Next generation serology: integrating cross-sectional and capture-recapture approaches to infer disease dynamics

Amandine Gamble, Romain Garnier, Thierry Chambert, Olivier Gimenez, Thierry Boulinier

Running head: Next generation serology field design

Next generation serology: integrating cross-sectional and capture-recapture approaches to infer disease dynamics

Amandine Gamble[1,2,*], Romain Garnier[3], Thierry Chambert[1], Olivier Gimenez[1] and Thierry Boulinier[1]

[1] CEFE, CNRS, University of Montpellier, EPHE, University Paul Valéry Montpellier 3, IRD, Montpellier, France

[2] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, United States

[3] Department of Biology, Georgetown University, Washington, DC, United States

* Corresponding author: Amandine Gamble. Email: amandine.gamble@gmail.com.

Abstract. Two approaches have been classically used in disease ecology to estimate epidemiological parameters from field studies: cross-sectional sampling from unmarked individuals and longitudinal capture-recapture setups, which generally involve more limited numbers of marked individuals due to cost and logistical constrains. Although the benefits of longitudinal setups are increasingly acknowledged in the disease ecology community, cross-sectional data remain largely over-represented in the literature, probably because of the inherent costs of longitudinal surveys. In this context, we used simulated data to compare the performances of cross-sectional and longitudinal designs to estimate the force of infection (*i.e.*, the rate at which susceptible individuals become infected). Then, inspired from recent method developments in quantitative ecology, we explore the benefits of integrating both cross-sectional (seroprevalences) and longitudinal (individuals histories) datasets. In doing so, we investigate the effects of host species life history, antibody persistence and degree of *a priori* knowledge and uncertainty on demographic and epidemiological parameters, as those are expected to affect in different ways the level of inference possible from the data. Our results highlight how those elements are important to consider to determine optimal sampling designs. In the case of long-lived species exposed to infectious agents resulting in persistent antibody responses, integrated designs are especially valuable as they benefit from the performances of longitudinal designs even with relatively small longitudinal sample sizes. As an illustration, we apply this approach to a combination of empirical and simulated data inspired from a case of bats exposed to a rabies virus. Overall, this work highlights that serology field studies could greatly benefit from the opportunity of integrating cross-sectional and longitudinal designs.

Key-words: eco-epidemiology, detectability, immunity persistence, sampling strategy, study design, wildlife

INTRODUCTION

Understanding the ecology and evolution of infectious diseases in wildlife has been highlighted as critical for public health (Jones et al. 2008) and biodiversity conservation (Smith et al. 2006). Natural host-parasite systems also offer useful models to obtain valuable insights on evolutionary ecology processes such as coevolution and local adaptation (Gandon 2002) or host and vector movements (Boulinier et al. 2016). However, investigations in the wild have been hampered by the difficulty of collecting data allowing efficient inference of eco-epidemiological dynamics (Plowright et al. 2019). For instance, the force of infection (*i.e.*, the rate at which susceptible individuals acquire an infectious disease), a key eco-epidemiological parameter (Hens et al. 2012), is difficult to estimate from field data as it requires assessing how many individuals went from susceptible (e.g., non-infected and non-immunized) to infected in a given time period, which is rarely observable. Estimating these parameters is however a critical step in the characterization of epidemiological dynamics and factors impacting them. Methods allowing their estimation from field data are thus needed.

The benefits of longitudinal setups, defined here as the repeated sampling of the same individuals across time, notably using capture-recapture designs, are increasingly acknowledged in the disease ecology community (e.g., Jenelle et al. 2007, Lachish et al. 2007, Chambert et al. 2012, Buzdugan et al. 2017, Marescot et al. 2018). However, cross-sectional data, defined here as the sampling of unmarked individuals at one or more points in time, remain largely over-represented in the literature, probably because of the inherent costs of longitudinal surveys. It requires much more time and skills to spot marked individuals and to recapture them than to capture a random sample of individuals in a target population (e.g., if a marked fur seal is spotted in the middle of a harem, field workers may have to postpone the

capture to limit disturbance and biting risks, while in a cross-sectional sampling design, the capture of another, more peripheral, individual would be much easier).

Recent advances in population ecology, such as the advent of integrated modeling, may open new perspectives for the estimation of eco-epidemiological parameters. Indeed, Integrated Population Modelling (IPM) has proven effective to improve demographic parameter estimations by integrating datasets of different natures (e.g., capture-recapture and counts) on the condition that they depend partly on the same set of (demographic) parameters (Besbeas et al. 2002, Schaub et al. 2007, Abadi et al. 2010, Fletcher et al. 2019). In disease ecology, a similar approach could thus be used to integrate low cost cross-sectional data with longitudinal data that provide key elements about processes underlying the dynamics of the considered variables (e.g., the kinetics of the immune response). IPM has been recently applied in an epidemiological context (McDonald et al. 2016), but to our knowledge approaches integrating cross-sectional and capture-recapture epidemiological data have never been explicitly used to estimate epidemiological parameters.

In some species, individuals can be marked and repeatedly (re)captured across time, allowing longitudinal sampling. This is particularly true for long-lived vertebrates showing seasonal and colonial breeding (such as seabirds, pinnipeds, and chiropterans) and which are often faithful to their breeding or roosting site (e.g., Chambert et al. 2012b, Robardet et al. 2017, Gamble et al. 2019a). In these systems, capture-recapture approaches have started to be used to estimate epidemiological state transition probabilities (e.g., from healthy to symptomatic) while accounting for recapture probabilities below unity, which are unavoidable in wild settings (Jennelle et al. 2007, Conn and Cooch 2009). However, longitudinal studies are usually based on relatively small sample sizes because field efforts needed to resight and recapture marked individuals tend to be intensive. In contrast, cross-sectional studies are

usually less costly and may also allow the estimation of epidemiological state transition probabilities. This type of data can generally be used to monitor variations of prevalences (*i.e.*, the proportion of infected individuals) or seroprevalences (*i.e.*, proportions of seropositive individuals). However, linking variations of prevalences or seroprevalences to epidemiological dynamics often requires additional data seldom available in wild populations, such as knowledge on the infectious period (e.g., Hénaux et al. 2010) and/or refined antibody kinetic curves (e.g., Borremans et al. 2016, Pepin et al. 2017), or strong assumptions on the host demography (e.g., Samuel et al. 2015). Both approaches (longitudinal and cross-sectional) thus present relative pros and cons. Because cross-sectional and longitudinal data are outcomes of the same eco-epidemiological processes based on the same demographic and epidemiological parameters (notably survival, force of infection, and antibody level persistence), their combination into an integrated model should improve the estimation of these parameters.

Serology has proven effective to detect patterns of exposure to many infectious agents and infer eco-epidemiological processes (Gilbert et al. 2013, Metcalf et al. 2016). Moreover, a wide range of approaches are now available to apply serology to wild settings (e.g., Garnier et al. 2017). However, the interpretation of serological data is not straightforward as they do not directly inform on the timing of infection. The reliability of the inference that can be made from serological data is thus dependent on the ecological and epidemiological characteristics of the considered system. Sampling schemes may need to be adjusted to reflect both these characteristics and what is possible in terms of field efforts. For instance, in some host-parasite systems, detectable antibody levels persist for many years after exposure (e.g., antibody level against the Newcastle disease virus vaccine in Ramos et al. 2014), while in other cases, they wane within a few weeks (e.g., antibody level against the avian cholera agent

in Samuel et al. 2003), complicating interpretation of serological data. Methods allowing the estimation of the force of infection from serological data when the kinetics of the immune response is not known are needed to better characterize the factors driving epidemiological dynamics.

