Assessing the dynamics of *Mycobacterium bovis* infection in three French badger populations

Clément Calenge[1], Ariane Payne[2], Édouard Réveillaud[3], Céline Richomme[4], Sébastien Girard[5] & Stéphanie Desvaux[6]

[1] Office Français de la Biodiversité – Direction Surveillance, Évaluation, Données – Unité Données et Appui Méthodologique, Saint Benoist, BP 20. 78612 Le Perray en Yvelines, France. Orcid: 0000-0001-7741-8016
[2] Office Français de la Biodiversité – Research and Scientific Support Direction, Service Santé Agri, 45100 Orléans.
[3] Regional Directorate for Food of Nouvelle-Aquitaine, Limoges, France.
[4] ANSES, Nancy Laboratory for Rabies and Wildlife, Malzéville, France. Orcid: 0000-0002-5632-7796
[5] Regional Directorate for Food of Bourgogne-Franche-Comté, Dijon, France.
[6] Office Français de la Biodiversité – Research and Scientific Support Direction , Service Santé Agri, 01330 Birieux, France.

Correspondence: clement.calenge@ofb.gouv.fr

Abstract

The Sylvatub system is a national surveillance programme set up in 2011 in France to monitor in wild species the infection by *Mycobacterium bovis*, the main aetiologic agent for bovine tuberculosis. This participatory programme, involving both national and local stakeholders, allowed to monitor from 2013 to 2019 the progression of the infection in three badger populations located in clusters covering between 3222 km^2 and 7698 km^2. In each cluster, badgers were trapped and tested for *M. bovis*. Our first aim was to describe the dynamics of the infection in these clusters. We developed a Bayesian model of the prevalence accounting for the spatial structure of the cases, the imperfect and variable sensitivity of the diagnostic tests, and the correlation of the infection status of badgers of the same commune caused by local factors (social structure, proximity to infected farms, etc.). This model revealed a prevalence increasing with time in one cluster (Dordogne/Charentes), decreasing in the second cluster (Burgundy), and remaining stable in the third cluster (Bearn). In all clusters, the infection was strongly spatially structured, whereas the mean correlation between the infection status of two animals trapped in the same commune was negligible. Our second aim was to develop indicators for the monitoring of the *M. bovis* infection by stakeholders of the programme. We used the fitted model to estimate, in each cluster, (i) the mean prevalence level at mid-period, and (ii) the proportion of the badger population becoming infected in one year. We then derived two indicators of these two key quantities from a much simpler regression model, and we showed how these two indicators could be easily used to monitor the infection in the three clusters. We showed with simulations that these two simpler indicators were a good approximation of these key quantities.

Keywords: participatory science; bovine tuberculosis; prevalence; indicators; spatial modelling; intraclass correlation

Introduction

Mycobacterium bovis is a bacterium that can be transmitted to several domestic and wild species, and to humans. It is the main aetiologic agent for bovine Tuberculosis (bTB), a regulated disease that is still detected in cattle in different European countries. When a farm is detected infected, different control measures can be applied depending on the country and the specific situation of the farm, including the slaughtering of the herd. France is officially free from bTB since 2001 (Delavenne, Pandolfi, et al., 2019), as less than 0.1% of cattle herds are infected annually. In certain parts of the country, infection is still regularly detected in cattle farms and in wild species, mainly wild boars and badgers. The main factor of persistence is the cattle-to-cattle transmission through between-herd contacts (Marsot et al., 2016; Palisson et al., 2016). However, in some areas, a complex multi-host system can explain the circulation of *M. bovis* between the different compartments (domestic species, wild species and environment, Réveillaud et al., 2018) though so far, even if badgers and wild boars are able to transmit *M. bovis* infection to cattle, these species are not considered long-term maintenance hosts in the bTB endemic areas in France (Payne, 2014).

However, due to an increasing number of *M. bovis* cases in wild species, a national surveillance programme of *M. bovis* in wildlife named 'Sylvatub' has been launched in September 2011 (Réveillaud et al., 2018; Rivière et al., 2012). This programme aims at detecting and monitoring *M. bovis* infection in wild species such as wild boar (*Sus scrofa*), red deer (*Cervus elaphus*), roe deer (*Capreolus capreolus*) and European badger (*Meles meles*) populations, by means of both event-based and targeted surveillance strategies. Sylvatub is a participatory monitoring programme (*sensu* Danielsen et al., 2003), i.e. carried out with the help of local stakeholders such as hunters associations, pest control officers, trapper associations, veterinary associations, livestock health defense associations and epidemiologists (Réveillaud et al., 2018). Briefly here, depending on the assessed bTB risk in a given department (French administrative division), three levels of surveillance can be implemented. Level 1 is implemented in a department if no domestic or wild animal has been found infected (relying on the post-mortem examination of hunted or found dead animals). Levels 2 and 3, which are of interest for us in this study, are implemented in departments with sporadic outbreaks in cattle (level 2) and in departments with several outbreaks in cattle and/or cases in wildlife (level 3). In level 3 departments, an at-risk area is defined. This at-risk area is composed of an infected area (communes where the infection has been detected in domestic and/or wild animals – a commune being the smallest French administrative subdivision) and a buffer zone (communes neighbouring the infected areas). Trapping is carried out in all the communes of the at-risk area. In level 2 departments, a prospection zone is defined within 2 km from the pastures of infected farms and trapping is restricted to this area (for details, see Réveillaud et al., 2018).

Three main clusters of *M. bovis* infection have been discovered in France during the last 20 years in badger and wild boar populations following an increased prevalence in cattle farms (Delavenne, Pandolfi, et al., 2019) and are being followed up by Sylvatub: Burgundy (initially discovered in wild boar in 2002, and in badgers afterwards), Dordogne/Charentes (initially discovered in red deer in 2010, and in wild boar and badgers afterwards), and Bearn (initially discovered in wild boar in 2005 and in badgers afterwards; Fig 2D). The data collected by this programme are used to monitor the spatial extension of the infection as well as its progression within these already infected wild populations, by estimating the prevalence level of the infection in badgers in the different clusters. Since the prevalence is simply the proportion of the population that is infected, it is easily understood by non-specialist local communities, which is important to keep stakeholders informed and involved in the programme. Of course, cattle bTB prevalence and incidence are key stone parameters to follow, especially to maintain the national official free status, but because of the multi-host system in place, the sole monitoring of cattle prevalence would not capture the complex epidemiological situation of bTB in an area. Therefore, estimating such a parameter in wildlife populations, easily comparable from one year to another, would also be essential to monitor the epidemiological situation and evaluate the impact of the control measures.

However, an ongoing issue in wildlife epidemiology is the difficulty to estimate prevalence in wild populations, as the sampling of animals used for this estimation cannot be entirely controlled. Indeed, the population is usually sampled using capture methods (e.g., traps for badgers), and the prevalence is usually estimated from the sample of captured animals, under the assumption that these animals are a random sample of the population (which therefore ignores the possible capture bias such as the uneven behavior of the animals towards the traps and the logistical constraints that can affect the placement of traps). Moreover, in the case of participatory monitoring programmes, the participating local communities generally already have their own objectives (e.g. wanting to trap more animals close to some given farms during certain years, and close to others during other years) in addition to the Sylvatub objectives. Thus, the monitoring protocols cannot be too rigid in participatory programmes implying volunteers (e.g. Pocock et al., 2015). However, the spatial structure of the infection must be accounted for in the estimation of the prevalence or any related indicator in a given population.

In addition, another estimation problem occurs when the sampled species is characterized by a social structure that makes trapped animals non-independent from each other. For example, badgers typically live in social groups sharing the same sett and mutually defend a group territory (Roper, 2010). As a consequence, a correlation of the infection status is expected among animals trapped at a given place (e.g. Delahay et al., 2000): when one trapped animal is infected, it is likely that other animals trapped at the same place belong to the same group, and therefore are also infected. Moreover, it has been shown that bTB infection in badgers and cattle are spatially associated (Bouchez-Zacria, Courcoul, et al., 2018; Bouchez-Zacria, Payne, et al., 2023) and therefore badgers trapped in the vicinity of an infected farm are more likely to be infected. Not accounting for this correlation when estimating the prevalence may lead to an overestimated precision (Hisakado et al., 2006).

