A modelling framework for the prediction of the herd-level probability of infection from longitudinal data

Aurélien Madouasse[1], Mathilde Mercat[1], Annika van Roon[2], David Graham[3], Maria Guelbenzu[3], Inge Santman Berends[2, 4], Gerdien van Schaik[2, 4], Mirjam Nielen[2], Jenny Frössling[5], Estelle Ågren[5], Roger Humphry[6], Jude Eze[6], George Gunn[6], Madeleine Henry[6], Jörn Gethmann[7], Simon J. More[8], and Christine Fourichon[1]

[1]BIOEPAR, INRA, Oniris, La Chantrerie, Nantes 44307, France
[2]Department of Farm Animal Health, Faculty of Veterinary Medicine, Utrecht University, PO Box 80151, 3508, TD Utrecht, the Netherlands
[3]Animal Health Ireland, Unit 4/5, The Archways, Bridge St., Carrick-on-Shannon, Co. Leitrim N41 WN27, Ireland
[4]GD Animal Health, PO Box 9, 7400 AA, Deventer, the Netherlands
[5]Department of Disease Control and Epidemiology, National Veterinary Institute (SVA), 751 89 Uppsala, Sweden
[6]Scotland's Rural College, Kings Buildings, West Mains Road, Edinburgh, EH9 3JG, United Kingdom
[7]Institute of Epidemiology, Friedrich-Loeffler-Institute, Südufer 10, 17493 Greifswald, Germany
[8]Centre for Veterinary Epidemiology and Risk Analysis, UCD School of Veterinary Medicine, University College Dublin, Belfield, Dublin D04 W6F6, Ireland

July 13, 2020

Abstract

For many infectious diseases of farm animals, there exist collective control programmes (**CP**s) that rely on the application of diagnostic testing at regular time intervals for the identification of infected animals or herds. The diversity of these CPs complicates the trade of animals between regions or countries because the definition of freedom from infection differs from one CP to another. In this paper, we describe a statistical model for the prediction of herd level probabilities of infection from longitudinal data collected as part of CPs against infectious diseases of cattle. The model was applied to data collected as part of a CP against infections by the bovine viral diarrhoea virus (**BVDV**) in Loire-Atlantique, France. The model represents infection as a herd latent status with a monthly dynamics. This latent status determines test results through test sensitivity and test specificity. The probability of becoming status positive between consecutive months is modelled as a function of risk factors (when available) using logistic regression. Modelling is performed in a Bayesian framework. Prior distributions need to be provided for the sensitivities and specificities of the different tests used, for the probability of remaining status positive between months as well as for the probability of becoming positive between months. When risk factors are available, prior distributions need to be provided for the coefficients of the logistic regression in place of the prior for the probability of becoming positive. From these prior distributions and from the longitudinal data, the model returns posterior probability distributions for being status positive in all herds on the current months. Data from the previous months are used for parameter estimation. The impact of using different prior distributions and model settings on parameter estimation was evaluated using the data. The main advantage of this model is its ability to predict a probability of being status positive on a month from inputs that can vary in terms of nature of test, frequency of testing and risk factor availability. The main challenge in applying the model to the BVDV CP data was in identifying prior distributions, especially for test characteristics, that corresponded to the latent status of interest, i.e. herds with at least one persistently infected (**PI**) animal. The model is available on Github as an R package (https://github.com/AurMad/STOCfree).

1 Introduction

For many infectious diseases of farm animals, there exist collective control programmes that rely the application of diagnostic testing at regular time intervals for the identification of infected animals or herds. In cattle, such diseases notably include infection by the bovine viral diarrhoea virus (**BVDV**) or by *Mycobacterium avium* subspecies *paratuberculosis* (MAP). These control programmes (**CP**)s are extremely diverse. Their objective can range from decreasing the prevalence of infection to eradication. Participation in the CP can be voluntary or compulsory. The qualification of herds regarding infection can be based on a wide variety of testing strategies in terms of the nature of the tests used (identification of antibodies vs. identification of the agent), the groups of animals tested (e.g. breeding herd vs. young animals), number of animals tested, frequency of testing (once to several times a year, every calf born...). Even within a single CP, surveillance modalities may evolve over time. Such differences in CPs were described by van Roon *et al.* (2020b) for programmes targeting BVDV infections and by Whittington *et al.* (2019) for programmes against MAP.

