Detecting superspreaders in wildlife reservoirs of disease

Evandro Konzen[1,2], Richard J. Delahay[3,4], Dave J. Hodgson[5], Robbie A. McDonald[3,5], Ellen Brooks Pollock[6], Simon E. F. Spencer[2], and Trevelyan J. McKinley*[1]

[1]University of Exeter Medical School, University of Exeter, Exeter, Devon, UK

[2]Department of Statistics, University of Warwick, Coventry, UK

[3]Environment and Sustainability Institute, University of Exeter, Penryn, Cornwall, UK

[4]National Wildlife Management Centre, Animal and Plant Health Agency, Sand Hutton, York, UK

[5]Centre for Ecology and Conservation, University of Exeter, Penryn, Cornwall, UK

[6]Bristol Medical School, University of Bristol, Bristol, UK

Abstract

Wild animals commonly harbour infectious diseases with risk of spillover to humans and livestock. Bovine tuberculosis (bTB) is one such disease, posing significant socio-economic and welfare threats to the cattle industry worldwide. Identifying superspreaders, those individuals most responsible for onward transmission of infection, is critical for disease management. In practice, superspreaders are hard to identify because monitoring relies on imperfect surveillance and imperfect diagnostic tests, hence key epidemiological events, including transmission, are only partially observed. To infer the hidden dynamics of disease spread in wildlife, we fitted an individual-level stochastic spatial meta-population model of bTB transmission to data from a longitudinal study of the European badger (*Meles meles*). We develop a novel estimator for the individual effective reproduction number, providing quantitative evidence for the presence of superspreader badgers, despite the population-level effective reproduction number being less than one. The efficiency of bTB control in badgers could be substantially increased by targeting interventions at the relatively small proportion of individuals responsible for most onward transmission. Our modelling framework provides a flexible, efficient and generalisable means of fitting state-space models to individual-level data, to identify high-risk individuals and explore important epidemiological questions about bTB and other diseases of wildlife, livestock and humans.

Keywords: Bovine tuberculosis; Bayesian modelling; badger ecology; individual forwards-filtering backwards-sampling; stochastic compartmental epidemic models.

*Corresponding author: t.mckinley@exeter.ac.uk

Despite extensive research efforts to evaluate risks of bTB transmission between cattle and badgers [e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9], there remains considerable debate about the relative impacts of within- and between-species spread, and the efficacy of control measures such as badger culling. In both badger and cattle populations the population-level reproduction number has been inferred to be close to one in both species [1, 6] implying endemic persistence of infection. However, focusing on population-level epidemiological dynamics ignores the influence of individual variation in transmission potential. 'Superspreading', where a small proportion of individuals contribute disproportionately to the number of secondary cases [10], is frequently observed for human infectious diseases but is rarely characterised in wildlife populations. Superspreading can facilitate explosive early spread and transient epidemics, and maintain the infection burden even in the absence of widespread transmission. Identifying spatial locations and/or individuals that represent a high transmission risk should inform targeted disease management strategies [e.g. 11]. For example, if individuals with a high probability of being infected could be removed selectively, then any unintended consequences of population-level disruption (caused for example by non-selective culling, a process which has been linked to enhanced movement and disease spread in culled badger populations [12]) might be ameliorated.

At Woodchester Park (WP) in Gloucestershire, south-west England, a population of wild badgers has been studied since 1978, providing a unique long-term empirical data set that has provided unprecedented insights into badger ecology, bTB infection in badgers and transmission to cattle [see e.g. 13, 14, 15, 16, 17, amongst others]. Nonetheless, conducting statistical inference on infectious disease models in this system is challenging because of the high dimensionality of the data coupled with a large number of hidden processes, requiring large computational resources to overcome. In this paper we utilise a recent advance in Bayesian modelling known as the individual forward filtering backward sampling (iFFBS) algorithm [18] to efficiently fit a discrete-time Susceptible-Exposed-Infectious-Dead ($SEID$) compartmental model of disease transmission to individual-level data from 2,391 badgers captured during the WP study between 1980–2020. The model employs a meta-population structure encompassing 34 social-groups.

A key challenge with wildlife systems is that disease surveillance relies on imperfect sampling. Testing relies on the capture of live animals or dead animals recovered from the field. Even with more intensive sampling approaches, such as capture-mark-recapture (CMR), which is employed quarterly in WP, not all badgers are captured at each time point and capture rates vary seasonally. Furthermore, when most animals die their carcasses are not recovered, leading to considerable missing data and uncertainty regarding the number of badgers alive at any given time point. In an endemic disease system such as this, transmission dynamics are impacted by the mortality process, which for a wild mammal population must incorporate age-dependent mortality risks [see e.g. 16]. At WP, captured badgers are tested for bTB using a suite of different diagnostic tests: for the detection of antibodies (Brock, Dual Path Platform [DPP] and StatPak), a cellular immune response (the γ-interferon [γ-IFN] test), or the bacteria itself (microbiological culture). Diagnostic test performance (i.e. sensitivity and specificity) vary with the type of test, and different tests were in use at various times during the period of study, with some being phased out (e.g. Brock/StatPak) and replaced by others (e.g. γ-IFN, DPP)—see Fig. 1. There was also a suspected temporal change in the performance of one test.

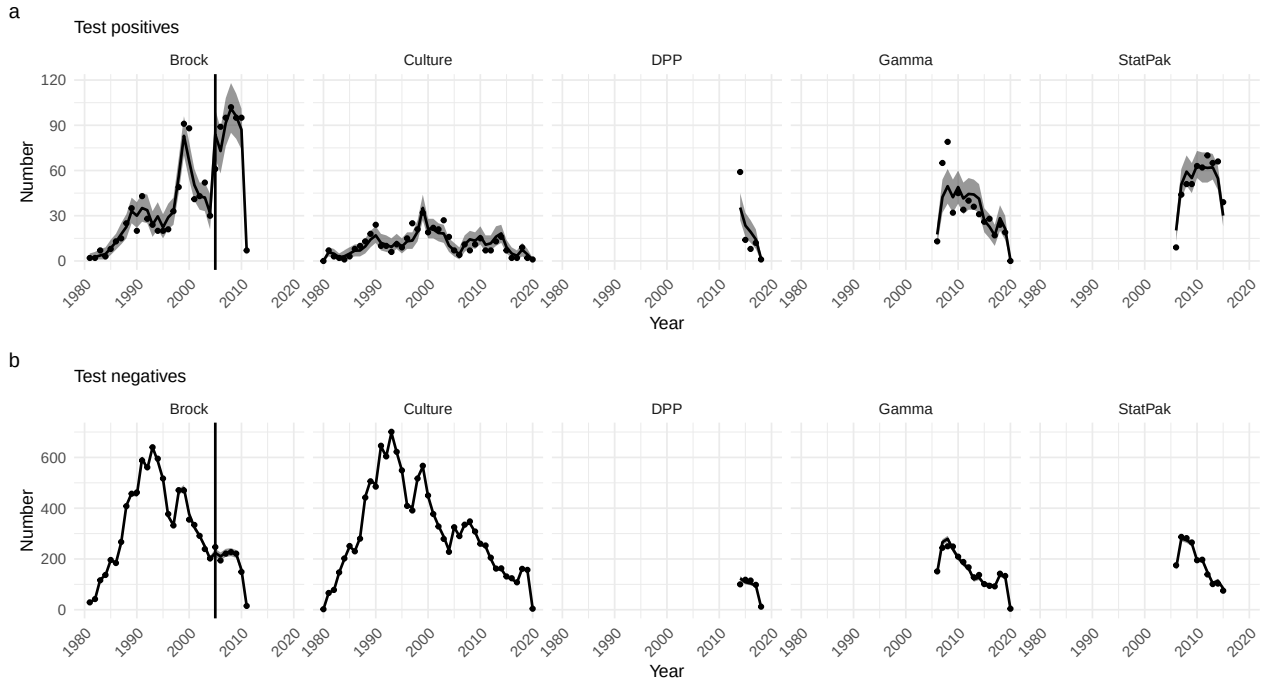


Figure 1: Posterior predictive distribution for the number of (a) test-positive and (b) test-negative results in each year against the observed data. Posterior means are indicated by black lines and 95% credible intervals by grey ribbons. Data are represented by black points. The vertical black line in the Brock panel corresponds to the posterior predictive mean of the changepoint for this test.

To deal with these challenges we developed a framework that simultaneously captures spatio-temporal disease dynamics, the CMR and mortality processes, and diagnostic test performance at the individual-level. The iFFBS algorithm provided a highly efficient and flexible inference methodology, enabling us to conduct Bayesian inference on all parameters and individual-level hidden states in a matter of a few hours on a desktop machine. For full details of the model, prior distributions and inference algorithm, see the Methods and Supplementary Notes.

Comparison of posterior predictive outputs against the observed data shows that the fitted model successfully captures the dynamics of test-positive and test-negative animals across the whole population for all test types over the complete 40+ year time period (Fig. 1). Analogous plots produced for each social group (Supplementary Notes S7.2) show similarly good fits, and we can therefore use the model to explore the key drivers of disease transmission in Woodchester Park.