In the present study, we use a simulation approach to compare the performances of different sampling designs to estimate the seroconversion probability, a proxy of the force of infection, when the kinetics of the immune response after exposure is not known. This parameter can be estimated either from the temporal variations of the seroprevalence based on cross-sectional data (e.g., Samuel et al. 2015) or as the transition probability from seronegative to seropositive states in a capture-recapture model based on longitudinal data (e.g., Conn and Cooch 2009). We moreover consider the possibility of integrating both sources of data in an integrated framework inspired from IPM. Based on data simulated under different scenarios, we notably account for several key parameters expected to have a strong impact on the observation process and the inference that can be made from serological data: host lifespan, temporal persistence of antibody levels, and detection and recapture probabilities. For instance, low annual survival will increase the turnover of individuals in the host population, which is expected to lower the benefit of longitudinal sampling designs, which rely on the repeated sampling of individuals. Finally, we illustrate how this method could be used on empirical data by considering the case of a serotine bat (*Eptesicus serotinus*) colony exposed to a rabies virus.

The results of the present study could have important implications regarding current practices in eco-epidemiology by (1) highlighting the benefits of longitudinal sampling designs compared to cross-sectional sampling designs, and (2) opening to possibility of

integrating the two types of approaches to design cost-efficient sampling protocols in study systems not yet subject to longitudinal monitoring programs.

MATERIALS AND METHODS

Eco-epidemiological model

Individual data resulting from an eco-epidemiological inter-annual process were simulated with a set of parameters fixed to different values in order to represent different demographic and epidemiological situations (Fig. 1 a): survival (ϕ), seroconversion (λ; *i.e.*, the probability for a seronegative individual to become seropositive, which usually corresponds to the mounting of an antibody response after exposure to an infectious agent) and seroreversion (ω; *i.e.*, the probability for a seropositive individual to become seronegative, which corresponds to the waning of the antibody response) probabilities. To illustrate how eco-epidemiological parameters could be quantified from serological data, we have chosen the simple situation of populations at the demographic and endemic equilibria with all individuals recruiting as seronegative and exposure having no impact on survival or detectability. Additional details and illustrations are given in Appendix S1-A.

Cross-sectional sampling

Each year, n_{CS} individuals are randomly captured and sampled for serological analyses. Seroprevalence at time t (π_t) is calculated as the proportion of seropositive individuals among the tested individuals. π_t thus corresponds to the probability for a sample randomly collected in a population to be seropositive at time t. Seroprevalences at times t and t+1 are linked by a function of survival, seroreversion, and seroconversion probabilities. Such approaches have previously been used to estimate seroconversion probabilities in wild populations (e.g., Hénaux et al. 2013, Samuel et al. 2015). Under the eco-epidemiological model assumptions (see above), this relation is given by equation 1:

$$\pi_{t+1} = \pi_t \, \phi \, (1 - \omega) + \pi_t \, \phi \, \omega \, \lambda + (1 - \pi_t) \, \phi \, \lambda + r \, \lambda \quad (1)$$

In equation 1, the first additive term [$\pi_t \, \phi \, (1 - \omega)$] corresponds to seropositive individuals at time t that survive and maintain detectable antibody levels between time t and t+1; the second [$\pi_t \, \phi \, \omega \, \lambda$] to seropositive individuals at time t that survive, lose their antibodies and seroconvert between t and t+1; the third [$(1 - \pi_t) \, \phi \, \lambda$] to seronegative individuals at time t that survive and seroconvert between t and t+1; and the last [$r \, \lambda$] to individuals that recruit (here with a probability r) and seroconvert between t and t+1.

Under the assumption of demographic equilibrium, recruitment exactly compensates for mortality and r can be written as $(1 - \phi)$; and under the assumption of endemic equilibrium, seroprevalence (π^*) is stable over time (equation 2; intermediary steps are clarified in Appendix S1, equations S1-3). Serological states of the samples thus follow the binomial distribution given in equation 3.

$$\pi^* = -\frac{\lambda}{\phi \, (1-\omega + \omega \, \lambda-\lambda)-1} \quad (2) \qquad y \sim B\left(n_{CS}, -\frac{\lambda}{\phi \, (1-\omega + \omega \, \lambda-\lambda)-1}\right) \quad (3)$$

The estimation of unknown parameters will be facilitated if some of these parameters are known *a priori*. In this study, we thus notably considered the case when the model was informed with some values for the survival and the seroreversion probabilities (true or erroneous, e.g., based on the literature). Additional details are given in Appendix S1-A.

Longitudinal sampling

On the first year of the observation process, n_{LG} random individuals are captured and marked with a tag allowing individuals to be identified without recapture (e.g., rings or PIT tags). Each of the following years, each alive marked individual is resighted with a probability p and its serological state is ascertained with a probability δ corresponding to the recapture probability after resighting (the serological state being ascertained at the same time from a

177 blood sample). A fixed number of individuals is captured each year, with a priority on marked
178 individuals and some newly marked individuals if necessary to complete the sample size to
179 n_{LG}. An observation event is then attributed each year to each marked individual of the study
180 and recorded in the matrix *m*: 0 if not seen (for an individual either dead, alive but not present
181 in the study site, or present but not detected), 1 if captured and ascertained as seronegative, 2
182 if captured and ascertained as seropositive or 3 if seen but not captured (uncertain serological
183 state; Appendix S1-A). Note that we considered no state misclassification (*i.e.*, test sensitivity
184 and specificity are equal to one). These assumptions are discussed in Appendix S1-A.
185 Multievent models allowing for state uncertainty (corresponding to event 3) were then fitted
186 on the individual histories (Pradel 2005), similarly to classical applications to demographic
187 studies (Gimenez et al. 2012). Such models are increasingly used in population ecology and in
188 eco-epidemiology (e.g., Conn and Cooch 2009, Robardet et al. 2017, Buzdugan et al. 2017,
189 Marescot et al. 2018).

190 *Integrated modelling*

191 For a given simulated population, the cross-sectional and the longitudinal datasets (*y* and *m*
192 respectively) can be integrated together (Fig. 1 b; Schaub et al. 2007). Under the assumption
193 of independence of the two datasets (only data from unmarked individuals are included in the
194 cross-sectional dataset), the combined likelihood function (L_{IPM}) can thus be expressed as the
195 product of the likelihood function of the cross-sectional (L_{CS}) and longitudinal (L_{LG}) models:

196 $$L_{IPM}(y, m \mid \phi, \lambda, \omega, p, \delta) = L_{CS}(y \mid \phi, \lambda, \omega) \times L_{LG}(m \mid \phi, \lambda, \omega, p, \delta) \quad (4)$$

197 These parameters can thus conjointly be estimated based on the cross-sectional and
198 longitudinal datasets (*y* and *m*). As both datasets result from processes sharing some similar
199 eco-epidemiological parameters, the integrated estimator of these parameters is expected to be
200 less biased and more precise (Schaub et al. 2007, Abadi et al. 2010). As we considered

situations in which a small proportion of the population is sampled (≤ 10 % unmarked individuals and ≤ 10 % marked individuals) and cross-sectional and longitudinal samples were chosen randomly, leading to only a potentially small overlap of the two datasets, we made the assumption that our cross-sectional and longitudinal datasets were independent. In addition to the assumption of independence of the two datasets typical to integrated models, the main assumptions are the ones made by the multievent capture-recapture model (see for instance Riecke et al. 2019) and when formalizing the temporal variations of the seroprevalence. These assumptions are discussed in more details in Appendix S1-A.

