



**HAL**  
open science

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

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

### ► To cite this version:

Amandine Gamble, Romain Garnier, Thierry Chambert, Olivier Gimenez, Thierry Boulinier. Next generation serology: integrating cross-sectional and capture-recapture approaches to infer disease dynamics. *Ecology*, 2019, 10.1002/ecy.2923 . hal-02330417

**HAL Id: hal-02330417**

**<https://hal.science/hal-02330417>**

Submitted on 24 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Running head: Next generation serology field design

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

4 Amandine Gamble<sup>1,2,\*</sup>, Romain Garnier<sup>3</sup>, Thierry Chambert<sup>1</sup>, Olivier Gimenez<sup>1</sup> and Thierry  
5 Boulinier<sup>1</sup>

6 <sup>1</sup> CEFE, CNRS, University of Montpellier, EPHE, University Paul Valéry Montpellier 3, IRD,  
7 Montpellier, France

8 <sup>2</sup> Department of Ecology and Evolutionary Biology, University of California, Los Angeles,  
9 CA, United States

10 <sup>3</sup> Department of Biology, Georgetown University, Washington, DC, United States

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

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

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

37 Understanding the ecology and evolution of infectious diseases in wildlife has been  
38 highlighted as critical for public health (Jones et al. 2008) and biodiversity conservation  
39 (Smith et al. 2006). Natural host-parasite systems also offer useful models to obtain valuable  
40 insights on evolutionary ecology processes such as coevolution and local adaptation (Gandon  
41 2002) or host and vector movements (Boulinier et al. 2016). However, investigations in the  
42 wild have been hampered by the difficulty of collecting data allowing efficient inference of  
43 eco-epidemiological dynamics (Plowright et al. 2019). For instance, the force of infection  
44 (*i.e.*, the rate at which susceptible individuals acquire an infectious disease), a key eco-  
45 epidemiological parameter (Hens et al. 2012), is difficult to estimate from field data as it  
46 requires assessing how many individuals went from susceptible (e.g., non-infected and non-  
47 immunized) to infected in a given time period, which is rarely observable. Estimating these  
48 parameters is however a critical step in the characterization of epidemiological dynamics and  
49 factors impacting them. Methods allowing their estimation from field data are thus needed.

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

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

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

73 In some species, individuals can be marked and repeatedly (re)captured across time,  
74 allowing longitudinal sampling. This is particularly true for long-lived vertebrates showing  
75 seasonal and colonial breeding (such as seabirds, pinnipeds, and chiropterans) and which are  
76 often faithful to their breeding or roosting site (e.g., Chambert et al. 2012b, Robardet et al.  
77 2017, Gamble et al. 2019a). In these systems, capture-recapture approaches have started to be  
78 used to estimate epidemiological state transition probabilities (e.g., from healthy to  
79 symptomatic) while accounting for recapture probabilities below unity, which are unavoidable  
80 in wild settings (Jennelle et al. 2007, Conn and Cooch 2009). However, longitudinal studies  
81 are usually based on relatively small sample sizes because field efforts needed to resight and  
82 recapture marked individuals tend to be intensive. In contrast, cross-sectional studies are

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

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

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

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

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

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

## 132 MATERIALS AND METHODS

### 133 *Eco-epidemiological model*

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

### 145 *Cross-sectional sampling*

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



$$154 \quad \pi_{t+1} = \pi_t \phi (1 - \omega) + \pi_t \phi \omega \lambda + (1 - \pi_t) \phi \lambda + r \lambda \quad (1)$$

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

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

$$166 \quad \pi^* = -\frac{\lambda}{\phi (1 - \omega + \omega \lambda - \lambda) - 1} \quad (2) \quad y \sim B \left( n_{CS}, -\frac{\lambda}{\phi (1 - \omega + \omega \lambda - \lambda) - 1} \right) \quad (3)$$

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

### 171 *Longitudinal sampling*

172 On the first year of the observation process,  $n_{LG}$  random individuals are captured and  
 173 marked with a tag allowing individuals to be identified without recapture (e.g., rings or PIT  
 174 tags). Each of the following years, each alive marked individual is resighted with a probability  
 175  $p$  and its serological state is ascertained with a probability  $\delta$  corresponding to the recapture  
 176 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:

