



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/104540/>

Version: Accepted Version

Article:

Walton, Laura, Marion, Glenn, Davidson, Ross S. et al. (2016) The ecology of wildlife disease surveillance: Demographic and prevalence fluctuations undermine surveillance. *Journal of Applied Ecology*. 1460–1469. ISSN: 0021-8901

<https://doi.org/10.1111/1365-2664.12671>

Reuse

Other licence.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

The Ecology of Wildlife Disease Surveillance: demographic and prevalence fluctuations undermine surveillance

Journal:	<i>Journal of Applied Ecology</i>
Manuscript ID	JAPPL-2015-00745.R1
Manuscript Type:	Standard Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Walton, Laura; Biomathematics and Statistics Scotland, Marion, Glenn; Biomathematics and Statistics Scotland, Davidson, Ross; SRUC, White, Piran; University of York, Smith, Lesley; SRUC, Gavier-Widen, Dolores; Swedish University of Agricultural Sciences, Yon, Lisa; University of Nottingham, Hannant, Duncan; University of Nottingham, Hutchings, Michael; SRUC,
Key-words:	wildlife disease systems, wildlife ecology, disease surveillance, demographic fluctuations, wildlife populations, disease transmission models, stochastic population models

1 **The Ecology of Wildlife Disease Surveillance: demographic and prevalence**
2 **fluctuations undermine surveillance**

3
4
5 **Laura Walton^{1,2,3}, Glenn Marion^{1*}, Ross S. Davidson², Piran C.L. White³, Lesley A. Smith², Dolores**
6 **Gavier-Widen⁴, Lisa Yon⁵, Duncan Hannant⁵ and Michael R. Hutchings²**

7 1. Biomathematics and Statistics Scotland, Edinburgh, United Kingdom

8 2. Disease Systems Team, SRUC, Edinburgh, United Kingdom

9 3. Environment Department, University of York, York, United Kingdom

10 4. Swedish University of Agricultural Sciences, Uppsala, Sweden

11 5. School of Veterinary Medicine and Science, University of Nottingham, United Kingdom

12
13 **Email Addresses:** Laura Walton: laura@bioass.ac.uk, Glenn Marion: glenn@bioass.ac.uk, Ross S.
14 Davidson: ross.davidson@sruc.ac.uk, Piran C.L. White: piran.white@york.ac.uk, Lesley A. Smith:
15 lesley.smith@sruc.ac.uk, Dolores Gavier-Widen: dolores.gavier-widen@slu.se, Lisa Yon:
16 lisa.yon@nottingham.ac.uk, Duncan Hannant: duncan.hannant@nottingham.ac.uk, Michael R.
17 Hutchings: mike.hutchings@sruc.ac.uk

18 ***Correspondence author:** Glenn Marion, Biomathematics and Statistics Scotland, Peter Guthrie Tait
19 Road, Edinburgh, EH9 3FD, UK. Telephone: +44 (0)131 650 4898. Fax: +44 (0)131 650 4901. E-mail:
20 glenn@bioass.ac.uk

21
22 **Running title: The Ecology of Wildlife Disease Surveillance**

23
24 **Word Count: 6979**

25

26 **Summary**

27 1. Wildlife disease surveillance is the first line of defence against infectious disease. Fluctuations in
28 host populations and disease prevalence are a known feature of wildlife disease systems.
29 However, the impact of such heterogeneities on the performance of surveillance is currently
30 poorly understood.

31 2. We present the first systematic exploration of the effects of fluctuations prevalence and host
32 population size on the efficacy of wildlife disease surveillance systems. In this study efficacy is
33 measured in terms of ability to estimate long term prevalence and detect disease risk.

34 3. Our results suggest that for many wildlife disease systems fluctuations in population size and
35 disease lead to bias in surveillance-based estimates of prevalence and over-confidence in
36 assessments of both the precision of prevalence estimates and the power to detect disease.

37 4. Neglecting such ecological effects may lead to poorly designed surveillance and ultimately to
38 incorrect assessments of the risks posed by disease in wildlife. This will be most problematic in
39 systems where prevalence fluctuations are large and disease fade-outs occur. Such fluctuations
40 are determined by the interaction of demography and disease dynamics and although
41 particularly likely in highly fluctuating populations typical of fecund short lived hosts, can't be
42 ruled out in more stable populations of longer lived hosts.

43 5. *Synthesis and Applications:* Fluctuations in population size and disease prevalence should be
44 considered in the design and implementation of wildlife disease surveillance and the framework
45 presented here provides a template for conducting suitable power calculations. Ultimately
46 understanding the impact of fluctuations in demographic and epidemiological processes will
47 enable improvements to wildlife disease surveillance systems leading to better characterisation
48 of, and protection against endemic, emerging and re-emerging disease threats.

49 **Key-words:** wildlife disease systems, wildlife ecology, disease surveillance, demographic
50 fluctuations, wildlife populations, disease transmission models, stochastic population models

51 Introduction

52 Surveillance is the first line of defence against disease, whether to monitor endemic cycles of
53 infection (Ryser-Degiorgis 2013) or to detect incursions of emerging or re-emerging diseases (Kruse,
54 Kirkemo & Handeland 2004; Lipkin 2013)). Identification and quantification of disease presence and
55 prevalence is the starting point for developing disease control strategies as well as monitoring their
56 efficacy (OIE 2013). Knowledge of disease in wildlife is of considerable importance for managing risks
57 to humans (Daszak, Cunningham & Hyatt 2000; Jones *et al.* 2008) and livestock (Gortázar *et al.*
58 2007), as well as for the conservation of wildlife species themselves (Daszak, Cunningham & Hyatt
59 2000).

60
61 Recent public health concerns e.g. Highly Pathogenic Avian Influenza (Artois *et al.* 2009b) , Alveolar
62 Echinococcosis (Eckert & Deplazes 2004) and West Nile Virus (Brugman *et al.* 2013)), have led to a
63 growing recognition that current approaches need to be improved (Mörner *et al.* 2002). For
64 example, there is no agreed wildlife disease surveillance protocol shared among the countries in the
65 European Union (Kuiken *et al.* 2011). Furthermore several authors have identified the need for
66 improvements to the structure, understanding and evaluation of wildlife disease surveillance (Bengis
67 *et al.* 2004; Gortázar *et al.* 2007).

68
69 Much current practice for wildlife disease surveillance (Artois *et al.* 2009a) is based on ideas
70 developed for surveillance in livestock, including calculation of sample sizes needed for accurate
71 prevalence estimation (Grimes & Schulz 1996; Fosgate 2005) and detection of disease within a
72 population (Dohoo, Martin & Stryhn 2005). A common feature of these methods is that they assume
73 constant host populations and disease prevalence. These assumptions lead naturally to sample size
74 calculations (for both disease detection and prevalence estimation) which are based on a binomial
75 distribution and associated corrections for populations of finite size, such as the hyper-geometric
76 distribution (Artois *et al.* 2009a). (Fosgate 2009) reviewed current approaches to sample size

77 calculations in livestock systems and emphasised the importance of basing analyses on realistic
78 assumptions about the system under surveillance.

79

80 Although constant population size and prevalence may often be reasonable assumptions for the
81 analysis of livestock systems, they are considerably less tenable in wildlife disease systems, which
82 are typically subject to much greater fluctuations in host population density and disease prevalence.

83 Both sampling practicalities and changes in population density make it much harder to obtain a
84 random sample of hosts of the desired sample size in wildlife disease surveillance programmes

85 (Nusser *et al.* 2008), compared with livestock systems. It is not uncommon for wildlife disease
86 surveillance to extend over several years and to test only a small fraction of the at risk population.

87 For example, McGarry and co-workers report overall prevalence of zoonotic helminths in 42 brown
88 rats (*Rattus norvegicus*) captured in a programme of active surveillance carried out in an urban area

89 in England between 2008 and 2011 (McGarry *et al.* 2014). These authors also present comparable
90 results from several studies in Europe and North America while another of the same host species

91 conducted over a two year period across a broad area of Northwestern England captured just 133
92 individuals (Pounder *et al.* 2013). A notable example of passive surveillance i.e. the testing of found

93 dead individuals, is that for zoonotic West Nile Virus (WNV) in wild birds across the whole of Great
94 Britain during 2002-2009 in which only 2072 individuals representing 240 species were tested

95 (Brugman *et al.* 2013).

96

97 The importance of temporal (Renshaw 1991; Wilson & Hassell 1997), spatial (Lloyd & May 1996;
98 Tilman & Kareiva 1997) and other forms of heterogeneity (Read & Keeling 2003; Vicente *et al.* 2007;

99 Davidson, Marion & Hutchings 2008) in population ecology has long been recognised (Anderson
100 1991; Smith *et al.* 2005), along with their role in the dynamics and persistence of infectious disease

101 (Fenton *et al.* 2015). Detailed field observations have provided valuable insights into the temporal
102 dynamics of wildlife disease systems. For example a study (Telfer *et al.* 2002) of cowpox virus in two

103 rodent host species at two sites over a four year period reveals strong temporal fluctuations in both
104 population size and disease prevalence including disease fade-out (local extinction and re-
105 emergence). Fade-outs are also observed in wildlife populations of longer lived mammals as shown
106 by a six year study (Hawkins *et al.* 2006) of Devil Facial Tumour Disease in *Sarcophilus harrisii*
107 (Tasmanian devil). One of the longest running and most intensive studies of disease in wildlife is the
108 surveillance from 1982 to the present of TB in badgers at Woodchester Park, England where around
109 80% of the population is trapped tested and released annually (Delahay *et al.* 2000). These long term
110 observations have revealed important insights into the dynamics of TB in badgers e.g. that infection
111 within social groups is persistent whereas transmission between social groups is limited (Delahay *et*
112 *al.* 2000). Parameter estimates derived from this study are used as a reference point for the
113 simulation studies conducted below.

114

115 Despite these theoretical and empirical studies of temporal heterogeneities in wildlife disease
116 systems, such effects have yet to be systematically accounted for, either in the design of surveillance
117 programmes for wildlife disease systems, or in the analysis of the data obtained from them. Here
118 we address this gap by using a non-spatial simulation model of a wildlife host population, subject to
119 demographic fluctuations and pathogen transmission, in order to explore the impact of stochastic
120 fluctuations in host demography and disease dynamics on the performance of surveillance. Two
121 measures of surveillance performance are considered; estimation of long term prevalence and the
122 ability (probability) to detect disease. Our results show that temporal fluctuations in wildlife disease
123 systems limit the ability of surveillance to achieve both.

124

125 **Methods**

126 We develop a generic modelling framework that represents key features of surveillance in wildlife
127 disease systems including essential aspects of demography, disease dynamics and surveillance
128 design. This framework is described below along with three simulation studies that enable us to

129 explore the performance of surveillance across a wide range of scenarios representative of real
130 world systems.

131 Stochastic Modelling framework

132 The model represents a host population subject to demographic fluctuations (births, deaths and
133 immigration) and the transmission of a single pathogen. At each point in time t , the state-space
134 represents the total population size $N(t)$, with $I(t)$ of these infected and $S(t) = N(t) - I(t)$
135 susceptible. The prevalence is then given by $p(t) = I(t)/N(t)$.

136 **Demography.** The birth rate of individuals is logistic, $rN(1 - N/k)$, with intrinsic growth rate r and
137 carrying capacity k , reflecting the assumptions that population growth is resource limited.
138 Individuals have a per capita death rate μ and immigration occurs at a constant rate ν .

139 **Disease dynamics.** A proportion γ of immigrants are infected, but otherwise all individuals enter the
140 population (through birth or immigration) as susceptible, since we assume vertical and pseudo-
141 vertical transmission are negligible. Susceptible individuals become infected at rate $\beta_0 S(t)$ through
142 primary transmission (contact with infectious environmental sources including individuals outside
143 the modelled population) and at rate $\beta S(t)I(t)$ by secondary transmission (contact with already
144 infected individuals from within the population).

145 **Disease surveillance.** During a single period of surveillance (a *surveillance bout*), individuals are
146 captured at per capita rate α , tested and released, and both the total number, and the number of
147 infected individuals caught are recorded. Perfect diagnostic tests are assumed although limited
148 sensitivities and specificities could be accounted for. A surveillance bout continues until a defined
149 sample size m is obtained or some upper time limit has been reached. Such surveillance is most
150 naturally considered in the context of active capture campaigns but could also be adapted to
151 samples obtained from hunting and passive surveillance by accounting for the losses and sources of
152 bias associated with such surveillance methods (see e.g. McElhinney *et al.* 2014).