Simulations and model fitting

For each set of parameters, 1000 populations with a size of 600 individuals were simulated using a specifically developed individual based model (see Appendix S2 for codes). To compare the performances of both designs under various scenarios, one cross-sectional sample and one longitudinal sample of 50 individuals ($n_{CS} = n_{LG}$) per year were then taken per simulated population following the designs described above. In the case of integrated modelling, several combinations of cross-sectional (n_{CS} = 20, 40 or 60) and longitudinal (n_{LG} = 20 or 40 or 60) sample sizes were tested. Unless otherwise stated, the resighting (p) and recapture (δ) probabilities were set to 0.80 and sampling was conducted over five years after having reached the endemic equilibrium (Fig. S1). Within a time step, samples were collected after exposure. The performances of the estimators were then compared based first on their bias, and second on their Mean Square Error (MSE = bias^2 + variance) in order to account for the bias and the precision of the estimators; the lower the bias or MSE, the more accurate the estimator. In the three cases (cross-sectional, longitudinal and integrated), eco-epidemiological parameters were estimated from the data by maximization of likelihood using a frequentist approach. This method was preferred due to reduced computation time compared

to Bayesian inference. Sensitivity analyses were conducted to explore the validity of the results for ranges of biological and observation parameters. All simulations and analyses were run within R 3.3.3. Simulation codes are provided in Appendix S2, including examples of frequentist and Bayesian estimations of the parameters.

Illustrative example

The integrated estimator was then applied to a real case study of serotine bats (*Eptesicus serotinus*) exposed to a bat rabies virus (European Bat Lyssavirus type 1; EBLV-1) in Pagny-sur-Moselle, France (Robardet et al. 2017). Because we were unable to find a dataset combining cross-sectional and capture-recapture setups in the literature, we chose to use this capture-recapture dataset and to simulate additional cross-sectional data using the simulation model presented above and parameterized based on the demographic and epidemiological parameters estimated using a multievent model. Juvenile serotine bats from the study site are known to be exposed to EBLV-1 (Robardet et al. 2017). Thus, instead of making the assumption that individuals recruit as seronegative (as in equation 1), we made the assumption that females recruit in the breeder pool with the same probability of being seropositive as former breeders (*i.e.*, seroprevalence is similar in the new recruit and former breeder pools), leading to equation 5 in which new recruits and former breeders are not distinguished:

$$\pi_{t+1\ bats,\ EBLV\text{-}1} = \pi_{t\ bats,\ EBLV\text{-}1}\ \phi\ (1-\omega) + \pi_{t\ bats,\ EBLV\text{-}1}\ \phi\ \omega\ \lambda + (1-\pi_{t\ bats,\ EBLV\text{-}1})\ \phi\ \lambda \quad (5)$$

And seroprevalence at the equilibrium can be written:

$$\pi^*_{bats,\ EBLV\text{-}1} = \frac{\lambda}{(\omega - \omega\ \lambda + \lambda)} \quad (6)$$

The simulation of the cross-sectional data was also modified to reflect this assumption.

We considered 102 marked individuals captured between one and five times over eight capture occasions (corresponding to the empirical longitudinal data). In parallel, during each of the eight capture occasions, n_{cs} (20, 40 or 60) unmarked individuals were randomly

captured and used to calculate the seroprevalence at each occasion (corresponding to the simulated cross-sectional data). We then estimated the survival, seroconversion and seroreversion probabilities using the integrated estimator based on the best model retained in Robardet et al. (2017), in which the resighting probability varies over time: $\phi(.)$, $\lambda(.)$, $\omega(.)$, $p(t)$, $\delta(.)$. Additional details are given in Appendix S1-B and codes in Appendix S3.

RESULTS

Cross-sectional estimator. The seroconversion probability (λ) was estimated without bias (*i.e.* absolute difference between the true and estimated value close to zero) when using the cross-sectional estimator informed with the true values of the survival (ϕ) and seroreversion (ω) probabilities (Fig. 2 a and b). In contrast, informing the cross-sectional estimator with slightly erroneous values for these parameters led to biases when lifespan and antibody persistence were long (when ϕ tends to one and ω tends to zero). The bias was smaller when lifespan or antibody persistence were short, which can easily be explained by the fact that when ϕ tends to zero and/or ω tends to one, π^* tends to λ (equation 2) and the seroconversion probability can thus be directly deducted from the observed seroprevalence. Hence, the cross-sectional estimator overall performed better (lower MSE independently of the *a priori* knowledge) when ϕ was low (*i.e.*, short-lived host species) and/or ω was high (*i.e.*, short-lived immune response; Fig. 2 a and b and sensitivity analyses presented in Fig. S4).

Longitudinal estimator. When using the longitudinal estimator, the seroconversion probability (λ) was estimated without bias without any *a priori* knowledge of the true survival (ϕ) and seroreversion (ω) probabilities, except when the survival probability was close to zero (Fig. 2 a). In addition to the higher bias, precision was also lower at low survival probabilities. The lower performances observed for low survival probabilities are expected when using capture-recapture models as fewer individuals can be recaptured over the years, reducing the

effective sample size. Precision was also slightly decreased when antibody level persistence was longer (low ω). This could be explained by the model not being able to distinguish individuals that maintained their antibody levels (at a probability $1 - \omega$) from individuals that were observed seropositive once and then seroreverted and got exposed again (at a probability $\omega \times \lambda$) as both situations fit with the observation of the individuals as seropositive during two consecutive occasions. This is supported by the fact that the precision was lower for higher seroconversion probabilities when the seroreversion was low but not when it was high (Fig. 2 c and d). Hence, the longitudinal estimator overall performed better when ϕ was high (*i.e.*, long-lived host species) and/or ω was high (*i.e.*, short-lived immune response; sensitivity analyses presented in Fig. S4).

Integrated estimator. Similarly to the longitudinal estimator, the integrated estimator of the seroconversion probability (λ) was unbiased without having to rely on any *a priori* knowledge on the survival (ϕ) and/or seroreversion (ω) probabilities (Fig. 3). In addition, integrating cross-sectional data to longitudinal data increased the precision of the estimator for any fixed longitudinal sample size. For instance, when antibody level persistence was long, adding 20 unmarked individuals to 20 marked individuals at each sampled occasion allowed the standard error of the estimated values to be divided by 1.7. The results are not trivial though: for instance, for intermediate antibody level persistence, sampling longitudinally 20 marked individuals and a novel batch of 20 unmarked individuals at each yearly sampling occasion gives a more accurate estimation than sampling longitudinally 40 marked individuals (Fig. 3 b), while this is not the case for persisting antibody levels (Fig. 3 a). In such comparisons, one need to keep in mind the relative field costs (in time spent and skills required) associated with (re)capturing marked versus unmarked individuals (see Fig. S9 for illustrative examples). Additional results are presented in the Appendix S1-C, notably considering the effects of

various biological (host survival, antibody persistence; Fig. S5) and observations parameters (resighting and recapture probabilities, study duration, sample sizes; Fig. S6-8).