Common measures for understanding the propensity of a disease to spread within a population are the *basic* and *effective* reproduction numbers [19, 20, 21, 22]. However, it has been widely discussed that population-level reproduction numbers ignore the propensity for some individuals to make a disproportionately large contribution to the spread of infection [10, 23, 24, 25, 26, 27, 28]. To explore this, we estimate individual effective reproduction numbers, R_i, from which a population-level effective reproduction number, R, can be derived [20, 21, 22] (for details, see Methods and Supplementary Notes S5). We define R_i to be the *expected number of secondary infections caused by an infected individual i across its lifetime.* Here we choose to calculate R_i only for animals that have died, and so each estimate adjusts for the latent and infectivity processes across the lifetime of each

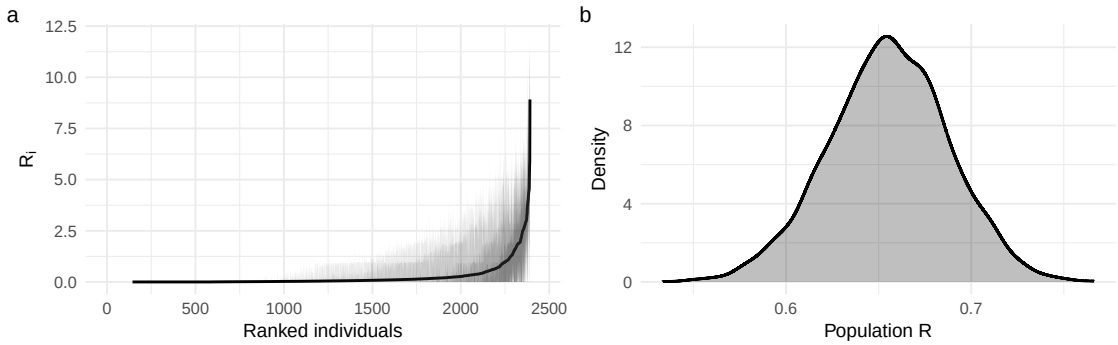


Figure 2: (a) Posterior means and credible intervals (50% and 95%) for the individual reproduction number for all individuals, ranked in increasing order of posterior means, (b) posterior distribution for the population effective reproduction number.

animal, and automatically accounts for the social-group specific background infection rates and changes in group sizes over time (and also any movements of badgers between social groups). Rather than being limited to the onset of clinical signs [e.g. 20], we integrate directly over the time-since-infection for each individual. Fig. 2a reveals substantial heterogeneity between individuals: some badgers have $R_i \approx 8$, whereas most have $R_i < 1$. This infers the presence of superspreader badgers, a result that emerges from the stochastic model after fitting it to the observed data, rather than being formally incorporated [10]. This highlights the importance of accounting for stochastic dynamics when modelling disease outbreaks in small- to medium-scale populations. Heterogeneity in the duration of the infectious period due to variations in infection time and survival seems to be the main (although not sole) driver of these differences (Fig. 3). Given that R_i values do not associate strongly with the average number of social groups that a badger has been associated with, or with average group sizes (Fig. 3), it is most likely that on average badgers infected at a younger age simply have more opportunity for onwards transmission. Since each badger has a different posterior probability of being infected, we also compare the posterior mean probability of infection against the posterior mean R_i values across individuals and show that superspreading is not just a function of the probability of infection (Extended Figure 7).

Despite some animals having high reproduction numbers, the population effective R [20] is on average < 1 (Fig. 2b) with a posterior mean estimate of 0.66 (95% credible interval [CI]: 0.59–0.72), implying an expected long-term decline in prevalence in the absence of additional sources of infection. This raises interesting questions around the mechanisms contributing to the endemicity of bTB in this population. It is possible that periodic reintroductions from external sources and/or persistent environmental contamination, such as shedding, serve to maintain infection in the population—captured through the background rate-of-infection terms. Our model is based on a subset of badger social groups, and it is likely that reintroduction of infection into this meta-population could occur through interactions with unmodelled neighbouring social groups in the wider landscape. With the currently available data we are not able to disentangle these mechanisms, but it is notable that since the population-level effective R is estimated to be < 1 with a high posterior probability, it seems likely that a suitable management strategy such as vaccination or selective removal of superspreaders would be an efficient means of driving infection rates down, since most transmission events are likely to occur by direct spread

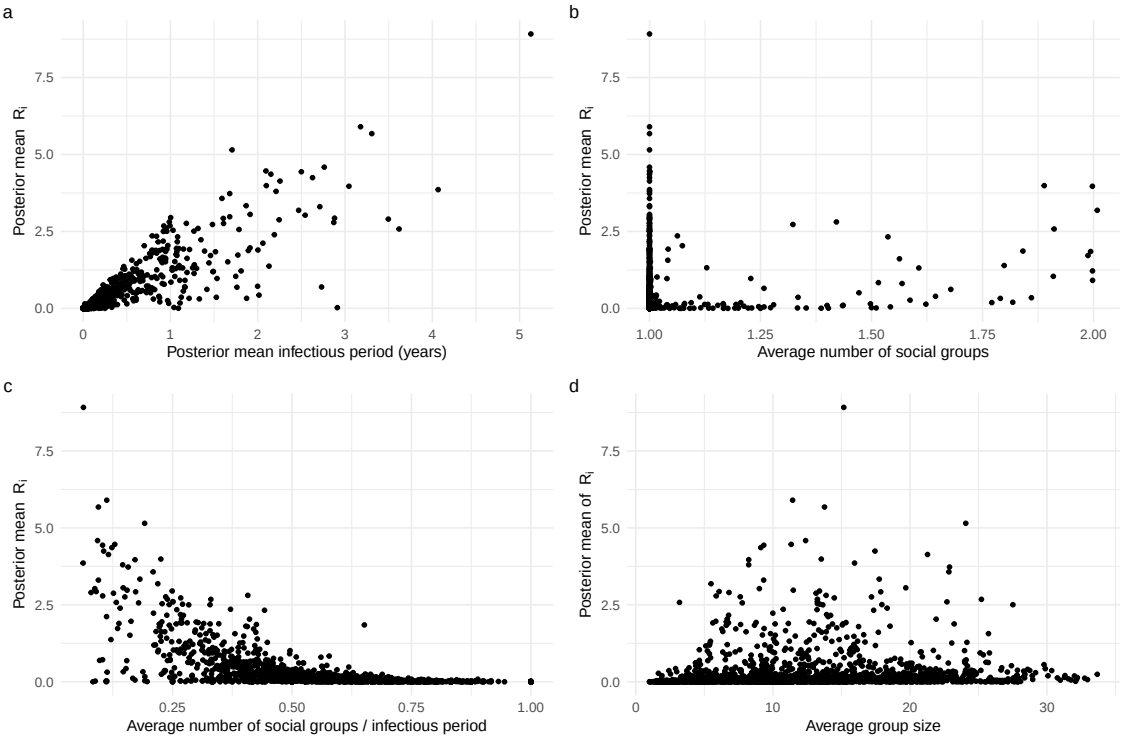


Figure 3: Posterior mean estimates of individual reproduction numbers relative to (a) posterior mean duration of the infectious period, (b) average number of social groups an individual belonged to, (c) average number of social groups an individual belonged to divided by the duration of the infectious period, and (d) the average size of social groups that the individual belonged to.

amongst badgers (Fig. 4b), and predominantly from superspreading individuals (Fig. 2a). A key question for further research is whether reintroduction of infection from external sources could be sufficient to maintain infection in the face of this type of management strategy within this study area.

Other modelling studies have assumed or estimated slightly higher *basic* reproduction numbers (R_0) for bTB in cattle and wildlife populations of just above 1 [e.g. 29, 30, 31, 6], though estimates are often made using simpler population-level (often deterministic) models with constant mortality assumptions. A recent paper [9] fitted a model that estimated the within- and between-species reproduction numbers for cattle and badgers from data in Kilkenny, Ireland, using an explicit spatial connectivity structure between badger social groups and cattle farms with an environmental decay term to capture persistence of the pathogen in the environment (see also [32]). Their analysis suggested that environmental transmission can play a major role in disease transmission, and determined what relative badger-to-cattle densities and badger vaccination coverage can cause the basic reproduction number to be above or below one in different cattle herds. Nevertheless, these previous estimates are not directly comparable to our population-level effective R, since R_0 characterises the expected number of secondary infections per primary infection introduced to a *fully susceptible population*. In contrast, our population-level R characterises the posterior expected number of secondary infections per primary infection occurring within our study population, and hence will be lower than R_0 due to factors such as varying numbers of susceptible badgers over time, the meta-population structure or changes in social group sizes. Furthermore, our estimate also adjusts for the background risk-of-infection (which if ignored will assign

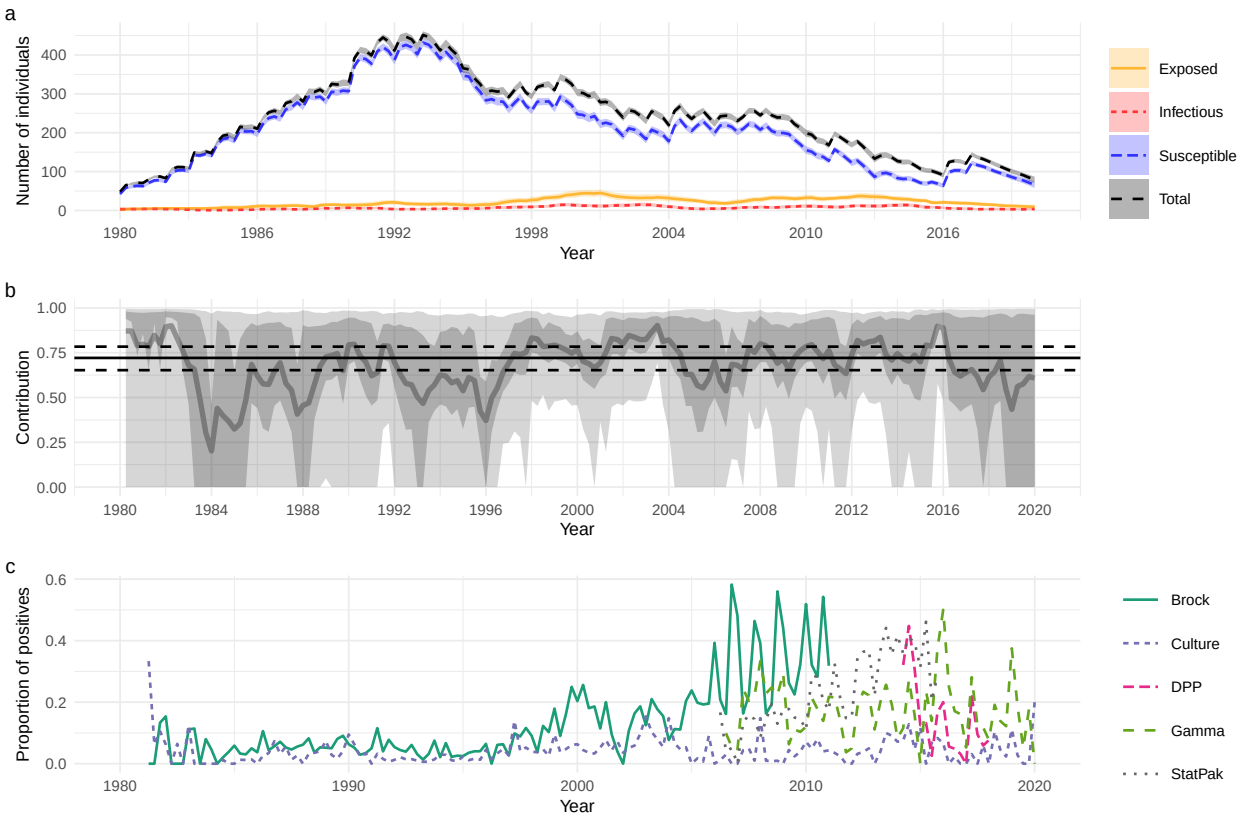


Figure 4: (a) Posterior predictive means and 95% prediction intervals for the number of individuals in each epidemiological state over time, (b) posterior predictive means and credible intervals (50% and 95%) for the relative contribution of badger-to-badger transmission vs. background rates-of-infection on new infections over time (the solid black line is the average contribution over all time points, with the dashed lines giving the corresponding 95% credible interval), (c) the observed proportion of positive test results over time for each test.

higher weight to badger-to-badger interactions) and age-dependent mortality. This study adds to the literature suggesting that the population level reproduction number for badgers is likely to be low [e.g. 1, 6].