153

154

155 **Model implementation.** The model framework is summarised in Table 1. Reported results are
156 temporal averages (e.g. expected mean $E[M]$ and variance $\text{Var}[M]$ in population size) based on long
157 run simulations following a burn-in period to allow the population to reach equilibrium where the
158 effects of initial conditions are negligible. Within each run repeated surveillance bouts are simulated
159 and the probability of detection PD is estimated as the proportion of bouts where disease is
160 detected. The mean $E[\hat{p}_{surv}]$ and variance $\text{Var}[\hat{p}_{surv}]$ of the prevalence estimates averaged over
161 repeated bouts are also recorded. We consider a continuous state-space implementation simulated
162 by numerically integrating a set of stochastic differential equations (SDEs) and a discrete state-space
163 implementation using the Gillespie algorithm (see Appendix S1 for details).

164 Simulation studies

165 **Study 1** (results shown in Fig.1 and Fig.3) uses the SDE implementation and is designed to explore a
166 generic but representative range of wildlife disease systems. Simulations were run for four values
167 (0.01, 0.04, 0.1, 1.0) of the secondary transmission rate β . In each case the population death rate μ
168 was varied over a wide range between 0.1 and 0.5, with the intrinsic growth rate set at $r=0.5$ so
169 that, at the upper end of this range, populations are highly unstable. This gives rise to typical
170 population sizes of 10-40 (see Fig.1a) and a wide range of disease prevalence. Similar results are
171 obtained from simulations (not shown) where β is varied for a set of fixed values of μ where
172 mortality rates span the interval $(0, r)$. Simulations not included here show that our results
173 generalise, holding for transmission rates relative to a recovery rate (governing an additional
174 transition from I to S) and death rates relative to birth rate, r . Different intensities of surveillance
175 were simulated using four capture rates α (0.01, 0.1, 1.0, 10), for a sample size $m=10$. Full
176 parameterisations for Fig.1 and Fig. 3 are shown in Tables S3 and S6 respectively.

177 **Study 1a** (results shown in Fig. 2) explores the effect of surveillance design using a subset of the
178 parameter sets considered in study 1, namely $(\beta, \mu) = (1.0, 0.43)$; $(1.0, 0.4)$; and $(0.1, 0.43)$. For
179 each, a range of capture rates $\alpha = 0 \dots 10$ (with $m=10$) and a range of sample sizes $m=1, \dots,$

180 10000 (with $\alpha=0.1$) are considered. The values of all model parameters used are shown in Tables
181 S4 and S5 (see Supporting Information).

182 *Relevance to real wildlife disease systems.* The intrinsic annual growth and death rates for badgers
183 have been estimate as $r=0.6$ and $\mu=0.4$ (Anderson & Trewhella 1985). Rescaling for $r=0.5$ as used
184 in simulation study 1 corresponds to a rescaled $\mu=0.33$. In addition the secondary transmission rate
185 for TB in badger populations was been estimated by the same authors to be $\beta=0.06-0.08$ assuming
186 a density of badgers necessary for disease persistence is ~ 5 badgers km^{-2} (Anderson & Trewhella
187 1985). The population size considered in simulation study 1 therefore corresponds to a surveillance
188 area of around 8 km^2 . The range of parameters considered in study 1 places badgers towards the
189 stable end of the spectrum. More fecund and shorter lived species would be expected to be less
190 stable e.g. have higher mortality and secondary transmission rates. As noted earlier surveillance of
191 badgers at Woodchester Park is relatively intensive leading to an annual probability of capture of
192 around 80% corresponding to capture rates of $\alpha=1.6-2.2$ (Delahay *et al.* 2000). The population of
193 *Sarcophilus harrisii* (Tasmanian devil) discussed earlier consisted of between 20-60 individuals and
194 was subject to annual capture rates between 0.5 and 1.7 (Hawkins *et al.* 2006). Estimates of capture
195 rates are not available for the larger scale studies referred to in the introduction (Brugman *et al.*
196 2013; Pounder *et al.* 2013; McGarry *et al.* 2014) but given the sample sizes obtained and the
197 temporal and geographic scales involved it seems reasonable to assume that they are considerably
198 lower. Simulation study 1 encompasses a wide range of real world wildlife disease surveillance.

199 **Study 2** (results shown in Fig. 4) is designed to test the robustness of study 1 by exploring a wider
200 range of scenarios: with intrinsic growth rates in the range (0,23); mortality rates in the range
201 (0.25,14), carrying capacities in the range (0,36) and secondary contact rates in the range (0.01,5).
202 Focussed on disease detection, results are conditioned on the presence of disease and simulations
203 based on the Gillespie implementation which explicitly handles the discrete nature of small
204 populations. The values of all model parameters used in Fig. 4 are shown in Table S7 (see Supporting
205 Information).

206 **Results**207 **Estimating Prevalence**

208 In order to develop an understanding of the properties of wildlife disease surveillance using the
 209 above model, we developed expressions describing prevalence estimates obtained by continuous
 210 surveillance, i.e. continuously deployed effort resulting in per capita capture rate α .

211

212 Consider the interval $[0, T]$ during which the population history is $\mathcal{H}[0, T] = \{(N(t), p(t)): t \in [0, T]\}$,
 213 where $N(t)$ and $p(t)$ represent the population size and disease prevalence at time $t \in [0, T]$
 214 respectively (see above). Let n_T represent the total number, and i_T the number of infected
 215 individuals sampled during this time interval. Conditional on the history $\mathcal{H}[0, T]$, the expectations of
 216 these quantities are:

217

$$218 \quad E[n_T | \mathcal{H}[0, T]] = \int_0^T \alpha N(t) dt \quad \text{and} \quad E[i_T | \mathcal{H}[0, T]] = \int_0^T \alpha N(t) p(t) dt.$$

219

220 The surveillance estimate of disease prevalence is simply the ratio $\hat{p}_{surv}(T) = i_T / n_T$. Since
 221 immigration prevents extinction of the population and disease then the long time limit of this
 222 estimate can be equated with its expectation over all histories as follows:

$$223 \quad \lim_{T \rightarrow \infty} \hat{p}_{surv}(T) = E[\hat{p}_{surv}] = \lim_{T \rightarrow \infty} \frac{\frac{1}{T} \int_0^T N(t) p(t) dt}{\frac{1}{T} \int_0^T N(t) dt} = \frac{E[N(t) p(t)]}{E[N(t)]}.$$

224

225 This can be re-expressed in the more suggestive form:

226

$$E[\hat{p}_{surv}] = E[p(t)] + \frac{Cov[N(t), p(t)]}{E[N(t)]} \quad (1)$$

227

228 Thus, when the covariance $Cov[N(t), p(t)] = E[N(t) p(t)] - E[N(t)] E[p(t)]$ between the
 229 population size and the prevalence is non-zero, the surveillance estimate of prevalence is a biased

230 estimate of the true prevalence, $E[p(t)]$. Since $Cov[N(t), p(t)]$ will be zero when either $N(t)$ or $p(t)$
231 are constant, we conclude that demographic fluctuations and stochasticity in disease dynamics
232 undermine the efficacy of surveillance.

233 *Effect of host demography and disease dynamics*

234 Fig. 1 is based on simulation study 1 (see methods) and illustrates how population fluctuations and
235 disease dynamics in the host-pathogen system affect the bias and variance of estimated prevalence.
236 These results are generated by simulating the system, in each case until it reaches equilibrium, for a
237 range of values of the death rate μ , with other parameters fixed. As the death rate increases, the
238 equilibrium expected population size decreases and the relative size of the population fluctuations
239 increase as measured by the coefficient of variation. For a given rate of disease transmission β ,
240 increasing the death rate reduces expected prevalence, and therefore simulating for different values
241 of μ generates the range of prevalence values shown. The resulting relationship between
242 demography and expected prevalence for particular disease characteristics (here a fixed
243 transmission rate, β) is illustrated in Figs 1a & 1b. These figures show increasing population size and
244 lower demographic fluctuations as expected prevalence increases (i.e. as μ decreases).

245

246 Fig. 1c shows the bias in the surveillance estimate of prevalence $E[\hat{p}_{surv}] - E[p(t)]$ obtained from the
247 same set of simulations. Results shown are based on 10^6 surveillance bouts with sample size $m =$
248 10. The bias predicted by continuous sampling theory (which does not account for sample size) is
249 also shown, and in this case accurately predicts simulated bias. Fig. 1c shows the bias in surveillance
250 estimates of prevalence for four different transmission rates. For a given prevalence, populations
251 associated with higher transmission rate (β) are more variable than those with lower transmission
252 rate and therefore Fig. 1c shows that such variability increases the bias of surveillance estimates of
253 disease prevalence. Fig. 1d shows the standard deviation in surveillance estimates of prevalence
254 obtained from the same set of simulations. Comparison with the variability in prevalence estimates
255 expected under the zero fluctuation assumption reveals that fluctuations in our simulated wildlife

256 disease system reduce the precision (increase the variance) of estimates obtained by surveillance.
257 The variability of these estimates also increases with demographic fluctuations. Thus, in terms of
258 prevalence estimation, the dynamics of the host-pathogen interaction are integral in determining
259 the efficacy of surveillance. Assessment for a given system would require parameterisation of
260 demography and disease dynamic, but the bias and variance in prevalence estimates shown in Fig. 1
261 are representative of a wide range of wildlife disease systems (see methods).

262

263 Additional studies shown in the supporting information confirm the qualitative impact of
264 fluctuations in population and prevalence seen in Fig. 1 are robust to sample and population size and
265 mode of secondary transmission. Fig. S1 shows analogous results with sample size 100, where
266 environmental variability drives fluctuations in a population around 100 times larger than
267 considered above. Fig. S3 shows results for simulation study 1 but where secondary transmission is
268 frequency (as opposed to density) dependent. Fig. S5 and Fig. S6 show results from simulation study
269 1 with sample sizes 20 and 50 respectively.

270

271 *Surveillance design*

272 Based on simulation study 1a, Fig. 2 shows how the bias and variance of the estimate of prevalence
273 changes as the intensity of surveillance (measured by the capture rate α) increases for fixed sample
274 size (Figs 2a & 2c), and as the sample size, m , increases for a fixed capture rate (Figs 2b & 2d). For
275 low capture rates, as $\alpha \rightarrow 0$ (and based on a fixed sample size), the continuous sampling estimate
276 given in equation (1) provides an accurate prediction for the level of prevalence estimated from
277 surveillance. As shown above, this is a biased estimate of the true prevalence $E[p(\ell)]$. However,
278 increasing the capture rate reduces bias, and as α increases, this bias tends to zero. In addition, for
279 large capture rates, the precision of the surveillance estimate of prevalence matches the variability
280 of the underlying wildlife disease system (see Fig. 2c). Thus for low capture rates, the bias in
281 surveillance estimates of prevalence is well described by continuous sampling theory (equation 1).

282 However, for larger capture rates, the properties of the surveillance estimate of prevalence
283 increasingly reflect both the expected true prevalence (i.e. bias reduces), and the variability in the
284 prevalence of the underlying disease system. In contrast, increasing sample size improves precision,
285 but not bias (Fig. 2b). In comparison to the predictions from the standard binomial approach (which
286 neglects fluctuations), these have lower precision, and improve less quickly with increasing sample
287 size (see Fig. 2d). Additional simulation results (not shown) indicate that as the sample size
288 increases, the capture rate required to obtain unbiased estimates increases. However, even for
289 large sample sizes, when sampling is instantaneous sampling (i.e. $\alpha \rightarrow \infty$), the bias is zero and the
290 standard deviation in the surveillance estimate of prevalence corresponds to that of the underlying
291 wildlife disease system as shown above.

292

293 We previously noted that capture rates for relatively intensely monitored populations (Delahay *et al.*
294 2000; Hawkins *et al.* 2006) were between 0.5 and 2.2 with those of larger scale studies (Brugman *et*
295 *al.* 2013; Pounder *et al.* 2013; McGarry *et al.* 2014) lower still. Therefore, the results of Fig. 2 suggest
296 fluctuations will lead to bias in surveillance-based estimates of prevalence for a wide range of
297 wildlife disease systems. However, the size of these effects will be dependent on the details of host
298 species demography and disease dynamics.

299

300 **The Probability of Detection**

301 If prevalence is assumed constant and equal to the long term average prevalence $E[p]$ of the wildlife
302 disease system, then the probability that disease is detected in a sample of size m is given by:

303

$$304 \quad PD^{Bin} = f(E[p], m) = 1 - (1 - E[p])^m \quad (2)$$

305

306 This formula, based on simple binomial arguments, and variants that also assume constant
307 prevalence, are the standard basis for sample size calculations (see e.g. Fosgate 2009). However, if
308 prevalence fluctuates PD^{Bin} is a misleading estimate of the probability of detection.

309

310 When conducting surveillance prevalence will vary between the times when each of the m samples
311 are collected, but we assume prevalence within a given surveillance bout is constant, and denoted p .

