minibicornis*. The largest discrepancy between model infection rates and unadjusted prevalence was observed for gill metacercariae, which was the only pathogen detected in a single tissue, gill in this case.

Effects of replicates and hierarchical sampling on pathogen detection

The number of host samples required to have an 80% probability of detecting a pathogen in a sample, given it is present, varied among the parameter combinations simulated (Fig. 3). As expected, the required host sample size decreased with increasing host-level and organ-level infection rates. Increasing the number of within organ replicates decreased the host sample size required to detect the pathogen in the sample. This reduction in host sample size was especially
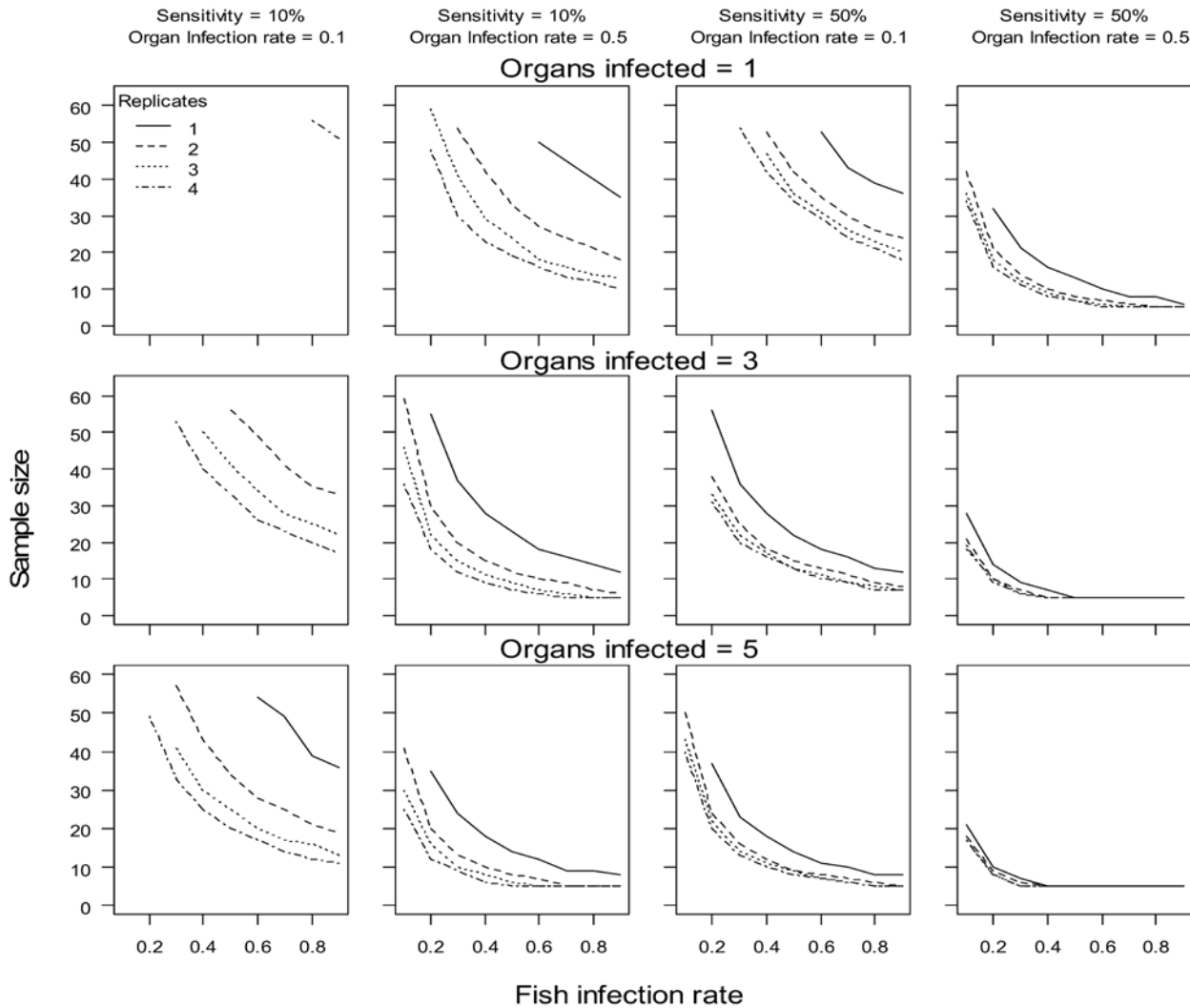
Fig 3. Number of host samples needed to detect a pathogen 80% of the time for varying host-level infection rates (Ψ). Panel rows illustrate the effect of pathogens being searched for in multiple organs and the columns illustrate the effect of varying sensitivity and organ-level infection rate (φ). Organ-level infection rate (φ) and sensitivity (s) assumed to be constant among organs in simulations. Sampling sizes for the top left panel required more than 60 and therefore the panel represents the few combinations where the success criteria was met.