Differences in surveillance modalities can be problematic when purchasing animals from areas with different CPs because the free status assigned to animals or herds might not be equivalent between CPs. A standardised method for both describing surveillance programmes and estimating confidence of freedom from surveillance data would be useful when trading animals across countries or regions. While inputs can vary between programmes, the output needs to be comparable across programmes. This is called output-based surveillance (Cameron, 2012). Probabilities measure both the chance of an event and the uncertainty around its presence/occurrence. If well designed, a methodology to estimate the probability of freedom from infection would meet the requirements of both providing a confidence of freedom from infection as well as of being comparable whatever the context.

Currently, the only quantitative method used to substantiate freedom from infection to trading partners is the scenario tree method (Martin *et al.*, 2007). The method is applied to situations where there is a surveillance programme in place, with no animals or herds confirmed positive on testing. Scenario trees are based on the premise that it is impossible to prove that a disease is totally absent from a territory unless the entire population is tested with a perfect test. What is estimated with the scenario tree method is the probability that the infection would be detected in the population if it

were present at a chosen *design prevalence*. The output from this approach is the probability that the infection prevalence is not higher than the design prevalence given the negative test results (Cameron, 2012). Therefore, this method is well suited for those countries that are free from infection and that want to quantify this probability of freedom from infection for the benefit of trading partners (Norström *et al.*, 2014).

The scenario tree method is not adapted to countries or regions where there is a CP against an infectious disease which is still present. In such a context, only herds that have an estimated probability of freedom from infection that is deemed sufficiently high or, equivalently, a probability of infection that is deemed sufficiently low, would be safe to trade with. Identifying these herds involves estimating a probability of infection for each herd in the CP and then defining a decision rule to categorise herds as uninfected or infected based on these estimated probabilities.

In this paper, we propose a method to estimate herd level probabilities of infection from heterogeneous longitudinal data generated by CPs. The method predicts herd-month level probabilities of being latent status positive from longitudinal data collected in CPs. The input data are test results, and associated risk factors when available. Our main objective is to describe this modelling framework by showing how surveillance data are related to the *probabilities of infection* (strictly speaking, *probabilities of being latent status positive*) and by providing details regarding the statistical assumptions that are made. A secondary objective is to estimate these probabilities of being latent status positive, using different definitions for the latent status, from surveillance data collected as part of a CP against the infection by the BVDV in Loire-Atlantique, France. The challenges of defining prior distributions and the implications of using different prior distributions are discussed. R functions to perform the analyses described in this paper are gathered in an R package which is available from GitHub (`https://github.com/AurMad/STOCfree`).

2 Materials and methods

2.1 Description of the model

2.1.1 Conceptual representation of surveillance programmes

Surveillance programmes against infectious diseases can be seen as imperfect repeated measures of a true status regarding infection. In veterinary epidemiology, the issue of imperfect testing has traditionally been addressed using latent class models. With this family of methods, the true status regarding infection is modelled as an unobserved quantity which is linked to test results through test sensitivity and specificity. Most of the literature on the subject is on estimating both test characteristics and infection prevalence (Collins & Huynh, 2014). For the estimations to work, the same tests should be used in different populations (Hui & Walter, 1980), the test characteristics should be the same among populations and test results should be conditionally independent given the infection status (Toft *et al.*, 2005; Johnson *et al.*, 2009). Latent class models can also be used to estimate associations between infection, defined as the latent class, and risk factors when the test used is imperfect (Fernandes *et al.*, 2019). In the study by Fernandes *et al.* (2019), the latent class was defined using a single test, through the prior distributions put on sensitivity and specificity. When using latent class models with longitudinal data, the dependence between successive test results in the same herds must be accounted for. In the context of estimating test characteristics and infection prevalence from 2 tests in a single population from longitudinal data, Nusinovici *et al.* (2015) proposed a Bayesian latent class model which incorporated 2 parameters for new infection and infection elimination. The model we describe below combines these different aspects of latent class modelling into a single model.

We propose to use a class of models called Hidden Markov Models (HMM, see Zucchini *et al.* (2017)). Using surveillance programmes for infectious diseases as an example, the principles of HMMs can be described as follows: the latent status (*class*) of interest is a herd status regarding infection. This status is evaluated at regular time intervals: HMMs are discrete time models. The status at a given time only depends on the status at the previous time (Markovian property). The status of interest is not directly observed, however, there exists some quantity (such as test results) whose distribution depends on the unobserved status. HMMs have been used for decades

in speech recognition (Rabiner, 1989) and other areas. They have also been used for epidemiological surveillance (Le Strat & Carrat, 1999), although not with longitudinal data from multiple epidemiological units such as herds. The model we developed is therefore a latent class model that takes into account the time dynamics in the latent status. The probability of new infection between consecutive time steps is modelled as a function of risk factors.