Furthermore, our model can produce estimates for the probability of infection/infectivity of any badger at any given time, and make predictions for the corresponding event time (e.g. Extended Figure 4). These predictions account not only for the current diagnostic test results for an individual badger, but also for any previous test information, and data from other badgers. These estimates can be aggregated up to higher spatial levels, such as the social group level, or population. For example Fig. 4a shows the predicted number of individuals in each epidemiological state across the whole population over the study period, giving an estimate of the overall hidden burden of infection [33], with analogous plots for each social group shown in Extended Figures 5 and 6. These may be a useful tool to guide social-group or even sett-specific surveillance efforts. Here the model predicts relatively low levels of underlying infection in the population as a whole over the study period. However, there is a steep decline in population size from about 1993 onwards, with the proportion of infected badgers increasing towards the end of the study period.

Our model also predicts that on average 72% (95% CI: 65%–78%) of new infections can be attributed to direct badger-to-badger transmission, with the rest attributed to the background rate-of-infection terms (Fig. 4b). We note that in addition to transmission arising from other species (e.g. cattle), the background infection risk can

also encompass indirect transmission from badgers, such as from environmental contamination, and hence the 72% estimate represents a lower bound for the relative rates of *all* badger-to-badger transmission in WP. Future work will explore ways in which some of these competing processes could be disentangled.

We obtain full posterior distributions for all epidemiological, CMR, mortality and diagnostic test parameters (Table 1, Extended Figures 1, 2 and Extended Table 1). We obtain an estimate of the average (discrete-time) latent period (the time between infection and infectiousness) of 3.7 (95% CI: 2.9–4.7) years. Historic models of bTB in badgers produced estimates for the latent period of around 1–1.5 years [34], but these were estimated from population-level deterministic models with constant mortality rates. Other approaches estimated shorter latent periods of between 95–158 days, but these were estimated in a laboratory setting [35]. In cattle, estimates vary more widely; typical assumptions are between 6–20 months [see e.g. 33], but some estimates are much longer ($>$ 11 years) [32].

We note that our estimate of the average latent period is similar to the estimated median survival time of badger hosts (Extended Figure 3). These mortality curves are again in line with earlier estimates [16], with a median survival time of 3.6 (95% CI: 3.5–3.8) years (though slightly longer than earlier estimates for expected life-expectancy of between 2.2–2.9 years assuming constant mortality rates [36]). It should be noted that our survival curves are conditional on *at least one capture event*, as we were unable to observe early-life mortality events that occur in underground burrows (setts). Taken together, these estimates suggest that there is a high probability that badgers can become infected but then die before they become infectious, which introduces natural heterogeneity in the degree to which individual badgers can contribute to the spread of bTB.

We also obtain estimates for the infectious period distribution, which is dependent on the age-dependent mortality curves, as well as the time-of-infection. To understand the transmission potential of individual badgers, we define an *effective infectious period* distribution as the time between when an animal becomes *infected* (i.e. when they enter the E class) and when they die. We can define the *infectious period* as the time between when an animal becomes *infectious* (i.e. when they enter the I class) and when they die (in this model the latter distribution is equivalent to the conditional survival distribution given survival beyond infectiousness). Since we simulate the time-of-infectiousness and the time-of-death for each animal, we can generate empirical estimates for these distributions for each animal that average over the unknown event times. We present summaries of the effective infectious period distribution for individual badgers in Fig. 5a. These highlight that a large proportion of badgers die before becoming infectious, and so have an effective infectious period of zero. This is shown even more clearly if we average over the population (Fig. 5b). Considering the infectious period distribution conditional on progression to infectiousness (i.e. being in the I state), then we obtain a posterior mean of 1.4 (95% CI: 0.25–4) years across the population (Fig. 5c). In comparison, [9] assume a mean infectious period of 1 year, with a range between 0.32–3.6 years (derived from [35, 36]). In cattle accurate estimation of the latent and infectious periods is challenging due to censoring and selection induced by the routine slaughter of test-positives, but these challenges should be mitigated in this study since no artificial control measures were implemented in this population over the study period.

Transmission also depends on a parameter q, which mediates the extent to which contact rates are governed

Type	Parameter	Mean	95% CI
Transmission	β	0.019	(0.009, 0.037)
	q	0.59	(0.23, 0.91)
	τ (years)	3.8	(3, 4.8)
Mortality	a	0.036	(0.032, 0.039)
	b	0.033	(0.029, 0.037)
	c	0.0016	(0.000034, 0.0056)
Sensitivity (E)	Brock1	0.56	(0.49, 0.63)
	Brock2	0.73	(0.65, 0.8)
	Culture	0.073	(0.05, 0.098)
	DPP	0.41	(0.31, 0.53)
	Gamma	0.47	(0.41, 0.52)
	StatPak	0.82	(0.75, 0.88)
Sensitivity (I)	Brock1	0.91	(0.84, 0.96)
	Brock2	0.93	(0.83, 0.99)
	Culture	0.8	(0.73, 0.86)
	DPP	0.92	(0.76, 1)
	Gamma	0.85	(0.76, 0.93)
	StatPak	0.95	(0.89, 0.99)
Specificity	Brock1	0.98	(0.98, 0.99)
	Brock2	0.78	(0.76, 0.8)
	Culture	0.99	(0.99, 1)
	DPP	0.94	(0.92, 0.97)
	Gamma	0.92	(0.91, 0.93)
	StatPak	0.94	(0.93, 0.96)
Capture Probability	Winter	0.18	(0.17, 0.18)
	Spring	0.36	(0.35, 0.37)
	Summer	0.46	(0.45, 0.47)
	Autumn	0.28	(0.27, 0.28)
Changepoint	ξ	2005	(2005, 2005)

Table 1: Posterior means and 95% credible intervals for key model parameters (to 2 significant figures [4 s.f. for changepoint]). For full details of the model structures and parameters, see Supplementary Notes S2.

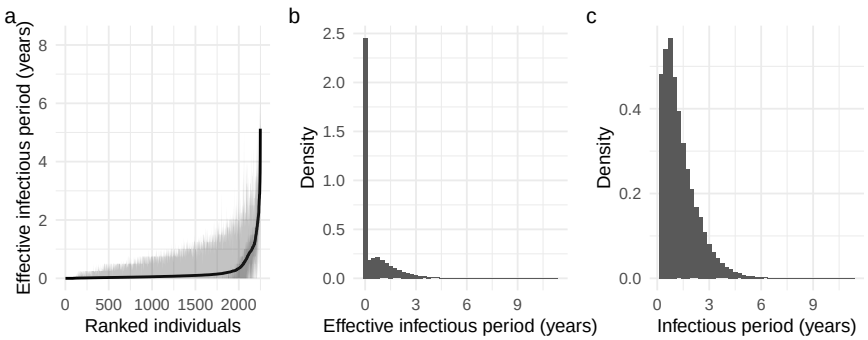


Figure 5: (a) Posterior means and credible intervals (50% and 95%) for the individual effective infectious period distribution across all individuals, ranked in increasing order of posterior means; (b) posterior distribution for the population-averaged *effective* infectious period distribution; (c) posterior distribution for the population-averaged infectious period distribution.

by badger density [37]. The posterior mean for q is 0.59 (95% CI: 0.23–0.91), which provides evidence consistent with the transmission process being somewhere in-between frequency- and density-dependent (which are defined when $q = 1$ and $q = 0$ respectively) [37, 33]. We note that our choice of $U(0, 1)$ prior distribution (following [33]) stops q from being exactly 0 or 1, but the area of high posterior density is nevertheless away from either of these two extremes (Extended Figure 1). Frequency-dependence is a common assumption for badger bTB models [e.g. 38, 39, 9], and our estimated q is not dissimilar from some earlier cattle studies [33]. This result therefore suggests that contact rates will increase with badger density at low densities, and then begin to saturate at higher densities [37]. This result also has implications for disease control strategies, because under a frequency-dependent transmission structure population-reduction is less viable as a means of disease eradication [40].

Our approach also allows us to estimate diagnostic test sensitivities and specificities relative to infection, rather than to a proxy measure such as visible lesions [see also 15, 33]. In the first version of our model we assumed that each of the five tests had constant sensitivity and specificity values over time. We obtained posterior mean estimates for the sensitivities (for infectious animals) of: Brock 0.88 (95% CI: 0.84–0.91), Culture 0.32 (0.29–0.35), StatPak 0.93 (0.89–0.96), γ-IFN 0.57 (0.53–0.62) and DPP 0.63 (0.53–0.73). We note that the sensitivity of the γ-IFN test was much lower than expected from cattle studies [41]. Previous estimates in badgers are more variable: [15] estimated a γ-IFN sensitivity of about 0.44, whereas [13] estimated a value of 0.799; although the former did not differentiate between latently infected and infectious badgers (which may drive the estimate to be lower if sensitivity is lower during the latent period), and the latter used a different method but had a more informative prior distribution with a median sensitivity of 0.796. Model fits and main parameter estimates for this initial model are shown in the Supplementary Notes S7.3.

This original low sensitivity estimate for the γ-IFN test led us to explore other possible explanations. Plotting the proportion of test positive results over time for the five tests, we see that from about 2005 onwards the proportion of test-positives for the Brock test increases rapidly, whilst the other tests (notably culture) drop or remain constant (Fig. 4c). The exception is the StatPak (an antibody test like the Brock), and similar inconsistencies with the other tests have previously been noted [42]. Consequently, we developed our current model, where the Brock test sensitivity and specificity were permitted to change at some specific changepoint (to be estimated). Following this modification the posterior distribution for the changepoint is contained between the final quarter of 2004 and the first two quarters of 2005 (Table 1), which is consistent with the data in Fig. 4c. This has several consequences. Firstly, the pre- and post-changepoint Brock test sensitivities (denoted Brock1 and Brock2 respectively) are both high (0.91 and 0.93), but the specificity of the Brock post-changepoint drops to 0.78 (from 0.98 pre-changepoint). The inclusion of the changepoint also results in the γ-IFN sensitivity estimate rising to about 0.85, which is more consistent with earlier literature (except [15] as discussed above). This suggests some change in the Brock test itself or how it was delivered after 2005, and this is an ongoing area of investigation. The clear reduction in Brock test specificity in later years also explains another phenomenon in the observed data, which is that various individual badgers consistently tested positive by the Brock test, but negative by the other tests over time. These findings are also consistent with [42], which observed a marked increase in the frequency of positive Brock test results relative to the other tests.