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

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

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

### 209 *Simulations and model fitting*

210 For each set of parameters, 1000 populations with a size of 600 individuals were simulated  
211 using a specifically developed individual based model (see Appendix S2 for codes). To  
212 compare the performances of both designs under various scenarios, one cross-sectional  
213 sample and one longitudinal sample of 50 individuals ( $n_{CS} = n_{LG}$ ) per year were then taken per  
214 simulated population following the designs described above. In the case of integrated  
215 modelling, several combinations of cross-sectional ( $n_{CS} = 20, 40$  or  $60$ ) and longitudinal ( $n_{LG}$   
216  $= 20$  or  $40$  or  $60$ ) sample sizes were tested. Unless otherwise stated, the resighting ( $p$ ) and  
217 recapture ( $\delta$ ) probabilities were set to 0.80 and sampling was conducted over five years after  
218 having reached the endemic equilibrium (Fig. S1). Within a time step, samples were collected  
219 after exposure. The performances of the estimators were then compared based first on their  
220 bias, and second on their Mean Square Error ( $MSE = \text{bias}^2 + \text{variance}$ ) in order to account for  
221 the bias and the precision of the estimators; the lower the bias or MSE, the more accurate the  
222 estimator. In the three cases (cross-sectional, longitudinal and integrated), eco-  
223 epidemiological parameters were estimated from the data by maximization of likelihood using  
224 a frequentist approach. This method was preferred due to reduced computation time compared

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

229 *Illustrative example*

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

$$242 \pi_{t+1 \text{ bats, EBLV-1}} = \pi_{t \text{ bats, EBLV-1}} \phi (1 - \omega) + \pi_{t \text{ bats, EBLV-1}} \phi \omega \lambda + (1 - \pi_{t \text{ bats, EBLV-1}}) \phi \lambda \quad (5)$$

243 And seroprevalence at the equilibrium can be written:

$$244 \pi_{\text{bats, EBLV-1}}^* = \frac{\lambda}{(\omega - \omega \lambda + \lambda)} \quad (6)$$

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

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

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

## 254 RESULTS

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

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

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

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

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

299 *Illustrative example.* The estimation of seroconversion probability was improved (smaller  
300 confidence interval) when longitudinal and cross-sectional data were integrated together  
301 (compared to using only longitudinal data; Table 1). For instance, the seroconversion  
302 probability [95% confidence interval] was estimated at 0.085 [0.033; 0.201] using the  
303 longitudinal design and 0.079 [0.043; 0.139] using the integrated design including data from  
304 60 unmarked individuals each year. The estimates of survival, resighting and recapture  
305 probabilities were unchanged, as expected considering that these parameters were not  
306 expected to impact seroprevalence (equation 5).

307

308

## DISCUSSION

309 Based on an eco-epidemiological model and simulations under different sampling  
310 scenarios, our results suggest that longitudinal data analyzed in capture-recapture frameworks  
311 are preferable to cross-sectional data when poor *a priori* knowledge (for instance on the  
312 survival and seroreversion probabilities) is available on the system, which is the case with  
313 most wildlife-parasites systems. The cross-sectional estimator can nonetheless be accurate for  
314 hosts with short lifespan and/or short antibody level persistence or when informed with  
315 reliable *a priori* knowledge on these parameters. In contrast, the longitudinal approach  
316 provided accurate estimates and also allowed survival and seroreversion probabilities to be  
317 estimated along with observation parameters (resighting and recapture probabilities; e.g., the  
318 serotine bat example). Finally, the integrated estimator benefited from the performances of  
319 longitudinal designs, notably it did not rely on any *a priori* known parameters, even with  
320 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 (Alekseev et al. 2014, Hammouda et al. 2014, Gamble et al. 2019b; discussed in Appendix S1-D). Further simulation work could aim at optimizing designs (e.g., sample sizes, sampling frequencies, study duration...) for various scenarios, similar to work performed for occupancy models (Mackenzie and Royle 2005, Guillera-Aroita 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-

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

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

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

419 LITERATURE CITED

420 Abadi, F., O. Gimenez, R. Arlettaz, and M. Schaub. 2010. An assessment of integrated  
421 population models: bias, accuracy, and violation of the assumption of independence. *Ecology*  
422 91:7–14.

