

ORIGINAL RESEARCH

Disease-structured N -mixture models: A practical guide to model disease dynamics using count data

Graziella V. DiRenzo¹  | Christian Che-Castaldo² | Sarah P. Saunders^{1,3}  |
 Evan H. Campbell Grant⁴  | Elise F. Zipkin^{1,5}

¹Department of Integrative Biology,
 College of Natural Science, Michigan State
 University, East Lansing, Michigan

²Department of Ecology and
 Evolution, Stony Brook University, Stony
 Brook, New York

³National Audubon Society, East Lansing,
 Michigan

⁴SO Conte Anadromous Fish Research Lab,
 Patuxent Wildlife Research Center, U.S.
 Geological Survey, Turners Falls,
 Massachusetts

⁵Ecology, Evolutionary Biology, and Behavior
 Program, Michigan State University, East
 Lansing, Michigan

Correspondence

Graziella V. DiRenzo, Department of
 Ecology, Evolution, and Marine Biology,
 University of California, Santa Barbara, CA.
 Email: grace.direnzo@gmail.com

Abstract

Obtaining inferences on disease dynamics (e.g., host population size, pathogen prevalence, transmission rate, host survival probability) typically requires marking and tracking individuals over time. While multistate mark-recapture models can produce high-quality inference, these techniques are difficult to employ at large spatial and long temporal scales or in small remnant host populations decimated by virulent pathogens, where low recapture rates may preclude the use of mark-recapture techniques. Recently developed N -mixture models offer a statistical framework for estimating wildlife disease dynamics from count data. N -mixture models are a type of state-space model in which observation error is attributed to failing to detect some individuals when they are present (i.e., false negatives). The analysis approach uses repeated surveys of sites over a period of population closure to estimate detection probability. We review the challenges of modeling disease dynamics and describe how N -mixture models can be used to estimate common metrics, including pathogen prevalence, transmission, and recovery rates while accounting for imperfect host and pathogen detection. We also offer a perspective on future research directions at the intersection of quantitative and disease ecology, including the estimation of false positives in pathogen presence, spatially explicit disease-structured N -mixture models, and the integration of other data types with count data to inform disease dynamics. Managers rely on accurate and precise estimates of disease dynamics to develop strategies to mitigate pathogen impacts on host populations. At a time when pathogens pose one of the greatest threats to biodiversity, statistical methods that lead to robust inferences on host populations are critically needed for rapid, rather than incremental, assessments of the impacts of emerging infectious diseases.

KEYWORDS

Bayesian, Dail-Madsen model, disease ecology, emerging infectious diseases, generalized N -mixture model, hierarchical models, host-pathogen interaction, mark-recapture models, multistate models, occupancy model

1 | INTRODUCTION

Emerging infectious diseases threaten human health, food security, and global biodiversity (Daszak, Cunningham, & Hyatt, 2000; Fisher et al., 2012; Jones et al., 2008). Disease ecologists have used theoretical models (e.g., susceptible, infected, recovered [SIR] models, Kermack & McKendrick, 1927, Anderson & May, 1979; individual-based models, Briggs, Knapp, & Vredenburg, 2010) to understand and predict pathogen spread and disease dynamics (reviewed in Cooch, Conn, Ellner, Dobson, & Pollock, 2012; Joseph et al., 2013; Langwig et al., 2015). Despite the important theory generated by such models, they are often not practical for guiding disease management because they require large amounts of data to parameterize (Barlow, 1995; Boersch-Supan, Ryan, & Johnson, 2016; Lloyd-Smith et al., 2005; Smith et al., 2009). To obtain parameter estimates for disease models, ecologists have relied on mark-recapture which require marking and tracking individuals over time (Table 1; Cooch

et al., 2012). However, tracking animals can be difficult and logistically infeasible, especially for cryptic/secretive organisms or in small populations (Conn & Cooch, 2009; Faustino et al., 2004; Harmsen, Foster, & Doncaster, 2011; Lachish, Knowles, Alves, Wood, & Sheldon, 2011; Pryde, O'Donnell, & Barker, 2005). Advanced statistical methods that provide similar inferences as mark-recapture models for unmarked host populations (e.g., populations where individuals are not individually tracked over time) are critically needed to understand disease dynamics, assess pathogen impacts on populations, and develop pathogen mitigation strategies. We highlight the utility of *N*-mixture models (Dail & Madsen, 2011; Hostetler & Chandler, 2015; Royle, 2004; Zipkin, Sillett, et al., 2014; Zipkin, Thorson, et al., 2014) to study disease dynamics using count data. *N*-mixture models use data from repeated count surveys of multiple sites within a period of population closure to estimate detection probability of individuals and thus population abundance (Royle, 2004). Our guide is aimed toward disease ecologists with

TABLE 1 Set of individual, population, and site-level parameters of interest to disease ecologists

Process and scale	Count vs. detection/non-detection data		Number of seasons		Parameter
	Count	Detection/non-detection	Dynamic (≥ 2 seasons)	Single season	
Individual-level		X	X		Host survival probability
			X		Host life expectancy
	X		X		Host reproduction
	X		X		Host immigration
	X		X		Host emigration
	X		X		Transmission probability
		X	X		Expected time to first infection
		X	X		Duration of illness
		X	X		Recovery probability
		X		X	Host-pathogen load
Population-level		X	X	X	Host infection status
	X		X		Host population size
	X	X	X	X	Host population growth rate
	X		X	X	Pathogen prevalence
Site-level		X	X	X	Average host infection intensity
		X	X		Site-occupancy probability
		X	X		Host extinction probability
Sampling process		X	X		Pathogen extinction probability
		X	X	X	Host detection probability
		X	X	X	Pathogen detection probability
		X	X	X	Observation probability (i.e., seen alive; but unknown disease state)

Note. We summarize the minimal type of observation data needed for inference (i.e., counts/abundance or detection–non-detection/presence–absence). These quantities may be estimable using multistate mark–recapture, multistate dynamic site-occupancy models, or disease-structured *N*-mixture models. Note that to estimate some parameters, multiple seasons of data are required, while others only require a single season of data.

some modeling experience. However, the Supporting Information provides all the tutorials and code needed to implement the models outlined in the main text, including additional variations. Our motivation for this paper is threefold. First, count data are currently underutilized in disease modeling but offer an opportunity to estimate critical parameters (Table 1) from commonly collected data. Second, wider adoption of count data models will enhance field-based testing of epidemiological theory and enable quantitative evaluation of pathogen mitigation strategies at a time when emerging infectious diseases pose one of the greatest threats to biodiversity. Lastly, recent advances in N -mixture models provide new opportunities to study disease dynamics at large spatial and long temporal scales or in remnant host populations decimated by pathogens, when tracking individuals is not feasible. In cases where count data are unavailable, detection/non-detection data can be used under a site-occupancy framework as a cost-effective alternative for assessing disease dynamics (e.g., extirpation, colonization) and spatial patterns, which can be useful for rapid pathogen assessment and development of mitigation strategies. We focus on N -mixture models and their variants because they provide demographic estimates and detailed disease dynamics (Table 1), whereas detection/non-detection approaches only estimate

a fraction of that information (i.e., infected vs. uninfected sites, colonization/extirpation probability). The Supporting Information is integral to the main text and provides additional details on the development and application of our modeling framework.