Estimates of the hidden states change dramatically when the change in Brock sensitivity and specificity is included, with the original model showing much higher levels of underlying infection than the changepoint model (see Supplementary Notes S7.3, Fig. S4), with higher rates of badger-to-badger transmission. Nevertheless, we note that both models suggest the presence of superspreader badgers with a population-level $R < 1$. However, the changepoint model is preferred here, for several reasons: the model fits were better; more credible estimates for the other test sensitivities and specificities were obtained; and the model acknowledges previously highlighted concerns with the Brock test specificity in later years.

The iFFBS algorithm is a flexible and efficient methodology for fitting infectious disease models to individual-level data that could be extended to other disease systems. Nevertheless, there are various areas that could usefully be developed further. For example, exploring more nuanced infectivity-profile dependent sensitivity and specificity processes, rather than the coarse step-function that we currently employ could be insightful [43]. It may also be valuable to include individual-level covariate dependencies in different components of the model, for example, sex-specific differences in mortality [16]. Furthermore, since we have an explicit transmission model, we can also directly investigate the effect of infection on mortality risk [29, 16], or sex-/age-specific differences in transmission or susceptibility. Furthermore, with suitable data it would be of interest to integrate different choices of spatial connectivity structure and/or environmental transmission, or to deal with multiple species [32, 9]. We also propose further exploration of the characteristics of superspreader badgers, which may inform operational targeting. The insights described above have been gleaned from a single plausible model, but a key area of future work is to extend this framework to allow for systematic model comparison and/or model averaging, which could provide even more robust predictions and help disentangle complex model interactions [e.g. 44, 45].

To summarise, we have developed an individual-level stochastic meta-population compartmental model of bTB in badgers, and fitted this to a dataset derived from an intensive long-term capture-mark-recapture study using an individual forward filtering backward sampling algorithm. The model fitted in a few hours, enabling us to conduct efficient Bayesian inference simultaneously on all epidemiological parameters, as well as those relating to age-dependent mortality, the CMR surveillance process and imperfect diagnostic testing.

The model provides novel estimates of individual effective reproduction numbers, suggesting that a small proportion of infected badgers, or superspreaders, contribute disproportionately to the transmission potential of the disease. This is despite the population-level effective reproduction number being low, suggesting that management strategies could be improved by targeting these animals or associated social groups with interventions or more intensive surveillance. Given systematic surveillance data, the model produces predictions for the hidden epidemiological states of individual badgers over time, which can be aggregated to different spatio-temporal resolutions. This could be used to help target surveillance or management strategies since we can estimate the likelihood of individuals being infected or infectious, and the expected infectious period given infection, that accounts for the age-dependent mortality process as well as the time-of-infection. These estimates are not just based on the diagnostic test history for individual animals, but also borrow weight from other animals through the transmission model structure. Detailed predictive information like this is only feasible in a population with

systematic surveillance, such as in Woodchester Park. The iFFBS algorithm has demonstrated its value in modelling such data efficiently, and could be employed in other settings, for example for modelling the within-herd spread of bTB in livestock.

References

[1] David R. Cox, Christl A. Donnelly, F. John Bourne, George Gettinby, John P. McInerney, W. Ivan Morrison, and Rosie Woodroffe. Simple model for tuberculosis in cattle and badgers. *Proceedings of the National Academy of Sciences of the United States of America*, 102(49):17588–17593, 2005.

[2] John Bourne, Christl Donnelly, David Cox, George Gettinby, John McInerney, Ivan Morrison, and Rosie Woodroffe. Bovine TB: the scientific evidence. Technical report, Independent Scientific Group on Cattle TB, 2007.

[3] Christl A. Donnelly, Gao Wei, W. Thomas Johnston, David R. Cox, Rosie Woodroffe, F. John Bourne, C. L. Cheesman, Richard S. Clifton-Hadley, George Gettinby, Peter Gilks, Helen E. Jenkins, Andrea M. Le Fevre, John P. McInerney, and W. Ivan Morrison. Impacts of widespread badger culling on cattle tuberculosis: concluding analyses from a large-scale field trial. *International Journal of Infectious Diseases*, 11(4):300–308, 2007.

[4] Joanne L. Hardstaff, Glenn Marion, Michael R. Hutchings, and Piran C.L. White. Evaluating the tuberculosis hazard posed to cattle from wildlife across Europe. *Research in Veterinary Science*, 97:S86–S93, 2014.

[5] Christl A. Donnelly and Pierre Nouvellet. The contribution of badgers to confirmed tuberculosis in cattle in high-incidence areas in England. *PLoS Currents*, 2013.

[6] Ellen Brooks-Pollock and James LN Wood. Eliminating bovine tuberculosis in cattle and badgers: insight from a dynamic model. *Proceedings of the Royal Society B: Biological Sciences*, 282(1808):20150374, 2015.

[7] Roman Biek, Anthony O'Hare, David Wright, Tom Mallon, Carl McCormick, Richard J Orton, Stanley McDowell, Hannah Trewby, Robin A Skuce, and Rowland R Kao. Whole genome sequencing reveals local transmission patterns of *Mycobacterium bovis* in sympatric cattle and badger populations. *PLoS Pathogens*, 8(11):e1003008, 2012.

[8] Gianluigi Rossi, Joseph Crispell, Daniel Balaz, Samantha J Lycett, Clare H Benton, Richard J Delahay, and Rowland R Kao. Identifying likely transmissions in *Mycobacterium bovis* infected populations of cattle and badgers using the Kolmogorov Forward Equations. *Scientific Reports*, 10(1):1–13, 2020.

[9] You Chang, Nienke Hartemink, Andrew W. Byrne, Eamonn Gormley, Guy McGrath, Jamie A. Tratalos, Philip Breslin, Simon J. More, and Mart C. M. de Jong. Inferring bovine tuberculosis transmission between cattle and badgers via the environment and risk mapping. *Frontiers in Veterinary Science*, 10, 2023.

[10] James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438:355–359, 2005.

[11] Piran C.L. White, Monika Böhm, Glenn Marion, and Michael R. Hutchings. Control of bovine tuberculosis in British livestock: there is no 'silver bullet'. *Trends in Microbiology*, 16(9):420–427, 2008.

[12] Rosie Woodroffe, Christl A. Donnelly, Helen E. Jenkins, W. Thomas Johnston, David R. Cox, F. John Bourne, Chris L. Cheeseman, Richard J. Delahay, Richard S. Clifton-Hadley, George Gettinby, Peter Gilks, R. Glyn Hewinson, John P. McInerney, and W. Ivan Morrison. Culling and cattle controls influence tuberculosis risk for badgers. *Proceedings of the National Academy of Sciences of the United States of America*, 103(40):14713–14717, 2006.

[13] Julian A. Drewe, Alexandra J. Tomlinson, Neil J. Walker, and Richard J. Delahay. Diagnostic accuracy and optimal use of three tests for tuberculosis in live badgers. *PLoS ONE*, 5(6):e11196, 2010.

[14] Jenni L. McDonald, Andrew Robertson, and Matthew J. Silk. Wildlife disease ecology from the individual to the population: Insights from a long-term study of a naturally infected European badger population. *Journal of Animal Ecology*, 87(1):101–112, 2017.

[15] Svetlana N Buzdugan, Timothée Vergne, Vladimir Grosbois, Richard J Delahay, and Julian A Drewe. Inference of the infection status of individuals using longitudinal testing data from cryptic populations: Towards a probabilistic approach to diagnosis. *Scientific Reports*, 7(1):1–11, 2017.

[16] Dave W. Hudson, Richard Delahay, Robbie McDonald, Trevelyan J. McKinley, and Dave J. Hodgson. Analysis of lifetime mortality trajectories in wildlife disease research: BaSTA and beyond. *Diversity*, 11(10):182, 2019.

[17] Joseph Crispell, Clare H. Benton, Daniel Balaz, Nicola De Maio, Assel Ahkmetova, Adrian Allen, Roman Biek, Eleanor L. Presho, James Dale, Glyn Hewinson, Samantha J. Lycett, Javier Nunez-Garcia, Robin A. Skuce, Hannah Trewby, Daniel J. Wilson, Ruth N. Zadoks, Richard J. Delahay, and Rowland Raymond Kao. Combining genomics and epidemiology to analyse bi-directional transmission of *Mycobacterium bovis* in a multi-host system. *eLife*, 8:e45833, 2019.

[18] Panayiota Touloupou, Bärbel Finkenstädt, and Simon E. F. Spencer. Scalable Bayesian inference for coupled hidden Markov and semi-Markov models. *Journal of Computational and Graphical Statistics*, 29(2):238–249, 2020.

[19] R.M. Anderson and R.M. May. *Infectious Diseases of Humans*. Oxford University Press, 1991.

[20] Jacco Wallinga and Peter Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, 160(6):509–516, 2004.

[21] Christophe Fraser. Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE*, 2(8):e758, 2007.

[22] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, 2013.

[23] Paul C Cross, Philip LF Johnson, James O Lloyd-Smith, and Wayne M Getz. Utility of R_0 as a predictor of disease invasion in structured populations. *Journal of the Royal Society Interface*, 4(13):315–324, 2007.

[24] Jing Li, Daniel Blakeley, et al. The failure of R_0. *Computational and Mathematical Methods in Medicine*, 2011, 2011.

[25] Andrea Apolloni, Chiara Poletto, José J Ramasco, Pablo Jensen, and Vittoria Colizza. Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theoretical Biology and Medical Modelling*, 11(1):1–26, 2014.

[26] Laurent Hébert-Dufresne, Benjamin M Althouse, Samuel V Scarpino, and Antoine Allard. Beyond R_0: heterogeneity in secondary infections and probabilistic epidemic forecasting. *Journal of the Royal Society Interface*, 17(172):20200393, 2020.

[27] Clara L Shaw and David A Kennedy. What the reproductive number R_0 can and cannot tell us about COVID-19 dynamics. *Theoretical Population Biology*, 137:2–9, 2021.

[28] Brandon Lieberthal and Allison M Gardner. Connectivity, reproduction number, and mobility interact to determine communities' epidemiological superspreader potential in a metapopulation network. *PLOS Computational Biology*, 17(3):e1008674, 2021.

[29] D. Wilkinson, G. C. Smith, R. J. Delahay, L. M. Rogers, C. L. Cheeseman, and R. S. Clifton-Hadley. The effects of bovine tuberculosis (*Mycobacterium bovis*) on mortality in a badger (*Meles meles*) population in England. *Journal of Zoology*, 250:389–395, 2000.