doi:10.1371/journal.pone.0116605.g003

apparent for 2 or more replicates when sensitivities were low (10%). However, gains in terms of host sample size reduction with increasing number of replicates diminished with increasing number of organs surveyed.

Discussion

We developed an approach to use hierarchically structured pathogen detection data from imperfect diagnostic tests to estimate test sensitivity and use that information to achieve unbiased estimation of pathogen infection rate in host populations and among organs within infected hosts. The approach provided a potentially cost-efficient means to improve the effectiveness of lower sensitivity diagnostic tests, such as histology, which is advantageous when pathogen infections are unknown prior to examination [17]. Histology also provides information on detections that are biologically meaningful, representing organ or tissue damage at a histological

level, which may be preferable to highly sensitive molecular diagnostic tests that can detect pathogen presence, but disease status may be uncertain (i.e., pathogen is present but not causing disease or pathogen may be dead).

All pathogens found in this study were detected imperfectly (i.e., present in host and organ but not detected in every tissue sample). Imperfect pathogen detection within infected hosts and organs can arise as a result of pathogen distribution and abundance. Non-detection can occur if tissues are sampled prior to pathogen dispersal through the host and subsequent colonization of organ tissues and due to localized areas of focal infection that may not be contained in histological sections. In this case, once pathogens infect a host, detection depends on the pathogens dispersing, infecting, and distributing within host organ tissues. For example, proliferative pathogens, such as *R. salmoninarum* are likely distributed in relation to the area of initial infection in early stages. Detection of the pathogen will be difficult unless the area of initial infection is contained in the sampled tissue. Related to infection distribution, pathogen detection will generally be related to intensity of infection. In particular, as pathogen intensity increases due to proliferation or continuous colonization of host tissues, the probability of encountering and detecting a pathogen generally increases. Test sensitivity is likely to be low for sublinical infections because pathogen intensities are low and their distribution is likely to be focalized to few locations and therefore difficult to detect.

Host-level unadjusted prevalence was always higher than organ-level unadjusted prevalence, due to the inclusion of multiple organ-level observations where a pathogen was detected in one organ but not others. Sampling multiple organs within an individual constitutes an ad hoc approach to account for imperfect detection and inform individual level infection status, similar to spatial resampling [44–46]. In other words, evaluating multiple organs and aggregating detections to the host provided multiple opportunities to detect the pathogen. Discrepancies between host-level unadjusted and model estimated infection rates were lower for pathogens that infected multiple organs, such as *N. salmincola*, due to these additional detections from multiple organs. Because sampling multiple organ tissues provided additional opportunities to detect a pathogen, discrepancies between prevalence estimates was the largest for gill metacercariae, which were only detected in gill tissue.

The occupancy model used in this study provided estimates of diagnostic test sensitivity. Pathogen-specific test sensitivities estimates reported in our earlier study [17] and this study were similar for *N. salmincola*, but higher for *R. salmoninarum*. This is not unexpected as sensitivity is likely dependent on pathogen abundance and distribution within tissues and will vary among studies, unless pathogen abundances and distribution are similar within tissues. Therefore, the discrepancy with *R. salmoninarum* sensitivity estimates reported in Kent et al. [17] is likely due to pathogen and associated lesion abundance and distribution differences within organ tissues. The samples from Kent et al. [17] include adult Chinook salmon from rivers, including moribund fish. These likely had higher *R. salmoninarum* levels relative to the source of hatchery fish used in this study, which received prophylactic antibiotic injections. The relatively few detections and high variability of sensitivity estimates for *R. salmoninarum* in the present study were likely influenced by relatively low *R. salmoninarum* abundances in individuals. Another explanation for variability between sensitivity estimates is due to the differences in how sensitivity was calculated. Kent et al. [17] used enzyme linked immunosorbent assay (ELISA) to determine whether *R. salmoninarum* was present in the host. While the highly sensitivity ELISA can readily detect *R. salmoninarum* presence, it does not necessarily translate to infection as observed by histology because this test detects bacterial proteins that may persist long after viable infections have resolved [47]. Because we estimated organ tissue-specific sensitivity for histology by repeated detections of infection using multiple tissue samples, it provided

more realistic sensitivity estimates relative to comparing infection detections to pathogen presence, which may not be present at sufficient levels to cause clinical infection.