2 | CHALLENGES TO PARAMETER ESTIMATION IN DISEASE MODELING

Obtaining inference for disease dynamics is challenging because (a) demographic rates (e.g., survival, transmission; Table 1) are difficult to quantify if either population size or recapture rates are low and (b) there are multiple ways in which sampling error affects the observed data (Figure 1). For example, sampling error relevant to disease studies can manifest in two forms: (a) uncertainty of host occurrence or abundance (i.e., imperfect host detection), and (b) uncertainty of pathogen occurrence or abundance (i.e., imperfect pathogen detection; DiRenzo, Campbell Grant, et al., 2018; Lachish, Gopalaswamy, Knowles, & Sheldon, 2012; Miller, Talley, Lips, & Grant, 2012). Historically, sampling error has been either ignored or acknowledged but not modeled (reviewed by McClintock et al., 2010). Yet, failure to account for sampling error

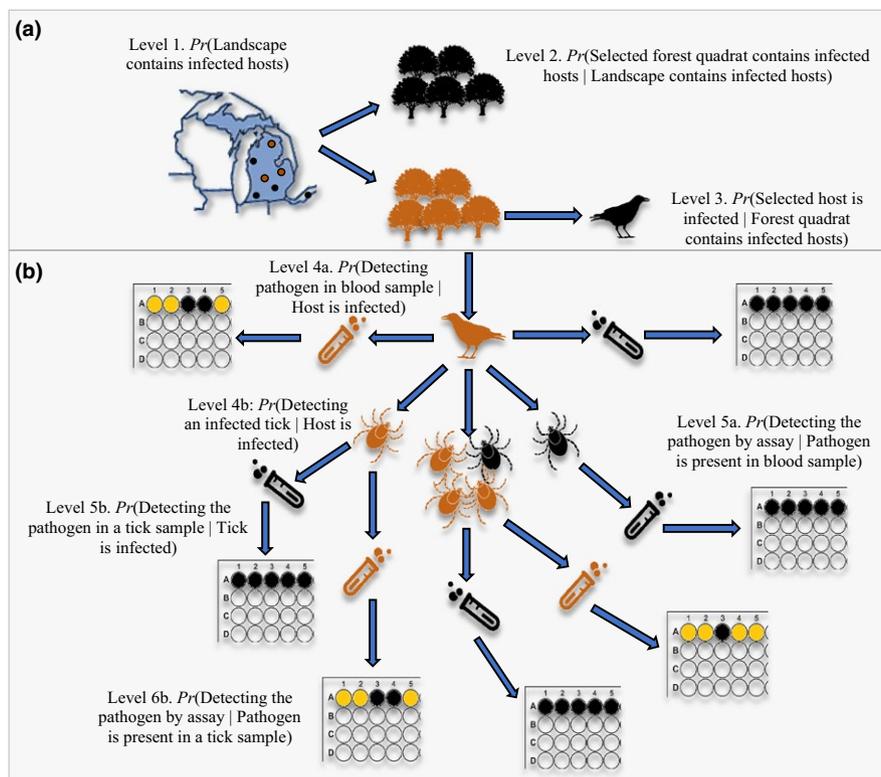


FIGURE 1 Hierarchical formulation illustrating how imperfect (a) host and (b) pathogen detection manifest in wildlife disease ecology. In this example, we depict *Borrelia burgdorferi*, the causative agent of Lyme disease, infections that are transferred from ticks to birds that inhabit forests. The outer images represent wells of a qPCR plate, and the test tubes represent blood samples. The numbers 1 to 7 illustrate the nested hierarchy of the pathogen within the landscape. Orange shapes indicate infection, including (1) forest sites inhabited by infected birds, (2) infected birds with infected ticks, (3) ticks infected by *B. burgdorferi*, (4) blood samples with *B. burgdorferi* drawn from birds and ticks, and (5) qPCR wells with *B. burgdorferi* DNA. Black shapes indicate no infection. Multiple arrows from a single figure represent repeated samples. To illustrate the concept of nested probabilities, we embed probability statements into the figure. Note that at each level of the hierarchy measurement error may be accommodated using conditional probability statements if count data are subject to sampling bias

can result in biased parameter estimates and thus erroneous management decisions and potentially ineffective disease mitigation strategies (Grant et al., 2017; Langwig et al., 2015; Russell, Katz, Richgels, Walsh, & Grant, 2017).

2.1 | Estimating demographic rates

Critical life cycle processes, such as birth and death, are not readily observable, which makes estimating demographic rates difficult for most populations. In the context of disease ecology, transmission and recovery events can be cryptic and go unnoticed. For example, pathogen transmission is the product of two events: (a) a contact between an infected and uninfected host and (b) the probability that the infected host transfers the pathogen to the uninfected host. Multistate mark-recapture models have long been used in disease ecology to estimate disease dynamics, but such approaches can be time and labor intensive (Conn & Cooch, 2009; Faustino et al., 2004; Harmsen et al., 2011; Lachish et al., 2011; Pryde et al., 2005; Williams, Nichols, & Conroy, 2002). Mark-recapture studies are therefore often limited to a small number of locations, providing information for a single species or region (e.g., Parmenter et al., 1998, Briggs et al., 2010). The spread and rise of emerging infectious diseases (Jones et al. 2008, Fisher et al., 2012) render mark-recapture methods impractical for rapid pathogen assessment and development of broad-scale mitigation strategies (Grant et al., 2017; Langwig et al., 2015).

2.2 | Host detection

Infected and uninfected hosts may have different detection rates based on data collection procedures, and there is no a priori predictable pattern as to which may be easier to observe during sampling. If the detection probability of infected and uninfected hosts differs, then estimates of pathogen prevalence can be biased toward the more detectable host (e.g., Senar & Conroy, 2004; Jennelle, Cooch, Conroy, & Senar, 2007; Conn & Cooch, 2009; Schmidt, 2010; Poulin, 2010). For example, house finches (*Carpodacus mexicanus*) infected by *Mycoplasma gallisepticum* are less detectable than uninfected hosts; not accounting for this bias results in the underestimation of infected individuals in the population (Faustino et al., 2004). In other cases, infected hosts are easier to detect than uninfected hosts, such as European serins (*Serinus serinus*) infected by avian pox (Senar & Conroy, 2004).