Illustrative example. The estimation of seroconversion probability was improved (smaller confidence interval) when longitudinal and cross-sectional data were integrated together (compared to using only longitudinal data; Table 1). For instance, the seroconversion probability [95% confidence interval] was estimated at 0.085 [0.033; 0.201] using the longitudinal design and 0.079 [0.043; 0.139] using the integrated design including data from 60 unmarked individuals each year. The estimates of survival, resighting and recapture probabilities were unchanged, as expected considering that these parameters were not expected to impact seroprevalence (equation 5).

DISCUSSION

Based on an eco-epidemiological model and simulations under different sampling scenarios, our results suggest that longitudinal data analyzed in capture-recapture frameworks are preferable to cross-sectional data when poor *a priori* knowledge (for instance on the survival and seroreversion probabilities) is available on the system, which is the case with most wildlife-parasites systems. The cross-sectional estimator can nonetheless be accurate for hosts with short lifespan and/or short antibody level persistence or when informed with reliable *a priori* knowledge on these parameters. In contrast, the longitudinal approach provided accurate estimates and also allowed survival and seroreversion probabilities to be estimated along with observation parameters (resighting and recapture probabilities; e.g., the serotine bat example). Finally, the integrated estimator benefited from the performances of longitudinal designs, notably it did not rely on any *a priori* known parameters, even with relatively small longitudinal sample sizes. Based on these results, we hope to encourage

321 researchers to think about the benefits of implementing longitudinal setups, potentially of
322 relatively small scope, in parallel to already existing cross-sectional studies. We also propose
323 a method to integrate these two types of data, which we believe could be useful in the future
324 to motivate researchers to switch from cross-sectional to integrated designs. The method we
325 present here also offers the possibility to integrate datasets that were previously analyzed
326 independently, and thus to improve the inference of eco-epidemiological processes made from
327 these data. For instance, multi-site cross-sectional data could be integrated with single-site
328 longitudinal data (e.g., Picard-Meyer et al. 2011 and Robardet et al. 2017) to overcome the
329 need of *a priori* knowledge on the host kinetics of the immune response, which is likely
330 conserved within a species sampled across sites.
331 Although the benefits of longitudinal setups are increasingly acknowledged in the disease
332 ecology community, our study is the first to our knowledge to explore the conditions in which
333 these benefits are actually found. Overall, the results highlight that the key elements to
334 determine an optimal sampling design are: (1) host species life history, (2) the degree of
335 antibody persistence and (3) the degree of *a priori* knowledge and uncertainty on
336 demographic and epidemiologic parameters. This work also stresses the potential benefits of
337 incorporating data from capture-recapture sampling designs in eco-epidemiological analyses,
338 often largely based on cross-sectional field surveys. In practice, this integrated approach
339 would be particularly beneficial in systems in which (1) individuals can be recaptured over
340 several years (relatively long lifespan and high site faithfulness) and (2) large numbers of
341 unmarked individuals can be sampled without increasing too much the cost of the study. This
342 is for instance the case when samples can be collected when accidental capture is frequent
343 (e.g., when using non-targeted capture methods such as mist nets, harp or Sherman traps: e.g.,
344 Robardet et al. 2017, Mariën et al. 2018), or as part of harvesting practices (e.g., Rossi et al.

2005), or from the offspring of colonial breeders (e.g., Chambert et al. 2012b). In such cases, seroprevalence data from unmarked individuals may be collected with minimal additional effort in parallel to capture-recapture setups. For instance, particularly efficient cross-sectional sampling designs may not even require the capture of adults if the sampling of offspring, or eggs, can be used as a reliable alternative to adult blood sampling (Alekseeev et al. 2014, Hammouda et al. 2014, Gamble et al. 2019b; discussed in Appendix S1-D). Further simulation work could aim at optimazing designs (e.g., sample sizes, sampling frequencies, study duration…) for various scenarios, similar to work performed for occupancy models (Mackenzie and Royle 2005, Guillera-Arroita and Lahoz-Monfort 2012).

The present study illustrates that setting up a capture-recapture program, potentially in parallel to extensive cross-sectional sampling, to estimate epidemiological parameters may be particularly rewarding in long-lived host species and when specific antibody level persistence is unknown, which is often the case for non-model species (e.g., seabirds, Chambert et al. 2012b; or marine mammals, Chambert et al. 2012a). Conversely, in a species expected to be subjected to high yearly mortality probabilities (e.g., small passerines, Grosbois et al. 2006; or rodents, Mariën et al. 2018), cross-sectional surveys may be the most efficient way to explore inter-annual processes. Nevertheless, implementing longitudinal, or integrated, setups can still be valuable in short-lived species to study processes occurring at smaller time scales (e.g., monthly; Mariën et al. 2018). In case of doubt about annual survival and/or the temporal persistence of antibody levels, it is always advisable to implement a capture-recapture program at a time scale adapted to the host species phenology. The inter-annual time scale we considered here may be particularly suited to the long-term monitoring of seasonally breeding species or to investigate the potential impact of diseases on long-lived populations (e.g., Lachish et al. 2007, Robardet et al. 2017). In disease systems with strong expected within-

and between-year dynamics, the approach would need to incorporate some temporal hierarchy in considered eco-epidemiological parameters and in the corresponding timing of sampling.

Overall, given the relatively realistic situations we considered and the possibility to tailor the approach to more specific cases, the present study could have important implications regarding current practices in eco-epidemiology. For instance, the presented approach could be adapted to consider the time variations of the force of infection to account for epidemic cases or to incorporating parameters to account for a potential disease-induced mortality (discussed in Appendix S1-A). Our study continues to expand the currently proposed framework to improve inference of the circulation of infectious agents in wild populations using serological data (see Appendix S1-E). The sampling design will of course have to be adapted to the main objective of the survey (Yoccoz et al. 2001). For instance, if the main objective of the study is to estimate the seroconversion probability in a long-lived host species, putting important efforts on recapture (to insure a high δ) as part of a longitudinal setting, and integrating additional cross-sectional data could greatly improve the precision of the seroconversion estimators (Figure S6 b, top panel). In contrast, if the main interest is on the survival probability, putting more effort on resighting (independently of recapture) could improve the precision of the estimates (Lahoz-Monfort et al. 2014, Lieury et al. 2017), but integrating cross-sectional data will provide no added benefit (Figure S6 a, middle panel). In any case, as already advocated in other papers (Albert et al. 2010, Garnett et al. 2011, Restif et al. 2012), but still seldom done (Herzog et al. 2017), we recommend *a priori* modelling based on available knowledge when designing eco-epidemiological studies, notably to account for host demography, immune response characteristics and sampling costs (Fig. S9). In addition to the assumption of independence of the datasets, the approach we used relies on the same assumptions as the ones classically made by the chosen capture-recapture and

393 compartmented epidemiological models, and thus the same limitations apply. Notably, it is
394 important to note that, because we used a simulated dataset, the performances of the three
395 presented approaches could have been overestimated. For instance, we did not consider the
396 effect of potential heterogeneities between individuals included in the cross-sectional and
397 longitudinal datasets (e.g., mean age differences or differences in age variances between the
398 marked and unmarked individuals). If they cannot be avoided, these sources of
399 heterogeneities could be accounted for in the modelling process. Finally, considering the
400 recent advances made in quantitative ecology, this approach could be applied to more
401 complex scenarios than the one we considered here, by being combined with methods
402 accounting for state misclassification by repeating sampling (McClintock et al. 2010, Lahoz-
403 Monfort et al. 2016), using the information contained in quantitative measurements (Choquet
404 et al. 2013), combining assays such as serology and direct detection (Viana et al. 2016,
405 Buzdugan et al. 2017) or by integrating individual traits more explicitly (Plard et al. 2019).