312 Fig. 3a indicates that accounting only for fluctuations between surveillance bouts is an accurate
313 approximation. Therefore, the expected probability of detection for sample size m is defined as

314

$$315 \quad PD = E[f(p, m)] = E[1 - (1 - p)^m] \quad (3)$$

316

317 where the expectation is over the between bout prevalence distribution $P(p)$ which accounts only
318 for prevalence fluctuations between surveillance bouts. For a single sample $m = 1$, equation (3)
319 reduces to a linear form, so that $PD = PD^{Bin} = E[p]$. However, if $m > 1$, then equation (3) is non-
320 linear, and therefore $PD \neq PD^{Bin}$. Further analysis of equation (3) e.g. suggesting $PD < PD^{Bin}$, is
321 shown in Appendix S4 (see supporting information).

322

323 *Effect of host demography and transmission dynamics*

324 The results shown in Fig. 3 demonstrate the effect of host demography, transmission dynamics and
325 surveillance design on the probability of detection. These results are obtained from the simulations
326 described in Fig. 1, except for those in Fig. 3d where these simulations are rerun for different values
327 of the capture rate (see study 1a in methods).

328

329 Fig. 3b illustrates an analytic calculation of PD based on approximating the between bout prevalence
330 distribution $P(p)$ as a gamma distribution (see supporting information). Although, not completely
331 successful, this does provide a more accurate prediction than PD^{Bin} . This approach could be used to

332 improve sample size calculations in situations where simulation is not possible, but information
333 about prevalence fluctuations is available. Moreover, the results of Fig. 3a show that such
334 approximations could be improved by assuming a more accurate representation of the prevalence
335 distribution $P(p)$. Crucially, these calculations support the conclusion that the true probability of
336 detection is less than that obtained when ignoring fluctuations i.e. less than PD^{Bin} . Fig. 3b also
337 shows the impact of biased prevalence estimation on disease detection for the case $\beta = 0.1$. Fig. 1
338 demonstrates that in this case, surveillance results in inflated estimates of prevalence $E[\hat{p}_{surv}] >$
339 $E[p(\hat{t})]$. Ignoring the effect of fluctuations would therefore lead to an estimated detection
340 probability greater than PD^{Bin} , which is based on the true average prevalence $E[p]$.

341

342 Fig. 3c shows the effect of interactions between disease dynamics and demography. As in the case of
343 prevalence estimation, conditioned on a given expected prevalence, larger contact rates β are
344 associated with greater fluctuations in the underlying wildlife disease system (i.e. greater
345 transmission rates are needed to sustain a given prevalence). Here larger fluctuations translate into
346 reduced probability of detection. In Fig. 3c, for $\beta = 1.0$, the probability of detection is only a little
347 above the line $PD = E[p]$; this corresponds to a single sample $m = 1$. Thus, in contrast to the zero
348 fluctuation approximation PD^{Bin} , fluctuations reduce the effective sample size, for the $\beta = 1.0$ case
349 from $m = 10$ to close to $m = 1$. Results not shown indicate that the reduction in effective sample
350 size increases with sample size (and see Fig. 4). Fig. 3d shows the effect of capture rate on the
351 probability of detection; counter intuitively, more intense surveillance effort actually reduces the
352 probability of detection. This is consistent with the above observations regarding β ; less intense
353 effort means that the required sample size takes longer to gather, which reduces between-bout
354 fluctuations in prevalence.

355

356

357

358 *Limits to disease detection in wildlife disease systems*

359 The nature of host demography and disease dynamics in wildlife disease systems will often be poorly
360 understood especially in cases of emerging disease. Fig. 4 is based on simulation study 2 (see
361 methods) and shows the probability of detection associated with surveillance subject to
362 demographic and disease fluctuations and the zero fluctuation approximation PD^{bin} . This is done for
363 two different sampling levels, and across a broader range of wildlife disease systems than
364 considered above, each represented by one of the points on the graph. Depending on the level of
365 fluctuations in the system, the effective sample size can range from the actual number of samples
366 taken to $m \approx 1$. These results suggest that, when designing surveillance, ignoring the effect of
367 fluctuations could lead to studies that are underpowered in their ability to detect disease. These
368 results are consistent with those of Fig. 3 based on the SDE implementation.

369

370 **Discussion**

371 This paper represents the first systematic exploration of the impact of pathogen transmission
372 dynamics and demographic aspects of host ecology on wildlife disease surveillance efficacy. We have
373 introduced a framework within which surveillance design is characterised by the capture rate (α), in
374 addition to the standard sample size (m). In this extended framework, the performance of
375 surveillance is assessed in light of the ecology of the wildlife disease system of interest i.e. for
376 particular population and disease parameters. The framework introduced here can thus serve as a
377 template for performing power calculations that account for fluctuations in populations and disease
378 prevalence for specific hosts and pathogens.

379

380 Our results show that surveillance design (choice of m and α) can have a large impact on bias and
381 precision of prevalence estimation, and on the power to detect disease. With more unstable
382 populations and greater fluctuations in disease, bias in prevalence estimates increases, and the
383 precision of such estimates decreases. Such bias can be reduced by increasing capture rate, but for

384 fixed sample size this also reduces the ability to detect disease. However, results suggest that even
385 in the most intensive wildlife disease surveillance programs (Delahay *et al.* 2000; Hawkins *et al.*
386 2006) typical capture rates are not sufficient to eliminate bias. In contrast, increasing sample size
387 does not affect bias, but does improve statistical power in terms of both precision of prevalence
388 estimates and disease detection. However, as sample size increases, such improvements in power
389 are not as fast as would be expected if fluctuations were ignored, as they are in current surveillance
390 design and analysis (Grimes & Schulz 1996; Dohoo, Martin & Stryhn 2005).

391

392 Surveillance is a critical prerequisite for defining and controlling wildlife disease risks, and our results
393 suggest that ignoring significant temporal fluctuations in the design of wildlife disease surveillance
394 generates inadequate assessments of risk. Moreover, the ecology of many wildlife species and the
395 pathogens to which they are exposed lead to significant temporal fluctuations in both population
396 size and disease prevalence (Anderson & May 1979; Anderson 1991; Renshaw 1991; Wilson &
397 Hassell 1997; Telfer *et al.* 2002; Hawkins *et al.* 2006). The studies reported here were designed to
398 explore these effects in a wide range of scenarios representative of actual surveillance in wildlife
399 disease systems (see methods), and suggest that such issues are likely to be widespread. A key
400 aspect not accounted for in the work presented here is disease induced mortality which preliminary
401 results (not shown) suggest is likely to accentuate the effects shown here. Moreover, frequency
402 dependent transmission and fluctuations driven by environmental variation, studied only briefly
403 here, also reduced the efficacy of surveillance. The framework presented could also be extended to
404 account for known extrinsic sources of bias, such as imperfect disease diagnostics, variation in
405 habitat quality (Nusser *et al.* 2008; Walsh & Miller 2010) and biased capture rates (Tuytens *et al.*
406 1999) including aspects associated with passive surveillance.

407

408 There is much current interest in quantifying risks from wildlife disease (Daszak, Cunningham &
409 Hyatt 2000; Jones *et al.* 2008), and this is stimulating debate on the need to improve wildlife disease

410 surveillance (Bengis *et al.* 2004; Butler 2006; Gortázar *et al.* 2007; Béneult, Ciliberti & Artois 2014).
411 This paper will help to further inform this debate, highlighting the need to consider the ecology of
412 wildlife disease systems when designing or analysing surveillance programs (Béneult, Ciliberti &
413 Artois 2014). This assessment emphasizes the importance of accounting for temporal
414 heterogeneities induced by population fluctuations and disease dynamics. Further research is
415 needed to assess the impacts of ecology on wildlife disease surveillance including alternative and
416 complimentary heterogeneities such as intrinsic and extrinsic forms of spatial heterogeneity, and
417 other population structures. There is a wealth of literature describing the effects of such
418 heterogeneity in ecology and epidemiology (Lloyd & May 1996; Tilman & Kareiva 1997; Keeling,
419 Wilson & Pacala 2000; Read & Keeling 2003; Keeling 2005; Vicente *et al.* 2007), and our results
420 suggest that these are likely to have important, but as yet unexplored, impacts on the efficacy of
421 wildlife disease surveillance.

422

423 **Acknowledgements**

424 This work was supported by the European Commission under the Food, Agriculture and Fisheries,
425 and Biotechnology Theme of the 7th Framework Programme for Research and Technological
426 Development, grant agreement no. 222633. GM, RSD, LAS and MRH are grateful for funding from
427 the Scottish Government's RESAS. We are grateful to the referees whose comments helped to
428 significantly improve the manuscript.

429

430

431 **Supporting Information**

432 Additional Supporting Information may be found in the online version of this article :

433 **Appendix S1:** Model implementation

434 **Appendix S2:** Additional scenarios

435 **Appendix S3:** Analysis of disease detection probability

436

437 **Data Accessibility**

438 Model code is available at: DRYAD entry doi: tbc

439

For Peer Review

440 **References**

441

442 Anderson, R. (1991) Populations and infectious diseases: ecology or epidemiology? *The Journal of*
443 *Animal Ecology*, **60**, 1–50.

444 Anderson, R. & May, R. (1979) Population biology of infectious diseases: Part I. *Nature*, **280**, 361–
445 367.

446 Anderson, R.M. & Trewhella, W. (1985) Population Dynamics of the Badger (*Meles meles*) and the
447 Epidemiology of Bovine Tuberculosis (*Mycobacterium bovis*). *Philosophical Transactions of the*
448 *Royal Society B: Biological Sciences*, **310**, 327–381.

449 Artois, M., Bengis, R., Delahay, R.J., Duchêne, M.-J., Duff, J.P., Ferroglio, E., Gortazar, C., Hutchings,
450 M.R., Kock, R.A., Leighton, F.A., Mörner, T. & Smith, G.C. (2009a) *Management of Disease in*
451 *Wild Mammals* (eds RJ Delahay, GC Smith, and MR Hutchings). Springer Japan, Tokyo.

452 Artois, M., Bicoût, D., Doctrinal, D., Fouchier, R., Gavier-Widen, D., Globig, a, Hagemeyer, W.,
453 Mundkur, T., Munster, V. & Olsen, B. (2009b) Outbreaks of highly pathogenic avian influenza in
454 Europe: the risks associated with wild birds. *Revue scientifique et technique (International*
455 *Office of Epizootics)*, **28**, 69–92.

456 Béneult, B., Ciliberti, A. & Artois, M. (2014) A Generic Action Plan against the Invasion of the EU by
457 an Emerging Pathogen in Wildlife-A WildTech Perspective. *Planet@ Risk*, **2**, 174–181.

458 Bengis, R.G., Leighton, F.A., Fischer, J.R., Artois, M. & Mörner, T. (2004) The role of wildlife in
459 emerging and re-emerging zoonoses Recent emerging zoonoses Viral zoonoses. , **23**, 497–511.

460 Brugman, V.A., Horton, D.L., Phipps, L.P., Johnson, N., Cook, A.J.C., Fooks, A.R. & Breed, A.C. (2013)
461 Epidemiological perspectives on West Nile virus surveillance in wild birds in Great Britain.
462 *Epidemiology and Infection*, **141**, 1134–1142.

463 Butler, D. (2006) Disease surveillance needs a revolution. *Nature*, **440**, 6–7.

464 Daszak, P., Cunningham, a a & Hyatt, a D. (2000) Emerging infectious diseases of wildlife--threats to
465 biodiversity and human health. *Science (New York, N.Y.)*, **287**, 443–9.

466 Davidson, R.S., Marion, G. & Hutchings, M.R. (2008) Effects of host social hierarchy on disease
467 persistence. *Journal of theoretical biology*, **253**, 424–33.

468 Delahay, R.J., Langton, S., Smith, G.C., Clifton-Hadley, R.S. & Cheeseman, C.L. (2000) The spatio-
469 temporal distribution of *Mycobacterium bovis* (bovine tuberculosis) infection in a high-density
470 badger population. *Journal of Animal Ecology*, **69**, 428–441.

471 Dohoo, I., Martin, W. & Stryhn, H. (2005) Veterinary Epidemiologic Research. *Preventive Veterinary*
472 *Medicine*, **68**, 289–292.

473 Eckert, J. & Deplazes, P. (2004) Biological , Epidemiological , and Clinical Aspects of Echinococcosis , a
474 Zoonosis of Increasing Concern. *clinical microbiology reviews*, **17**, 107–135.

475 Fenton, A., Streicker, D.G., Petchey, O.L., Pedersen, A.B., Fenton, A., Streicker, D.G., Petchey, O.L. &
476 Pedersen, A.B. (2015) Are All Hosts Created Equal ? Partitioning Host Species Contributions to
477 Parasite Persistence in Multihost Communities. , **186**, 610–622.

478 Fosgate, G.T. (2005) Modified exact sample size for a binomial proportion with special emphasis on
479 diagnostic test parameter estimation. *Statistics in medicine*, **24**, 2857–66.