Organ-level infection rate estimates were conditional on the host infection state (infected or not) and represent the probability that an organ is infected given the host is infected. Therefore, these estimates provide useful information regarding pathogen specificity among organs and can inform sample designs and pathogen dynamics. For example, the estimated *R. salmoninarum* organ-level infection rates were greatest for kidney tissues, indicating that if an individual was infected, the pathogen is most likely found in the kidney. This makes biological sense because *R. salmoninarum* frequently occurs in the kidney and is the causative agent of bacterial kidney disease. Similarly, organ-specific conditional infection rates can provide insight into pathogen tissue preferences, where differences among conditional organ-level infection rates can provide evidence for hypothesized pathogen specific tissue preferences. These organ-level infection rates also provide information on what organs to sample if resources are limiting. For example, given conditional organ-level infection rates, *N. salmincola* was most likely found in gill and kidney tissues relative to heart. This information can help to inform sample design and evaluate tradeoffs among designs when resources are limiting and sampling of multiple organs is not feasible.

Sampling enough individual hosts to have sufficient power to detect a pathogen in a population sample, given it is present, is necessary for potential analysis of host-pathogen interactions and establishing sampling designs. Surveying multiple organs and using replicate tissue sampling within those organs increased the likelihood of encountering a pathogen and provided at least 2 benefits over traditional single samples. First, replicate sampling allowed for the estimation of sensitivity and unbiased infection rates and associated uncertainty. Second, by accounting for test sensitivity using within host replicate tissue samples, fewer individuals need to be handled, providing a beneficial reduction of animal use, an essential component of ethical animal use [24]. This effect was illustrated in the simulation analysis, where increasing the number of organs surveyed (assuming the pathogen infects multiple organs) and tissue replicates within host reduced the host sample size required to achieve a power of 80%. Additional general guiding rules that emerged from the simulation analysis indicated that the effect of increasing within organ replicates was minimal when the pathogen can be detected in multiple organs. However, this result does not imply that within organ replicates are not an important component of the sample design. Within organ replicates are critical for accurately estimating organ-level infection rates and reducing associated uncertainties [48]. Additional studies have demonstrated the tradeoffs involved with allocating efforts and provide guidance on occupancy-detection type survey designs [49–51]

Encountering and detecting a pathogen within host organ tissue is related, in part, to pathogen abundance and distribution and therefore sensitivities can vary with intensity of infection. Infection levels also will vary among the individuals sampled and this variation can influence sensitivity estimates. In particular, aggregation of high pathogen levels in few individuals is a common phenomenon encountered when sampling host populations, particularly with macroparasites [52]. Accounting for variation in pathogen abundance will improve estimates of infection rates by accounting for heterogeneous sensitivities. The analytical framework we used can account for this by predicting sensitivity as a function of auxiliary data, if those data are available. In other words, this approach could model sensitivity as function of additional measures of pathogen abundance. For example, the use of quantitative assays such as ELISA or qPCR that provide quantitative measures associated with pathogen abundance could be used to account for variability of pathogen abundances among hosts. Similarly, quantitative metacercariae counts for worm pathogens (e.g., *N. salmincola*) performed from wet mount preparations of pre-weighed tissue samples [18,53] could be used to provide additional quantitative information to predict the probability of detecting the pathogen given pathogen abundance or density.

Using models to predict infection probabilities provides a formal approach to evaluate hypotheses. For example, *R. salmoninarum* infections are well recognized in kidney tissues, and so it could be hypothesized that infection of organs other than the kidney may be dependent on kidney infection. This hypothesis can be evaluated by predicting organ-level infection rate as a function of kidney status (i.e., infected, uninfected). Similarly, the effects of host-level information can be used to evaluate hypotheses relating to host-level infection rates. For example, host-level infection rates could be predicted as a function of host weight or length to evaluate the potential for size dependent effects on host-level infection rates. Similarly, the relation between infection rate and host-specific characteristics such as sex, age, and condition also could be evaluated using this approach. Thus, the approach used in this study provides a formal analytical method to incorporate information from multiple diagnostic tests to additionally inform estimates of infection rate and test sensitivity and potentially evaluate hypotheses.