2.3 | Pathogen detection

Pathogen detection probability, which is the probability that a pathogen is detected on a host when it is present, is a significant concern in both veterinary and medical fields and is likely an issue in most sampling and diagnostic methods (reviewed in Enoe, Georgiadis, & Johnson, 2000; Greiner & Gardner, 2000; Mosher et al., 2018; Toft, Jørgensen, & Højsgaard, 2005). The probability of detecting a pathogen on (or within) an infected host is the product of two sequential processes: (a)

the sampling of an individual (e.g., blood sample, swab) and (b) laboratory analyses of the sample (e.g., DNA extraction method, qPCR efficiency; Figure 1). Ignoring imperfect detection can result in disease state misclassification, where infected hosts are misclassified as uninfected (i.e., false negatives). This misclassification can lead to biases in estimates of state-specific survival and transition probabilities and lowers the power to detect differences between survival probabilities of disease states. Across multiple host-pathogen systems, pathogen detection probability using routine laboratory assays shows positive correlations with host infection intensity, such as *Batrachochytrium dendrobatidis* (*Bd*) zoospores on amphibian skin (DiRenzo, Campbell Grant, et al., 2018; Miller et al., 2012); the causative agent of malaria, *Plasmodium* sp., in birds (Knowles et al., 2011; Lachish et al., 2012); the causative agent of tuberculosis, *Mycobacterium bovis*, in cattle (Drewe, Dean, Michel, & Pearce, 2009); and the causative agent of Lyme disease, the bacterium *Borrelia*, in *Ixodes uriae* ticks (Gómez-Díaz, Doherty, Duneau, & McCoy, 2010). This suggests that infected individuals with low infection intensities are likely to be misclassified as uninfected in many systems using a variety of methods.

3 | DISEASE-STRUCTURED N-MIXTURE MODELS

The *N*-mixture model (Royle, 2004) is a type of state-space model where the true state of the system (i.e., animal abundance) is assumed to be imperfectly observed during the sampling process. *N*-mixture models attribute observation error to imperfect detection of individuals (i.e., false negatives) which can be estimated from repeated surveys of a site over a period in which the population is closed to demographic changes. A site can be defined as a spatial location, an individual, or a sample; a repeated survey can be defined as temporal or spatial replicates, or resampling of an individual.

In this section, we introduce the disease-structured *N*-mixture model framework for closed (single season) and open (multiseason) populations (see Supporting Information for additional details). Disease-structured *N*-mixture models provide a rapid, inexpensive, and efficient approach for estimating demographic rates, while accounting for imperfect host and/or pathogen detection (e.g., Dail & Madsen, 2011; Zipkin, Sillett, et al., 2014; Zipkin, Thorson, et al., 2014; Hostetler & Chandler, 2015; Zhao, Royle, & Boomer, 2017; Brintz, Fuentes, & Madsen, 2018). Using data collected across time and space, disease-structured *N*-mixture models can estimate state-dependent abundance and related demographic rates (e.g., survival, recruitment, immigration), as well as transition probabilities between disease states (Table 1).

Generalized *N*-mixture models have recently been applied to two disease-specific case studies (Brintz et al., 2018; DiRenzo, Zipkin, et al., 2018). Brintz et al. (2018) estimate an increase in the number of chlamydia cases in Oregon over time and highlight how imperfect pathogen detection may be due to latency of the infection or possibly hard-to-reach populations, including drug users, sex workers, and others having the infection. In a different model formulation,

DiRenzo, Zipkin, et al. (2018) evaluated support for three host-pathogen coexistence hypotheses (i.e., source-sink, eco-evolutionary rescue, and spatial variation in pathogen transmission) in a Neotropical amphibian community decimated by *Bd* in 2004. They found that the primary driver of host-pathogen coexistence was eco-evolutionary rescue, as evidenced by similar amphibian survival and recruitment rates between infected and uninfected hosts.

3.1 | Single-season closed population model

We outline the general modeling framework for a closed population disease-structured N -mixture model to estimate the abundance of infected and uninfected hosts using one of the most frequently collected data types: host and pathogen counts. The goal is to estimate pathogen prevalence and site-level abundance of infected and uninfected hosts, while accounting for both imperfect host and pathogen detection. In the single-season study design, we assume that individual sites (e.g., survey plots or transects) are sampled on multiple occasions (survey replicates) during a time frame when the host population is closed to changes (i.e., birth, death, immigration, emigration, disease transmission). The observed number of infected and uninfected hosts is counted at each site during each survey event. Each observed host is assigned a disease state (infected or not infected), based on diagnostic symptoms of disease or laboratory analysis (e.g., swabs, blood, or tissue samples), and used to calculate the average infection intensity at site i during survey j . We assume that a fraction of infected hosts are misidentified as uninfected because of low pathogen infection intensities (e.g., DiRenzo, Campbell Grant, et al., 2018; Lachish et al., 2012; Miller et al., 2012), but misidentification may also be unrelated to infection intensity and attributable to other factors (e.g., field sampling and laboratory diagnostic testing).

We specify the observation model to account for imperfect host and pathogen detection during the sampling process by assuming that pathogen detection is related to infection intensities (e.g., DiRenzo, Campbell Grant, et al., 2018; Lachish et al., 2012; Miller et al., 2012). We denote $g_{s,i,j}$ as the number of hosts detected in each disease state s (where $s = 1$ for uninfected and $s = 2$ for infected) at site i during survey replicate j . We assume that there are a number of misidentified infected hosts, $m_{i,j}$ (i.e., the number of infected hosts classified as uninfected when they are actually infected). The corrected number of detected uninfected and infected hosts, $y_{s,i,j}$, in disease state s at site i during survey j is part of a deterministic relationship between the number of observed hosts, $g_{s,i,j}$, and the number of misspecified hosts, $m_{i,j}$:

$$g_{1,i,j} = y_{1,i,j} + m_{i,j}, \quad (1)$$

$$g_{2,i,j} = y_{2,i,j} - m_{i,j}. \quad (2)$$

We assume that the number of misspecified individuals, $m_{i,j}$, is a binomial random variable dependent on the average pathogen detection probability for hosts at site i during survey replicate j , $\theta_{i,j}$:

$$m_{i,j} \sim \text{bin}(y_{2,i,j}, 1 - \theta_{i,j}). \quad (3)$$

To estimate $\theta_{i,j}$, we model the relationship between pathogen detection probability and average site-specific pathogen infection intensity, $Z_{i,j}$: $\text{logit}(\theta_{i,j}) = \delta 0 + \delta 1 Z_{i,j}$, where $\delta 0$ and $\delta 1$ are the intercept and slope parameters on the effect (see Supporting Information for details). $Z_{i,j}$ is estimated using a lognormal distribution:

$$Z_{i,j} \sim \text{log normal}(\log(\mu_i + 0.001), \sigma^2) \quad (4)$$

where, μ_i is the average infection intensity at site i and σ^2 is the variation in infection intensity among individuals at site i . The observed data, $x_{i,j}$, is then modeled as a lognormal distribution with mean $Z_{i,j}$ and sampling error σ_e^2 :

$$x_{i,j} \sim \text{log normal}(\log(Z_{i,j} + 0.001), \sigma_e^2). \quad (5)$$

The parameters $\delta 0$, $\delta 1$, μ_i , and σ^2 are estimable by either (a) collecting multiple (>1) samples for diagnostic analysis from each observed host or (b) repeated diagnostic runs for individual samples, to explicitly model the relationship between pathogen infection intensity and pathogen detection probability. In cases where such data are unavailable, information obtained from outside studies can be used. The relationship between pathogen intensity and detection probability has been established for many disease systems (Drewet et al., 2009; Gómez-Díaz et al., 2010; Knowles et al., 2011; Lachish et al., 2012; Miller et al., 2012, DiRenzo, Campbell Grant, et al., 2018), which can be readily translated into informative priors (e.g., DiRenzo, Zipkin, et al., 2018).