480 Fosgate, G.T. (2009) Practical Sample Size Calculations for Surveillance and Diagnostic Investigations.
481 *Journal of Veterinary Diagnostic Investigation*, **21**, 3–14.

482 Gortázar, C., Ferroglio, E., Höfle, U., Frölich, K. & Vicente, J. (2007) Diseases shared between wildlife
483 and livestock: a European perspective. *European Journal of Wildlife Research*, **53**, 241–256.

484 Grimes, D.A. & Schulz, K.F. (1996) Determining sample size and power in clinical trials: the forgotten
485 essential. *Seminars in reproductive endocrinology*, **14**, 125–31.

486 Hawkins, C.E., Baars, C., Hesterman, H., Hocking, G.J., Jones, M.E., Lazenby, B., Mann, D., Mooney,
487 N., Pemberton, D., Pyecroft, S., Restani, M. & Wiersma, J. (2006) Emerging disease and
488 population decline of an island endemic, the Tasmanian devil *Sarcophilus harrisii*. *Biological*
489 *Conservation*, **131**, 307–324.

490 Jones, K.E., Patel, N.G., Levy, M. a, Storeygard, A., Balk, D., Gittleman, J.L. & Daszak, P. (2008) Global

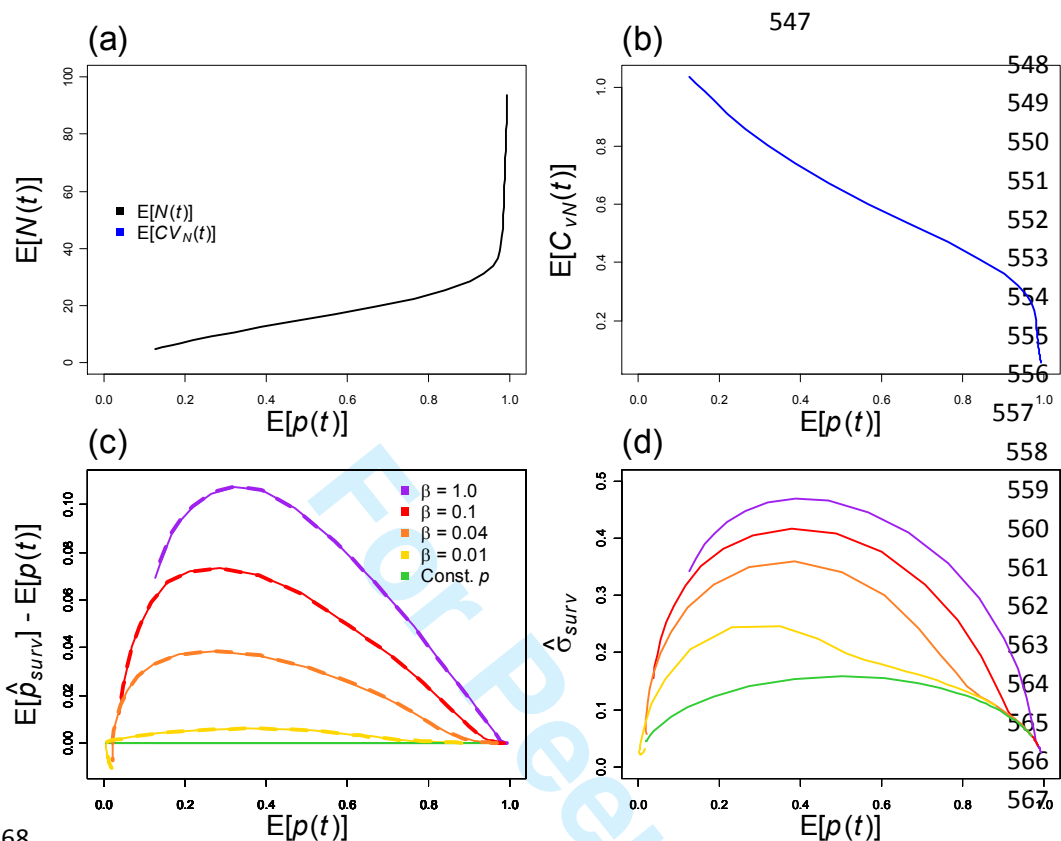
- 491 trends in emerging infectious diseases. *Nature*, **451**, 990–3.
- 492 Keeling, M. (2005) The implications of network structure for epidemic dynamics. *Theoretical*
493 *population biology*, **67**, 1–8.
- 494 Keeling, M.J., Wilson, H.B. & Pacala, S.W. (2000) Space , Reinterpreting Lags , Functional Responses
495 Models Ecological. , **290**, 1758–1761.
- 496 Kruse, H., Kirkemo, A.-M. & Handeland, K. (2004) Wildlife as source of zoonotic infections. *Emerging*
497 *infectious diseases*, **10**, 2067–72.
- 498 Kuiken, T., Ryser-Degiorgis, M.P., Gavier-Widen, D. & Gortázar, C. (2011) Establishing a European
499 network for wildlife. , **30**, 755–761.
- 500 Lipkin, W.I. (2013) The changing face of pathogen discovery and surveillance. *Nature reviews.*
501 *Microbiology*, **11**, 133–41.
- 502 Lloyd, A.L. & May, R.M. (1996) Spatial Heterogeneity in Epidemic Models. *Journal of theoretical*
503 *biology*, **179**, 1–11.
- 504 McElhinney, L.M., Marston, D. a., Brookes, S.M. & Fooks, A.R. (2014) Effects of carcase
505 decomposition on rabies virus infectivity and detection. *Journal of Virological Methods*, **207**,
506 110–113.
- 507 McGarry, J.W., Higgins, A., White, N.G., Pounder, K.C. & Hetzel, U. (2014) Zoonotic Helminths of
508 Urban Brown Rats (*Rattus norvegicus*) in the UK: Neglected Public Health Considerations?
509 *Zoonoses and Public Health*, 44–52.
- 510 Mörner, T., Obendorf, D.L., Artois, M. & Woodford, M.H. (2002) Surveillance and monitoring of
511 wildlife diseases. *Revue scientifique et technique (International Office of Epizootics)*, **21**, 67–76.
- 512 Nusser, S.M., Clark, W.R., Otis, D.L. & Huang, L. (2008) Sampling Considerations for Disease
513 Surveillance in Wildlife Populations. *Journal of Wildlife Management*, **72**, 52–60.
- 514 OIE. (2013) Terrestrial Animal Health Code
- 515 Pounder, K.C., Begon, M., Sironen, T., Henttonen, H., Watts, P.C., Voutilainen, L., Vapalahti, O.,
516 Klempa, B., Fooks, A.R. & McElhinney, L.M. (2013) Novel hantavirus in field vole, United
517 Kingdom. *Emerging infectious diseases*, **19**, 673–5.
- 518 Read, J.M. & Keeling, M.J. (2003) Disease evolution on networks: the role of contact structure.
519 *Proceedings. Biological sciences / The Royal Society*, **270**, 699–708.
- 520 Renshaw, E. (1991) *Modelling Biological Populations in Space and Time*. Cambridge University Press.
- 521 Ryser-Degiorgis, M.-P. (2013) Wildlife health investigations: needs, challenges and
522 recommendations. *BMC veterinary research*, **9**, 223.
- 523 Smith, K.F., Dobson, A.P., Mckenzie, F.E., Real, L.A., Smith, D.L. & Wilson, M.L. (2005) Ecological
524 theory to enhance infectious disease control and public health policy.
- 525 Telfer, S., Bennett, M., Bown, K., Cavanagh, R., Crespín, L., Hazel, S., Jones, T. & Begon, M. (2002) The
526 effects of cowpox virus on survival in natural rodent populations: Increases and decreases.
527 *Journal of Animal Ecology*, **71**, 558–568.
- 528 Tilman, D. & Kareiva, P. (1997) *Spatial Ecology: The Role of Space in Population Dynamics and*
529 *Interspecific Interactions*. Princeton University Press.
- 530 Tuytens, F.A.M., Macdonald, D.W., Delahay, R., Rogers, L.M., Mallinson, P.J., Donnelly, C.A. &
531 Newman, C. (1999) Differences in trappability of European badgers *Meles meles* in three
532 populations in England. *Journal of Applied Ecology*, 1051–1062.
- 533 Vicente, J., Delahay, R., Walker, N. & Cheeseman, C.L. (2007) Social organization and movement
534 influence the incidence of bovine tuberculosis in an undisturbed high-density badger *Meles*
535 *meles* population. *Journal of Animal ...*, **76**, 348–360.
- 536 Walsh, D.P. & Miller, M.W. (2010) A weighted surveillance approach for detecting chronic wasting
537 disease foci. *Journal of wildlife diseases*, **46**, 118–35.
- 538 Wilson, H.B. & Hassell, M.P. (1997) Host – parasitoid spatial models : the interplay of demographic
539 stochasticity and dynamics. *Proc. R. Soc. Lond. B*, **264**, 1189–1195.

540

541 **Table 1: Model structure.** Event, Rate and Effect on the State Space of the model. Conceptually the
 542 effect of each event affects an individual and this is reflected in the discrete nature of the
 543 corresponding changes in the state space. However, given this underlying conception of the model
 544 there are a number of different implementations which can be considered including via the Gillespie
 545 algorithm and stochastic differential equations (see text for details).

546

Event	Rate	Effect
Birth	$rN(1 - N/k)$	$S \rightarrow S + 1$
Death of Susceptible	μS	$S \rightarrow S - 1$
Death of Infected	μI	$I \rightarrow I - 1$
Susceptible Immigration	$(1 - \gamma) \nu$	$S \rightarrow S + 1$
Infected Immigration	$\gamma \nu$	$I \rightarrow I + 1$
Primary Transmission	$\beta_0 S$	$S \rightarrow S - 1$ $I \rightarrow I + 1$
Secondary Transmission	βIS	$S \rightarrow S - 1$ $I \rightarrow I + 1$
Susceptible Active Capture and Release	αS	$S \rightarrow S$
Infected Active Capture and Release	αI	$I \rightarrow I$