A last difficulty occurs when the sensitivity and specificity of the tests used for the diagnostic are not perfect: not all infected (resp. non-infected animals) are identified as such by these tests; there may be false-positives and false-negatives. Ignoring this imperfect measure of the infection status can lead to the biased estimation of the prevalence (Dohoo et al., 2003). Moreover, if the tests used for this diagnostic (and the corresponding sensitivity/specificity) change with time, the assessment of the infection progression based on the uncorrected prevalence estimation may also be biased.

In this study, we focus on the targeted surveillance of badgers, which is carried out in communes characterized by surveillance level 2 and 3 (representing 80% of the data collected in the framework of Sylvatub between 2013 and 2019): in each identified bTB cluster, traps are set up by members of Sylvatub in the communes of the at-risk areas, and a *M. bovis* infection is searched on a subsample of the trapped badgers (the proportion and spatial distribution of tested animals depends on the number of trapped animals, trap location and annual sampling objectives). We use these trapping data to develop a complex Bayesian model and provide insight into how the proportion of infected badgers vary in space and time in the three French bTB clusters, accounting for the complex spatial structure of the infection, the correlation between infection status of animals trapped at the same place and the limited sensitivity of the diagnostic tests. Then, we use this model as a basis to develop simpler indicators of the prevalence that also account for all the aforementioned difficulties. These simpler indicators can be easily understood by all the stakeholders and used to monitor both the mean prevalence level and the mean prevalence trend over a given period. Our work in this paper is summarized on Fig 1.

Material and methods

Sylvatub programme and database

The national surveillance system is described in details in Réveillaud et al. (2018). Briefly here, in the communes from level 2 and 3 departments (i.e., communes from infected areas), trapping and culling badgers is implemented as a control measure aiming at reducing badgers abundance. To do so, licensed field stakeholders (hunters, trappers, pest control officers) trap badgers, mostly between March and August. Trapped badgers are culled, and sent to the local veterinary laboratory for necropsy and *M. bovis* testing in the framework of Sylvatub. Not all dead animals are tested: national prescription is to analyze at most 2 animals in each commune and each year, which we suppose in the following to be a random sample of the sent animals. The result of the test for each analyzed animal is stored in a local database and then compiled in the national Sylvatub database. Trappers are encouraged to place their traps in the vicinity of bovine farms and to cover the entire infected areas.

As Sylvatub was launched in 2011 and was not yet well-established before 2013, our study period therefore covers 2013 to 2019. The set of communes where targeted surveillance was authorized at least one year between 2013 and 2019 defines three main spatially connected sets, which are hereafter called *M. bovis* clusters (Fig 2D). The Dordogne/Charentes cluster covers 7698 km^2 and is composed of 413 communes; the Burgundy cluster covers 4254 km^2 and is composed of 254 communes; the Bearn cluster covers 3222 km^2 and is composed of 196 communes. The median surface of a commune is 12 km^2 (interquartile range: 7.2 km^2 to 18.3 km^2).

Following the necropsy, two types of first-line tests were carried out on animal samples, depending on the period: (i) from 2013 to 2015, the first-line test was the bacterial culture performed on sampled tissues of all tested animals, following the protocol established by the French NRL (NFU 47-104) for isolation of *M. bovis*; (ii) since 2016, the first-line test was real-line PCR performed after DNA extraction from a pool of lymph nodes (retropharyngeal, pulmonary and mesenteric) and from organs with gross lesions; molecular typing (spoligotyping) was performed either on MTBC isolates or directly on PCR-positive sample DNA (see Réveillaud et al., 2018, for technical details on these two procedures). The sensitivity of the two techniques differs: the sensitivity of the microbiological cultures is estimated at 50%, whereas the sensitivity of the PCRs is estimated at 75% (Réveillaud et al., 2018; Riviere et al., 2015). The specificity is supposed to be equal to 100% for these two tests (i.e., no false positives).

During the study period, 4590 badgers were trapped and sent to the lab in Dordogne/Charentes, among which 4379 badgers were actually tested. Interpretable results were obtained for 4323 of them (i.e. in average 1.5 badgers per commune and per year; interquartile range: 0 animals to 2 animals tested per commune and per year). In Burgundy, 3042 badgers were trapped and sent to the lab, among which 2900 were actually tested, and interpretable results were obtained for 2786 of them (in average 1.56 animals tested per commune and per year; interquartile range: 0 to 2 animals tested per commune and per year). Finally, in Bearn, 2223 badgers were trapped and sent, among which 1999 were tested, and interpretable results were obtained for 1970 of them (in average 1.43 animals tested per commune and per year; interquartile range: 0 to 2 animals tested per commune and per year).

For each trapped animal, the following data were stored: date of trapping, name of the commune where the animal has been trapped, results of the test (*M. bovis* positive, *M. bovis* negative), type of first-line test carried out (bacterial culture; PCR), date of the analyses, surveillance level of the commune of trapping, and sex and age class (young; adult) of the animals (though this latter information is not systematically reported by the field partners).

A Bayesian model of the infection

Model fit

For each of the three *M. bovis* clusters, we fitted a Bayesian model describing the dynamics of the infection process. Consider one particular cluster. Let N_{it} be the known number of badgers trapped and tested in the commune i during year t. Let y_{it} be the unknown number of badgers actually infected among those N_{it} animals. Let z_{it} be the known number of animals for which the test indicated a *M. bovis* infection among those y_{it} infected animals. We fitted the following hierarchical Bayesian model:

$$z_{it} \sim \text{Binomial}(y_{it}, s_t) \quad (1)$$

$$y_{it} \sim \text{Beta-Binomial}(N_{it}, p_{it}, \rho) \quad (2)$$

$$\log \frac{p_{it}}{1-p_{it}} = \alpha + \beta \times t + u_i \quad (3)$$

$$u_i | \mathbf{u}_{-i} \sim \mathcal{N}(\frac{1}{d_i}\sum_{j\sim i} u_j, \frac{1}{d_i}\frac{1}{\tau}) \quad (4)$$

The equation (1) accounts for the known sensitivity s_t of the tests used during the year t (i.e., $s_t = 0.5$ for microbiological culture, and $s_t = 0.75$ for PCR): the number z_{it} of animals for which a *M. bovis* infection was diagnosed is a random subset of the unknown number y_{it} of animals actually infected (which is a latent variable in this model). Each infected animal is detected as such with a known probability s_t.

We supposed a beta-binomial distribution for the unknown number of infected animals y_{it} (equation (2)). This distribution accounts for a possible correlation between the infection status of two animals trapped the same year in the same commune, and is parameterized by the known number N_{it} of badgers trapped in commune i during year t, the unknown prevalence p_{it} of *M. bovis* infection in commune i and year t, and the unknown correlation coefficient ρ (estimated by the model fit) between the infection status of two animals trapped in the same commune. The parameterization of the beta-binomial distribution as a function of a probability (here, the prevalence) and a correlation coefficient was proposed by Hisakado et al. (2006) as a means to account for the correlation between binary variables in binomial counts. Appendix A give the formal expression of this distribution with this parameterization, and how it relates to the parameterization classically used by statistical software such as R.

The prevalence p_{it} is itself modeled by a logistic regression depending on a commune effect and a linear year effect (also estimated by the model fit; equation (3)). The effects u_i of the communes on the prevalence are not independent from each other. Indeed, because of the strong spatial structure of the infection in the clusters, there is a high probability that the prevalence is high in a commune if it is high in neighbouring communes. We account for this spatial autocorrelation of the commune effects by modelling these random effects u_i with an intrinsic Conditional AutoRegressive (iCAR) model (equation (4), see also Rue and Held, 2005). Thus, the random effect u_i of the commune i is supposed to be drawn from a Gaussian distribution with mean equal to the mean of the random effects of neighbouring communes. In equation (4), $i \sim j$ means that commune i shares a boundary with commune j; $\mathbf{u}_{-i}$ is the vector of commune random effects excluding the effect u_i, and d_i is the number of communes sharing a boundary with commune i. The parameter τ is estimated by the model fit, describing the precision (inverse of the variance) of the random effects u_i.