We use the corrected number of uninfected and infected hosts $y_{s,i,j}$ in our model to estimate the true number of uninfected and infected hosts at each site i , $N_{s,i}$, with the binomial distribution:

$$y_{s,i,j} \sim \text{bin}(N_{s,i}, p_{s,i,j}), \quad (6)$$

where $p_{s,i,j}$ is the host detection probability in disease state s at site i during survey j .

We model true host abundance, $N_{s,i}$, in disease class s and site i using a count distribution, such as the Poisson:

$$N_{s,i} \sim \text{Poisson}(\lambda_s) \quad (7)$$

where λ_s is the expected host abundance for disease state s across all sites, which can also be modeled with site-level habitat covariates using a log-link function. We use the Poisson distribution but others such as the zero-inflated Poisson or negative binomial, can be used in the case of excess zeros or when data are skewed (Kéry & Schaub, 2012).

Total host abundance at a site i is the sum of the latent number of infected and uninfected hosts: $\tilde{N}_i = N_{1,i} + N_{2,i}$. Site-level pathogen prevalence, P_i is then calculated as:

$$P_i = \frac{N_{2,i}}{\tilde{N}_i} \quad (8)$$

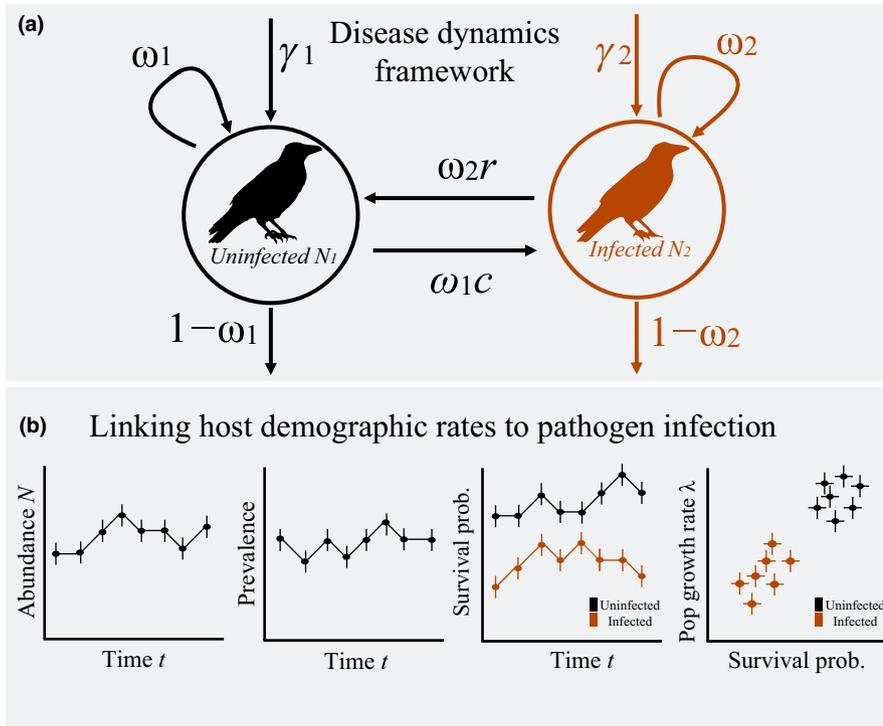


FIGURE 2 Disease dynamics of a host population governed by (a) the uninfected and infected sub-populations (N_1 and N_2 , respectively) that experience state transitions (i.e., c = infection, r = recovery), recruitment (γ), and survival (ω). This framework has traditionally been used in mark-recapture models, but recent advancements in unmarked data models allow for a similar parameter estimation. Parameters are defined in Table 1. To study the link between host demographic rates and pathogen infection, we can (b) correlate the estimates of annual population growth rates with those of estimated demographic rates in both infected and uninfected states

3.2 | Multiseason open population model

In the multiseason formulation of the disease-structured N -mixture model, our goal is to estimate the number of individuals (e.g., infected, uninfected), demographic parameters (e.g., survival, recruitment rates), and disease dynamics (i.e., transmission, recovery probabilities; Figure 2). Using a robust multiseason study design, sites are repeatedly surveyed during multiple primary sampling periods (e.g., seasons or years; Dail & Madsen, 2011; Zipkin, Sittell, et al., 2014; Zipkin, Thorson, et al., 2014). We assume that the population is open to changes in abundance via birth, survival, immigration, and changes in disease states between primary sampling periods. Within each primary sampling period, sites are repeatedly sampled during a period of population closure, and all hosts are counted and assigned to a disease state (i.e., infected or not infected) as described in the closed population model. The observation model linking the observed survey data (i.e., repeated counts of infected and uninfected hosts at sites) for the open population model is the same as in the closed population model with the only difference being that the data and true latent abundance of hosts in disease state s at each site j during primary season, $N_{s,i,t}$, are all also indexed by season/year t .

We model abundance for each disease state s in the first year, $N_{s,i,1}$, of sampling using a count distribution, such as a Poisson:

$$N_{s,i,1} \sim \text{Poisson}(\lambda_s), \quad (9)$$

where expected host abundance, λ_s , differs by disease state s . Expected abundance could also change by location if site-level covariates (indexed by i) are included. We model subsequent seasons ($t \geq 2$) by considering the number of hosts at each site i that survive

in disease state (S), transition between disease states (T) and are recruited via immigration and reproduction (G).

To model the number of hosts that survive from season $t - 1$ to t at site i , we define parameter ω_s as the disease state-specific apparent survival probability for an uninfected ($s = 1$) and infected ($s = 2$) hosts, such that:

$$S_{s,i,t} \sim \text{bin}(N_{s,i,t-1}, \omega_s). \quad (10)$$

We specify the number of hosts that transition from disease state s to ss at site i from season $t - 1$ to t ($T_{s(ss),i,t}$) based on site-specific transmission risk ($c_{i,t}$) and recovery probability ($r_{i,t}$):

$$T_{1(2),i,t} \sim \text{bin}(S_{1,i,t}, c_{i,t}), \quad (11)$$

$$T_{2(1),i,t} \sim \text{bin}(S_{2,i,t}, r_{i,t}). \quad (12)$$

With this specification, hosts must first survive with the probability associated with their disease state ω_s in season $t - 1$ before transitioning to another disease state. Finally, we model the number of hosts gained, $G_{s,i,t}$, to each disease state s at site i from season $t - 1$ to t :

$$G_{s,i,t} \sim \text{Poisson}(\gamma_{s,i,t}), \quad (13)$$

where $\gamma_{s,i,t}$ is the expected number of uninfected ($s = 1$) and infected ($s = 2$) hosts recruited (either by immigration or reproduction) to site i between seasons.