Precise and unbiased prevalence estimates with diagnostic tests are a crucial component of monitoring, research management, or quarantine efforts [33,54,55]. The occupancy-detection estimation approach developed in this study provides a formal method to utilize diagnostic tests with varying sensitivity to analyze pathogen detection data and avoid negative biases due to imperfect detections, which has received recent attention as being an important component of biological surveys [56]. It also provides a framework for analyzing pathogen detection data from multiple organ tissue samples within hosts. This approach is not limited to evaluating pathogens and could use other diagnostic endpoints, such as histological manifestations (e.g., neoplasms) or to evaluate contaminant or toxicant presence in multiple organ tissues. Additionally, the approach also could be extended to evaluate pathogen assemblages and co-occurrence by simultaneous analysis in a multi-pathogen occupancy models similar to those used in ecological studies [57]. Such an approach appears especially appropriate when considering the parasite community as potential population regulatory factors for wild hosts.

Supporting Information

S1 Information. Model Code. WinBUGS code provided to replicate analysis and R code to simulate data for model verification.
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: MEC JTP MLK CBS. Performed the experiments: MEC JTP MLK CBS. Analyzed the data: MEC JTP. Contributed reagents/materials/analysis tools: MLK CBS. Wrote the paper: MEC JTP MLK CBS.

References

1. Starliper CE, Schill WB (2011) Columnaris disease, coldwater disease, and bacterial gill disease. In: Woo PTK, Bruno DW, editors. Fish diseases and disorders, volume 3: viral, bacterial and fungal Infections. 2nd ed. Cambridge, MA: CABI International. pp. 606–631.

2. Kock RA, Woodford MH, Rossiter PB (2010) Disease risks associated with the translocation of wildlife. Revue Scientifique Et Technique-Office International Des Epizooties 29: 329–350. PMID: 20919586

3. Cunningham AA (1996) Disease risks of wildlife translocations. Conservation Biology 10: 349–353.

4. Leighton FA (2002) Health risk assessment of the translocation of wild animals. Revue Scientifique Et Technique De L Office International Des Epizooties 21: 187–195.

5. Nehring RB, Walker PG (1996) Whirling disease in the wild: the new reality in the intermountain West. Fisheries 21: 28–30.

6. Vincent ER (1996) Whirling disease and wild trout: The Montana experience. Fisheries 21: 32–34.

7. Hiner M, Moffitt CM (2001) Variation in infections of Myxobolus cerebralis in field-exposed cutthroat and rainbow trout in Idaho. Journal of Aquatic Animal Health 13: 124–132.

8. Allen TJ (1986) Evaluation of movements, harvest rate, vulnerability and survival of translocated racoons in southern West Virginia.. Transactions of the Northeastern Section of the Wildlife Society 43: 64.

9. Anthony JA, Childs JE, Glass GE, Korch GW, Ross L, et al. (1990) Land-Use Associations and Changes in Population Indexes of Urban Raccoons during a Rabies Epizootic. Journal of Wildlife Diseases 26: 170–179. PMID: 2338721

10. Warner RE (1968) The role of introduced disease in the extinction of the endemic Hawaiian avifauna. The Condor 70: 101–120.

11. Adams MJ, Chelgren ND, Reinitz D, Cole RA, Rachowicz LJ, et al. (2010) Using occupancy models to understand the distribution of an amphibian pathogen, *Batrachochytrium dendrobatidis*. Ecological Applications 20: 289–302. PMID: 20349848

12. Hanna SE, Connor CJ, Wang HH (2005) Real-time polymerase chain reaction for the food microbiologist: Technologies, applications, and limitations. Journal of Food Science 70: R49–R53.

13. Meyer CP, Paulay G (2005) DNA barcoding: Error rates based on comprehensive sampling. Plos Biology 3: 2229–2238.

14. Wilcox TM, McKelvey KS, Young MK, Jane SF, Lowe WH, et al. (2013) Robust Detection of Rare Species Using Environmental DNA: The Importance of Primer Specificity. Plos One 8. doi: 10.1371/journal.pone.0082806 PMID: 24482673

15. Rodnick KJ, St.-Hilaire S, Battiprolu PK, Seiler SM, Kent ML, et al. (2008) Habitat Selection Influences Sex Distribution, Morphology, Tissue Biochemistry, and Parasite Load of Juvenile Coho Salmon in the West Fork Smith River, Oregon. Transactions of the American Fisheries Society 137: 1571–1590.