We defined weakly informative priors for the parameters of the model. We fitted this model by MCMC, using 4 chains of 1 million iterations each, after a burn-in period of 3000 iterations. To save some memory space, we thinned the chains by selecting one sample every 1000 iterations. We checked the mixing properties of the chains both visually and using the diagnostic of Gelman and Rubin (1992). We checked the goodness of fit of our model, using the approach recommended by Gelman and Meng (1996): each MCMC iteration r generated a sampled value θ^r of the vector of parameters of interest $\theta = (\tau, \alpha, \beta, \rho, \mathbf{u})^t$. For each simulated

value θ^r, we simulated a replication of the Sylvatub dataset (i.e., we simulated a random infection status for each trapped animal of the dataset with the fitted model parameterized by the vector simulated by the r-th MCMC iteration). We then compared summary statistics calculated on the observed Sylvatub dataset with the distribution of the same statistics derived from the simulated datasets. All these checks indicated a satisfactory fit of the model (see appendix D for details on these checks and on model fit).

Estimation of the prevalence level and trend from the model

First, we used the fitted model to estimate the trend over time of the prevalence in each cluster. On a logit scale, the average change with time of the prevalence is reflected by the coefficient β in equation (3). It is well known that in a logistic regression, the exponential of a coefficient (here β) is equal to the odds-ratio of the corresponding variable (here the time t), i.e. $\{p_t/(1-p_t)\}/\{p_{t-1}/(1-p_{t-1})\}$, which in our model measures the amount by which the odds $p/(1-p)$ of the infection is multiplied in one year (Hosmer and Lemeshow, 2000, p. 50). However, odds-ratios are difficult to understand by stakeholders, which can be problematic in a participatory programme context. As noted by King and Zeng (2002), "*we have found no author who claims to be more comfortable communicating with the general public using an odds ratio*. Similarly, Gelman and Hill (2006, p. 83) reported "*we find that the concept of odds can be somewhat difficult to understand, and odds ratios are even more obscure. Therefore, we prefer to interpret coefficients on the original scale of the data*". In this section, we follow this last recommendation, by calculating the average rate of change of the prevalence in a cluster using the fitted model.

Due to both the nonlinearity of the logit transform used in the model and the strong spatialization of the infection, the estimation from the model of the average proportion of animals becoming infected in one year can be tricky. Gelman and Pardoe (2007) proposed an approach to estimate this rate of change, based on the concept of predictive comparison. For a given commune v and a given value of the vector of parameters θ of the model, the predictive comparison measures the expected rate of change of the prevalence p when the year changes from $t^{(1)}$ to $t^{(2)}$:

$$\delta_t(t^{(1)} \to t^{(2)}, v, \theta) = \frac{E(p|t^{(2)}, v, \theta) - E(p|t^{(1)}, v, \theta)}{t^{(2)} - t^{(1)}}$$

This quantity, easily calculated with our model, varies as a function of these inputs (the years compared, the commune, and the value of the vector of parameters). To summarize the overall effect of the year on the prevalence in a given dataset, Gelman and Pardoe (2007) proposed to calculate the mean value Δ_t of the predicted comparisons over the probability distribution of the inputs (years and communes) estimated with the data, and over the posterior distribution of the parameters. This averaging is equivalent to consider all pairs of animals (i, j) in the data, corresponding to pairs of transition of (t_i, v_i) to (t_j, v_j) in which the commune $v_i = v_j$ is held constant. Technical details on the calculation of the Average Predictive Comparison (APC) are given in appendix B. When positive, the APC estimated the proportion of the population animals becoming infected in one year in each cluster. When negative, the APC estimated the proportion of the population becoming sane in one year due to the death of infected animals, the birth of uninfected animals, and a decrease of the infection rate, which collectively lead to a decreasing prevalence in the population.

The APC gives an index of the dynamics of the infection in a cluster. We also estimated another statistic summarizing the mean prevalence level in a cluster during the study period. Because the prevalence varies in space and time, we used the equation (3) to estimate the expected prevalence in each commune during the middle year of our study period (i.e. for the year t =2016), and we averaged it over the communes of the cluster. This gave an idea of the importance of the infection in each cluster during the studied period.

The calculation of both the APC and the mean prevalence level during the middle year was restricted to the set of highly infected communes (i.e. communes for which $u_i > 0$ in equation (4)), to allow the comparison with the simpler indicators developed in the next section.

Development of simple indicators of *M. bovis* prevalence level and trend

Although the model developed in the previous sections is useful to understand the spatialization and dynamics of the infection process, it is too complex to be used on a regular basis by the stakeholders of Sylvatub to assess how the level of prevalence changes with time. Instead, we propose in this section two new indicators that can be estimated with the trapping data collected by the network. These indicators estimate in a simpler way the same quantities that were introduced in the last subsection, i.e. the mean prevalence level in the middle year of the study period and the mean proportion of animals becoming infected in one year.

Consider a given *M. bovis* cluster during a given study period with years $t = 1, 2, ..., T$, during which n animals have been collected by the Sylvatub network. For each animal i, let B_i be the infection status returned by the test (coded as 0/1) and s_i be the sensitivity of the test used for this diagnostic. We can derive two useful indicators with the classical simple linear regression fitted to the set of animals trapped during the study period:

$$B_i/s_i = a + b \times \tilde{t_i} + \epsilon_i \tag{5}$$

where $\tilde{t_i}$ is the centred year (i.e. $t_i - \bar{t}$, where $\bar{t}$ is the middle year of the study period), and ϵ_i is a residual. It can be easily demonstrated that, in this model, the coefficient a corresponds to the average prevalence observed in the middle year of the study period, and the coefficient b corresponds to the proportion of the badger population that becomes infected during a year in average during the study period (i.e., the same quantity as the APC calculated for the Bayesian model, see appendix C for a detailed explanation of the rationale).

This approach accounts for the imperfect sensitivity of the tests used for the *M. bovis* diagnostic, but does not account for the spatial structure of the infection in the cluster under study. We will show (see results section) that there is a very strong spatial structure of the infection in the three *M. bovis* clusters. Therefore, not accounting for this structure can lead to biased estimates if the trapping pressure in highly infected areas varies between years. We therefore suggest to calculate this prevalence indicators by focusing only on highly infected communes (i.e. communes characterized by an estimated random effect u_i greater than 0 in equation (4)), so that the remaining unaccounted spatial variability of infection can be ignored. This approach also ignores the correlation possibly caused by the social structure of the badger population and by other local factors (e.g. proximity to an infected farm), but we will show that this correlation is negligible in the three clusters (see results).

Assessing the indicators with the Bayesian model

The complex Bayesian model described by equations (1) to (4) and the simpler regression model described by equation (5) are two models of the same process, though the latter is much simpler. Both can be used to estimate the mean prevalence level during the middle year of the study period and the mean proportion of the population becoming infected in one year in a cluster. The simple regression model imperfectly accounts for the spatial structure of the infection and ignores the correlation caused by local factors (e.g. social structure, proximity to infected farms), but it is much easier to understand and to implement by stakeholders. This latter model is therefore proposed to stakeholders as a means to monitor the infection in a cluster.