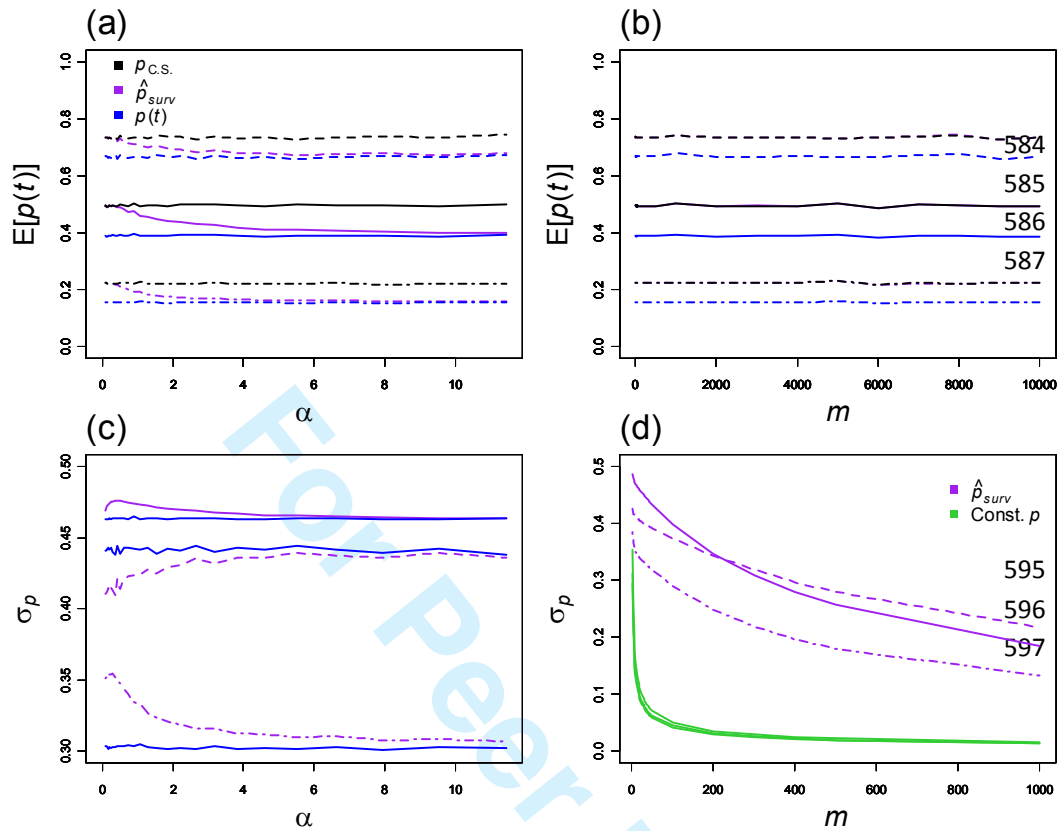


568

569

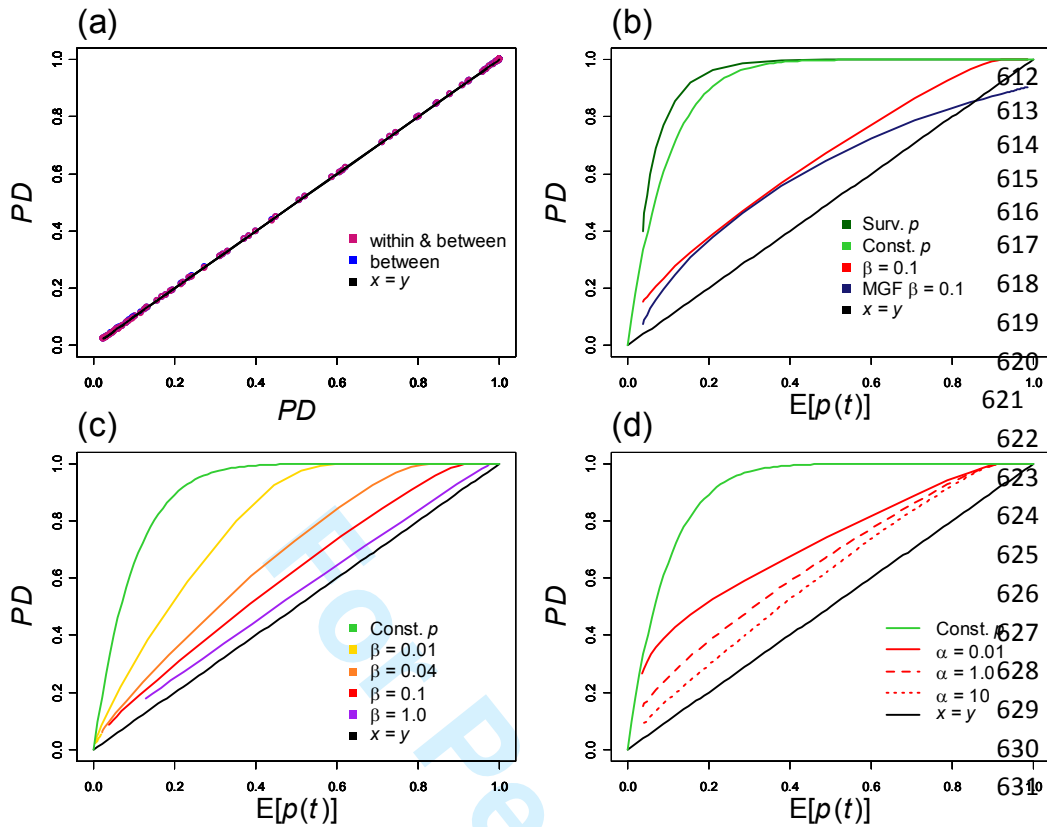
570 **Figure 1: Effect of host demography and disease transmission.** Data are shown for a range of values
 571 of the death rate μ which controls the stability and size of the population, and thus determines
 572 disease prevalence for a given transmission rate, β . For $\beta=1$ plot 1.a shows that expected
 573 population size increases with expected prevalence $E[p(t)]$ (i.e. as μ decreases) whilst plot 1.b
 574 shows that the coefficient of variation of the population size decreases. For the four values of β
 575 indicated and fixed sample size $m=10$, plot 1.c shows the bias $E[\hat{p}_{surv}] - E[p(t)]$, and plot 1.d the
 576 standard deviation in surveillance estimates of prevalence, versus the expected value of true disease
 577 prevalence in the system, $E[p(t)]$. Results shown are based on 10^6 surveillance bouts using the SDE
 578 implementation of the model (see text) using the set of parameter values described in Appendix S2.

579
580
581
582
583



588
589
590
591
592
593
594
598
599
600
601
602

603 **Figure 2: Effect of surveillance design.** In all plots results are shown for three wildlife disease
604 systems with (β, μ) : (1, 0.43) solid lines; (1, 0.4) dashed; and (0.1, 0.43) dot-dashed. Plots (a) and
605 (b) show expected values of the surveillance estimate of prevalence (purple), the true prevalence
606 (blue) and the continuous sampling theory prediction (black, see text for details). Plots (c) and (d)
607 show the expected standard deviation (denoted, σ_p) in both the true (blue) and the surveillance
608 estimated (purple) prevalence. (a) and (c) are plotted against a range of values of the capture rate α ,
609 for $m = 10$, and (b) and (d) versus a range of sample sizes m for $\alpha = 0.1$. Plot (d) also shows the
610 constant prevalence estimate of the standard deviation based on the binomial (green). Parameter
611 values used are as described in Table S3 (see Supporting Information).



632

633

634

Figure 3: Effect of host-pathogen and surveillance dynamics on probability of detection. Results

635

based on simulations used for Figure 1 (for details see Table S4, Appendix S2). (d) estimated PD

636

versus approximations based on modifications of equation (3) accounting for fluctuations in

637

prevalance (i) within and between bouts and (ii) between bouts only. (c) shows PD^{Bin} based on both

638

$E[p]$ (green) and $E[\hat{p}_{surv}]$ (black) and (for $\beta = 0.1$) PD and the approximation (equation 4) based on

639

an assumed gamma distribution. (a) shows PD^{Bin} (green) and PD for various values of β (as shown

640

yellow ($\beta = 0.01$); orange ($\beta = 0.04$); red ($\beta = 0.1$); purple ($\beta = 1.0$)) versus actual prevalence $E[p]$.

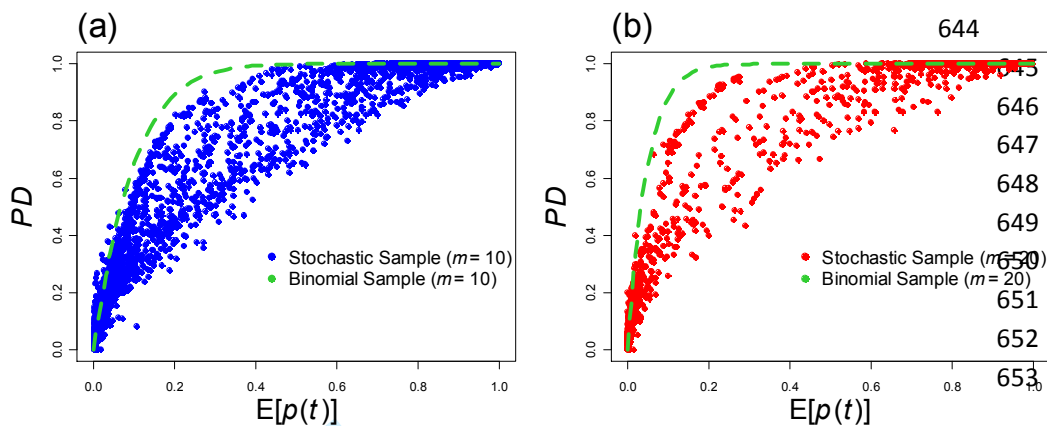
641

(b) shows PD^{Bin} (green) and PD for $\beta = 0.1$ and the three capture rates $\alpha = 0.01, 1.0, 10$. In (a), (b)

642

and (c) the black line indicates $PD = E[p(t)]$.

643



654

655

656 **Figure 4:** Fluctuations reduce power to detect disease: The two panels show the probability that
 657 disease is detected (conditional on non-zero prevalence) for target sample sizes 10 and 20. Each
 658 coloured dot represents the average of 100-1000 realisations of the model implemented using the
 659 Gillespie algorithm that met the sample target for a particular combination of parameters
 660 representing a distinct host-pathogen system (for details see Table S5, Appendix S2). The green
 661 dashed line in both graphs represents PD^{Bin} the probability of detection assuming constant
 662 prevalence (see equation 2). It can be seen that PD^{Bin} generally over-estimates the power of the
 663 sample in that it predicts a larger probability of detection than is realised in the stochastic
 664 simulations.

Supporting Information:

The Ecology of Wildlife Disease Surveillance:

Demographic and prevalence fluctuations undermine surveillance

Laura Walton^{1,2,3}, Glenn Marion^{1*}, Ross S. Davidson², Piran C.L. White³, Lesley A. Smith², Dolores Gavier-Widen⁴, Lisa Yon⁵, Duncan Hannant⁵ and Michael R. Hutchings²

1. Biomathematics and Statistics Scotland, Edinburgh, United Kingdom
2. Disease Systems Team, SRUC, Edinburgh, United Kingdom
3. Environment Department, University of York, York, United Kingdom
4. Swedish University of Agricultural Sciences, Uppsala, Sweden
5. School of Veterinary Medicine and Science, University of Nottingham, United Kingdom

Appendix S1. Model implementation

Model implementation

The model is implemented as a set of coupled Stochastic Differential Equations, (SDEs) (see e.g. Mao 1997) and simulated using the Euler-Maruyama algorithm (e.g. see Higham 2001) which is essentially a generalisation of the Euler discretisation for Ordinary Differential Equations to SDEs. The model is also implemented (for simulation study 2) as a continuous-time discrete-state space Markov process, simulated using Gillespie's algorithm (Gillespie 1976). The Gillespie algorithm is an event-based method that makes use of the fact that in the underlying discrete state-space Markov process at any point in time the waiting time between events is exponential and parameterised by the total rate of all possible events i.e. the sum of all possible events. The Gillespie algorithm proceed from time t by drawing a waiting time τ from this distribution, advancing time to $t+\tau$, and then selects the nature of the event at random but weighted according to the relative rates of the possible events. The SDE implementation has been constructed so that it is the diffusion limit of the Gillespie implementation, ensuring that the results are consistent between the two implementations (see below). The Gillespie algorithm is computationally more intensive; by contrast, using SDEs is faster and therefore

facilitates both more accurate estimation of model statistics (i.e. a greater number of surveillance bouts can be run) and more extensive exploration of parameter space. However, the discrete nature of the state-space under the Gillespie algorithm is a more direct implementation of the model described in Table 1, and provides a more accurate representation of population dynamics especially for small populations.

Relationship between discrete and continuous (SDE) state-space model implementations.

In this appendix we describe the relationship between the continuous time discrete state-space Markov process and the stochastic differential equation (SDE) implementations of the model described in the main text.

Our starting point is the SI model described in Table 1 (main text) implemented as a continuous time discrete state-space Markov process in which the number of infected individuals $I(t)$ and total population size $N(t) = S(t) + I(t)$, are represented as integer variables. The Gillespie algorithm exploits the fact that the time between events is distributed exponentially with parameter $R(t)$ given by the sum of all the event rates in Table 1 and the probability that a given event occurs is given by the associated event rate divided by $R(t)$.

However, under this implementation one can also consider the expectation and variance-covariance of the change in the state-space variables $I(t)$ and $N(t)$ during a small time interval. For convenience denote the state of the system at time t by $X(t) = \{I(t), N(t)\}$. Then for example, conditional on the state of the system at time t , the expected change in the population size associated with birth events from time t to $t + \delta t$ is given by $E_B[\delta N(t) | X(t)] = rN(t) (1 - N(t)/k) \delta t$. Similarly, the variance in δN associated with birth events is $\text{Var}_B[\delta N(t)] = rN(t) (1 - N(t)/k) \delta t + O(\delta t^2)$, and henceforth we will assume δt is sufficiently small to ignore the higher order terms. In the model described in the main text (see Table 1 and surrounding text) all individuals are born susceptible and therefore birth does not affect the infective population size $I(t)$ i.e. $E_B[\delta I(t) | X(t)] = 0$, $\text{Var}_B[\delta I(t)] = 0$, and $\text{Cov}_B[\delta I(t), \delta N(t) | X(t)] = 0$. However, migration of infectives affects both $I(t)$ and $N(t)$ and to first order in δt we find that $E_{mI}[\delta N(t) | X(t)] = \gamma \nu \delta t$, $\text{Var}_{mI}[\delta N(t)] = \gamma \nu \delta t$, $E_{mI}[\delta I(t) | X(t)] = \gamma \nu \delta t$, $\text{Var}_{mI}[\delta I(t)] = \gamma \nu \delta t$ and $\text{Cov}_{mI}[\delta I(t), \delta N(t) | X(t)] = \gamma \nu \delta t$. The full set of first- and second-order statistics describing changes in the state-space associated with each event type are given (up to first order in δt) in Table S1.