16. Kent ML, Traxler GS, Kieser D, Richard J, Dawe SC, et al. (1998) Survey of salmonid pathogens in ocean-caught fishes in British Columbia, Canada. Journal of Aquatic Animal Health 10: 211–219.

17. Kent M, Benda S, St-Hilaire S, Schreck CB (2013) Sensitivity and specificity of histology for diagnoses of four common pathogens and detection of non-target pathogens in adult Chinook salmon in freshwater. Journal of Veterinary Diagnostic Investigation 25: 341–351. doi: 10.1177/1040638713482124 PMID: 23536613

18. Ferguson JA, St-Hilaire S, Peterson TS, Rodnick KJ, Kent ML (2011) Survey of parasites in threatened stocks of coho salmon (*Oncorhynchus kisutch*) in Oregon by examination of wet tissues and histology. Journal of Parasitology 97: 1085–1098. doi: 10.1645/GE-2757.1 PMID: 21668345

19. Blaustein AR, Romansic JM, Scheessele EA, Han BA, Pessier AP, et al. (2005) Interspecific variation in susceptibility of frog tadpoles to the pathogenic fungus Batracbochytrium dendrobatidis. Conservation Biology 19: 1460–1468.

20. Saksida SM, Marty GD, Jones SRM, Manchester HA, Diamond CL, et al. (2012) Parasites and hepatic lesions among pink salmon, Oncorhynchus gorbuscha (Walbaum), during early seawater residence. Journal of Fish Diseases 35: 137–151. doi: 10.1111/j.1365-2761.2011.01330.x PMID: 22233513

21. Koponen K, Myers MS, Ritola O, Huuskonen SE, Lindstrom-Seppa P (2001) Histopathology of feral fish from a PCB-Contaminated freshwater lake. Ambio 30: 122–126. PMID: 11436658

22. Stentiford GD, Longshaw M, Lyons BP, Jones G, Green M, et al. (2003) Histopathological biomarkers in estuarine fish species for the assessment of biological effects of contaminants. Marine Environmental Research 55: 137–159. PMID: 12502035

23. NRC, (National Research Council) (2008) Brucellosis in the Greater Yellowstone Area. Washington, DC: The National Academy Press.

24. NRC, (National Research Council) (2011) Guide for the care and use of laboratory animals. Washington, DC: The National Academy Press.

25. Greiner M, Gardner IA (2000) Application of diagnostic tests in veterinary epidemiologic studies. Preventive Veterinary Medicine 45: 43–59. PMID: 10802333

26. Tamhane A (2008) Matched-sample analysis: Sensitivity, specificity, and P values. Clinical Infectious Diseases 46: 474–475. doi: 10.1086/524893 PMID: 18181745

27. AFS-FHS, (American Fisheries Society—Fish Health Section),, (2010) AFS-FHS blue book: suggested procedures for the detection and identification of certain finfish and shellfish pathogens. Bethesda, MD: American Fisheries Society.

28. Courcoul A, Moyen JL, Brugere L, Faye S, Henault S, et al. (2014) Estimation of Sensitivity and Specificity of Bacteriology, Histopathology and PCR for the Confirmatory Diagnosis of Bovine Tuberculosis Using Latent Class Analysis. Plos One 9. doi: 10.1371/journal.pone.0115916 PMID: 25551830

29. Williams CJ, Moffitt CM (2005) Estimation of pathogen prevalence in pooled samples using maximum likelihood methods and open-source software. Journal of Aquatic Animal Health 17: 386–391.

30. Williams CJ, Moffitt CM (2010) Estimation of fish and wildlife disease prevalence from imperfect diagnostic tests on pooled samples with varying pool sizes. Ecological Informatics 5: 273–280.

31. Lachish S, Gopalaswamy AM, Knowles SCL, Sheldon BC (2012) Site-occupancy modelling as a novel framework for assessing test sensitivity and estimating wildlife disease prevalence from imperfect diagnostic tests. Methods in Ecology and Evolution 3: 339–348.

32. Thompson KG (2007) Use of site occupancy models to estimate prevalence of *Myxobolus cerebralis* infection in trout. Joural of Aquatic Animal Health 19: 8–13. doi: 10.1577/H06-016.1 PMID: 18236627

33. McClintock BT, Nichols JD, Bailey LL, MacKenzie DI, Kendall WL, et al. (2010) Seeking a second opinion: uncertainty in disease ecology. Ecology Letters 13: 659–674. doi: 10.1111/j.1461-0248.2010.01472.x PMID: 20426794

34. Miller DAW, Talley BL, Lips KR, Grant EH (2012) Estimating patterns and drivers of infection prevalence and intensity when detection is imperfect and sampling error occurs. Methods in Ecology and Evolution 3: 850–859.