We carried out two sets of simulations to assess the ability of the simpler regression model to estimate the two target quantities. For these two sets of simulations, we simulated different infection situations, using the Dordogne/Charentes cluster as an example. We simulated datasets covering 7 years in this cluster. We generated a number of trapped animals for each year and each commune of the cluster from a binomial negative distribution with mean μ and dispersion parameter $\theta = 0.48$ (value estimated from our dataset by maximum likelihood). In each set of simulations, we simulated four possible values of μ corresponding to four levels of trapping pressure: $\mu = 0.5$ animals trapped per commune and per year in average, $\mu = 1$, $\mu = 3$ and $\mu = 10$ (as a point of comparison, remember that in our dataset, $\mu \approx 1.5$ in all clusters). For each simulated animal, we simulated a probability of infection with the help of equation (3). Different values of the slope β and intercept α were specified for the different simulated situations (see below). We simulated an iCAR process to generate random commune effects u_i using equation (4), setting $\tau = 0.73$ (corresponding to the mean value estimated by the model with the Sylvatub dataset in the Dordogne/Charentes Cluster, see results). For each animal, we could calculate the probability of infection p_{it} from the vector of $(\alpha, \beta, \{u_i\})$ with equation (3). We then simulated a random infection status for each animal using equation (2), fixing the correlation coefficient $\rho = 0.04$ (also corresponding to the value estimated in the Dordogne/Charentes cluster using the Sylvatub dataset, see results). Finally, we used equation (1) to simulate an imperfect but variable sensitivity (sensitivity equal to 0.5 during the first three years and 0.75 during the four last years).

In the first set of simulations, we wanted to assess the ability of our regression model to estimate the trend over time of the prevalence in two different situations with regard to its change with time: (i) low but increasing prevalence: we simulated a *M. bovis* infection rarely present on the study area during the first year ($\approx$ 5% of the animals are infected in a typical commune of the cluster) with a prevalence increasing with time. More precisely, we set the intercept $\alpha = -3.1$ in model (3) and the slope β of the year was randomly drawn from a uniform distribution bounded between 0 and 0.4; (ii) high prevalence, increasing or decreasing: we simulated a frequent infection during the first year of the study period ($\approx$ 20% of the animals are infected in a typical commune) with a prevalence either increasing or decreasing. More precisely, we set the intercept $\alpha = -1.38$ and the slope β randomly drawn from a uniform distribution bounded between -0.4 and 0.4. For each combination of trapping pressure μ and simulated situation (either low but increasing prevalence or high prevalence), we simulated 1000 datasets. For each dataset, we estimated the true proportion Δ_u of animals of the area becoming infected in one year in the highly infected communes (i.e. those with simulated random effect greater than 0) with the APC procedure. We applied the linear regression (5) to the data simulated in these communes. We then compared the estimated slope b with the APC Δ_u of the simulated model, which should in theory be equal if the two models are equivalent.

In the second set of simulations, we wanted to assess the ability of our regression model to estimate the mean prevalence level during the middle year. We simulated data with our Bayesian model, using different values of the intercept $\alpha = -4, -3, -2, -1, 0$, describing different mean prevalence levels. We then randomly sampled a slope β from a uniform distribution bounded between -0.4 and 0.4. We simulated 1000 datasets for each combination of value of intercept α and of trapping pressure μ. For each simulated dataset, we considered only the highly infected communes (i.e., with $u_i > 0$) and we calculated the true mean prevalence over the area during the middle year of the study period. We then applied the linear regression (5) to the data simulated in these communes. We then compared the estimated slope a with this true mean prevalence, which should be equal if the two models are equivalent.

Computational aspects

All our analyses and simulations were carried out with the R software (R Core Team, 2023). We used the package `nimble` for model fit (Valpine et al., 2017), `coda` for the analysis of the fit (Plummer et al., 2006), and

`tidyverse` (Wickham and Grolemund, 2017) and `ggplot2` (Wickham, 2016) for data manipulation and graphical display respectively. We have programmed an R package named `badgertub`, available at https://github.com/ClementCalenge/badgertub, containing all the code and data used to fit the model. It can be installed in R with the package `devtools` , using the function `devtools::install_github("ClementCalenge/badgertub", ref="main")`. This package includes a vignette describing how the user can reproduce easily the model fit and simulations (vignette available with the command `vignette("badgertub")` once the package has been installed). This vignette also serves as the supplementary material of the paper and contains additional information on our model (e.g., precisions on the parameterization of the beta-binomial distribution, formal description of the iCAR model, etc.).

Results

Model fit

The estimated parameters of the model for each cluster are presented in Tab 1. The situation was contrasted between the three clusters: the infection was strongly decreasing in Burgundy, strongly increasing in Charentes/Dordogne and seemed stable in Bearn, as revealed by both the slope β of the year in the model and the APC (i.e. proportion of animals becoming infected in one year). The correlation ρ between the infection status of animals trapped in the same commune was actually rather small in all clusters ($\approx$ 0.03), revealing that the correlation caused by local factors (social structure, local environment, etc.) was not causing a strong dependency between animals of a commune. On the other hand, there was a strong spatial structure in all the studied clusters, with highly infected areas and low risk areas in every cluster (see Fig 2).

Table 1. Main results derived from the model fit to the three *M. bovis* clusters. We present here: (i) the parameters of interest in the model (first three rows are the parameters of the model: slope β associated to the year, correlation coefficient ρ between infection status of animals trapped in the same commune, standard deviation $1/\sqrt{\tau}$ of the commune effects), (ii) the average predictive comparison (APC) estimating the proportion of the population getting infected in one year as estimated by the complex Bayesian model and by the simpler regression in the highly infected communes (see text), (iii) the mean prevalence level in the highly infected communes in the middle year of the study period (see text) as estimated by the complex Bayesian model and by the regression. For each parameter and each cluster, we give the point estimate (mean of the posterior distribution) and the 90% credible interval

Parameter	Dordogne/Charentes	Burgundy	Bearn
β	0.18 [0.11, 0.25]	-0.29 [-0.39, -0.2]	0.05 [-0.04, 0.14]
ρ	0.04 [0.02, 0.08]	0.02 [0.01, 0.04]	0.04 [0.02, 0.08]
$1/\sqrt{\tau}$	1.17 [0.87, 1.5]	1.58 [1.11, 2.04]	1.08 [0.69, 1.54]
APC (model)	0.018 [0.009, 0.027]	-0.028 [-0.039, -0.017]	0.005 [-0.008, 0.018]
APC (regression)	0.029 [0.02, 0.037]	-0.034 [-0.043, -0.024]	0.008 [-0.002, 0.018]
Mean Prevalence (model)	0.126 [0.109, 0.143]	0.08 [0.065, 0.097]	0.112 [0.092, 0.134]
Mean Prevalence (regression)	0.157 [0.141, 0.174]	0.117 [0.098, 0.136]	0.133 [0.113, 0.153]

In the three clusters, there was a close agreement between the parameters estimated by the Bayesian model and the same parameters estimated by the simple linear regression (Tab 1), though the mean prevalence seems either slightly overestimated by the regression approach in the three clusters, or slightly underestimated by the Bayesian model.

The two sets of simulations revealed that the two indicators estimate correctly the mean prevalence and

the mean proportion of animals becoming infected fixed in our simulated situations. On one hand, the first set of simulations of two contrasted situations (high prevalence or low and increasing prevalence) showed that the slope of the year in the regression agreed closely with the true simulated proportion of animals becoming infected in one year, whatever the simulated trapping pressure (Fig 3). Of course, the uncertainty was larger when the trapping pressure was lower (the cloud of points was more dispersed around the line $y = x$ when μ was low), but this indicator estimated correctly the target proportion.

On the other hand, the second set of simulations of different prevalence levels under different trapping pressures showed that there was a close agreement between the intercept of the linear regression and the true mean prevalence level during middle year in highly infected communes (Fig 4). Similarly, the uncertainty was larger for low trapping pressures.

Table 2. Coverage probability of the 95% confidence interval on the proportion of animals of a cluster getting infected in one year estimated with the simple linear regression, estimated by simulations for the two tested settings (either high prevalence or low but increasing prevalence) and the 4 trapping pressure. The value of μ corresponds to the mean number of animals trapped in each commune.

Situation	Trapping Pressure	Coverage Probability
High	μ = 0.5	0.93
High	μ = 1	0.94
High	μ = 3	0.91
High	μ = 10	0.84
Low Increasing	μ = 0.5	0.94
Low Increasing	μ = 1	0.94
Low Increasing	μ = 3	0.89
Low Increasing	μ = 10	0.83

Since we use a linear regression to estimate our two indicators, we can derive confidence intervals on these two parameters using the classical formulas derived from the normal theory. We calculated the coverage probability of the 95% confidence intervals for the different simulated situations (Tab 2 and Tab 3). In both sets of simulations, the coverage probability of 95% confidence intervals on the two indicators was closer to 90% than to 95% for moderate trapping pressure. When the trapping pressure was extremely high (i.e. 10 animals trapped in average in each commune of a cluster), the coverage probability of the 95% confidence interval decreased to $\approx$ 80% for the proportion of animals becoming infected in one year, and to $\approx$ 60% for the mean prevalence level during the middle year.