The state-specific host abundance, $N_{s,i,t}$, for disease state s at site i during season t is then the sum of the number of hosts that are gained at a site, survive at a site, and remain there, and those

TABLE 2 Assumption violations, problems, and solutions to *N*-mixture models

Violation	Why is it a problem?	What can be done?	Citation
Double counting	When less than one animal in twenty is double counted, then the model biases estimates of abundance by 21%	Use <i>N</i> -mixture model for relative abundances or allocate additional resources for mark-recapture data	Link et al. (2018)
Unmodeled variation in population size over time	Estimation of average abundance is biased, bias increases as the proportion of variation in population size that occurs among sites decreases	Use <i>N</i> -mixture model for relative abundances or allocate additional resources for mark-recapture data	Link et al. (2018)
Unmodeled variation in detection probability over time	2% variation in detection results in 19% to 21% additional bias in estimation of average abundance	Use <i>N</i> -mixture model for relative abundances or allocate additional resources for mark-recapture data	Link et al. (2018)
Unmodeled variation in detection probability over time	Alternative models are indistinguishable, and no reliable estimate for abundance can be obtained	Use <i>N</i> -mixture model for relative abundances or allocate additional resources for mark-recapture data	Barker et al. (2018)
When detection probability and number of sampling occasions are small	An infinite estimate of abundance can arise	Use a sample covariance as a diagnostic test to identify this problem	Dennis, Morgan, and Ridout (2015)

that transition into disease state *s* minus those that transition out of disease state *s*:

$$N_{s,i,t} = G_{s,i,t} + S_{s,i,t} + T_{ss(s),i,t} - T_{s(ss),i,t} \quad (14)$$

3.3 | Assumptions and limitations

While we highlight the utility and potential for the use of *N*-mixture model to studying disease dynamics, the benefits of cheaper sampling designs (i.e., count data compared to mark-recapture) comes at inferential costs, such as restrictive assumptions and less power (Table 2; Barker, Schofield, Link, & Sauer, 2018; Link, Schofield, Barker, & Sauer, 2018). Like all statistical models, *N*-mixture models are an oversimplification of the biological world, but they provide the opportunity to extrapolate demographic information from count data and account for sampling error. Kéry (2018) concluded that parameter identifiability problems associated with *N*-mixture models arise when the number of sites sampled is small, but that most models are well identifiable. Failure to meet modeling assumptions for structured *N*-mixture models (e.g., disease, age, sex, size, etc.) will likely result in similar problems to single-season *N*-mixture models (Table 2), but such models have not yet been explicitly evaluated. Structured *N*-mixture models may be more susceptible to unidentified parameter biases, depending on circumstances, because of their complicated framework. Assumption violations may also have cascading effects. For example, if the closure assumption were violated such that there was high immigration of infected hosts within a primary sampling period, estimates of disease transmission and infected host survival probability could be overestimated, and recovery and uninfected host survival probability could be underestimated. Likewise, estimates of infected and uninfected host abundance would be

biased, overestimating numbers of infected hosts, and therefore, pathogen prevalence and average infection intensity. Structured *N*-mixture models are also likely to require a greater number of sites, more repeated surveys, and a larger number of primary seasons in order for parameter identifiability. Thus, a wider breadth of count data is required to compensate for the lack of depth of demographic information that is otherwise present in mark-recapture data.

4 | FUTURE DIRECTIONS

Emerging infectious diseases are challenging to forecast, which impedes development of optimal pathogen control strategies (Daszak et al., 2000; Jones et al., 2008; Fisher et al., 2012; but see Russell et al., 2017). One of the primary goals for management of emerging infectious diseases is to minimize pathogen spread and their impacts on host populations (Bielby, Cooper, Cunningham, Garner, & Purvis, 2008; Kilpatrick, Daszak, Goodman, Rogg, et al., 2006; Kilpatrick, Daszak, Jones, Marra, & Kramer, 2006; Kilpatrick, Kramer, Jones, Marra, & Daszak, 2006; Langwig et al., 2015; Smith, Waller, Russell, Childs, & Real, 2005). Given the typically small population sizes of wildlife disease ecology studies, *N*-mixture models can be a powerful tool to estimate disease dynamics, especially when combined with other sources of information (e.g., informative priors) or data (e.g., mark-recapture). Below, we suggest potential avenues of research to further exploit unmarked datasets for disease analyses.

4.1 | False positives

Ignoring false positives also results in disease state misclassification, where uninfected hosts are misclassified as infected (i.e., false positive). False positives (i.e., incorrectly diagnosing uninfected hosts

as infected) can occur during pathogen diagnostic tests because of cross-contamination of samples or lack of diagnostic test specificity. Consideration of false positives in disease ecology is scarce (but see Abad-Franch, Valença-Barbosa, Sarquis, & Lima, 2014) as most studies assert that either strict protocols were used to prevent cross-contamination or that negative controls on PCR plates were used to identify cases where false positives occurred (e.g., Lachish et al., 2012; Miller et al., 2012; Schmidt, Kéry, Ursenbacher, Hyman, & Collins, 2013; Zelé et al., 2014). However, strict protocols alone may not be enough to eliminate false positives. For example, false positives can occur from carryover of amplified DNA sequences, reagent contamination, positive displacement by pipettes, and aerosol effects (Kwok, 1990). Modeling approaches that estimate both false negatives and false positives could accommodate ambiguous state assignment by supplying prior information on the probability of a false positive (e.g., Miller et al., 2011). The prior information can take the form of an inequality constraint (e.g., false negatives < false positives; see Royle & Link, 2006) or a prior distribution informed by a priori knowledge, expert opinion, or data (Chambert, Miller, & Nichols, 2015; Miller et al., 2011). To date, there have only been detection/non-detection models of false positives (Chambert et al., 2015; Miller et al., 2011). If *N*-mixture models were combined with false-positive models, then that would aid in discerning double counts (see Table 2 for consequences of double counts).

4.2 | Spatially explicit models

An emerging frontier for *N*-mixture models is spatially explicit dynamic models (Zhao et al., 2017), in which emigration and immigration are separated from survival and reproduction, respectively. Spatially explicit models have the advantage of providing biologically realistic models across heterogeneous space. Spatial heterogeneity can be added into models through covariates, with random spatial effects (e.g., Yackulic et al., 2012), or by making assumptions about the ability of individuals to move between sites (e.g., assuming that movement can only occur among adjacent habitat patches; Zhao et al., 2017). Within a disease context, spatially explicit models could be used to identify spatial metapopulations or habitats that serve as pathogen hotspots (i.e., areas where pathogen survival and transmission are high) or host refugia (i.e., areas where host survival is high and pathogen transmission is low; Paull et al., 2012). In this case, conservation managers can target disease prevention to specific metapopulations or habitat patches rather than an entire region.

4.3 | Integration of untapped data sources

Count data can be combined with other data types (e.g., mark-recapture data, opportunistic presence-only data) to model disease dynamics. Recently developed approaches, such as integrated population models and integrated distribution models, combine different types of data including count, detection–non-detection, mark–recapture data, or opportunistic presence only (e.g., Koshkina

et al., 2017; Morris, Reich, Pacifici, & Lei, 2017; Pacifici et al., 2017; Zipkin et al., 2017), improving parameter accuracy and precision, and thus inferences on population-level processes. New modeling frameworks are regularly developed to take advantage of cheaply collected data to estimate demographic rates. For example, the dynamic *N*-occupancy model uses only detection–non-detection data to estimate abundance, population gains (i.e., via immigration and reproduction), and apparent survival probabilities (Rossman et al., 2016). Extending such models to accommodate disease structures might be useful in the integration of multiple data types, provided that certain assumptions and data requirements are met.