[30] G. C. Smith. Models of *Mycobacterium bovis* in wildlife and cattle. *Tuberculosis*, 81:51–64, 2001.

[31] Fiona Mathews, David W. Macdonald, G. Michael Taylor, Merryl Gelling, Rachel A. Norman, Paul E. Honess, Rebecca Foster, Charlotte M. Gower, Susan Varley, Audrey Harris, Simonette Palmer, Glyn Hewinson, and Joanne P. Webster. Bovine tuberculosis (*Mycobacterium bovis*) in British farmland wildlife: the importance to agriculture. *Proceedings of the Royal Society B: Biological Sciences*, 273(1584):357–365, 2006.

[32] Ellen Brooks-Pollock, Gareth O Roberts, and Matt J Keeling. A dynamic model of bovine tuberculosis spread and control in Great Britain. *Nature*, 511(7508):228–231, 2014.

[33] Andrew JK Conlan, Trevelyan J McKinley, Katerina Karolemeas, Ellen Brooks Pollock, Anthony V Goodchild, Andrew P Mitchell, Colin PD Birch, Richard S Clifton-Hadley, and James LN Wood. Estimating the hidden burden of bovine tuberculosis in Great Britain. *PLOS Computational Biology*, 8(10), 2012.

[34] D. E. Bentil and J. D. Murray. Modelling bovine tuberculosis in badgers. *Journal of Animal Ecology*, 62:239–250, 1993.

[35] T. W. Little, P. F. Naylor, and J. W. Wilesmith. Laboratory study of *Mycobacterium bovis* infection in badgers and calves. *The Veterinary Record*, 111(24):550–557, 1982.

[36] Roy M. Anderson and W. Trewhella. Population dynamics of the badger (*Meles meles*) and the epidemiology of bovine tuberculosis (*Mycobacterium bovis*). *Philosophical Transactions of the Royal Society Series B*, 310:327–381, 1985.

[37] Matthew J Smith, Sandra Telfer, Eva R Kallio, Sarah Burthe, Alex R Cook, Xavier Lambin, and Michael Begon. Host–pathogen time series data in wildlife support a transmission function between density and frequency dependence. *Proceedings of the National Academy of Sciences*, 106(19):7905–7909, 2009.

[38] Rosie Woodroffe, Christl A. Donnelly, Gao Wei, D. R. Cox, F. John Bourne, Terry Burke, Roger K. Butlin, C. L. Cheeseman, George Gettinby, Peter Gilks, Simon Hedges, Helen E. Jenkins, Thomas Johnston, John P. McInerney, W. Ivan Morrison, and Lisa C. Pope. Social group size affects *Mycobacterium bovis* infection in European badgers (*Meles meles*). *Journal of Animal Ecology*, 78:818–827, 2009.

[39] Clare H. Benton, Richard J. Delahay, Andrew Robertson, Robbie A. McDonald, Alastair J. Wilson, Terry A. Burke, and Dave Hodgson. Blood thicker than water: kinship, disease prevalence and group size drive divergent patterns of infection risk in a social mammal. *Proceedings of the Royal Society B: Biological Sciences*, 283:20160798, 2016.

[40] James O. Lloyd-Smith, Paul C. Cross, Cheryl J. Briggs, Matt Daugherty, Wayne M. Getz, John Latto, Maria S. Sanchez, Adam B. Smith, and Andrea Swei. Should we expect population thresholds for wildlife disease? *Trends in Ecology and Evolution*, 20(9):511–519, 2005.

[41] T.J. Ryan, B.M. Buddle, and G.W. De Lisle. An evaluation of the gamma interferon test for detecting bovine tuberculosis in cattle 8 to 28 days after tuberculin skin testing. *Research in Veterinary Science*, 69(1):57–61, 2000.

[42] Richard J Delahay, N Walker, GS Smith, D Wilkinson, RS Clifton-Hadley, CL Cheeseman, AJ Tomlinson, and MA Chambers. Long-term temporal trends and estimated transmission rates for *Mycobacterium bovis* infection in an undisturbed high-density badger (*Meles meles*) population. *Epidemiology & Infection*, 141(7):1445–1456, 2013.

[43] Mark A Chambers, Sue Waterhouse, Konstantin Lyashchenko, Richard Delahay, Robin Sayers, and R. Glyn Hewinson. Performance of TB immunodiagnostic tests in Eurasian badgers (*Meles meles*) of different ages and the influence of duration of infection on serological sensitivity. *BMC Veterinary Research*, 5(42), 2009.

[44] Panayiota Touloupou, Naif Alzahrani, Peter Neal, Simon E. F. Spencer, and Trevelyan J. McKinley. Model comparison with missing data using MCMC and importance sampling. *Bayesian Analysis*, 13(2):437–459, 2018.

[45] Trevelyan J. McKinley, Peter Neal, Simon E. F. Spencer, Andrew J. K. Conlan, and Laurence Tiley. Efficient Bayesian model choice for partially observed processes: with application to an experimental transmission study of an infectious disease. *Bayesian Analysis*, 15(3):839–870, 2020.

Methods

For full details of the data, please see the Supplementary Notes S1. The data are collected and curated by the Animal and Plant Health Agency on behalf of the Department for the Environment, Food and Rural Affairs. For data access requests, please contact Dez.Delahay@apha.gov.uk.

We use a discrete-time $SEID$-type compartmental infection model, defined over quarterly periods, where at any point in time individuals are either susceptible-to-infection (S), infected but not yet infectious (E), infected and infectious (I) or dead (D) (and can only transition through these states in that order). The transmission rate in a given social group s at time t depends on the number of susceptibles, infectives and the total social group size, with a scaling parameter q used to estimate the degree of density vs. frequency dependence in transmission [37]. We assume that the social group structure is known, and that each badger had only moved to a new social group if it was captured there. Individuals born during the study period are assumed to be susceptible at birth, and those born before the monitoring period begins in any social group are given a non-zero probability of being in either the S, E or I states once monitoring commenced.

The mortality process in any mammal population is age-dependent, and conceptually a bathtub-shaped hazard function (such as a Siler model) would be appropriate [46], where high early-life mortality drops to a relatively constant level during mid-life and then increases exponentially in later-life. However, in wild badgers early-life mortality is often hidden, since deaths in newborns occur underground and so are never observed [e.g. 16]. Hence, fitting a Siler model to available CMR data would likely underestimate early-life mortality and/or cause an identifiability challenge. Therefore we instead model the *conditional* mortality distribution *given survival to first capture*, as a Gompertz-Makeham distribution [16, 47, 48]. We assume that badgers surviving this period of early-life mortality and that are still alive at some time t are thus able to be captured with a probability that depends on the sampling quarter (to represent seasonality in capture probabilities). As such, mortality curves are typically right-censored at the last capture time for almost all badgers. However, the capture process plays an important role in identifying the mortality parameters, since under the model assumptions the probability that an animal is alive but evades capture at successive time points becomes negligible quickly if the capture probabilities are not close to zero. Hence the mortality and capture processes together provide identifiability despite rarely observing dead badgers in the data [e.g. 16].

Environmental contamination (e.g. from bacteria shed in urine and faeces) may be a major source of bTB infection in badgers and cattle [49]. Furthermore, studies have also shown that external sources of infection can contribute to the persistence of bTB in badger populations [50] in addition to badger-to-badger transmission. Several studies analyse the spatial organisation of badger populations and its implications for disease transmission at different spatial scales [51, 52, 53, 54, 55, 56]. Thus here we consider social-group specific background

rates-of-infection, to represent indirect transmission between badgers and also infection from other sources such as cattle. Here we do not consider a direct between-social group badger-to-badger transmission process, outside of explicit movements of individual animals between groups (which is captured in the data).

We assume that the diagnostic tests are conditionally independent *given the (hidden) infection status* of each animal, and thus the observation process can be governed by the sensitivities and specificities of the different tests, which are included as parameters to estimate here [see also 15]. This approach can be thought of as an extension of the earlier multi-state model of [15] to incorporate a fully mechanistic model of disease transmission and progression, instead of a simpler time-varying estimate of infection probability between any successive capture events. This means that we are able to extract key quantities-of-interest that result from the mechanistic model, such as individual effective reproduction numbers. Since there is no recovered state in this system, it is necessary to also model the mortality process, since this limits the length of the possible infectious period for infected badgers. We also fit to a much longer time period (1980–2020), which incorporates the use of two additional types of test (Brock and DPP) which were not used during the 2006–2013 time period examined by [15]. We found it necessary to also estimate a change-point for the Brock test sensitivities and specificities. Full mathematical details of the model and parameters are given in the Supplementary Notes S2. We used vague or weakly informative prior distributions for the parameters, full details of which are given in Supplementary Notes S2.3.

Performing robust statistical inference on such a model is highly challenging, since model dynamics rely on a set of (often high-dimensional) unobserved epidemiological states which must be inferred as part of the model fitting process. Gold-standard approaches to this problem [45, 57] are computationally expensive, and common approximation approaches are usually less suitable when considering highly-resolved (such as individual-level) data [e.g. 58]. Instead we leverage a recent advance called the individual forward filtering backward sampling (iFFBS) algorithm [18], which operates within a data-augmented Markov chain Monte Carlo (MCMC) routine. Standard FFBS approaches scale exponentially in the number of individuals, but the iFFBS algorithm (in certain situations) can be made to scale linearly with the number of individuals, substantially reducing computational overheads. The iFFBS only requires transition probabilities for the hidden states and probabilities of observing outcomes (captures and test results) given the hidden states, and it simulates the hidden states for each individual conditional on the observed data and the hidden states of the other individuals, from which we can generate full likelihood estimates. Suitable MCMC techniques can be used to update the parameters given the hidden states, and here we employed a combination of Hamiltonian Monte Carlo, Gibbs Sampling and Metropolis-Hastings updates. A total of 25,000 iterations, ran in a matter of hours, was sufficient to achieve good mixing of parameters (see Supplementary Notes Figs. S1–S2) and event times. We ran two chains, discarded the first 5,000 as burn-in and then thinned the remaining samples to give 5,000 iterations in total across both chains. For full details of the MCMC/iFFBS algorithm, see Supplementary Notes S4.

For each iteration of the MCMC, posterior predictive samples for the test results from individual badgers can be easily obtained by taking a random sample from the testing process for any given observed test, conditioning on the hidden states of the model at the current iteration. These test results can then be aggregated to any

spatio-temporal level, and repeating this process for different MCMC samples allows for a predictive distribution to be built empirically (see e.g. Fig. 4a).