Discussion

We developed a complex Bayesian model to describe how the infection status of badgers changed in space and time in three *M. bovis* clusters in France, accounting for the resolution of the data (commune scale), the spatial structure of the infection, the imperfect and variable sensitivity of the diagnostic tests, and the possible correlation of the infection status of badgers within the same commune. This model allowed to estimate both the mean prevalence level and the mean proportion of badgers becoming infected in one year. We also developed an alternative, much simpler model of the infection process, based on a classical linear regression, which also allowed to easily estimate these two quantities in the highly infected communes only. Simulations of the complex model showed that the two simpler indicators were a good approximation of the true quantities, and could easily be used by stakeholders to estimate the key parameters of the infection process in the most

Table 3. Coverage probability of the 95% confidence interval on the mean prevalence during the middle year in a *M. bovis* cluster estimated with the simple linear regression, for the different tested prevalence levels (Intercept α, see text) and for the different trapping pressure μ. The value of μ corresponds to the mean number of animals trapped in each commune.

Intercept	TrapPress	Probability
$\alpha =$ -4	μ = 0.5	0.91
$\alpha =$-4	μ = 1	0.90
$\alpha =$-4	μ = 3	0.78
$\alpha =$-4	μ = 10	0.58
$\alpha =$-3	μ = 0.5	0.94
$\alpha =$-3	μ = 1	0.90
$\alpha =$-3	μ = 3	0.79
$\alpha =$-3	μ = 10	0.56
$\alpha =$-2	μ = 0.5	0.93
$\alpha =$-2	μ = 1	0.90
$\alpha =$-2	μ = 3	0.80
$\alpha =$-2	μ = 10	0.64
$\alpha =$-1	μ = 0.5	0.92
$\alpha =$-1	μ = 1	0.90
$\alpha =$-1	μ = 3	0.77
$\alpha =$-1	μ = 10	0.58
$\alpha =$0	μ = 0.5	0.92
$\alpha =$0	μ = 1	0.92
$\alpha =$0	μ = 3	0.86
$\alpha =$0	μ = 10	0.66

infected communes.

Basically, if the tests used to diagnose *M. bovis* were characterized by a sensitivity of 100%, our regression approach would be equivalent to a simple linear regression of the *M. bovis* infection status of each animal coded as a binary variable as a function of the year (the form of the response variable B_i/s_i in equation (5) is just a way to account for the imperfect sensitivity of the tests). The suggestion to use a classical linear regression to model what is basically a binary variable can seem surprising, given that such variables are usually modelled with logistic regressions. We preferred to fit a classical linear regression, since its coefficients (intercept and slope of the year) are directly interpretable as the mean prevalence level and proportion of animals becoming infected in one year respectively. Of course, using a classical linear regression to predict a binary variable leads to the violation of several hypotheses underlying this method. However, this violation is not really a problem when the aim is to estimate the regression parameters, as long as we do not want to use the regression model to predict the infection status of each animal. Thus, as long as we are only interested in the slope and intercept of the regression, it does not matter that the linear regression can in theory predict probabilities of infection greater than 1 or lower than 0. Similarly, as noted by Gelman and Hill (2006, p. 46), "*for the purpose of estimating the regression line (as compared to predicting individual data points), the assumption of normality is barely important at all*". Finally, the violation of the homoscedasticity assumption (equal variance of the residuals for all the predicted values) is also a minor issue in this case (Gelman and Hill, 2006, p. 46). The greater interpretability of the regression coefficients and the easier application of the linear regression has led several authors to recommend this method instead of the logistic regression for binary variables (Gomila, 2021; Hellevik, 2009), as long as the model is not used to predict new data points. Note however that the departure from the normal distribution led to low coverage probabilities for the two parameters (and especially the mean prevalence level

at mid-period) when the sample size was large. Indeed, in these conditions, the departure from normality has a stronger effect on the estimation of the precision on the parameters. But as long as the mean sample size in a commune is not too large (say, less than 3 animals per commune and per year), the coverage probability of the 95% confidence intervals derived from the linear regression for these parameters is close to the nominal level, and can provide rough first approximation of the uncertainty of the target quantities.

The correlation between infection status of badgers trapped in the same commune during a given year was low (≈ 0.03), and we showed that indicators ignoring it were not characterized by strongly biased measures of precision. Other authors have found that different badgers of the same sett have a larger chance to be infected (e.g. Delahay et al., 2000; Weber et al., 2013). However, our spatial resolution is much coarser than that in the study of these authors: we work at the commune scale (median area of 12 km^2), whereas the badger home-range rarely exceeds 4 km^2 and is often much smaller (Elmeros et al., 2005; Payne, 2014). The traps set up in a commune often allowed to capture badgers from different social groups, thereby limiting the resulting correlation between infection status. Moreover, the local environmental context may be very variable around different traps within a given commune (e.g. some places can be very close to an infected farm whereas other can be much further), which also limits this correlation. In addition, on a larger scale, in the complex multi-host system encountered in France, the source of *M. bovis* infection for badgers might be various and may also come from other wild hosts such as wild boar (whose movements may exceed the commune scale). If the traps are set where these interspecies transmission may occur, it limits the correlation at a commune scale.

Our complex model identified a very marked spatial structure of the infection in the three studied *M. bovis* clusters, and both our complex model and the simpler regression approach assumed that this structure was stable in time (i.e., the areas with the highest prevalence remain the same every year; even if the mean prevalence increases or decreases in time, it changes in the same way everywhere). In statistical terms, we supposed the additivity of the time and space effects on the prevalence. If the spatial distribution of the infection had changed over time, which can occur for some disease (e.g. with some clusters becoming larger with time, see Wobeser, 1994, p. 29), this assumption would be violated. However, this assumption of additivity is reasonable for the *M. bovis* infection, as demonstrated by both a preliminary exploratory analysis of our dataset and by epidemiological properties of this infection. On one hand, the preliminary fit of a simplistic generalized additive model to predict the infection status of trapped badgers as a function of space and time showed that space-time interaction could be ignored in all clusters and that the spatial distribution of the infection in badgers was stable over time during our study period (see appendix E for more details). On the other hand, this stability can also be explained by the infection dynamic of M. bovis in relation to the structure of the multi-host system. Indeed, infection of the badger population may result from two different dynamics: a within-species transmission related to the social structure of the badgers population, and a between-species transmission caused by the contacts with infected animals of other species – in our context, mainly cattle and wild boar. The relative importance of those two dynamics varies according to the context. For instance, in Burgundy, in a recent study, we found that the spatial structure of the infected badgers population was highly related to the spatial structure of the pastures of infected cattle (Bouchez-Zacria, Payne, et al., 2023), suggesting a between-species transmission dynamic still very active 20 years after infection was detected in both cattle and badgers populations. In any case, within- and between-species infection dynamics logically lead to a strong and stable spatial structure of badger infection because of (i) the strong social structure of the badger population associated with a small number of dispersing animals that usually move between adjacent groups (Rogers et al., 1998)), (ii) the strong spatial structure of the main external source of infection, i.e. the cattle population, which is relatively stable over the years, and (iii) the *M. bovis* transmission mode, which involves direct or indirect contacts between animals as well as an infection resulting frequently in a chronic disease (with animals being infectious for a long time). Thus, these elements suggest that the diffusion in space of the *M. bovis* infection is rather slow so that it is reasonable to suppose that the spatial structure of the

infection in a cluster is stable over a period of a few years (say, 5 to 10 years). The two proposed indicators can therefore be used at this time scale to monitor the changes in the infection pattern. In particular, a few informal tests of the indicators seem to indicate that a 5-year scale is an interesting scale to assess the effect of management measures implemented to control the *M. bovis* infection. When the study period covers more than 10 years, a sliding window in time can be used to fit the linear regression.