Applying *N*-mixture models to long-term and/or spatially expansive datasets may allow for assessments of historic patterns of disease dynamics and predictive capacity to forecast impacts of novel pathogen invasions. Citizen science programs, such as the FeederWatch project (<http://feederwatch.org/>), Saving Salamanders with Citizen Science (<http://www.amphibians.org/>), ZomBee Watch (<https://www.zombeewatch.org/>), and Monarch Health (<http://monarchparasites.org/>), have the capacity to contribute valuable data to such analyses. Combining citizen science data with data obtained from designed, researcher-collected surveys is likely to yield more precise estimates of demographic and disease parameters than can be obtained solely from the individual sources (e.g., van Strien, Swaay, & Termaat, 2013).

5 | CONCLUSIONS

The rise of emerging infectious diseases requires rapid creation of disease mitigation programs (e.g., Langwig et al., 2015). Yet, to develop effective preventions and controls, wildlife managers need estimates of disease prevalence, pathogen infection intensity, transmission, and recovery rates—all of which may be biased by sampling error. The *N*-mixture modeling framework that we outline here can be modified for specific control scenarios and can additionally accommodate outside sources of data—including expert opinion, borrowed information from better-studied disease systems, and laboratory or pilot studies (Russell et al., 2017). Model extensions can include more complicated disease structures such as multiple host species, multiple pathogen infections, age structure, stage structure, or heterogeneity in rates, although inevitably more data will be needed to estimate additional parameters in such models. The template that we provide (including code within the Supporting Information) can be customized based on specific study objectives. Unmarked data offer excellent opportunities to understand complex host–pathogen interactions, bridging the long-standing gap between disease ecology and theory-based wildlife disease management.

ACKNOWLEDGMENTS

We thank the Zipkin Lab for valuable input, J. Andy Royle, Luke Browne, Adrienne Brand, two anonymous reviewers, Beth Gardner, Robert Robinson, and Bob O'Hara for comments that improved the

quality of the manuscript. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government. This is contribution number 680 of the U.S. Geological Survey Amphibian Research and Monitoring Initiative (ARMI).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

GVD wrote the first draft of the manuscript. All authors contributed to the ideas, editing, and review of subsequent drafts.

DATA ACCESSIBILITY

This manuscript does not use any data, but all code and R scripts to simulate data can be found in the Supporting Information or online on the Github repositories of Grace89 (<https://doi.org/10.5281/zenodo.2223304>).

ORCID

Graziella V. DiRenzo  <https://orcid.org/0000-0001-5264-4762>

Sarah P. Saunders  <https://orcid.org/0000-0002-2688-9528>

Evan H. Campbell Grant  <https://orcid.org/0000-0003-4401-6496>

REFERENCES

- Abad-Franch, F., Valença-Barbosa, C., Sarquis, O., & Lima, M. M. (2014). All that glitters is not gold: Sampling-process uncertainty in disease-vector surveys with false-negative and false-positive detections. *PLoS Neglected Tropical Diseases*, 8, e3187. <https://doi.org/10.1371/journal.pntd.0003187>
- Anderson, R. M., & May, R. M. (1979). Population biology of infectious disease: Part I. *Nature*, 280, 361–367.
- Barker, R. J., Schofield, M. R., Link, W. A., & Sauer, J. R. (2018). On the reliability of N-mixture models for count data. *Biometrics*, 74, 369–377. <https://doi.org/10.1111/biom.12734>
- Barlow, N. D. (1995). Critical evaluation of wildlife disease models. In B. T. Grenfell, & A. P. Dobson (Eds.), *Ecology of infectious diseases in natural populations* (pp. 230–259). Cambridge, UK: Cambridge University Press.
- Bielby, J., Cooper, N., Cunningham, A. A., Garner, T. W. J., & Purvis, A. (2008). Predicting susceptibility to future declines in the world's frogs. *Conservation Letters*, 1, 82–90. <https://doi.org/10.1111/j.1755-263X.2008.00015.x>
- Boersch-Supan, P. H., Ryan, S. J., & Johnson, L. R. (2016). deBInfer: Bayesian inference for dynamical models of biological systems in R. *Methods in Ecology and Evolution*, 8, 511–518.
- Briggs, C. J., Knapp, R. A., & Vredenburg, V. T. (2010). Enzootic and epizootic dynamics of the chytrid fungal pathogen of amphibians. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 9695–9700. <https://doi.org/10.1073/pnas.0912886107>
- Brintz, B., Fuentes, C., & Madsen, L. (2018). An asymptotic approximation to the N-mixture model for the estimation of disease prevalence. *Biometrics*. <https://doi.org/10.1111/biom.12913>
- Chambert, T. C., Miller, D. A. W., & Nichols, J. D. (2015). Modeling false positive detections in species occurrence data under different study designs. *Ecology*, 96, 332–339. <https://doi.org/10.1890/14-1507.1>
- Conn, P. B., & Cooch, E. G. (2009). Multi-state capture-recapture analysis under imperfect state observation: An application to disease models. *Journal of Applied Ecology*, 46, 486–492. <https://doi.org/10.1111/j.1365-2664.2008.01597.x>
- Cooch, E. G., Conn, P. B., Ellner, S. P., Dobson, A. P., & Pollock, K. H. (2012). Disease dynamics in wild populations: Modeling and estimation: A review. *Journal of Ornithology*, 152, S485–S509. <https://doi.org/10.1007/s10336-010-0636-3>
- Dail, D., & Madsen, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*, 67, 577–587. <https://doi.org/10.1111/j.1541-0420.2010.01465.x>
- Daszak, P., Cunningham, A., & Hyatt, A. (2000). Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science*, 287, 443–449. <https://doi.org/10.1126/science.287.5452.443>
- Dennis, E. B., Morgan, B. J. T., & Ridout, M. S. (2015). Computational aspects of N-mixture models. *Biometrics*, 71, 237–246. <https://doi.org/10.1111/biom.12246>
- DiRenzo, G. V., Campbell Grant, E. H., Longo, A. V., Che-Castaldo, C., Zamudio, K. R., & Lips, K. R. (2018). Imperfect pathogen detection from non-invasive skin swabs biases disease inference. *Methods in Ecology and Evolution*, 9, 380–389.
- DiRenzo, G. V., Zipkin, E. F., Grant, E. H. C., Royle, J. A., Longo, A. V., Zamudio, K. R., & Lips, K. R. (2018). Eco-evolutionary rescue promotes host-pathogen coexistence. *Ecological Applications*, 28(8), 1948–1962. <https://doi.org/10.1002/eap.1792>
- Drewe, J. A., Dean, G. S., Michel, A. L., & Pearce, G. P. (2009). Accuracy of three diagnostic tests for determining *Mycobacterium bovis* infection status in live-sampled wild meerkats (*Suricata suricatta*). *Journal of Veterinary Diagnostic Investigation*, 21, 31–39.
- Enoe, C., Georgiadis, M. P., & Johnson, W. O. (2000). Estimation of sensitivity and Sp of diagnostic tests and disease prevalence when the true disease status is unknown. *Preventive Veterinary Medicine*, 45, 61–81.
- Faustino, C. R., Jennelle, C. S., Connolly, V., Davis, A. K., Swarthout, E. C., Dhondt, A. A., & Cooch, E. G. (2004). *Mycoplasma gallisepticum* infection dynamics in a house finch population: Seasonal variation in survival, encounter and transmission rate. *Journal of Animal Ecology*, 73, 651–669. <https://doi.org/10.1111/j.0021-8790.2004.00840.x>
- Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., & Gurr, S. J. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature*, 484, 186–194. <https://doi.org/10.1038/nature10947>
- Gómez-Díaz, E., Doherty, P. F., Duneau, D., & McCoy, K. D. (2010). Cryptic vector divergence masks vector-specific patterns of infection: An example from the marine cycle of Lyme borreliosis. *Evolutionary Applications*, 3, 391–401. <https://doi.org/10.1111/j.1752-4571.2010.00127.x>
- Grant, E. H. C., Muths, E., Katz, R. A., Canessa, S., Adams, M. J., Ballard, J. R., ... White, C. L. A. (2017). Using decision analysis to support proactive management of emerging infectious wildlife diseases. *Frontiers in Ecology and the Environment*, 15, 214–221. <https://doi.org/10.1002/fee.1481>
- Greiner, M., & Gardner, I. A. (2000). Epidemiologic issues in the validation of veterinary diagnostic tests. *Preventive Veterinary Medicine*, 45, 3–22. [https://doi.org/10.1016/S0167-5877\(00\)00114-8](https://doi.org/10.1016/S0167-5877(00)00114-8)
- Harmsen, B. J., Foster, R. J., & Doncaster, C. P. (2011). Heterogeneous capture rates in low density populations and consequences for capture-recapture analysis of camera-trap data. *Population Ecology*, 53, 253–259. <https://doi.org/10.1007/s10144-010-0211-z>