406 ## ACKNOWLEDGEMENTS

407 We are thankful to Rémi Choquet and Roger Pradel for discussions, to Emmanuelle
408 Robardet, Evelyne Picard-Meyer and Florence Cliquet for the serotine bat data, and to three
409 anonymous reviewer's for their suggestions. This work used computational and storage
410 services associated with the shared clusters provided by CEFE-CNRS and UCLA Institute for
411 Digital Research and Education's Research Technology Group (Hoffman2). This paper is a
412 contribution to the French Polar Institute IPEV programs ECOPATH 1151 and PARASITO-
413 ARCTIQUE 333 and to the ECOPOP observation service of the OREME scientific
414 observatory. AG was supported by a PhD fellowship from French Ministry of Research and
415 the DARPA, project PREEMPT # D18AC00031. The content of the article does not
416 necessarily reflect the position or the policy of the U.S. government, and no official

endorsement should be inferred. TC was supported by a CeMEB LabEx post-doctoral fellowship and OG by the ANR, project DEMOCOM # 16-CE02-0007.

LITERATURE CITED

Abadi, F., O. Gimenez, R. Arlettaz, and M. Schaub. 2010. An assessment of integrated population models: bias, accuracy, and violation of the assumption of independence. Ecology 91:7–14.

Albert, C. H., N. G. Yoccoz, T. C. Edwards, C. H. Graham, N. E. Zimmermann, and W. Thuiller. 2010. Sampling in ecology and evolution - bridging the gap between theory and practice. Ecography 33:1028–1037.

Alekseeev, A. Y., K. A. Sharshov, V. Y. Marchenko, Z. Li, J. Cao, F. Yang, A. M. Shestopalov, V. A. Shkurupy, and L. Li. 2014. Antibodies to Newcastle Disease Virus in egg yolks of great cormorant (*Phalacrocorax carbo*) at Qinghai Lake. Advances in Infectious Diseases 04:194–197.

Besbeas, P., S. N. Freeman, B. J. T. Morgan, and E. A. Catchpole. 2002. Integrating Mark-Recapture-Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters. Biometrics 58:540–547.

Borremans, B., N. Hens, P. Beutels, H. Leirs, and J. Reijniers. 2016. Estimating time of infection using prior serological and individual information can greatly improve incidence estimation of human and wildlife infections. PLOS Computational Biology 12:e1004882.

Boulinier, T., S. Kada, A. Ponchon, M. Dupraz, M. Dietrich, A. Gamble, V. Bourret, O. Duriez, R. Bazire, J. Tornos, T. Tveraa, T. Chambert, R. Garnier, and K. D. McCoy. 2016. Migration, prospecting, dispersal? What host movement matters for infectious agent circulation? Integrative and Comparative Biology 56:330–342.

Buzdugan, S. N., T. Vergne, V. Grosbois, R. J. Delahay, and J. A. Drewe. 2017. Inference of the infection status of individuals using longitudinal testing data from cryptic populations: towards a probabilistic approach to diagnosis. Scientific Reports 7:1111.

Chambert, T., J. J. Rotella, and R. A. Garrott. 2012a. Environmental extremes versus ecological extremes: impact of a massive iceberg on the population dynamics of a high-level Antarctic marine predator. Proceedings of the Royal Society B: Biological Sciences 279:4532–4541.

Chambert, T., V. Staszewski, E. Lobato, R. Choquet, C. Carrie, K. D. McCoy, T. Tveraa, and T. Boulinier. 2012b. Exposure of black-legged kittiwakes to Lyme disease spirochetes: dynamics of the immune status of adult hosts and effects on their survival. Journal of Animal Ecology 81:986–995.

Choquet, R., C. Carrié, T. Chambert, and T. Boulinier. 2013. Estimating transitions between states using measurements with imperfect detection: application to serological data. Ecology 94:2160–2165.

Conn, P. B., and E. G. Cooch. 2009. Multistate capture-recapture analysis under imperfect state observation: an application to disease models. Journal of Applied Ecology 46:486–492.

Fletcher, R. J., T. J. Hefley, E. P. Robertson, B. Zuckerberg, R. A. McCleery, and R. M. Dorazio. (2019). A practical guide for combining data to model species distributions. Ecology e02710, in press.

Gamble, A., R. Garnier, A. Jaeger, H. Gantelet, E. Thibault, P. Tortosa, V. Bourret, J.-B. Thiebot, K. Delord, H. Weimerskirch, J. Tornos, C. Barbraud, and T. Boulinier. 2019a. Exposure of breeding albatrosses to the agent of avian cholera: dynamics of antibody levels and ecological implications. Oecologia 189:939–949.

Gamble, A., R. Ramos, Y. Parra-Torres, A. Mercier, L. Galal, J. Pearce-Duvet, I. Villena, T. Montalvo, J. González-Solís, A. Hammouda, D. Oro, S. Selmi, and T. Boulinier. 2019b.

Exposure of yellow-legged gulls to *Toxoplasma gondii* along the Western Mediterranean coasts: Tales from a sentinel. International Journal for Parasitology: Parasites and Wildlife 8:221–228.

Gandon, S. 2002. Local adaptation and the geometry of host–parasite coevolution. Ecology Letters 5:246–256.

Garnett, G. P., S. Cousens, T. B. Hallett, R. Steketee, and N. Walker. 2011. Mathematical models in the evaluation of health programmes. The Lancet 378:515–525.

Garnier, R., R. Ramos, A. Sanz-Aguilar, M. Poisbleau, H. Weimerskirch, S. Burthe, J. Tornos, and T. Boulinier. 2017. Interpreting ELISA analyses from wild animal samples: some recurrent issues and solutions. Functional Ecology 31:2255–2262.

Gilbert, A. T., A. R. Fooks, D. T. S. Hayman, D. L. Horton, T. Müller, R. Plowright, A. J. Peel, R. Bowen, J. L. N. Wood, J. Mills, A. A. Cunningham, and C. E. Rupprecht. 2013. Deciphering serology to understand the ecology of infectious diseases in wildlife. EcoHealth 10:298–313.

Gimenez, O., J.-D. Lebreton, J.-M. Gaillard, R. Choquet, and R. Pradel. 2012. Estimating demographic parameters using hidden process dynamic models. Theoretical Population Biology 82:307–316.

Grosbois, V., P.-Y. Henry, J. Blondel, P. Perret, J.-D. Lebreton, D. W. Thomas, and M. M. Lambrechts. 2006. Climate impacts on Mediterranean blue tit survival: an investigation across seasons and spatial scales. Global Change Biology 12:2235–2249.

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

Hammouda, A., J. Pearce-Duvet, T. Boulinier, and S. Selmi. 2014. Egg sampling as a possible alternative to blood sampling when monitoring the exposure of yellow-legged gulls (*Larus michahellis*) to avian influenza viruses. Avian Pathology 43:547–551.

Hénaux, V., M. D. Samuel, and C. M. Bunck. 2010. Model-based evaluation of highly and low pathogenic avian influenza dynamics in wild birds. PLOS ONE 5:e10997.