During our study period, we observed different tendencies in the 3 main *M. bovis* clusters in France. In Burgundy, there was an annual decrease of the proportion of infected badgers between 2013 and 2019, and the mean prevalence in 2016 was estimated at 0.08 (0.065-0.097) with the model whereas in the 2 other *M. bovis* clusters the tendency was either an annual increase proportion of infected badgers (Dordogne/Charente) or a stabilisation (Bearn) with slightly higher mean prevalence than in Burgundy: respectively 0.126 (0.109-0.143) and 0.112 (0.092-0.134). The observations in the captured badgers' population are in line with the bTB situation in the bovine population. Indeed, in Burgundy, the incidence in cattle farms decreased during the same period, which is not the case for the 2 other clusters (Delavenne, Desvaux, et al., 2021). Burgundy strengthened the bTB control measures earlier than in the other regions, especially in terms of early detection of the infected cattle farms and in badgers culling pressure, at least for some years. This is most likely the main reason explaining such differences, even if differences in the badger population and multi-hosts structures may also have played a role. South West of France (covering the 2 clusters with the higher proportion of infected badgers), is now concentrating the highest number of *M. bovis* cases (80% of the cattle bTB cases and 94% of the wildlife cases – all species included– in 2018, see Delavenne, Desvaux, et al., 2021) and would still need some years of effort to see an improvement of epidemiological indicators.

Having a follow-up of such indicators is therefore crucial to assess the efficiency of the measures being applied. In Sylvatub, it will now be easier to reevaluate the developed indicators regularly in the at-risk areas. We demonstrated that our indicators need to be calculated in the most infected communes. In our study, the complex Bayesian model that we used allowed to identify the highly infected communes (i.e., those with a commune random effect greater than the average), so that those communes can be used in later monitoring for the calculation of the indicators.

If the present indicators are to be used in other situations (e.g., in newly discovered clusters, or in other countries), there are several options to identify those highly infected places. One possibility would be to fit the complex model once, a few years after the time of discovery of the cluster, to identify those communes. But other approaches could also be used. Thus, given the reasonable additivity of the infection at a time scale of a few years, one could try to describe the spatial distribution of the infection risk using data collected over a short period, by ignoring the time dimension. For example, the nonparametric approach of Kelsall and Diggle (1995), which estimate the spatial distribution of the risk by calculating the ratio of two probability densities of positive and negative tests in space, could be used to identify the more infected places.

We developped this regression approach focusing on the badger populations in the infected areas in France, but it could in theory be used more generally for any infection characterized by an additivity of space and time effects on the prevalence. Thus, preliminary results indicate that this regression approach could also be used for the wild boar in the three main French *M. bovis* clusters. In this case too, the same Bayesian model provides a good description of the infection (though the spatial structure is much less clear, C. Calenge pers. com.), which suggests that the linear regression indicators proposed for the badger could also be used for the wild boar monitoring.

Acknowledgements

We warmly thank all the stakeholders of Sylvatub, including the French Reference Laboratory for Tuberculosis (Anses, Maisons-Alfort, France), local state veterinary services, departmental hunters federations, the local services of the French Agency for Biodiversity, departmental laboratories for analysis, local animal health defense associations, pest control officers, trappers and their associations. We are also indebted to Julie Rivière, the national coordinator of Sylvatub from 2011 to 2013, and the members of the National steering commitee and coordination subcommitee.

Fundings

Sample collections and analysis were funded by the direction générale de l'alimentation of the French ministry of agriculture.

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Data, script, code, and supplementary information availability

All the data, script and codes, as well as the supplementary information have been packaged in an R package named `badgertub`, available on Github, at https://github.com/ClementCalenge/badgertub. We have also stored this package on Zenodo, where it has been given the following Digital Object Identifier: https://dx.doi.org/10.5281/zenodo.7986442

References

Bouchez-Zacria M, A Courcoul, and B Durand (2018). The Distribution of Bovine Tuberculosis in Cattle Farms Is Linked to Cattle Trade and Badger-Mediated Contact Networks in South-Western France, 2007–2015. *Frontiers in Veterinary Science* 5. https://doi.org/10.3389/fvets.2018.00173.

Bouchez-Zacria M, A Payne, S Girard, C Richomme, ML Boschiroli, M Marsot, B Durand, and S Desvaux (2023). Spatial association of Mycobacterium bovis infection in cattle and badgers at the pastures interface in an endemic area in France. *Under review*.

Danielsen F, MM Mendoza, P Alviola, DS Balete, M Enghoff, MK Poulsen, and AE Jensen (2003). Biodiversity monitoring in developing countries: what are we trying to achieve? *Oryx* 37, 407–409. https://doi.org/10.1017/S0030605303000735.

Delahay R, S Langton, G Smith, R Clifton-Hadley, and C Cheeseman (2000). The spatio-temporal distribution of Mycobacterium bovis (bovine tuberculosis) infection in a high-density badger population. *Journal of Animal Ecology* 69, 428–441. https://doi.org/10.1046/j.1365-2656.2000.00406.x.

Delavenne C, F Pandolfi, S Girard, E Réveillaud, P Jabert, M Boschiroli, L Dommergues, F Garapin, N Keck, F Martin, M Moussu, S Philizot, J Rivière, I Tourette, D Calavas, C Dupuy, B Dufour, and F Chevalier (2019). Tuberculose bovine : bilan et évolution de la situation épidémiologique entre 2015 et 2017 en France métropolitaine. *Bulletin épidémiologique* 91, 1–22.

Delavenne C, S Desvaux, ML Boschiroli, S Carles, P Chaigneau, B Dufour, B Durand, K Gache, F Garapin, S Girard, P Jabert, N Keck, É Réveillaud, J Rivière, C Dupuy, and F Chevalier (2021). Surveillance de la tuberculose due à mycobacterium bovis en France métropolitaine en 2018 : résultats et indicateurs de fonctionnement. *Bulletin Épidémiologique* 94, 1–9.

Dohoo I, W Martin, and H Stryhn (2003). *Veterinary epidemiologic research*. University of Prince Edward Island, p. 706.

Elmeros M, AB Madsen, and A Prang (2005). Home range of the badger (*Meles meles*) in a heterogeneous landscape in Denmark. *Lutra* 48, 35.

Gelman A and X Meng (1996). Model checking and model improvement. In: *Markov chain Monte Carlo in practice*. Ed. by Gilks W and Richardson S. Chapman & Hall/CRC. Chap. 11, pp. 189–201.

Gelman A and D Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–472. https://doi.org/10.1214/ss/1177011136.

Gelman A and J Hill (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman A and I Pardoe (2007). 2. Average predictive comparisons for models with nonlinearity, interactions, and variance components. *Sociological Methodology* 37, 23–51. https://doi.org/10.1111/j.1467-9531.2007.00181.x.

Gomila R (2021). Logistic or Linear? Estimating Causal Effects of Experimental Treatments on Binary Outcomes Using Regression Analysis. *Journal of Experimental Psychology* 150, 700–709. https://doi.org/10.1037/xge0000920.

Hellevik O (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity* 43, 59–74. https://doi.org/10.1007/s11135-007-9077-3.

Hisakado M, K Kitsukawa, and S Mori (2006). Correlated binomial models and correlation structures. *Journal of Physics A: Mathematical and General* 39, 15365. https://doi.org/10.1088/0305-4470/39/50/005.

Hosmer D and S Lemeshow (2000). *Applied logistic regression. Second Edition.* John Wiley & Sons.

Kelsall JE and PJ Diggle (1995). Non-parametric estimation of spatial variation in relative risk. *Statistics in medicine* 14, 2335–2342. https://doi.org/10.1002/sim.4780142106.

King G and L Zeng (2002). Estimating risk and rate levels, ratios and differences in case-control studies. *Statistics in medicine* 21, 1409–1427. https://doi.org/10.1002/sim.1023.