Similarly, for each iteration of the MCMC, the relative contribution of direct badger-to-badger transmission on any new infection i at time t in group g_i can be calculated through the ratio of the badger-to-badger transmission rate over the total transmission rate in social group g_i at time t. These can then be aggregated at any spatio-temporal resolution, and repeating this process for different MCMC samples allows for a predictive distribution to be built empirically (e.g. Fig. 4b; see Supplementary Notes S6 for more details).

Our individual reproduction number estimator uses a variation of the likelihood-based approach of [20], in which the authors consider pairs of cases and use the serial (or generation) interval distribution (the time from symptom onset in a primary case to symptom onset in a secondary case) to derive a relative likelihood that a given case i was infected by case j. The individual-level reproduction number for individual j is thus a weighted sum of these relative likelihoods over all future cases. The generation-time interval is non-trivial to define, since the infectious period in our model is governed by the age-dependent mortality distribution. However, since we directly sample the hidden states for each individual at each point in time, then for any given set of samples it is possible to derive a relative likelihood that any individual was infected by any other source (including the background infection rate). These relative likelihoods can then be summed across all future infections for each infected individual. Since we can do this for any set of posterior samples, we can thus derive an empirical posterior distribution for the individual-level reproduction number for each infected individual, which can again be aggregated to different spatio-temporal levels as required (see Supplementary Notes S5 for details).

Methods references

[46] William Siler. A competing-risk model for animal mortality. *Ecology*, 60:750–757, 1979.

[47] Benjamin Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society*, 115:513–585, 1825.

[48] William Matthew Makeham. On the law of mortality and the construction of annuity tables. *J. Inst. Actuaries and Assur. Mag.*, 8(6):301–310, 1860.

[49] Piran CL White, Julian A Brown, and Stephen Harris. Badgers *(Meles meles)*, cattle and bovine tuberculosis (*Mycobacterium bovis*): a hypothesis to explain the influence of habitat on the risk of disease transmission in southwest England. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 253(1338):277–284, 1993.

[50] Joanne L Hardstaff, Mark T Bulling, Glenn Marion, Michael R Hutchings, and Piran CL White. Impact of external sources of infection on the dynamics of bovine tuberculosis in modelled badger populations. *BMC Veterinary Research*, 8(92):1–10, 2012.

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

[52] Kate L Palphramand, Geraldine Newton-Cross, and Piran CL White. Spatial organization and behaviour of badgers (*Meles meles*) in a moderate-density population. *Behavioral Ecology and Sociobiology*, 61(3):401–413, 2007.

[53] Kate L. Palphramand and Piran C.L. White. Badgers, *Meles meles*, discriminate between neighbour, alien and self scent. *Animal Behaviour*, 74(3):429–436, 2007.

[54] Geraldine Newton-Cross, Piran CL White, and Stephen Harris. Modelling the distribution of badgers *Meles meles*: Comparing predictions from field-based and remotely derived habitat data. *Mammal Review*, 37(1):54–70, 2007.

[55] Monika Böhm, Kate L Palphramand, Geraldine Newton-Cross, Michael R Hutchings, and Piran CL White. The spatial distribution of badgers, setts and latrines: the risk for intra-specific and badger-livestock disease transmission. *Ecography*, 31(4):525–537, 2008.

[56] Monika Böhm, Michael R Hutchings, and Piran CL White. Contact networks in a wildlife-livestock host community: identifying high-risk individuals in the transmission of bovine tb among badgers and cattle. *PloS one*, 4(4):e5016, 2009.

[57] Chris P. Jewell, Theodore Kypraios, Peter Neal, and Gareth O. Roberts. Bayesian analysis for emerging infectious diseases. *Bayesian Analysis*, 4(4):465–496, 2009.

[58] Trevelyan J. McKinley, Alex R. Cook, and Robert Deardon. Inference in epidemic models without likelihoods. *The International Journal of Biostatistics*, 5(1), 2009.

Supplementary information

Supplementary Notes are provided.

Acknowledgments

We would like to thank Andrew Conlan for useful discussions. This work was funded by the Natural Environment Research Council grant number NE-V000616-1. We thank the many field and laboratory staff who have contributed to the collection of the unique Woodchester Park database. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

E.K., T.J.M. and S.E.F.S. developed the model and statistical inference algorithms; E.K. wrote the code, ran the models and produced the outputs; T.J.M., R.A.M., R.J.D. and D.J.H. conceived the project and developed the framework; R.J.D. curated and provided the WP data set; E.K., E.BP and T.J.M. developed the individual reproduction number approach; all authors contributed to the writing of the manuscript.

Competing interests

None.

Materials and correspondence

T.J.M. is the corresponding author. For queries about the Woodchester Park data, please contact R.J.D.

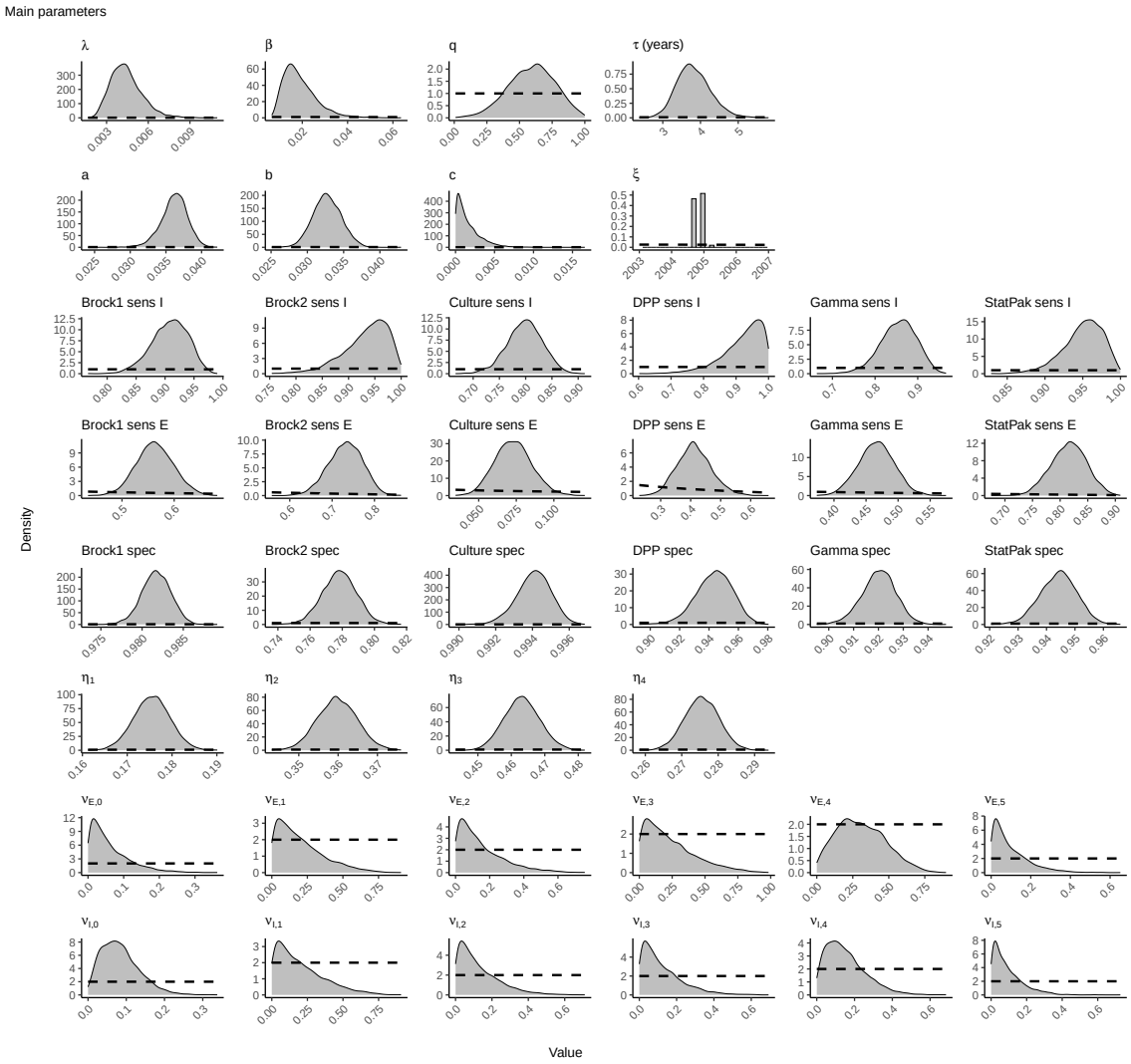
Code availability

Code used to fit this specific model is available at https://github.com/evandrokonzen/WP_bTB_code and through the Environmental Information Data Centre repository (LINK TO FOLLOW).