Hens, N., Z. Shkedy, M. Aerts, C. Faes, P. Van Damme, and P. Beutels. 2012. Modeling infectious disease parameters based on serological and social contact data: a modern statistical perspective. Springer Science & Business Media, New York.

Herzog, S. A., S. Blaizot, and N. Hens. 2017. Mathematical models used to inform study design or surveillance systems in infectious diseases: a systematic review. BMC Infectious Diseases 17.

Jennelle, C. S., E. G. Cooch, M. J. Conroy, and J. C. Senar. 2007. State-specific detection probabilities and disease prevalence. Ecological Applications 17:154–167.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. 2008. Global trends in emerging infectious diseases. Nature 451:990–993.

Lachish, S., M. Jones, and H. McCallum. 2007. The impact of disease on the survival and population growth rate of the Tasmanian devil. Journal of Animal Ecology 76:926–936.

Lahoz-Monfort, J. J., G. Guillera-Arroita, and R. Tingley. 2016. Statistical approaches to account for false-positive errors in environmental DNA samples. Molecular Ecology Resources 16:673–685.

Lahoz-Monfort, J. J., M. P. Harris, B. J. T. Morgan, S. N. Freeman, and S. Wanless. 2014. Exploring the consequences of reducing survey effort for detecting individual and temporal variability in survival. Journal of Applied Ecology 51:534–543.

Lebreton, J.-D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. Ecological monographs 62:67–118.

Lieury, N., S. Devillard, A. Besnard, O. Gimenez, O. Hameau, C. Ponchon, and A. Millon. 2017. Designing cost-effective capture-recapture surveys for improving the monitoring of survival in bird populations. Biological Conservation 214:233–241.

Mackenzie, D. I., and J. A. Royle. 2005. Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology 42:1105–1114.

Marescot, L., S. Benhaiem, O. Gimenez, H. Hofer, J.-D. Lebreton, X. A. Olarte-Castillo, S. Kramer-Schadt, and M. L. East. 2018. Social status mediates the fitness costs of infection with canine distemper virus in Serengeti spotted hyenas. Functional Ecology 32:1237–1250.

Mariën, J., V. Sluydts, B. Borremans, S. Gryseels, B. Vanden Broecke, C. A. Sabuni, A. A. S. Katakweba, L. S. Mulungu, S. Günther, J. G. de Bellocq, A. W. Massawe, and H. Leirs. 2018. Arenavirus infection correlates with lower survival of its natural rodent host in a long-term capture-mark-recapture study. Parasites & Vectors 11:90.

McClintock, B. T., J. D. Nichols, L. L. Bailey, D. I. MacKenzie, W. L. Kendall, and A. B. Franklin. 2010. Seeking a second opinion: uncertainty in disease ecology. Ecology Letters 13:659–674.

McDonald, J. L., T. Bailey, R. J. Delahay, R. A. McDonald, G. C. Smith, and D. J. Hodgson. 2016. Demographic buffering and compensatory recruitment promotes the persistence of disease in a wildlife population. Ecology Letters 19:443–449.

Metcalf, C. J. E., J. Farrar, F. T. Cutts, N. E. Basta, A. L. Graham, J. Lessler, N. M. Ferguson, D. S. Burke, and B. T. Grenfell. 2016. Use of serological surveys to generate key insights into the changing global landscape of infectious disease. The Lancet 388:728–730.

534 Pepin, K. M., S. L. Kay, B. D. Golas, S. S. Shriner, A. T. Gilbert, R. S. Miller, A. L. Graham,
535 S. Riley, P. C. Cross, M. D. Samuel, M. B. Hooten, J. A. Hoeting, J. O. Lloyd-Smith, C. T.
536 Webb, and M. G. Buhnerkempe. 2017. Inferring infection hazard in wildlife populations by
537 linking data across individual and population scales. Ecology Letters 20:275–292.
538 Plard, F., D. Turek, M. U. Grüebler, and M. Schaub. (2019). IPM2: Towards better
539 understanding and forecasting of population dynamics. Ecological Monographs 89:e01364.
540 Plowright, R. K., D. J. Becker, H. McCallum, and K. R. Manlove. 2019. Sampling to elucidate
541 the dynamics of infections in reservoir hosts. Philosophical Transactions of the Royal Society
542 B: Biological Sciences 374:20180336.
543 Pradel, R. 2005. Multievent: an extension of multistate capture-recapture models to uncertain
544 states. Biometrics 61:442–447.
545 Ramos, R., R. Garnier, J. González-Solís, and T. Boulinier. 2014. Long antibody persistence
546 and transgenerational transfer of immunity in a long-lived vertebrate. The American Naturalist
547 184:764–776.
548 Restif, O., D. T. S. Hayman, J. R. C. Pulliam, R. K. Plowright, D. B. George, A. D. Luis, A. A.
549 Cunningham, R. A. Bowen, A. R. Fooks, T. J. O'Shea, J. L. N. Wood, and C. T. Webb. 2012.
550 Model-guided fieldwork: practical guidelines for multidisciplinary research on wildlife
551 ecological and epidemiological dynamics. Ecology Letters 15:1083–1094.
552 Riecke, T. V., P. J. Williams, T. L. Behnke, D. Gibson, A. G. Leach, B. S. Sedinger, P. A.
553 Street, and J. S. Sedinger. 2019. Integrated population models: model assumptions and
554 inference. Methods in Ecology and Evolution 10: 1072–1082.
555 Robardet, E., C. Borel, M. Moinet, D. Jouan, M. Wasniewski, J. Barrat, F. Boué, E. Montchâtre-
556 Leroy, A. Servat, O. Gimenez, F. Cliquet, and E. Picard-Meyer. 2017. Longitudinal survey of
557 two serotine bat (*Eptesicus serotinus*) maternity colonies exposed to EBLV-1 (European Bat

Lyssavirus type 1): Assessment of survival and serological status variations using capture-recapture models. PLOS Neglected Tropical Diseases 11:e0006048.

Rossi, S., M. Artois, D. Pontier, C. Crucière, J. Hars, J. Barrat, X. Pacholek, and E. Fromont. 2005. Long-term monitoring of classical swine fever in wild boar (*Sus scrofa sp.*) using serological data. Veterinary Research 36:27–42.

Samuel, M. D., J. S. Hall, J. D. Brown, D. R. Goldberg, H. Ip, and V. V. Baranyuk. 2015. The dynamics of avian influenza in lesser snow geese: implications for annual and migratory infection patterns. Ecological Applications 25:1851–1859.

Samuel, M. D., D. J. Shadduck, D. R. Goldberg, and W. P. Johnson. 2003. Comparison of methods to detect *Pasteurella multocida* in carrier waterfowl. Journal of Wildlife Diseases 39:125–135.

Schaub, M., O. Gimenez, A. Sierro, and R. Arlettaz. 2007. Use of Integrated Modeling to Enhance Estimates of Population Dynamics Obtained from Limited Data. Conservation Biology 21:945–955.

Staszewski, V., K. D. McCoy, T. Tveraa, and T. Boulinier. 2007. Interannual dynamics of antibody levels in naturally infected long-lived colonial birds. Ecology 88:3183–3191.

Smith, K. F., D. F. Sax, and K. D. Lafferty. 2006. Evidence for the role of infectious disease in species extinction and endangerment. Conservation Biology 20:1349–1357.