Marsot M, M Béral, A Scoizec, Y Mathevon, B Durand, and A Courcoul (2016). Herd-level risk factors for bovine tuberculosis in French cattle herds. *Preventive Veterinary Medicine* 131, 31–40. https://doi.org/10.1016/j.prevetmed.2016.07.006.

Palisson A, A Courcoul, and B Durand (Mar. 2016). Role of Cattle Movements in Bovine Tuberculosis Spread in France between 2005 and 2014. *PLOS ONE* 11, 1–19. https://doi.org/10.1371/journal.pone.0152578.

Payne A (2014). Rôle de la faune sauvage dans le système multi-hôtes de Mycobacterium bovis et risque de transmission entre faune sauvage et bovins: étude expérimentale en Côte d'Or. PhD thesis. Université Claude Bernard-Lyon I.

Plummer M, N Best, K Cowles, and K Vines (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* 6, 7–11.

Pocock MJ, SE Newson, IG Henderson, J Peyton, WJ Sutherland, DG Noble, SG Ball, BC Beckmann, J Biggs, and T Brereton (2015). Developing and enhancing biodiversity monitoring programmes: a collaborative assessment of priorities. *Journal of Applied Ecology* 52, 686–695. https://doi.org/10.1111/1365-2664.12423.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Réveillaud E, S Desvaux, ML Boschiroli, J Hars, É Faure, A Fediaevsky, L Cavalerie, F Chevalier, P Jabert, S Poliak, I Tourette, P Hendrikx, and C Richomme (2018). Infection of Wildlife by Mycobacterium bovis in France Assessment Through a National Surveillance System, Sylvatub. *Frontiers in Veterinary Science* 5, 262. https://doi.org/10.3389/fvets.2018.00262.

Riviere J, Y Le Strat, B Dufour, and P Hendrikx (2015). Sensitivity of bovine tuberculosis surveillance in wildlife in France: a scenario tree approach. *PLoS One* 10, e0141884. https://doi.org/10.1371/journal.pone.0141884.

Rivière J, J Hars, C Richomme, A Fediaevsky, D Calavas, E Faure, and P Hendrikx (2012). La surveillance de la faune sauvage : de la théorie à la pratique avec l'exemple du réseau Sylvatub. *Épidémiologie et Santé Animale* 61, 5–16.

Rogers L, R Delahay, C Cheeseman, S Langton, G Smith, and R Clifton-Hadley (1998). Movement of badgers (Meles meles) in a high–density population: individual, population and disease effects. *Proceedings of the Royal Society of London B: Biological Sciences* 265, 1269–1276. https://doi.org/10.1098/rspb.1998.0429.

Roper TJ (2010). *Badger*. HarperCollins UK.

Rue H and L Held (2005). *Gaussian Markov Random Fields. Theory and Applications.* Chapman & Hall/CRC.

Valpine P de, D Turek, CJ Paciorek, C Anderson-Bergman, DT Lang, and R Bodik (2017). Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics* 26, 403–413. https://doi.org/10.1080/10618600.2016.1172487.

Weber N, S Bearhop, SR Dall, RJ Delahay, RA McDonald, and SP Carter (2013). Denning behaviour of the European badger (Meles meles) correlates with bovine tuberculosis infection status. *Behavioral Ecology and Sociobiology* 67, 471–479. https://doi.org/10.1007/s00265-012-1467-4.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham H and G Grolemund (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 1st ed. O'Reilly Media.

Wobeser GA (1994). *Investigation and management of disease in wild animals*. Springer Science & Business Media.

Figure 1. Summary of the models fitted in this paper. For each cluster (here illustrated with the Dordogne/Charentes cluster), our dataset consists in a sample of badgers trapped in different communes during different years, tested for *M. bovis*. We first fit a complex Bayesian model to this dataset accounting for many characteristics of the infection (left part). We then focus on highly infected communes and use the average predictive comparison to estimate the mean proportion of the cluster population becoming infected in one year, and we use the model to estimate the mean prevalence in these infected communes during the year in the middle of the period. We then fit a much simpler linear regression (right part) on the data collected in the highly infected communes, which allow to directly produce estimates of the mean proportion of the cluster population becoming infected in one year, and of the mean prevalence in the cluster population during the year in the middle of the study period. Simulations indicate that the two approaches return nearly identical results.

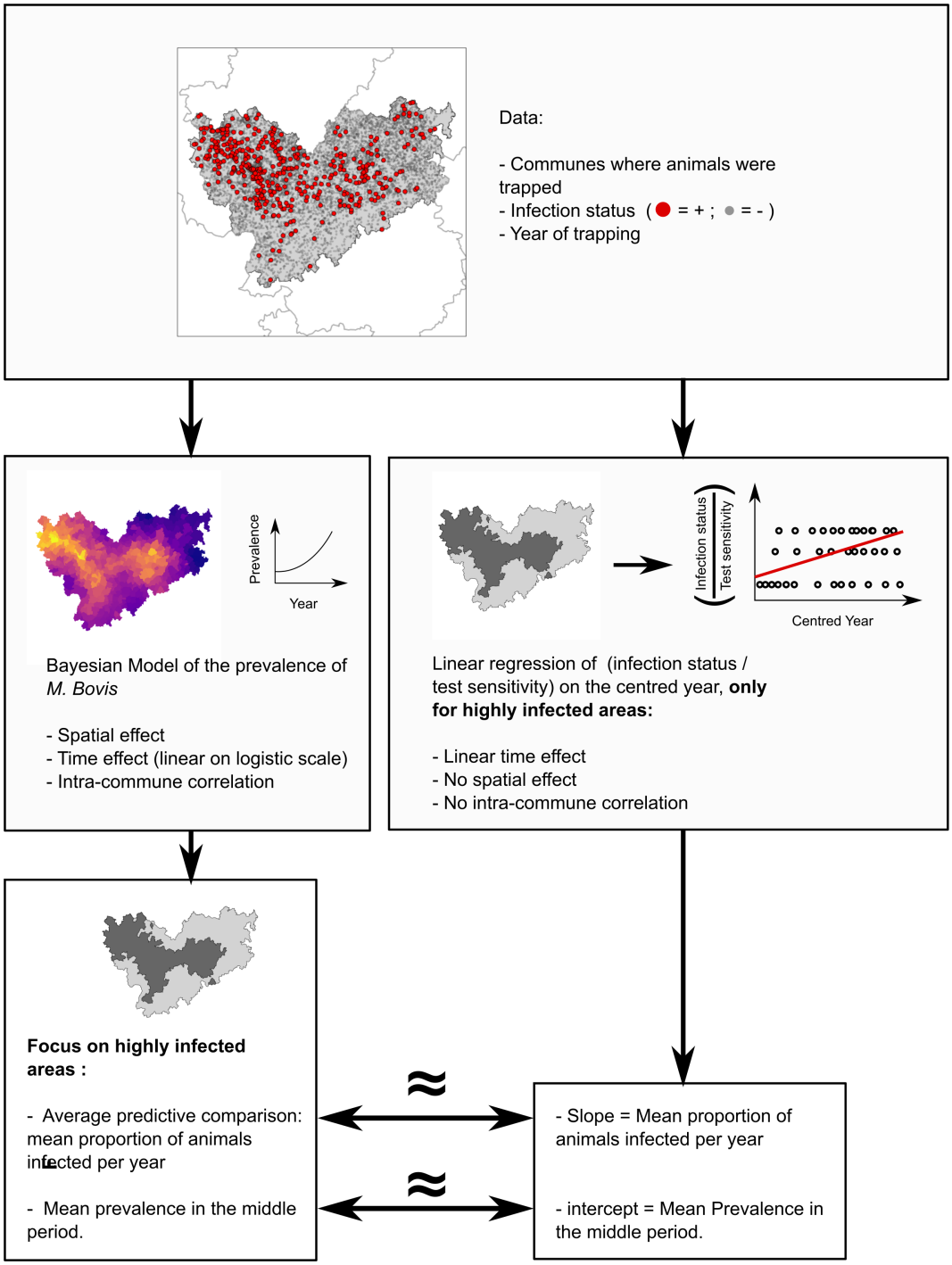

Figure 2. Location of the three *M. bovis* clusters in France (D) – the limits of the French departments are displayed on this map – as well as the median prevalence estimated by our Bayesian model for each commune in the Dordogne/Charentes cluster (A), the Burgundy cluster (B), and the Bearn cluster (C). A common colorscale is used for all clusters (inset in the Bearn map).