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

426 Alekseev, A. Y., K. A. Sharshov, V. Y. Marchenko, Z. Li, J. Cao, F. Yang, A. M. Shestopalov,  
427 V. A. Shkurupy, and L. Li. 2014. Antibodies to Newcastle Disease Virus in egg yolks of great  
428 cormorant (*Phalacrocorax carbo*) at Qinghai Lake. *Advances in Infectious Diseases* 04:194–  
429 197.

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

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

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

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

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

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

450 Choquet, R., C. Carrié, T. Chambert, and T. Boulinier. 2013. Estimating transitions between  
451 states using measurements with imperfect detection: application to serological data. *Ecology*  
452 94:2160–2165.

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

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

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

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

464 Exposure of yellow-legged gulls to *Toxoplasma gondii* along the Western Mediterranean  
465 coasts: Tales from a sentinel. *International Journal for Parasitology: Parasites and Wildlife*  
466 8:221–228.

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

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

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

474 Gilbert, A. T., A. R. Fooks, D. T. S. Hayman, D. L. Horton, T. Müller, R. Plowright, A. J. Peel,  
475 R. Bowen, J. L. N. Wood, J. Mills, A. A. Cunningham, and C. E. Rupprecht. 2013.  
476 Deciphering serology to understand the ecology of infectious diseases in wildlife. *EcoHealth*  
477 10:298–313.

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

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

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

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

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

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

495 Herzog, S. A., S. Blaizot, and N. Hens. 2017. Mathematical models used to inform study design  
496 or surveillance systems in infectious diseases: a systematic review. *BMC Infectious Diseases*  
497 17.

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

500 Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak.  
501 2008. Global trends in emerging infectious diseases. *Nature* 451:990–993.

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

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

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

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

513 Lieury, N., S. Devillard, A. Besnard, O. Gimenez, O. Hameau, C. Ponchon, and A. Millon.  
514 2017. Designing cost-effective capture-recapture surveys for improving the monitoring of  
515 survival in bird populations. *Biological Conservation* 214:233–241.

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

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

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

525 McClintock, B. T., J. D. Nichols, L. L. Bailey, D. I. MacKenzie, W. L. Kendall, and A. B.  
526 Franklin. 2010. Seeking a second opinion: uncertainty in disease ecology. *Ecology Letters*  
527 13:659–674.

528 McDonald, J. L., T. Bailey, R. J. Delahay, R. A. McDonald, G. C. Smith, and D. J. Hodgson.  
529 2016. Demographic buffering and compensatory recruitment promotes the persistence of  
530 disease in a wildlife population. *Ecology Letters* 19:443–449.

531 Metcalf, C. J. E., J. Farrar, F. T. Cutts, N. E. Basta, A. L. Graham, J. Lessler, N. M. Ferguson,  
532 D. S. Burke, and B. T. Grenfell. 2016. Use of serological surveys to generate key insights into  
533 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. Gruebler, 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

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

560 Rossi, S., M. Artois, D. Pontier, C. Crucière, J. Hars, J. Barrat, X. Pacholek, and E. Fromont.  
561 2005. Long-term monitoring of classical swine fever in wild boar (*Sus scrofa sp.*) using  
562 serological data. *Veterinary Research* 36:27–42.

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

566 Samuel, M. D., D. J. Shadduck, D. R. Goldberg, and W. P. Johnson. 2003. Comparison of  
567 methods to detect *Pasteurella multocida* in carrier waterfowl. *Journal of Wildlife Diseases*  
568 39:125–135.