Figure 1 shows how surveillance programmes are represented in the model as a succession of discrete time steps. The focus of this model is a latent status evaluated at the herd-month level. This latent status is not directly observed but inferred from its causes and consequences incorporated as data. The consequences are the test results. Test results do not have to be available at every time step for the model to work. The causes of infection are risk factors of infection. In the application presented below, the latent status will be either herd seropositivity or presence of a PI animal in the herd, depending on the testing scheme as well as on the prior distributions put on the characteristics of the tests used. The model estimates this latent status monthly, and predicts it for the last month of data. These herd-month latent statuses will be estimated/predicted from test results (BTM ELISA testing or confirmatory testing) and risk factors (cattle introductions or local seroprevalence) recorded in each herd.

2.1.2 Modelling framework, inputs and outputs

The model is designed to use longitudinal data collected as part of surveillance programmes against infectious diseases. In such programmes, each herd level status is re-evaluated when new data (most commonly test results, but may also be data related to risk factors) are available. The model mimics this situation by predicting the probability of a positive status for all herds in the CP on the last month of available data. Data from all participating herds up to the month of prediction are used as historical data for parameter estimation (Figure 1).

The estimation and prediction are performed within a Bayesian framework using Markov Chain Monte Carlo (MCMC) in the JAGS computer programme (Plummer, 2017). The model encodes the relationships between all the variables of interest in a single model. Each variable is modelled as drawn from a statistical distribution. The estimation requires prior distributions for all the parameters in the model. These priors are a way to incorporate either existing knowledge or hypotheses in the estimation. For example, we may

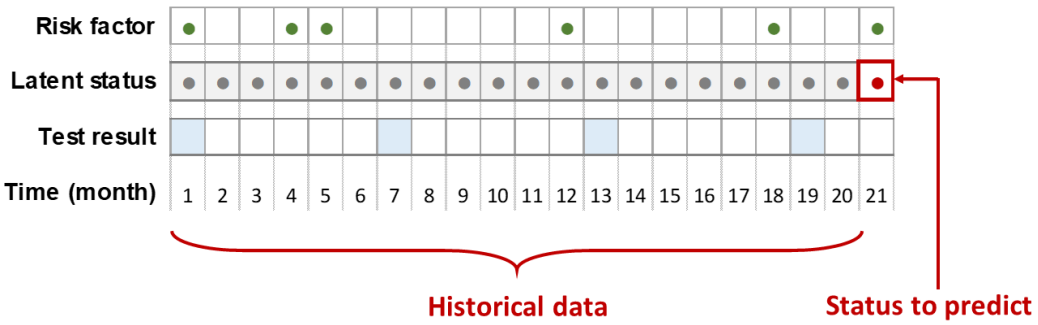

Figure 1: Conceptual representation of the implementation of a surveillance programme within a herd. The focus of the model is the latent status regarding infection, which is modelled at the herd-month level. This status partly depends on risk factors and determines test results. In this diagram, risk factors are represented as green dots when present and available test results as blue shaded squares. The model predicts a probability of infection for the most recent month in the surveillance programme using all the data collected for the estimation of model parameters.

know that the prevalence of herds infected with BVDV in our CP is probably lower than 20%, certainly lower than 30% and greater than 5%. Such constraints can be specified with a Beta distribution. The Beta distribution is bounded between 0 and 1, with 2 parameters α and β determining its shape. With the constraints specified above, we could use as a prior distribution a $Beta(\alpha = 15, \beta = 100)$[1]. If we do not know anything about this infection prevalence (which is rare), we could use a $Beta(\alpha = 1, \beta = 1)$ prior, which is uniform between 0 and 1. From the model specification, the prior distributions and the observed data, the MCMC algorithm draws samples from the posterior distributions of all the variables in the model. These posterior distributions are the probability distributions for the model parameters given the data and the prior distributions. MCMC methods are stochastic and iterative. Each iteration is a set of samples from the joint posterior distributions of all variables in the model. The algorithm is designed to reach the target joint posterior distribution, but at any moment, there is no guarantee that it has done. To overcome this difficulty, several independent instances of the algorithm (i.e. several chains) are run in parallel. For a variable, if all the MCMC draws from the different chains have the same distribution, it can be concluded that the algorithm has reached the posterior distribution. In this case, it is said that the model has converged.