Etype	Event	$E[\delta N X(t)]$	$E[\delta I X(t)]$	$\text{Var}[\delta N X(t)]$	$\text{Var}[I X(t)]$	$\text{Cov}[\delta N, \delta I X(t)]$
B	Birth	$rN(1 - N/k)\delta t$	0	$rN(1 - N/k)\delta t$	0	0
DS	Death of Susceptible	$-\mu S\delta t$	0	$\mu S\delta t$	0	0
DI	Death of Infected	$-\mu I\delta t$	$-\mu I\delta t$	$\mu I\delta t$	$\mu I\delta t$	$\mu I\delta t$
mS	Susceptible Immigration	$(1 - \gamma)v\delta t$	0	$(1 - \gamma)v\delta t$	0	0
mi	Infected Immigration	$\gamma v\delta t$	$\gamma v\delta t$	$\gamma v\delta t$	$\gamma v\delta t$	$\gamma v\delta t$
1ry	Primary Transmission	0	$\beta_0 S\delta t$	0	$\beta_0 S\delta t$	0
2ry	Secondary Transmission	0	$\beta I S\delta t$	0	$\beta_0 S\delta t$	0

Table S1: Expectations and variance-covariances in changes (during the time interval t to $t+\delta t$) to the state space $\{I(t), N(t)\}$ associated with each event type in the discrete state-space model described in the main text (see Table 1). All such quantities are shown to first order in δt . Note: capture and release events are omitted since they affect neither $I(t)$ or $N(t)$.

We now show how to construct a continuous time, continuous state-space (diffusion) version of the model which is consistent with above implementation in that it preserves the means and variance-covariance statistics shown in Table S1. To do so we construct a set of stochastic differential equations (SDEs) which we later solve numerically in discrete time steps (e.g. see Higham 2001). The following Itô stochastic differential equations represent the change in the system state variables during an infinitesimally small time interval dt

$$\begin{aligned}
 dN(t) = & \left(f_{N,B}(X(t)) + f_{N,DS}(X(t)) + f_{N,DI}(X(t)) + f_{N,mS}(X(t)) + f_{N,mi}(X(t)) \right. \\
 & \left. + f_{N,1ry}(X(t)) + f_{N,2ry}(X(t)) \right) dt \\
 & + g_{N,B}(X(t))dB_B(t) + g_{N,DS}(X(t))dB_{DS}(t) + g_{N,DI}(X(t))dB_{DI}(t) \\
 & + g_{N,mS}(X(t))dB_{mS}(t) + g_{N,mi}(X(t))dB_{mi}(t) \\
 & + g_{N,1ry}(X(t))dB_{1ry}(t) + g_{N,2ry}(X(t))dB_{2ry}(t)
 \end{aligned}$$

$$\begin{aligned}
 dI(t) = & \left(f_{I,B}(X(t)) + f_{I,DS}(X(t)) + f_{I,DI}(X(t)) + f_{I,mS}(X(t)) + f_{I,mi}(X(t)) \right. \\
 & \left. + f_{I,1ry}(X(t)) + f_{I,2ry}(X(t)) \right) dt
 \end{aligned}$$

$$\begin{aligned}
& + g_{I,B}(X(t))dB_B(t) + g_{I,DS}(X(t))dB_{DS}(t) + g_{I,DI}(X(t))dB_{DI}(t) \\
& \quad + g_{I,mS}(X(t))dB_{mS}(t) + g_{I,mI}(X(t))dB_{mI}(t) \\
& \quad + g_{I,1ry}(X(t))dB_{1ry}(t) + g_{I,2ry}(X(t))dB_{2ry}(t)
\end{aligned}$$

Here the quantities $B_B(t)$, $B_{DS}(t)$, $B_{DI}(t)$, $B_{mS}(t)$, $B_{mI}(t)$, $B_{1ry}(t)$, $B_{2ry}(t)$ are independent Brownian motions corresponding to each of the seven event types and the correct interpretation of these equations requires consideration of associated stochastic integrals (Mao, 1997). For small but finite dt the quantities $dB_B(t)$, $dB_{DS}(t)$, $dB_{DI}(t)$, $dB_{mS}(t)$, $dB_{mI}(t)$, $dB_{1ry}(t)$, $dB_{2ry}(t)$ can be interpreted as independent draws from a zero mean Gaussian with variance dt for each event type and each time point $0, dt, 2dt, \dots, T \in (0, T)$. Thus e.g. $E[dB_B(t)] = 0$, $E[dB_B(t)dB_B(t)] = 0$ and $E[dB_B(t)dB_{DS}(t)] = 0$. This discretisation is the basis for the numerical simulation of these SDEs used in this paper.

The so-called drift, $f_{N,B}(X(t))$, $f_{N,DS}(X(t))$, $f_{N,DI}(X(t))$, $f_{N,mS}(X(t))$, $f_{N,mI}(X(t))$, $f_{N,1ry}(X(t))$, $f_{N,2ry}(X(t))$ and diffusion, $g_{N,B}(X(t))$, $g_{N,DS}(X(t))$, $g_{N,DI}(X(t))$, $g_{N,mS}(X(t))$, $g_{N,mI}(X(t))$, $g_{N,1ry}(X(t))$, $g_{N,2ry}(X(t))$, terms representing changes in the variable $N(t)$ and the corresponding quantities representing changes in $I(t)$ are deterministic functions of the state-space $X(t)$ determined as follows.

Given the nature of the Brownian motions taking the expectation of the above equations yields

$$\begin{aligned}
& E[dN(t)|X(t)] \\
& = \left(f_{N,B}(X(t)) + f_{N,DS}(X(t)) + f_{N,DI}(X(t)) + f_{N,mS}(X(t)) + f_{N,mI}(X(t)) \right. \\
& \quad \left. + f_{N,1ry}(X(t)) + f_{N,2ry}(X(t)) \right) dt
\end{aligned}$$

$$\begin{aligned}
& E[dI(t)|X(t)] \\
& = \left(f_{I,B}(X(t)) + f_{I,DS}(X(t)) + f_{I,DI}(X(t)) + f_{I,mS}(X(t)) + f_{I,mI}(X(t)) \right. \\
& \quad \left. + f_{I,1ry}(X(t)) + f_{I,2ry}(X(t)) \right) dt
\end{aligned}$$

Which suggests that for each event type $Etype$ $f_{N,Etype}(X(t))$ and $f_{I,Etype}(X(t))$ should be interpreted as the mean update shown in Table S1 for $N(t)$ and $I(t)$ respectively. For example, $f_{N,1ry}(X(t))$ and

$f_{N,2ry}(X(t))$ are both zero since only birth, death and migration change the population size, i.e. neither primary nor secondary infection changes the population size.

The variance in the update for $N(t)$ is given by

$$\text{Var}[dN(t)|X(t)] = E[dN(t)^2|X(t)] - E[dN(t)|X(t)]^2$$

However, we have just shown that $E[dN(t)|X(t)]$ is of order dt and therefore to first order in dt we can write

$$\begin{aligned} \text{Var}[dN(t)|X(t)] &= E[dN(t)^2|X(t)] = \\ &g_{N,B}(X(t))^2 dt + g_{N,DS}(X(t))^2 dt + g_{N,DI}(X(t))^2 dt + g_{N,mS}(X(t))^2 dt \\ &+ g_{N,mI}(X(t))^2 dt + g_{N,1ry}(X(t))^2 dt + g_{N,2ry}(X(t))^2 dt \end{aligned}$$

and

$$\begin{aligned} \text{Var}[dI(t)|X(t)] &= E[dI(t)^2|X(t)] = \\ &g_{I,B}(X(t))^2 dt + g_{I,DS}(X(t))^2 dt + g_{I,DI}(X(t))^2 dt + g_{I,mS}(X(t))^2 dt \\ &+ g_{I,mI}(X(t))^2 dt + g_{I,1ry}(X(t))^2 dt + g_{I,2ry}(X(t))^2 dt \end{aligned}$$

Here we have made use of the independent nature of the Brownian motions described above.

These last two equations therefore suggest that for each event type E_{type} , $g_{N,E_{\text{type}}}(X(t))^2$ and $g_{I,E_{\text{type}}}(X(t))^2$ should be interpreted as the variance in update shown in Table S1 for $N(t)$ and $I(t)$ respectively.

The above calculations are summarised in Table S2. Comparison with Table S1 allows the functional form for each drift and diffusion term to be identified.

Finally, the covariance

$$\text{Cov}[dN(t)dI(t)|X(t)] = E[dN(t)dI(t)|X(t)] - E[dN(t)|X(t)]E[dI(t)|X(t)]$$

to first order in dt is given by

$$\begin{aligned} \text{Cov}[dN(t)dI(t)|X(t)] &= E[dN(t)dI(t)|X(t)] = \\ &+ g_{N,DI}(X(t))g_{I,DI}(X(t))dt + g_{N,ml}(X(t))g_{I,ml}(X(t))dt \end{aligned}$$

where we have shown only the non-zero terms. Comparison with the functional forms for the diffusion terms described above shows that this expression is consistent with the covariance terms shown in Table S1.

Etype	$E[dN X(t)]$	$E[dI X(t)]$	$\text{Var}[dN X(t)]$	$\text{Var}[dI X(t)]$	$\text{Cov}[dN,dI X(t)]$
B	$f_{N,B}(X(t))dt$	$f_{I,B}(X(t))dt$	$g_{N,B}(X(t))^2 dt$	$g_{I,B}(X(t))^2 dt$	0
DS	$f_{N,DS}(X(t))dt$	$f_{I,DS}(X(t))dt$	$g_{N,DS}(X(t))^2 dt$	$g_{I,DS}(X(t))^2 dt$	0
DI	$f_{N,DI}(X(t))dt$	$f_{I,DI}(X(t))dt$	$g_{N,DI}(X(t))^2 dt$	$g_{I,DI}(X(t))^2 dt$	$g_{N,DI}(X(t))g_{I,DI}(X(t))dt$
mS	$f_{N,mS}(X(t))dt$	$f_{I,mS}(X(t))dt$	$g_{N,mS}(X(t))^2 dt$	$g_{I,mS}(X(t))^2 dt$	0
ml	$f_{N,ml}(X(t))dt$	$f_{I,ml}(X(t))dt$	$g_{N,ml}(X(t))^2 dt$	$g_{I,ml}(X(t))^2 dt$	$g_{N,ml}(X(t))g_{I,ml}(X(t))dt$
1ry	$f_{N,1ry}(X(t))dt$	$f_{I,1ry}(X(t))dt$	$g_{N,1ry}(X(t))^2 dt$	$g_{I,1ry}(X(t))^2 dt$	0
2ry	$f_{N,2ry}(X(t))dt$	$f_{I,2ry}(X(t))dt$	$g_{N,2ry}(X(t))^2 dt$	$g_{I,2ry}(X(t))^2 dt$	0

Table S2: Expectation and variance-covariances in changes (during the time interval t to $t+dt$) to the state space $\{I(t), N(t)\}$ associated with each event type in the SDE model as described in Appendix S1. All such quantities are shown to first order in dt . Comparison with Table S1 enables both drift e.g. $f_{N,B}(X(t))$ and diffusion e.g. $g_{N,B}(X(t))$ functions to be identified. Note: capture and release events are omitted since they affect neither $I(t)$ or $N(t)$.

References

Higham, D. J. (2001) An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations. *SIAM REVIEW* **43**(3), 525–546

Mao, X., (1997) *Stochastic Differential Equations and Applications*. Horwood, New York.

For Peer Review

Appendix S2. Parameterisations used.

This section of the appendix describes in detail the parameter combinations used to produce the graphs in the main text. Values of the form: a,b,c,d etc refer to discrete values used for different lines shown on the Figures. Values of the form a;b;c refer to smallest value; largest value; step size describing the range of values (e.g. of the death rate) simulated to produce the Figures. Values of the form a – b refer to the range of values covered with a non-constant step size. All other parameters with single values are held constant in simulations.