35. NMFS, (National Marine Fisheries Service) (1999) Endangered and threatened species: threatened status for three Chinook salmon evolutionarily significant units (ESUs) in Washington and Oregon, and endangered status for one Chinook salmon ESU in Washington. Federal Register 64: 14308–14328.

36. Schreck CB, Kent ML, Colvin ME, Benda S, Sharpe C, et al. (2013) Potential causes and managemant of prespawn mortality in adult upper Willamette River spring Chinook. Annual Report to for the US Army Corps of Engineers, Portland District.

37. Luna LG (1968) Manual of histologic staining methods of the Armed Forces Institute of Pathology. New York: McGraw-Hill Book Company.

38. Link WA, Barker RJ (2009) Bayesian inference: with ecological applications. San Francisco: Academic Press.

39. Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10: 325–337.

40. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Statistical Science 7: 457–511.

41. Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge; New York: Cambridge; New York: Cambridge University Press. doi: 10.1093/jxb/erm028 PMID: 25506957

42. R Development Core Team (2010) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

43. Sturtz S, Ligges U., Gelman A (2005) R2WinBUGS: a package for running WinBUGS from R. Journal of Statistical Software 12: 1–16.

44. Kendall WL, White GC (2009) A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. Journal of Applied Ecology 46: 1182–1188.

45. Guillera-Arroita G (2011) Impact of sampling with replacement in occupancy studies with spatial replication. Methods in Ecology and Evolution 2: 401–406.

46. Nichols JD, Bailey LL, O'Connell AF, Talancy NW, Grant EHC, et al. (2008) Multi-scale occupancy estimation and modelling using multiple detection methods. Journal of Applied Ecology 45: 1321–1329.

47. Pascho RJ, Goodrich TD, McKibben CL (1997) Evaluation by enzyme-linked immunosorbent assay (ELISA) of Renibacterium salmoninarum bacterins affected by persistence of bacterial antigens. Journal of Aquatic Animal Health 9: 99–107.

48. MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, et al. (2002) Estimating site occupancy rates when detection probabilities are less than one. Ecology 83: 2248–2255.

49. Mackenzie DI, Royle JA (2005) Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology 42: 1105–1114.

50. Guillera-Arroita G, Ridout MS, Morgan BJT (2010) Design of occupancy studies with imperfect detection. Methods in Ecology and Evolution 1: 131–139.

51. Guillera-Arroita G, Lahoz-Monfort JJ (2012) Designing studies to detect differences in species occupancy: power analysis under imperfect detection. Methods in Ecology and Evolution 3: 860–869.

52. Poulin R (2013) Explaining variability in parasite aggregation levels among host samples. Parasitology 140: 541–546. doi: 10.1017/S0031182012002053 PMID: 23343821

53. Jacobson KC, Teel D, Van Doornik DM, Casillas E (2008) Parasite-associated mortality of juvenile Pacific salmon caused by the trematode Nanophyetus salmincola during early marine residence. Marine Ecology-Progress Series 354: 235–244.

54. Chen PC, Liu SC, Liu HI, Chen TW, Chiang KS (2011) Probability of detecting nematode infestations for quarantine sampling with imperfect extraction efficacy. Journal of Nematology 43: 16–24. PMID: 22791911

55. Cooch E, Conn P, Ellner S, Dobson A, Pollock K (2012) Disease dynamics in wild populations: modeling and estimation: a review. Journal of Ornithology 152: 485–509.

56. Guillera-Arroita G, Lahoz-Monfort JJ, McKenzie DI, Wintle BA, McCarthy MA (In press) Ignoring imperfect detection in biological surveys is dangerous: a response to 'Fitting and Interpreting Occupancy Models'.. PLoS ONE. doi: 10.1371/journal.pone.0117679 PMID: 25626018

57. Shea CP, Peterson JT, Conroy MJ, Wisniewski JM (2012) Evaluating the influence of land use, drought and reach isolation on the occurrence of freshwater mussel species in the lower Flint River Basin, Georgia (U.S.A.). Freshwater Biology: 382–395. doi: 10.1182/asheducation-2012.1.382 PMID: 23233608