- Hostetler, J. A., & Chandler, R. B. (2015). Improved state-space models for inference about spatial and temporal variation in abundance from count data. *Ecology*, *96*, 1713–1723. <https://doi.org/10.1890/14-1487.1>
- Jennelle, C., Cooch, E., Conroy, M. J., & Senar, J. C. (2007). State-specific detection probabilities and disease prevalence. *Ecological Applications*, *17*, 154–167. [https://doi.org/10.1890/1051-0761\(2007\)017\[0154:SDPADP\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2007)017[0154:SDPADP]2.0.CO;2)
- Jones, K., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., ... Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, *451*, 990–993. <https://doi.org/10.1038/nature06536>
- Joseph, M. B., Mihaljevic, J. R., Arellano, A. L., Kueneman, J. G., Preston, D. L., Cross, P. C., & Johnson, P. T. J. (2013). Taming wildlife disease: Bridging the gap between science and management. *The Journal of Applied Ecology*, *50*, 702–712. <https://doi.org/10.1111/1365-2664.12084>
- Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *11*, 700–721. <https://doi.org/10.1098/rspa.1927.0118>
- Kéry, M. (2018). Identifiability in N-mixture models: A large-scale screening test with bird data. *Ecology*, *99*, 281–288.
- Kéry, M., & Schaub, M. (2012). *Bayesian population analysis using WinBUGS*. Waltham, MA: Elsevier.
- Kilpatrick, A. M., Daszak, P., Goodman, S. J., Rogg, H., Kramer, L. D., Cedeño, V., ... Cunningham, A. A. (2006). Predicting pathogen introduction: West Nile virus spread to Galápagos. *Conservation Biology*, *20*, 1224–1231.
- Kilpatrick, A. M., Daszak, P., Jones, M. J., Marra, P. P., & Kramer, L. D. (2006). Host heterogeneity dominates West Nile virus transmission. *Proceedings of the Royal Society B. Biological Sciences*, *273*, 2327–2333.
- Kilpatrick, A. M., Kramer, L. D., Jones, M. J., Marra, P. P., & Daszak, P. (2006). West Nile virus epidemics in North America are driven by shifts in mosquito feeding behavior. *PLoS Biology*, *4*, 606–610.
- Knowles, S. C. L., Wood, M. J., Alves, R., Wilkin, T. A., Bensch, S., & Sheldon, B. C. (2011). Molecular epidemiology of malaria prevalence and parasitaemia in a wild bird population. *Molecular Ecology*, *20*, 1062–1076. <https://doi.org/10.1111/j.1365-294X.2010.04909.x>
- Koshkina, V., Wang, Y., Gordon, A., Dorazio, R. M., White, M., & Stone, L. (2017). Integrated species distribution models: Combining presence-background data and site-occupancy data with imperfect detection. *Methods in Ecology and Evolution*, *8*, 420–430. <https://doi.org/10.1111/2041-210X.12738>
- Kwok, S. (1990). Procedures to minimize PCR-product carry-over. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, & T. J. White (Eds.), *PCR protocols: A guide to methods and applications* (pp. 142–145). London, UK: Academic Press.
- Lachish, S., Gopalaswamy, A. M., Knowles, S. C. L., & Sheldon, B. C. (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. <https://doi.org/10.1111/j.2041-210X.2011.00156.x>
- Lachish, S., Knowles, S. C. L., Alves, R., Wood, M. J., & Sheldon, B. C. (2011). Fitness effects of endemic malaria infections in a wild bird population: The importance of ecological structure. *The Journal of Animal Ecology*, *80*, 1196–1206. <https://doi.org/10.1111/j.1365-2656.2011.01836.x>
- Langwig, K. E., Voyles, J., Wilber, M. Q., Frick, W. F., Murray, K. A., Bolker, B. M., ... Kilpatrick, A. M. (2015). Context-dependent conservation responses to emerging wildlife diseases. *Frontiers in Ecology and the Environment*, *13*, 195–202. <https://doi.org/10.1890/140241>
- Link, W. A., Schofield, M. R., Barker, R. J., & Sauer, J. R. (2018). On the robustness of N-mixture models. *Ecology*, *99*, 1547–1551. <https://doi.org/10.1002/ecy.2362>
- Lloyd-Smith, J. O., Cross, P. C., Briggs, C. J., Daugherty, M., Getz, W. M., Latto, J., ... Swei, A. (2005). Should we expect population thresholds for wildlife disease? *Trends in Ecology and Evolution*, *20*, 511–519. <https://doi.org/10.1016/j.tree.2005.07.004>
- McClintock, B. T., Nichols, J. D., Bailey, L. L., MacKenzie, D. I., Kendall, W. L., & Franklin, A. B. (2010). Seeking a second opinion: Uncertainty in disease ecology. *Ecology Letters*, *13*, 659–674. <https://doi.org/10.1111/j.1461-0248.2010.01472.x>
- Miller, D. A., Nichols, J. D., McClintock, B. T., Grant, E. H. C., Bailey, L. L., & Weir, L. A. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology*, *92*, 1422–1428. <https://doi.org/10.1890/10-1396.1>
- Miller, D. A. W., Talley, B. L., Lips, K. R., & Grant, E. H. C. (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. <https://doi.org/10.1111/j.2041-210X.2012.00216.x>
- Morris, S. A., Reich, B. J., Pacifici, K., & Lei, Y. (2017). A spatial model for rare binary events. *Environmental and Ecological Statistics*, *24*, 485–504. <https://doi.org/10.1007/s10651-017-0385-z>
- Mosher, B. A., Bernard, R. F., Miller, D. A. W., Richgels, K. L. D., White, C. L., Lorch, J., & Grant, E. H. C. (2018). Broadening the conversation molecular tools, conservation, and communication. *Frontiers in Ecology and the Environment*, in review.
- Pacifici, K., Reich, B. J., Miller, D. A. W., Gardner, B., Stauffer, G., Singh, S., ... Collazo, J. A. (2017). Integrating multiple data sources in species distribution modeling: A framework for data fusion. *Ecology*, *98*, 840–850. <https://doi.org/10.1002/ecy.1710>
- Parmenter, C. A., Yates, T. L., Parmenter, R. R., Mills, J. N., Childs, J. E., Campbell, M. L., ... Milner, J. (1998). Small mammal survival and trapability in mark-recapture monitoring programs for hantavirus. *Journal of Wildlife Diseases*, *34*, 1–12. <https://doi.org/10.7589/0090-3558-34.1.1>
- Paull, S. H., Song, S., McClure, K. M., Sackett, L. C., Kilpatrick, A. M., & Johnson, P. T. (2012). From superspreaders to disease hotspots: Linking transmission across hosts and space. *Frontiers in Ecology and the Environment*, *10*, 75–82. <https://doi.org/10.1890/110111>
- Poulin, R. (2010). *Parasite manipulation of host behavior: An update and frequency asked questions. Advances in the study of behavior* (1st ed., Vol. 41). New York, NY: Elsevier Science Publishing Co Inc.
- Pryde, M. A., O'Donnell, C. F. J., & Barker, R. J. (2005). Factors influencing survival and long-term population viability of New Zealand long-tailed bats (*Chalinolobus tuberculatus*): Implications for conservation. *Biological Conservation*, *126*, 175–185. <https://doi.org/10.1016/j.biocon.2005.05.006>
- Rossmann, S., Yackulic, C. B., Saunders, S. P., Reid, J., Davis, R., & Zipkin, E. F. (2016). Dynamic N-occupancy models: Estimating demographic rates and local abundances from detection-nondetection data. *Ecology*, *97*, 3300–3307.
- Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, *60*, 108–115.
- Royle, J. A., & Link, W. A. (2006). Generalized site occupancy models allowing for false positives and false negative errors. *Ecology*, *87*, 835–841.
- Russell, R. E., Katz, R. A., Richgels, K. L. D., Walsh, D. P., & Grant, E. H. C. (2017). A framework for modeling emerging diseases to inform management. *Emerging Infectious Diseases*, *23*, 1–6. <https://doi.org/10.3201/eid2301.161452>
- Schmidt, B. R. (2010). Estimating the impact of disease in species threatened by amphibian chytrid fungus: Comment on Murray et al. *Conservation Biology*, *24*, 897–899. <https://doi.org/10.1111/j.1523-1739.2010.01507.x>
- Schmidt, B. R., Kéry, M., Ursenbacher, S., Hyman, O. J., & Collins, J. P. (2013). Site occupancy models in the analysis of environmental DNA presence/absence surveys: A case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution*, *4*, 646–653. <https://doi.org/10.1111/2041-210X.12052>