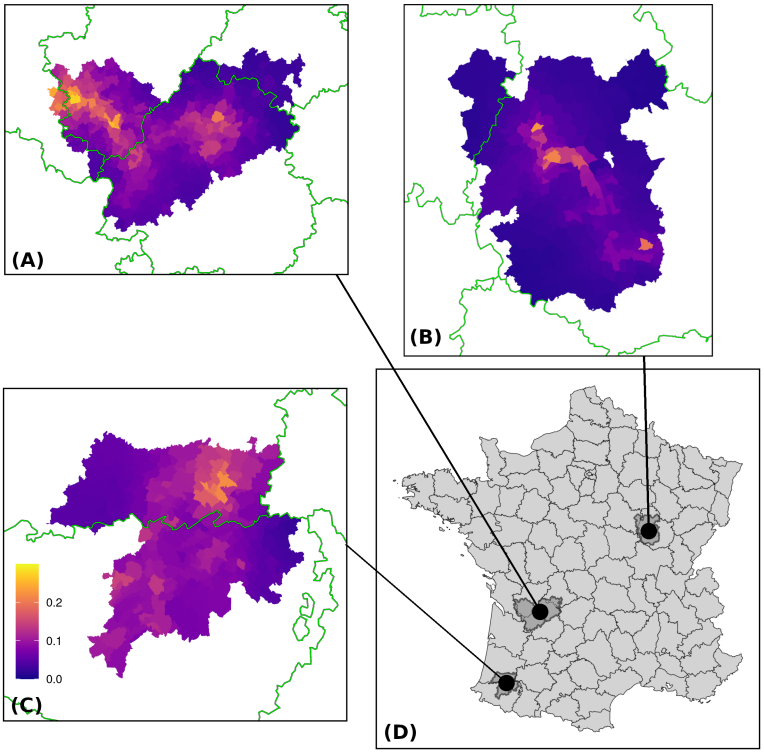

Figure 3. Comparison of the proportion of animals becoming infected in one year estimated using the regression indicator (see text) with the true value, estimated by simulations for the two different situations (High prevalence = top row; Low but increasing prevalence = bottom row) and the different trapping pressure (mu corresponds to the mean number of animals trapped per commune). The straight line is the line of equation $y = x$.

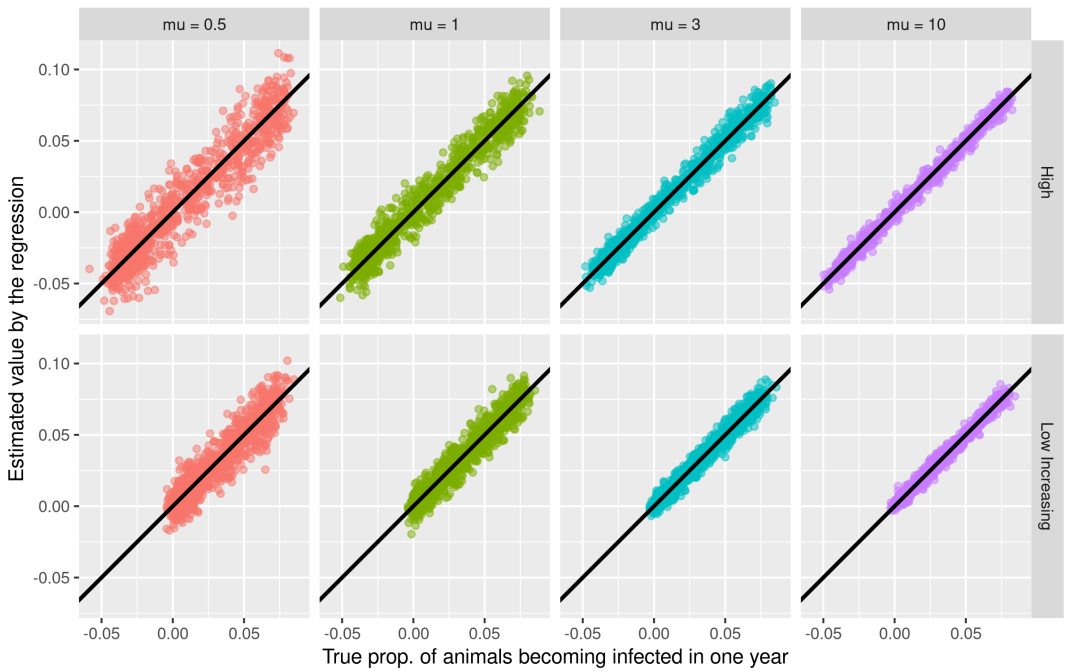
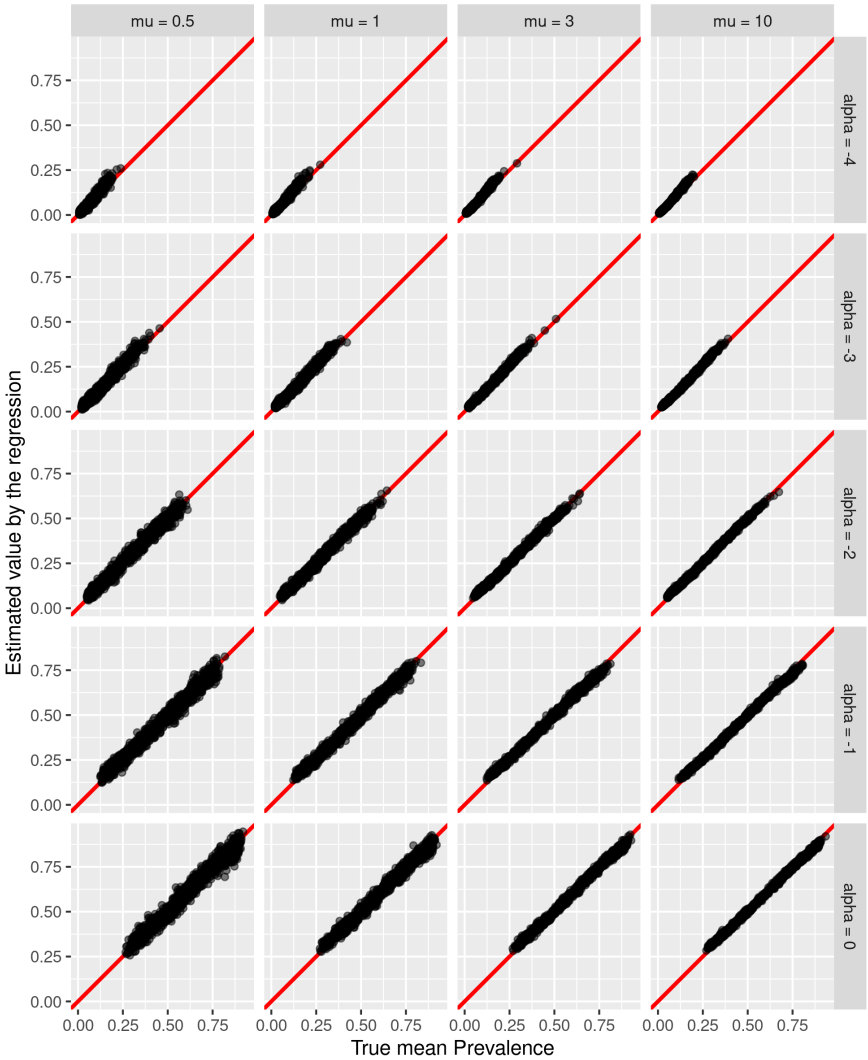
Figure 4. Comparison of the mean prevalence level during the middle year estimated using the regression indicator (see text) with the true value, estimated by simulations for five different prevalence levels (simulated by fixing different values of the intercept alpha) and the different trapping pressure (mu corresponds to the mean number of animals trapped per commune). The straight line is the line of equation $y = x$.