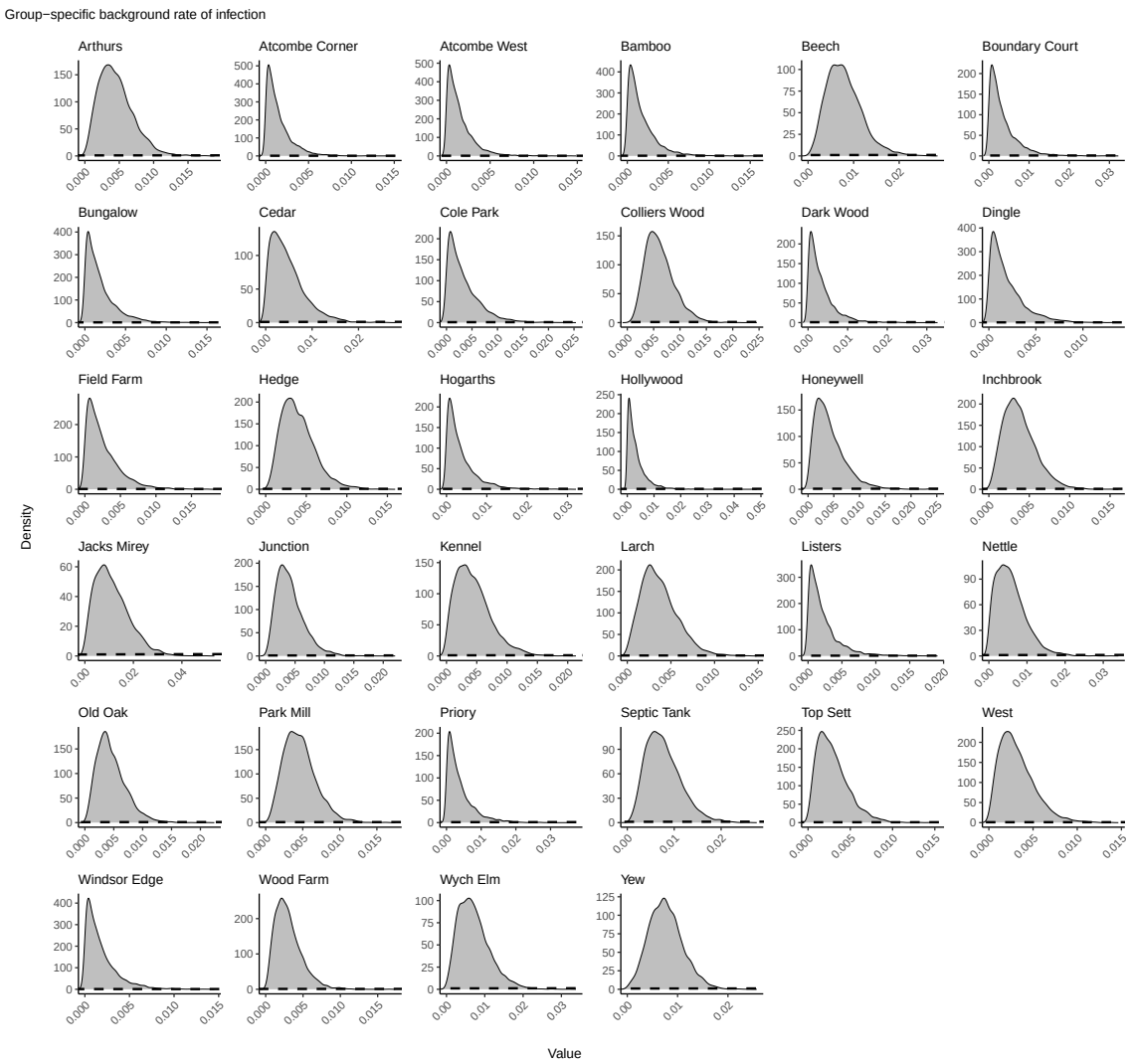
Ethical approval declarations

Badgers were trapped as part of a wider ongoing study at Woodchester Park under a Natural England Science and Conservation licence. Clinical samples were collected by experienced personnel holding relevant Personal Home Office licences and acting under a Home Office Project licence. All animals were examined and sampled under the supervision of a Named Veterinary Surgeon and a Named Animal Care and Welfare Officer. All work with live animals at Woodchester Park is subject to approval and periodic review by the Animal and Plant Health Agency Animal Welfare and Ethical Review Board.

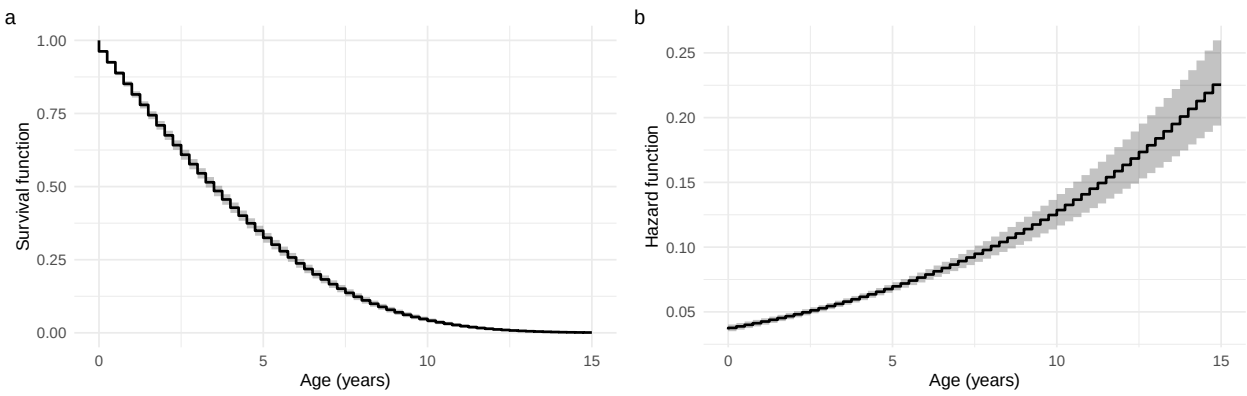
Extended Data



Extended Figure 1: Posterior densities for all parameters except the background rates of infection. Dashed lines represent the prior densities.



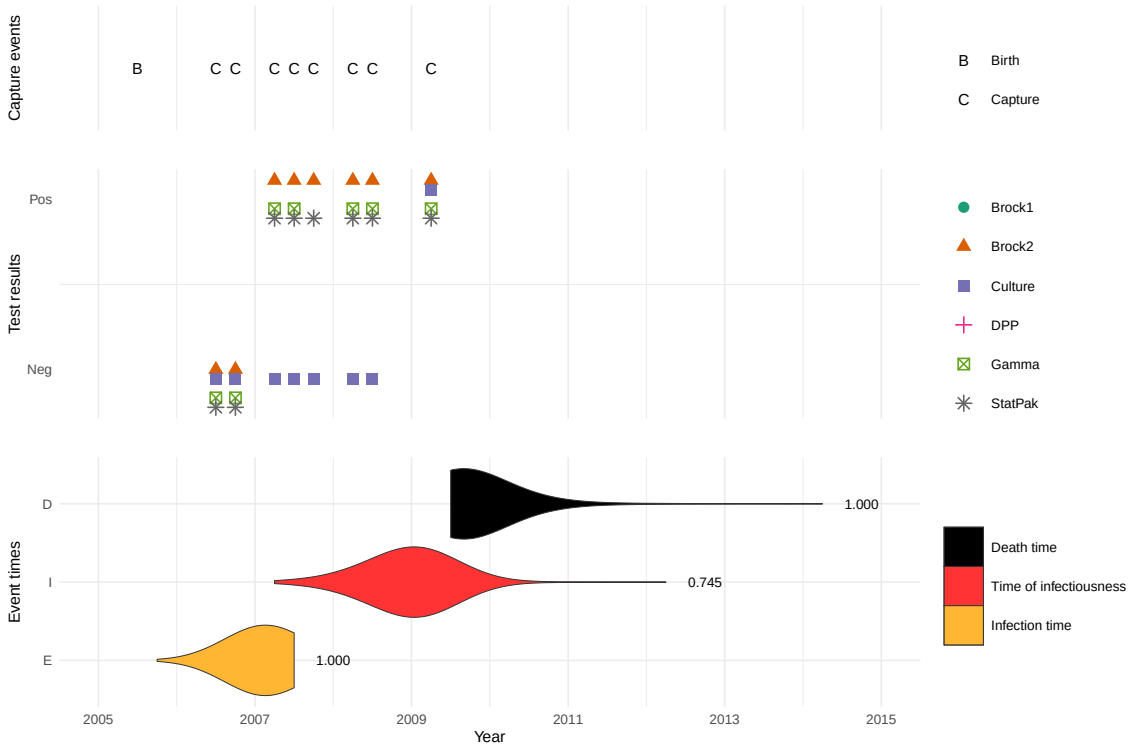
Extended Figure 2: Posterior densities of the background rates of infection. Dashed lines represent the prior densities.



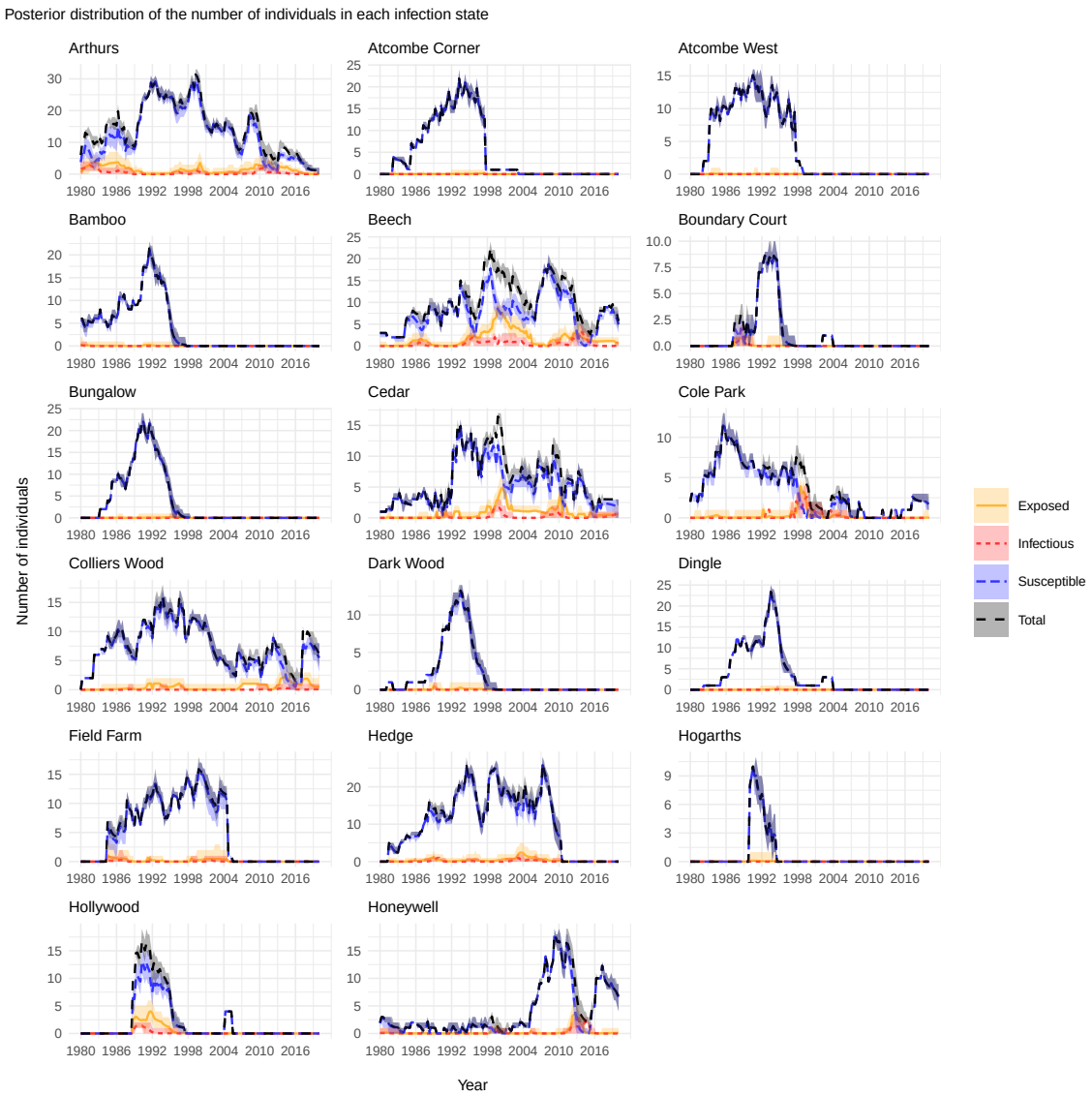
Extended Figure 3: Posterior predictive plots for (a) the survival function, and (b) the hazard function. Posterior means are indicated by black lines and 95% credible intervals by grey ribbons.