Viana, M., G. M. Shirima, K. S. John, J. Fitzpatrick, R. R. Kazwala, J. J. Buza, S. Cleaveland, D. T. Haydon, and J. E. B. Halliday. 2016. Integrating serological and genetic data to quantify cross-species transmission: brucellosis as a case study. Parasitology 143:821–834.

Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. Trends in Ecology & Evolution 16:446–453.

581 TABLES

582 TABLE 1. Eco-epidemiological parameters estimated from a bat colony exposed to a rabies
583 virus using the longitudinal or integrated design. The estimates are presented with their 95%
584 confidence interval between brackets. Note that the confidence interval of seroconversion
585 probability (in bold) is smaller when using the integrated design.

Parameter	Design			
	Longitudinal	Integrated n_{CS} = 20	Integrated n_{CS} = 40	Integrated n_{CS} = 60
Survival ϕ	0.750 [0.684; 0.807]	0.750 [0.684; 0.807]	0.750 [0.684; 0.807]	0.750 [0.684; 0.807]
Seroconversion λ	**0.085 [0.033; 0.201]**	**0.072 [0.038; 0.130]**	**0.085 [0.046; 0.151]**	**0.079 [0.043; 0.139]**
Seroreversion ω	0.145 [0.072; 0.271]	0.152 [0.080; 0.269]	0.145 [0.075; 0.261]	0.148 [0.078; 0.265]
Resighting t = 1 p_1	0.793 [0.733; 0.843]	0.793 [0.733; 0.843]	0.793 [0.733; 0.843]	0.793 [0.733; 0.843]
Resighting t = 2 p_2	0.152 [0.049; 0.383]	0.152 [0.049; 0.383]	0.152 [0.049; 0.383]	0.152 [0.049; 0.383]
Resighting t = 3 p_3	0.865 [0.630; 0.960]	0.865 [0.630; 0.960]	0.865 [0.630; 0.960]	0.865 [0.630; 0.960]
Resighting t = 4 p_4	0.138 [0.062; 0.277]	0.138 [0.062; 0.277]	0.138 [0.062; 0.277]	0.138 [0.062; 0.277]
Resighting t = 5 p_5	0.365 [0.218; 0.542]	0.365 [0.218; 0.542]	0.365 [0.218; 0.542]	0.365 [0.218; 0.542]
Resighting t = 6 p_6	0.702 [0.477; 0.859]	0.702 [0.477; 0.859]	0.702 [0.477; 0.859]	0.702 [0.477; 0.859]
Resighting t = 7 p_7	0.646 [0.421; 0.821]	0.646 [0.421; 0.821]	0.646 [0.421; 0.821]	0.646 [0.421; 0.821]
Recapture δ	0.659 [0.385; 0.856]	0.659 [0.385; 0.856]	0.659 [0.385; 0.856]	0.659 [0.385; 0.856]

586

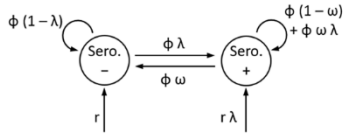
FIGURES

FIGURE 1. Methodological framework: eco-epidemiological process used for data simulation (a) and modelling framework for the estimation of the eco-epidemiological parameters (b).

FIGURE 2. The longitudinal design overall leads to no bias but low precision in the estimation of the seroconversion probability (λ) while small errors in *a priori* fixed seroreversion ($\tilde{\omega}$) or survival ($\tilde{\phi}$) probabilities can lead to strong biases in cross sectional designs, especially for long-lived host species and persisting antibody levels. Estimated values of the seroconversion probability and corresponding bias (a, b) or MSE (c, d) in relation to survival (a), seroreversion (b) and seroconversion (c, d) probabilities using cross-sectional or longitudinal estimators. For the cross-sectional design, results are shown for a realistic gradient of error on the *a priori* fixed value of seroreversion ($\tilde{\omega}$) or survival ($\tilde{\phi}$), while the longitudinal design does not require those parameters to be set *a priori* (not informed). The true seroconversion probability is represented by a black dashed line (a, b) or black diamonds (c, d). Notes: (a): a null seroreversion value corresponds to a lifelong persistence of antibody levels. (b): a survival value of $\phi \times p$ corresponds to an underestimated survival probability comparable to the raw return rate probability which is sometime used in the literature (the < 1 resighting probability being ignored).

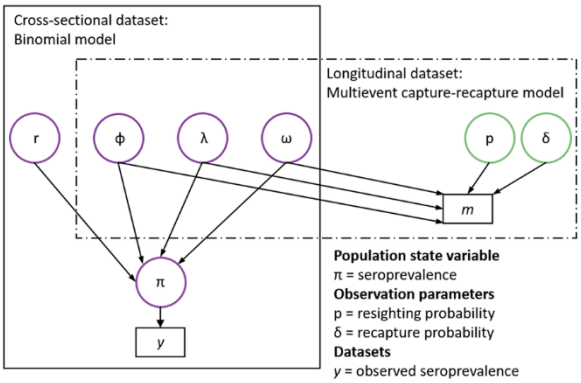
FIGURE 3. The integrated estimator leads to higher precision in the estimation of the seroconversion probability (λ) compared to the integrated estimator. Estimated values of the seroconversion probability and corresponding MSE for different combinations on datasets analyzed with the longitudinal ($n_{CS} = 0$) or the integrated ($n_{CS} > 0$) model. Two situations were explored: intermediate (a) or long (b) persistence of the antibody levels. The true seroconversion probability is represented by a black dashed line. The cross-sectional model is not represented on this figure as it requires *a priori* reliable knowledge on ϕ and ω.