- Senar, J. C., & Conroy, M. J. (2004). Multi-state analysis of the impacts of avian pox on a population of Serins (*Serinus serinus*): The importance of estimating recapture rates. *Animal Biodiversity and Conservation*, 27, 133–146.
- Smith, D. L., Waller, L. A., Russell, C. A., Childs, J. E., & Real, L. A. (2005). Assessing the role of long-distance translocation and spatial heterogeneity in the raccoon rabies epidemic in Connecticut. *Preventive Veterinary Medicine*, 71, 225–240. <https://doi.org/10.1016/j.prevetmed.2005.07.009>
- Smith, M. J., Telfer, S., Kallio, E. R., Burthe, S., Cook, A. R., Lambin, X., & Begon, M. (2009). Host-pathogen time series data in wildlife support a transmission function between density and frequency dependence. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 7905–7909. <https://doi.org/10.1073/pnas.0809145106>
- Toft, N., Jørgensen, E., & Højsgaard, S. (2005). Diagnosing diagnostic tests: Evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Preventive Veterinary Medicine*, 68, 19–33. <https://doi.org/10.1016/j.prevetmed.2005.01.006>
- van Strien, A. J., van Swaay, C. A. M., & Termaat, T. (2013). Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *The Journal of Applied Ecology*, 50, 1450–1458. <https://doi.org/10.1111/1365-2664.12158>
- Williams, B. K., Nichols, J. D., & Conroy, M. J. (2002). *Analysis and management of animal populations*, 1st ed. San Diego, CA: Academic Press.
- Yackulic, C. B., Reid, J., Davis, R., Hines, J. E., Nichols, J. D., & Forsman, E. (2012). Neighborhood and habitat effects on vital rates: Expansion of the barred owl in the Oregon Coast ranges. *Ecology*, 93, 1953–1966. <https://doi.org/10.1890/11-1709.1>
- Zelé, F., Vezilier, J., Lambert, G., Nicot, A., Gandon, S., Rivero, A., & Duron, O. (2014). Dynamics of prevalence and diversity of avian malaria infections in wild *Culex pipiens* mosquitoes: The effects of Wolbachia, filarial nematodes and insecticide resistance. *Parasites & Vectors*, 7, 437. <https://doi.org/10.1186/1756-3305-7-437>
- Zhao, Q., Royle, J. A., & Boomer, G. S. (2017). Spatially explicit dynamic N-mixture models. *Population Ecology*, 59, 293–300. <https://doi.org/10.1007/s10144-017-0600-7>
- Zipkin, E. F., Kraft, C. E., Cooch, E. G., & Sullivan, P. J. (2009). When can efforts to control nuisance and invasive species backfire? *Ecological Applications*, 19, 1585–1595. <https://doi.org/10.1890/08-1467.1>
- Zipkin, E. F., Rossman, S., Yackulic, C. B., Wiens, J. D., Thorson, J. T., Davis, R. J., & Grant, E. H. C. (2017). Integrating count and detection–nondetection data to model population dynamics. *Ecology*, 98, 1640–1650. <https://doi.org/10.1002/ecy.1831>
- Zipkin, E. F., Sillett, T. S., Grant, E. H. C., Chandler, R. B., & Royle, J. A. (2014). Inferences about population dynamics from count data using multistate models: A comparison to capture–recapture approaches. *Ecology and Evolution*, 4, 417–426.
- Zipkin, E. F., Thorson, J. T., See, K., Lynch, H. L., Grant, E. H. C., Kanno, Y., ... Royle, J. A. (2014). Modeling structured population dynamics using data from unmarked individuals. *Ecology*, 95, 22–29.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: DiRenzo GV, Che-Castaldo C, Saunders SP, Campbell Grant EH, Zipkin EF. Disease-structured *N*-mixture models: A practical guide to model disease dynamics using count data. *Ecol Evol*. 2019;9:899–909. <https://doi.org/10.1002/ece3.4849>