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

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

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

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

579 Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space  
580 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<br><i>n<sub>CS</sub></i> = 20 | Integrated<br><i>n<sub>CS</sub></i> = 40 | Integrated<br><i>n<sub>CS</sub></i> = 60 |
| Survival $\phi$                            | 0.750<br>[0.684; 0.807]               | 0.750<br>[0.684; 0.807]                  | 0.750<br>[0.684; 0.807]                  | 0.750<br>[0.684; 0.807]                  |
| <b>Seroconversion <math>\lambda</math></b> | <b>0.085</b><br><b>[0.033; 0.201]</b> | <b>0.072</b><br><b>[0.038; 0.130]</b>    | <b>0.085</b><br><b>[0.046; 0.151]</b>    | <b>0.079</b><br><b>[0.043; 0.139]</b>    |
| Seroreversion $\omega$                     | 0.145<br>[0.072; 0.271]               | 0.152<br>[0.080; 0.269]                  | 0.145<br>[0.075; 0.261]                  | 0.148<br>[0.078; 0.265]                  |
| Resighting t = 1 $p_1$                     | 0.793<br>[0.733; 0.843]               | 0.793<br>[0.733; 0.843]                  | 0.793<br>[0.733; 0.843]                  | 0.793<br>[0.733; 0.843]                  |
| Resighting t = 2 $p_2$                     | 0.152<br>[0.049; 0.383]               | 0.152<br>[0.049; 0.383]                  | 0.152<br>[0.049; 0.383]                  | 0.152<br>[0.049; 0.383]                  |
| Resighting t = 3 $p_3$                     | 0.865<br>[0.630; 0.960]               | 0.865<br>[0.630; 0.960]                  | 0.865<br>[0.630; 0.960]                  | 0.865<br>[0.630; 0.960]                  |
| Resighting t = 4 $p_4$                     | 0.138<br>[0.062; 0.277]               | 0.138<br>[0.062; 0.277]                  | 0.138<br>[0.062; 0.277]                  | 0.138<br>[0.062; 0.277]                  |
| Resighting t = 5 $p_5$                     | 0.365<br>[0.218; 0.542]               | 0.365<br>[0.218; 0.542]                  | 0.365<br>[0.218; 0.542]                  | 0.365<br>[0.218; 0.542]                  |
| Resighting t = 6 $p_6$                     | 0.702<br>[0.477; 0.859]               | 0.702<br>[0.477; 0.859]                  | 0.702<br>[0.477; 0.859]                  | 0.702<br>[0.477; 0.859]                  |
| Resighting t = 7 $p_7$                     | 0.646<br>[0.421; 0.821]               | 0.646<br>[0.421; 0.821]                  | 0.646<br>[0.421; 0.821]                  | 0.646<br>[0.421; 0.821]                  |
| Recapture $\delta$                         | 0.659<br>[0.385; 0.856]               | 0.659<br>[0.385; 0.856]                  | 0.659<br>[0.385; 0.856]                  | 0.659<br>[0.385; 0.856]                  |

586

587

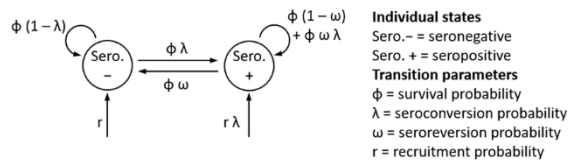
## FIGURES

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

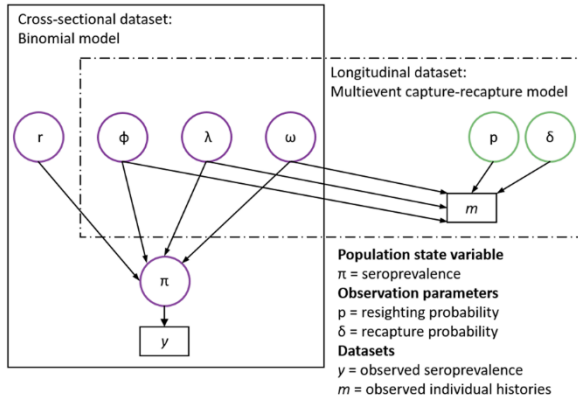
590 FIGURE 2. The longitudinal design overall leads to no bias but low precision in the estimation  
591 of the seroconversion probability ( $\lambda$ ) while small errors in *a priori* fixed seroreversion ( $\tilde{\omega}$ ) or  
592 survival ( $\tilde{\phi}$ ) probabilities can lead to strong biases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

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

a) Framework for the eco-epidemiological simulations



b) Framework for the integrated modelling for parameter estimation



611