Rate Name	Rate	Value
Secondary Transmission Rate	β	1.0, 0.1, 0.04, 0.01
Carrying Capacity	k	120
Growth Rate	r	0.5
Death Rate	μ	0.1;0.5;0.1
Immigration	ν	0.1
Infected Immigration Proportion	γ	0.1
Primary Transmission Rate	β_0	0
Susceptible Active Capture	α	0.1
Infected Active Capture	α	0.1
Sample Target	m	10.0

Table S3: Parameter values are shown for Figure 1 in the main text which demonstrates the effect of the death rate and transmission rate on the bias and variance of the prevalence estimate as well as the effect of the death rate on the population size and variance. 10^6 surveillance bouts are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	β	1.0, 0.1
Carrying Capacity	k	120
Growth Rate	r	0.5
Death Rate	μ	0.4, 0.43
Immigration	ν	0.1
Infected Immigration Proportion	γ	0.1
Primary Transmission Rate	β_0	0
Susceptible Active Capture	α	0 - 10
Infected Active Capture	α	0 - 10
Sample Target	m	10.0

Table S4: Parameter values are shown for Figure 2 in the main text which demonstrates the effect of the capture rate on the bias and variance of the prevalence estimate. 10^6 surveillance bouts are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	β	1.0, 0.1
Carrying Capacity	k	120
Growth Rate	r	0.5
Death Rate	μ	0.4, 0.43
Immigration	ν	0.1
Infected Immigration Proportion	γ	0.1
Primary Transmission Rate	β_0	0
Susceptible Active Capture	α	0.1
Infected Active Capture	α	0.1
Sample Target	m	1 - 10000

Table S5: Parameter values are shown for Figure 2 in the main text which demonstrates the effect of the sample size on the bias and variance of the prevalence estimate. 10^6 surveillance bouts are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	β	1.0, 0.1, 0.04, 0.01
Carrying Capacity	k	120
Growth Rate	r	0.5
Death Rate	μ	0.1;0.5;0.01
Immigration	ν	0.1
Infected Immigration Proportion	γ	0.1
Primary Transmission Rate	β_0	0
Susceptible Active Capture	α	10, 1.0, 0.1, 0.01
Infected Active Capture	A	10, 1.0, 0.1, 0.01
Sample Target	M	10

Table S6: Parameter values are shown for Figure 3 in the main text which demonstrates the effect of the death rate and transmission rate, as well as the sample size and capture rate, on the probability of detecting disease. 10^6 surveillance bouts are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	B	0.01,0.05,0.09,0.2,0.6, 1.0,2.0,5.0
Carrying Capacity	K	1;36.0;3.5
Growth Rate	R	0.5;23;2.5
Death Rate	μ	0.25;14.0;1.25
Immigration	N	1.0
Infected Immigration Proportion	Γ	0.01
Primary Transmission Rate	β_0	0.01
Susceptible Active Capture	A	0.5
Infected Active Capture	A	0.5
Sample Target	M	10.0, 20.0

Table S7: Parameter values are shown for Figure 4 in the main text which demonstrates the effect of the transmission, death rate, birth rate, carrying capacity, as well as the sample size, on the probability of detecting disease. 1000 simulations were run per parameter combination with a time limit of 45. If the simulation did not reach the sample target within the time limit, the run is discarded and not used in the statistical calculations. If out of 1000 realisations a parameter combination ceases to reach the sample target at least 15 times, that parameter combination is discarded totally as the results are deemed to be unreliable. Increasing the time limit bears little to no effect on the amount simulations which reach the target sample, so the precise value of the time limit does not affect the results obtained from the model. These parameters were implemented using the Gillespie version of the model.

Appendix S3. Additional scenarios.

This appendix shows results for a set of scenarios complimentary to those in the main text. It is shown that the effects described in the main text are robust to three factors: population size; mode of secondary transmission and sample size.

Population size

The simulations in the main text are based on relatively small populations where fluctuations are driven only by demographic stochasticity. Here we simulate disease dynamics and surveillance in a population driven by environmental stochasticity (see below for details). This enables consideration of fluctuations in a much larger population since demographic fluctuations reduce with population size whereas environmental fluctuations do not. We show that in a population larger by a factor of approximately 10-100 compared with that described in the main text (Fig. 1 and Fig 3.), and using a sample size that is 10 times larger, the effects described are if anything greater. When compared with calculations based on assuming constant prevalence we see that the probability of detecting disease is reduced and estimates of prevalence are both biased and less precise (see Fig. S1 and Fig. S2).

The model used is as described in the main text but here the death rate is subjected to a correlated random walk based on a mean reverting Ornstein-Uhlenbeck process. With finite time step dt this is represented as

$$\mu(t + dt) = \mu(t) + (\mu_0 - \mu(t)) b_\mu dt + \sigma_\mu dB_\mu(t)$$

where $dB_\mu(dt), dB_\mu(2dt), \dots$ are independent identically distributed Gaussian random variables with zero mean and variance dt . The above equation is integrated along with the equations described in Appendix S1. After a burn-in period the equilibrium dynamics of this equation fluctuate around the mean μ_0 . The parameter b_μ controls the correlation in time of $\mu(t)$ and in the long run the variance in $\mu(t)$ is given by $\sigma_\mu^2/2 b_\mu$. The resulting fluctuations in mortality rate represent a range of environmental conditions from harsh to mild which drive fluctuations in the population size. The results shown in Fig. S1 and Fig. S2 are based on this model and the parameter values shown in Table S8. They show qualitatively the same effects seen in Fig. 1 and Fig. 3 in the main text.

Rate Name	Rate	Value
Secondary Transmission Rate	B	1.0, 0.1, 0.04, 0.01
Carrying Capacity	K	6000
Growth Rate	R	1.0
Death Rate	μ_0	0.025;1.0;0.025
Immigration	N	0.1
Infected Immigration Proportion	Γ	0.1
Primary Transmission Rate	β_0	0
Susceptible Active Capture	A	0.001
Infected Active Capture	A	0.001
Sample Target	M	100.0
	b_μ	0.4
	σ_μ	0.5

Table S8: Parameter values are shown for Figures S1 and S2. 10^6 surveillance bouts are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model incorporating the stochastic variation in the death rate described above.

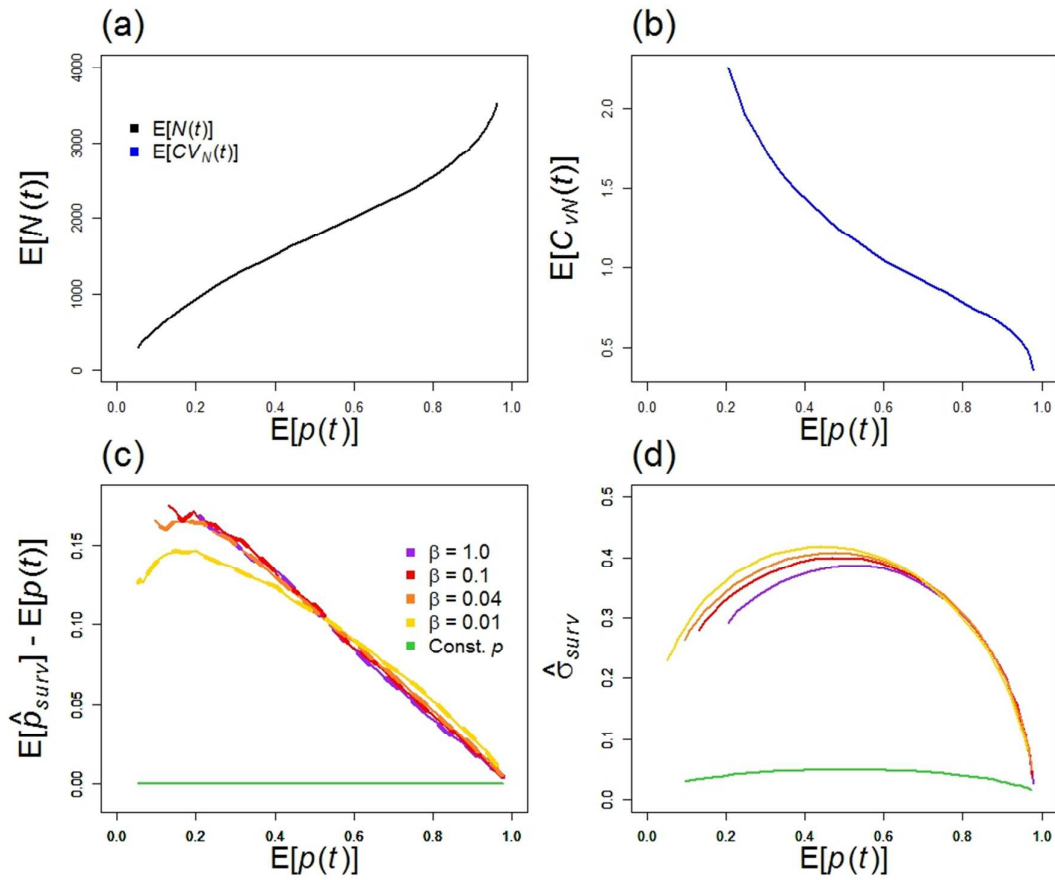


Figure S1: This figure is the counterpart to Fig. 1 in the main text but for the large population simulations with fluctuating death rate described above. The typical population sizes range from around 500-3000 and the sample size used is 100.

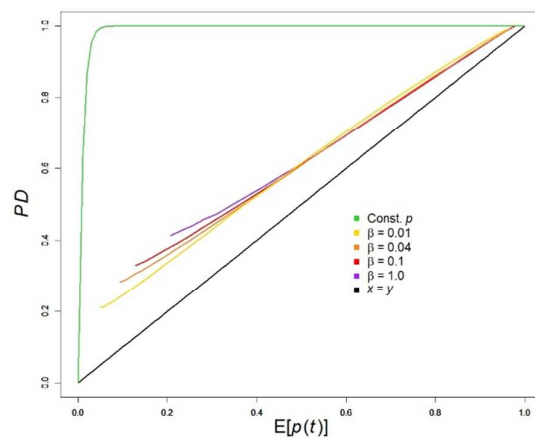


Figure S2 Probability of disease detection. This plot is the counterpart to Fig 3c in the main text but for the large population described above.

Frequency dependent transmission

The scenario simulated here is identical to that shown in Figs 1 and 3 in the main text except that here disease transmission is frequency dependent such that secondary infections occur at rate

$$\tilde{\beta} \frac{S(t) I(t)}{N(t)}$$

Recall that the total population size at time t is $N(t)$ and is made up of $S(t)$ susceptible and $I(t)$ infectives. Contrasting the above formulation with the density dependent transmission rate $\beta S(t) I(t)$ it is clear that to ensure comparable rates of transmission we require $\tilde{\beta} \approx \beta N$. Therefore to ensure comparability between the simulations of frequency and density dependent transmission the contact rate $\tilde{\beta}$ is given by

$$\tilde{\beta} = \beta K \frac{(r - \mu)}{r}$$

where β is the density dependent transmission rate and $K (r - \mu)/r$ is the equilibrium population size derived from the deterministic version of the model.

The results shown in Fig. S3 and Fig S4 show that the effects described in the main text are just as evident in the case of frequency dependent transmission as they are for density dependent transmission.

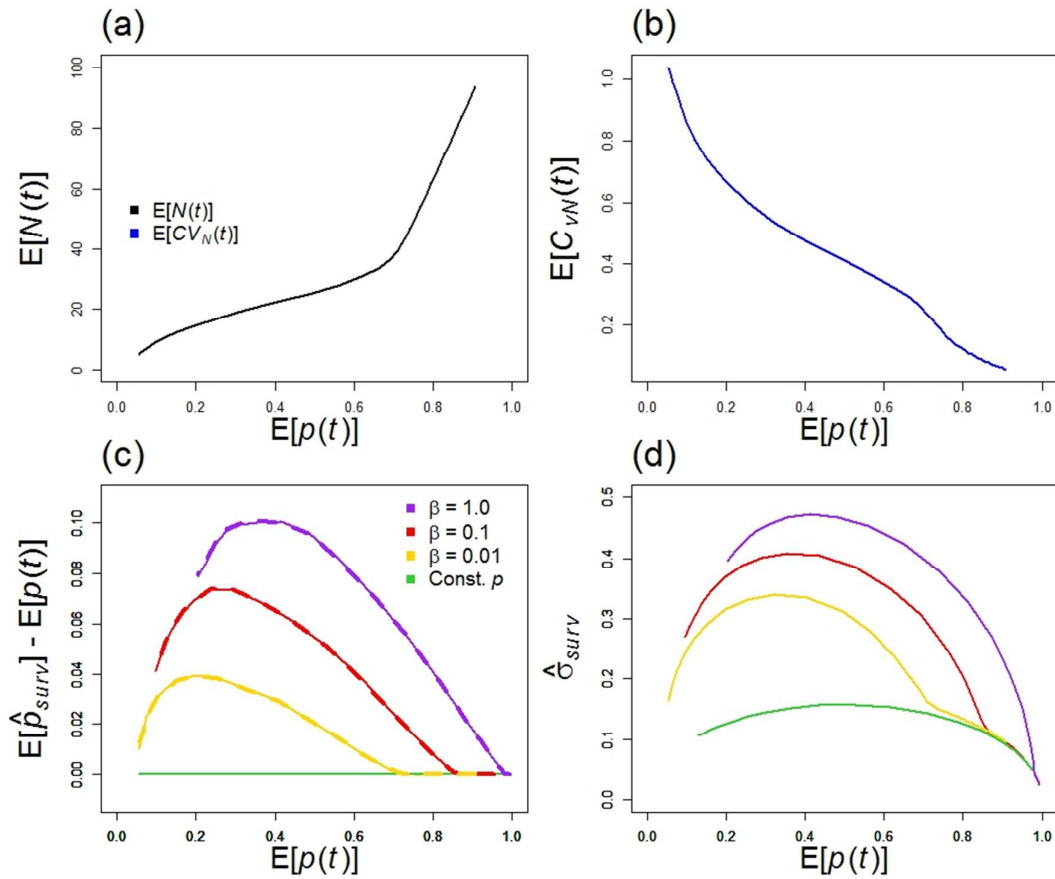


Figure S3: Equivalent to Fig 1 in the main text but for the frequency dependent transmission described above described above.