Group	Mean	95% CI
Arthurs	0.0045	(0.00088, 0.01)
Atcombe Corner	0.0015	(0.000036, 0.0054)
Atcombe West	0.0015	(0.000037, 0.0054)
Bamboo	0.0017	(0.000049, 0.0062)
Beech	0.008	(0.002, 0.017)
Boundary Court	0.0034	(0.000082, 0.013)
Bungalow	0.0019	(0.000046, 0.0068)
Cedar	0.0047	(0.00027, 0.014)
Cole Park	0.0034	(0.000084, 0.012)
Colliers Wood	0.0062	(0.002, 0.013)
Dark Wood	0.0033	(0.000085, 0.012)
Dingle	0.0019	(0.000051, 0.0068)
Field Farm	0.0025	(0.000077, 0.0086)
Hedge	0.004	(0.001, 0.0089)
Hogarths	0.0036	(0.000081, 0.014)
Hollywood	0.0032	(0.000063, 0.013)
Honeywell	0.0042	(0.00049, 0.012)
Inchbrook	0.0038	(0.00081, 0.0085)
Jacks Mirey	0.011	(0.0012, 0.028)
Junction	0.0042	(0.00087, 0.0098)
Kennel	0.0044	(0.00036, 0.012)
Larch	0.0037	(0.00055, 0.0084)
Listers	0.0022	(0.000061, 0.0084)
Nettle	0.0058	(0.00039, 0.015)
Old Oak	0.0046	(0.001, 0.011)
Park Mill	0.0046	(0.0012, 0.0095)
Priory	0.0039	(0.00011, 0.015)
Septic Tank	0.0077	(0.0021, 0.016)
Top Sett	0.0029	(0.00043, 0.0074)
West	0.0032	(0.00046, 0.008)
Windsor Edge	0.0018	(0.000048, 0.0066)
Wood Farm	0.003	(0.00058, 0.0073)
Wych Elm	0.0072	(0.0012, 0.017)
Yew	0.0077	(0.0019, 0.015)

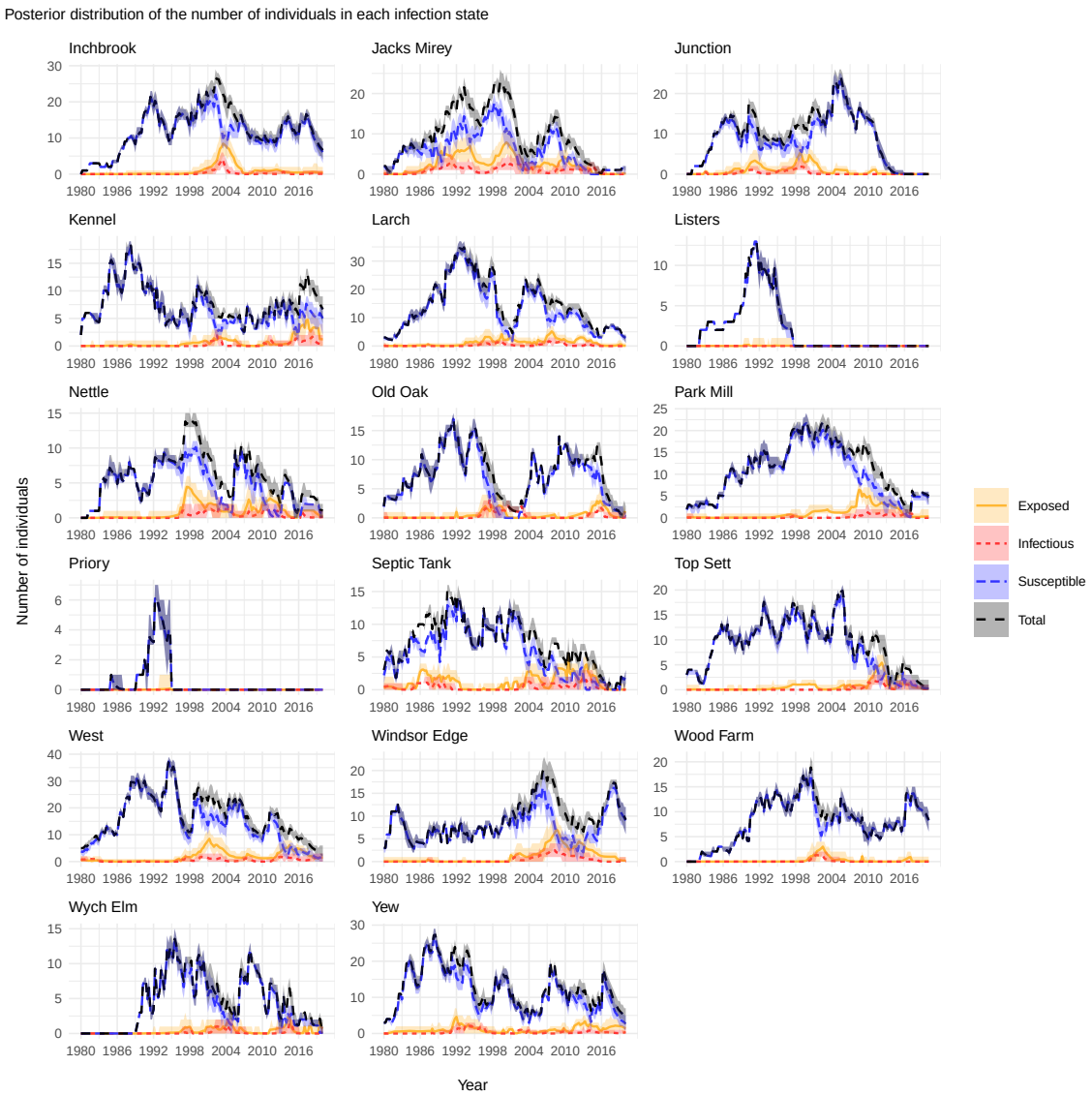
Extended Table 1: Posterior means and 95% credible intervals for group-specific background rate-of-infection terms (to 2 significant figures).



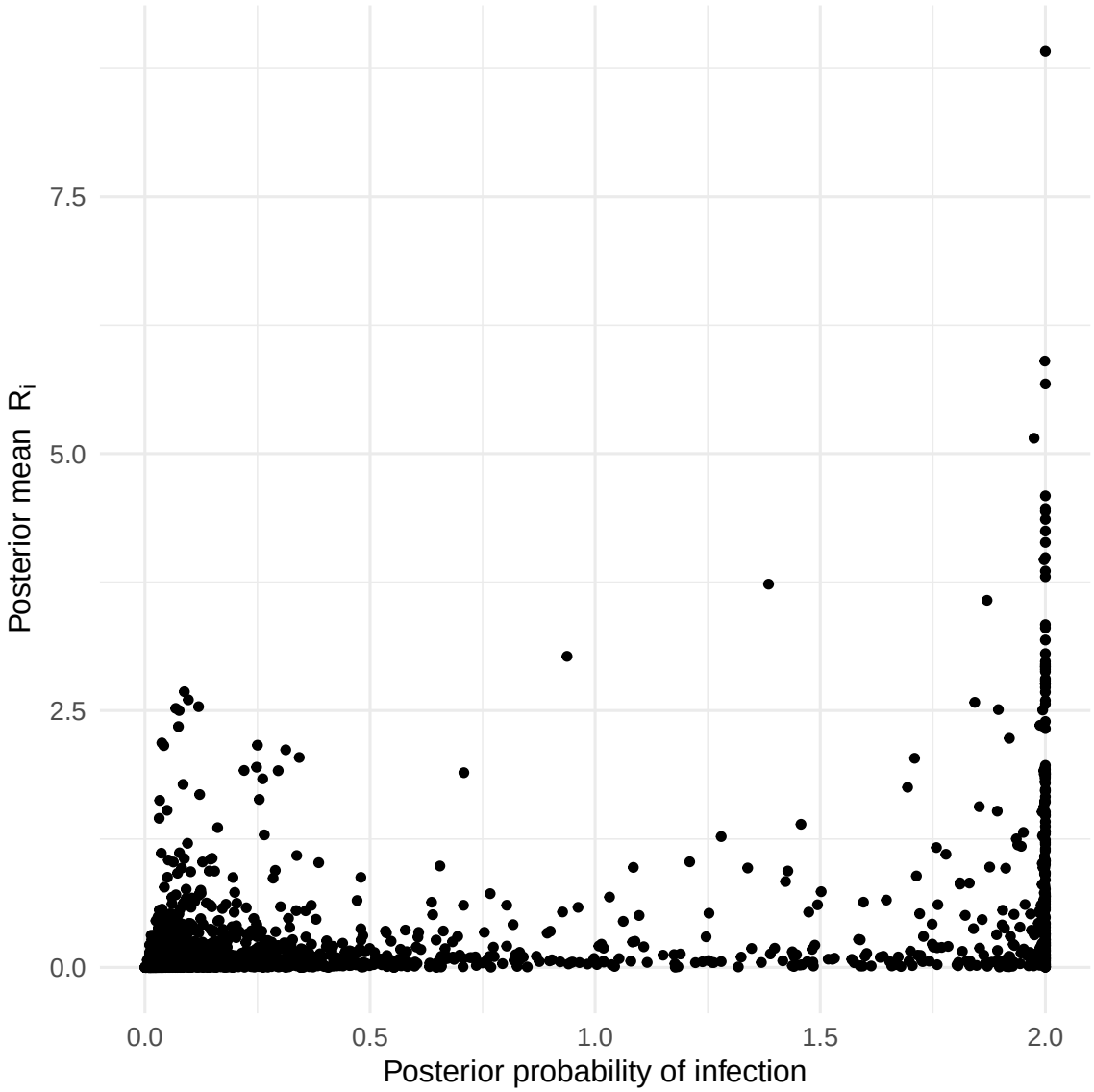
Extended Figure 4: Capture history, test results and posterior distribution of event times for a single individual badger. Densities are the conditional posteriors for the event time given that the event occurred.



Extended Figure 5: Posterior distribution (mean and 95% credible intervals) of the number of individual badgers in each infection state for social groups 1–17.



Extended Figure 6: Posterior distribution (mean and 95% credible intervals) of the number of individual badgers in each infection state for social groups 18–34.



Extended Figure 7: Posterior mean individual effective reproduction numbers against posterior probability of infection.