a) Framework for the eco-epidemiological simulations



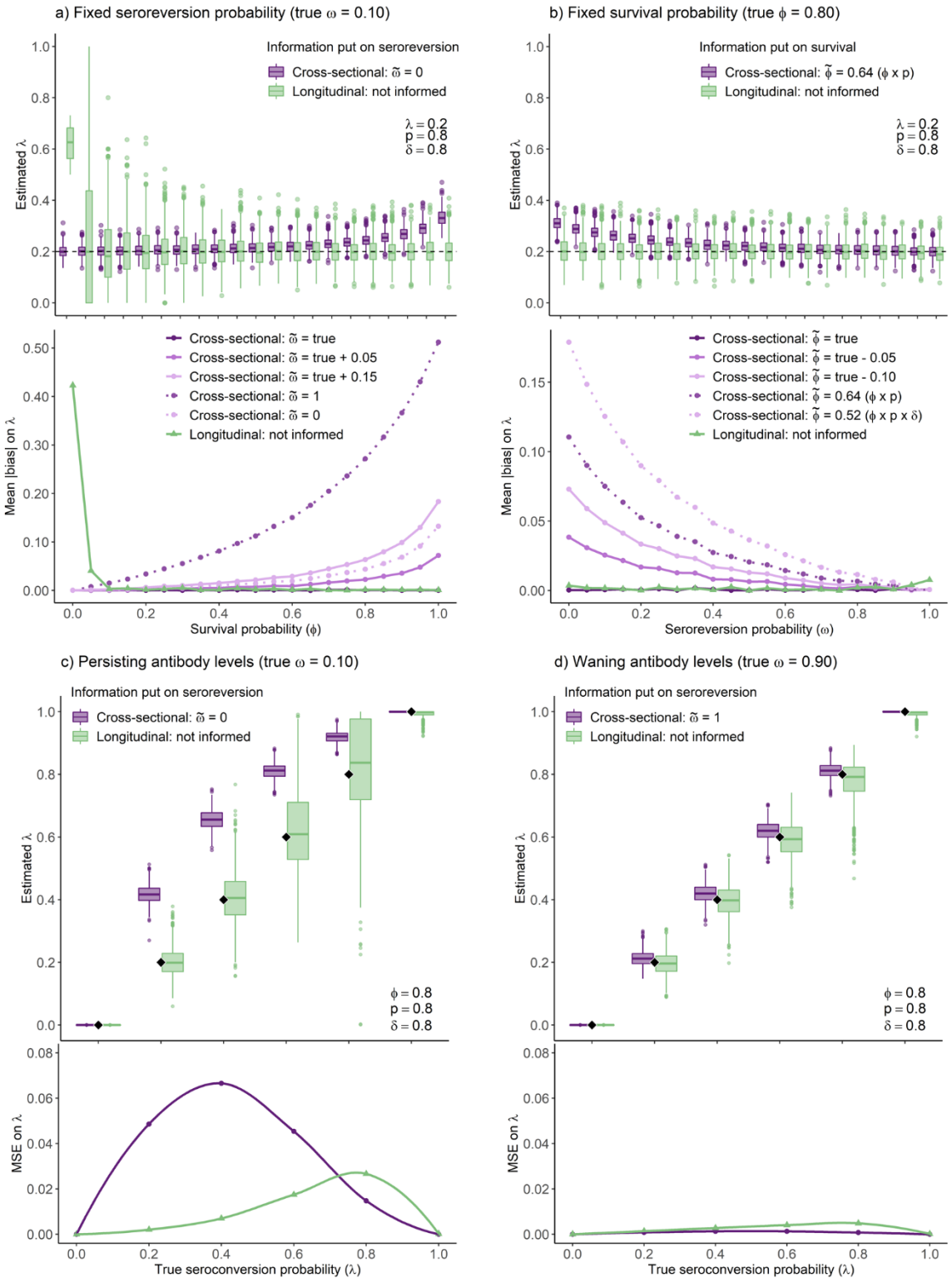
Individual states
Sero.– = seronegative
Sero. + = seropositive
Transition parameters
ϕ = survival probability
λ = seroconversion probability
ω = seroreversion probability
r = recruitment probability

b) Framework for the integrated modelling for parameter estimation

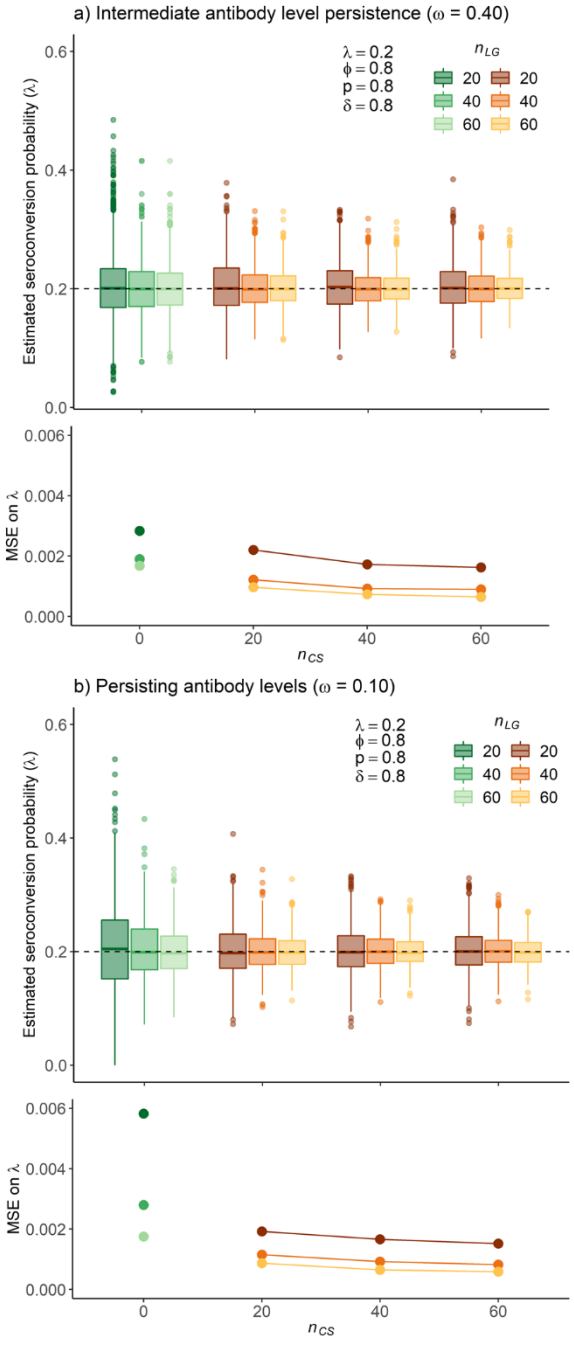


Population state variable
π = seroprevalence
Observation parameters
p = resighting probability
δ = recapture probability
Datasets
y = observed seroprevalence
m = observed individual histories

611


a) Fixed seroreversion probability (true ω = 0.10)
Information put on seroreversion
Cross-sectional: ω̃ = 0
Longitudinal: not informed
λ = 0.2
p = 0.8
δ = 0.8
Estimated λ
Cross-sectional: ω̃ = true
Cross-sectional: ω̃ = true + 0.05
Cross-sectional: ω̃ = true + 0.15
Cross-sectional: ω̃ = 1
Cross-sectional: ω̃ = 0
Longitudinal: not informed
Mean |bias| on λ
Survival probability (ϕ)
b) Fixed survival probability (true ϕ = 0.80)
Information put on survival
Cross-sectional: ϕ̃ = 0.64 (ϕ x p)
Longitudinal: not informed
λ = 0.2
p = 0.8
δ = 0.8
Estimated λ
Cross-sectional: ϕ̃ = true
Cross-sectional: ϕ̃ = true - 0.05
Cross-sectional: ϕ̃ = true - 0.10
Cross-sectional: ϕ̃ = 0.64 (ϕ x p)
Cross-sectional: ϕ̃ = 0.52 (ϕ x p x δ)
Longitudinal: not informed
Mean |bias| on λ
Seroreversion probability (ω)
c) Persisting antibody levels (true ω = 0.10)
Information put on seroreversion
Cross-sectional: ω̃ = 0
Longitudinal: not informed
ϕ = 0.8
p = 0.8
δ = 0.8
Estimated λ
MSE on λ
True seroconversion probability (λ)
d) Waning antibody levels (true ω = 0.90)
Information put on seroreversion
Cross-sectional: ω̃ = 1
Longitudinal: not informed
ϕ = 0.8
p = 0.8
δ = 0.8
Estimated λ
MSE on λ
True seroconversion probability (λ)

612


a) Intermediate antibody level persistence (ω = 0.40)
λ = 0.2
ϕ = 0.8
p = 0.8
δ = 0.8
n_{LG}
20 20
40 40
60 60
Estimated seroconversion probability (λ)
0.6
0.4
0.2
0.0
MSE on λ
0.006
0.004
0.002
0.000
0
20
40
60
n_{CS}
b) Persisting antibody levels (ω = 0.10)
λ = 0.2
ϕ = 0.8
p = 0.8
δ = 0.8
n_{LG}
20 20
40 40
60 60
Estimated seroconversion probability (λ)
0.6
0.4
0.2
0.0
MSE on λ
0.006
0.004
0.002
0.000
0
20
40
60
n_{CS}

613