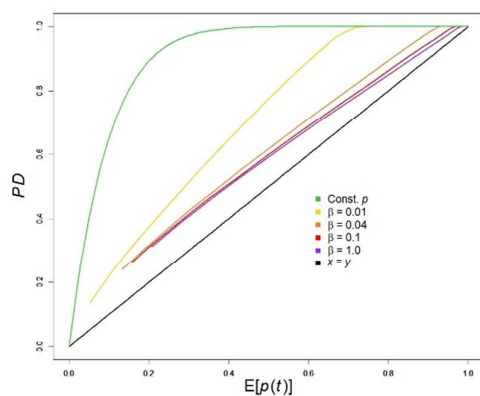


Figure S4 Probability of disease detection. This plot is the counterpart to Fig 3c in the main text but for the frequency dependent transmission described above described above.

Sample size:

Here we show results from a scenario identical to that shown in Figures 1 and 3 of the main text except that the sample size is increased from 10 to 20 and 50. In this scenario the population is typically between 10 and 40 individuals so although these sample sizes may seem low they represent a large fraction of the population. The figures below demonstrate that sample size has little effect on the degradation in the performance of surveillance. Thus these results support the conclusion drawn from Fig. 2 in the main text.

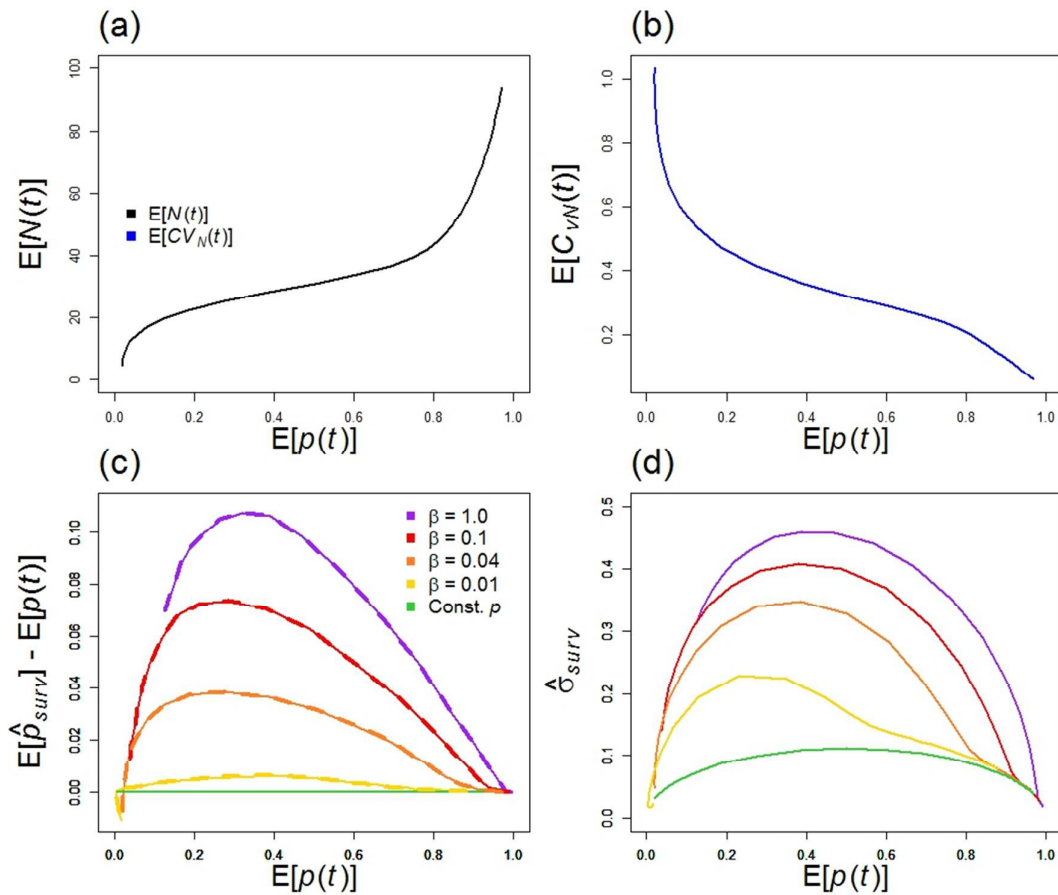


Figure S1: This figure depicts the scenario shown in Figure 1 of the main text but with sample size 20.

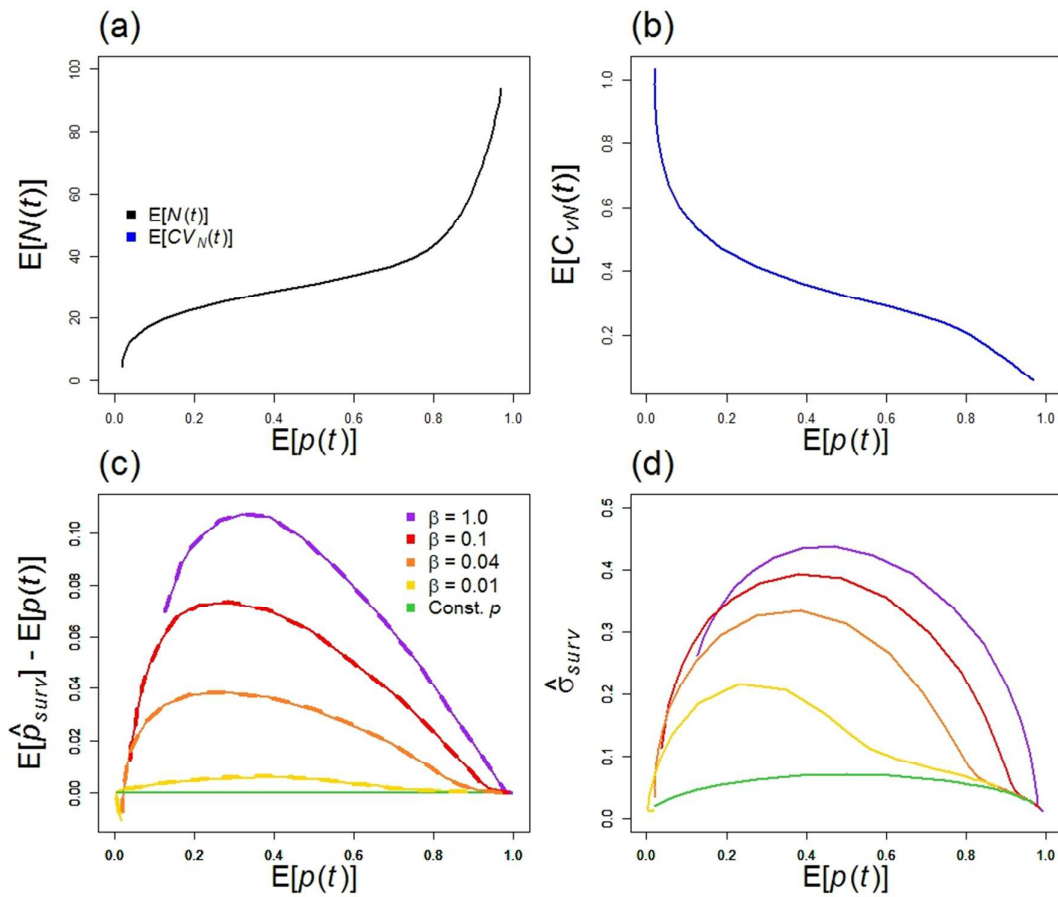


Figure S2: This figure depicts the scenarios shown in Figure 1 of the main text but with sample size 50.

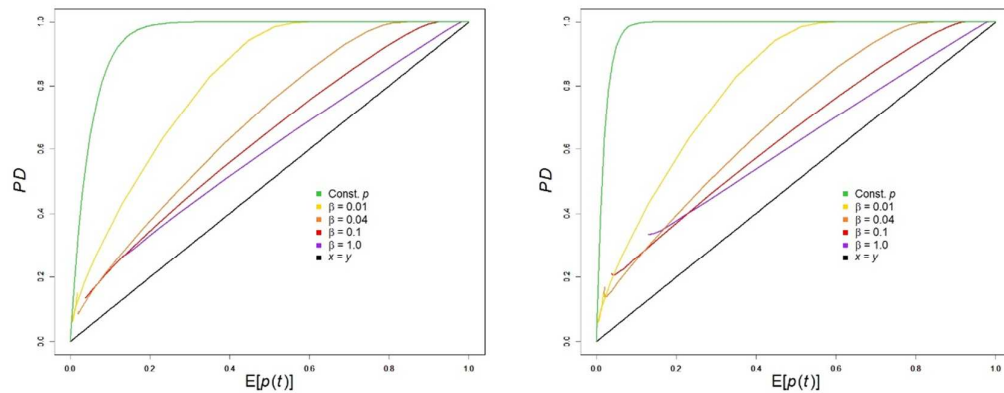


Figure S3: Probability of detection. This plot is the counterpart to Fig 3c in the main text but for increased sample sizes. The plot on the right shows sample size 20 and that the right 50 whereas Fig 3c is based on sample size 10.

Appendix S4. Analysis of disease detection probability

In many cases the primary goal of wildlife disease surveillance is detection of disease rather than quantification of prevalence. This is true, for example, for emerging or re-emerging disease, where detection is a precursor to further action, which would include heightened surveillance. If prevalence is assumed constant and equal to the long term average prevalence $E[p]$ of the wildlife disease system, then the probability that disease is detected in a sample of size m is given by:

$$PD^{Bin} = f(E[p], m) = 1 - (1 - E[p])^m$$

This formula, based on simple binomial arguments, and variants that also assume constant prevalence, are the standard basis for sample size calculations (see e.g. Fosgate 2009). However, if prevalence fluctuates PD^{Bin} is a misleading estimate of the probability of detection.

In real systems, prevalence varies with time; therefore, when conducting surveillance, the prevalence values will vary at the times when each of the m samples are collected. Nonetheless, for simplicity here we assume that the prevalence during a given surveillance bout (i.e. the collection of m consecutive samples) is constant, and denoted p . Fig. 3a (see main text) compares the probability of detection measured from simulations with two approximations. The first approximation accounts for fluctuations both within and between surveillance bouts and the second only that between surveillance bouts. These results indicate that accounting only for fluctuations between surveillance bouts is an accurate approximation. Therefore, the expected probability of detection for sample size m is defined as

$$PD = E[f(p, m)] = E[1 - (1 - p)^m]$$

where the expectation is over the between bout prevalence distribution $P(p)$ which accounts only for prevalence fluctuations between surveillance bouts. For a single sample $m = 1$, the above equation for PD reduces to a linear form, so that $PD = PD^{Bin} = E[p]$. However, if $m > 1$, then the equation for PD is non-linear, and therefore $PD \neq PD^{Bin}$.

To illustrate this, we Taylor expanded PD by assuming that the difference between the bout prevalence (p) and the long term average prevalence is small i.e. $p = E[p] + \Delta p$. Then, noting that $E[\Delta p] = 0$ and $var[p] = E[\Delta p^2]$ and ignoring terms containing higher powers of Δp , this yields

$$PD \cong PD^{Bin} + \frac{1}{2}var[p] \left. \frac{\partial^2 f(p, m)}{\partial p^2} \right|_{p=E[p]}$$

This suggests (to leading order in the expansion) that the true probability of detection will be lower than PD^{Bin} , since the second derivative $\partial^2 f(p, m)/\partial p^2 = -m(m-1)(1-p)^{m-2}$ is negative for sample size $m > 1$ and $p = E[p]$. In addition, the size of this deviation depends on the sample size and the variance in prevalence. Although these conclusions are broadly correct, when compared with simulation results, the above Taylor expansion does not provide an accurate approximation of the probability of detection. However, analytic progress can be made, with the following alternative approach. The approximation $(1-p)^m \approx e^{-pm}$ holds for m large (and is already accurate even for $m = 10$) and enables us to write the probability of detection as:

$$PD = 1 - E_p[(1-p)^m] \cong 1 - E[e^{-pm}] = 1 - M_p(m)$$

where $M_p(m) \equiv E[e^{-pm}]$ is the moment generating function associated with the between bout prevalence distribution $P(p)$. This suggests that if we could parameterise a suitable distribution to approximate $P(p)$ then we could use the corresponding moment generating function to calculate the probability of detection.

Fig. 3a (main text) suggests that a moment-generating function approximation (see last equation above) based on the actual distribution of prevalence between surveillance bouts would be an accurate approximation. Fig. 3b illustrates this approximation using an assumed gamma distribution, parameterised with the mean and variance of $P(p)$. Although the gamma approximation is not completely successful, it does provide a more accurate prediction of PD than PD^{Bin} . This could be used to improve sample size calculations in situations where simulation is not possible, but information about prevalence fluctuations is available. Moreover, the results of Fig. 3a show that such approximations could be improved by assuming a more accurate representation of the prevalence distribution $P(p)$.