The focus of our model is the monthly latent status of each herd. This latent status depends on the data on occurrence of risk factors and it affects test results. The data used by the model are the test results and risk factors. At each iteration of the MCMC algorithm, given the data and priors, a herd status (0 or 1) and the coefficients for the associations between risk factors, latent status and test results are drawn from their posterior distribution.

In the next 3 sections, the parameters for which prior distributions are required, i.e. test characteristics, status dynamics and risk factor parameters, are described. The outputs of Bayesian models are posterior distributions for all model parameters. Specifically, in our model, the quantities of interest are the herd level probabilities of being latent status positive on the last test month in the dataset as well as test sensitivity, test specificity, infection dynamic parameters and parameters for the strengths of association between risk factors and the probability of new infection. This is described in the

[1]The $Beta(\alpha = 15, \beta = 100)$ distribution has a mean of 0.13 and a standard deviation of 0.03. In R, it can be plotted using the following instructions `curve(dbeta(x, 15, 100))`

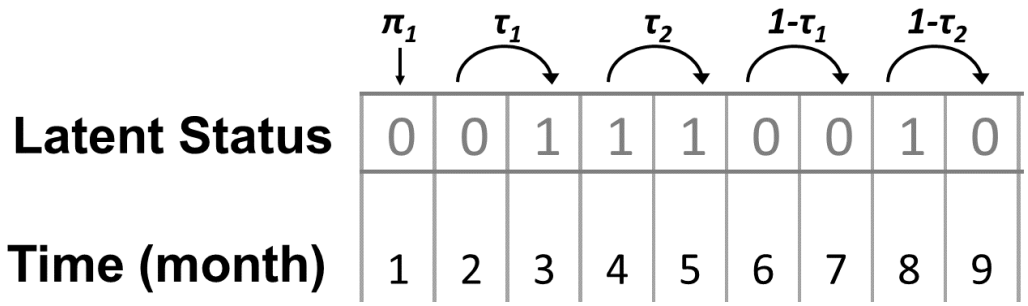

Figure 2: Modelling of infection dynamics. The diagram shows hypothetical latent statuses (0 for negative; 1 for positive) as a function of time in month, with examples of all possible transitions. $\pi_1 = p(S_1^+)$ is the probability of being status positive at the first point in time, $\tau_1 = p(S_t^+|S_{t-1}^-)$ is the probability of becoming status positive and $\tau_2 = p(S_t^+|S_{t-1}^+)$ is the probability of remaining status positive.

corresponding sections.

2.1.3 Latent status dynamics

Between test events, uninfected herds can become infected and infected herds can clear the infection. The model represents the probability of having a positive status at each time step as a function of the status at the previous time step (Figure 2). For the first time step when herd status is assigned, there is no previous status against which to evaluate change. From the second time step when herd status is assigned, and onwards, herds that were status negative on the previous time step have a certain probability of becoming status positive and herds that were status positive have a certain probability of remaining status positive.

These assumptions can be summarised with the following set of equations[2]. The status on the first time step (S_1^+) is a Bernoulli event with a Beta prior on its probability of occurrence:

$$S_1^+ \sim Bernoulli(p(S_1^+)) \tag{1}$$

[2]Statuses are estimated/predicted at the herd-month level. Herd is omitted from the notation to facilitate reading. S_t^+ should be read as S_{ht}^+ where h represents the herd.

$$p(S_1^+) \sim Beta(\pi_{1a}, \pi_{1b}) \tag{2}$$

From the second time step when herd status is assigned, and onwards, a positive status is also a Bernoulli event (S_t^+) with a probability of occurrence that depends on the status at the previous time step as well as on the probability of becoming status positive and the probability of remaining status positive. In this case, the probability of becoming status positive is $\tau_1 = p(S_t^+|S_{t-1}^-)$ and the probability of remaining positive is $\tau_2 = p(S_t^+|S_{t-1}^+)$.

$$S_t^+ \sim Bernoulli(p(S_t^+)) \tag{3}$$

$$p(S_t^+) = (1 - S_{t-1}^+)\tau_1 + S_{t-1}^+\tau_2 \tag{4}$$

$$\tau_1 \sim Beta(\tau_{1a}, \tau_{1b}) \tag{5}$$

$$\tau_2 \sim Beta(\tau_{2a}, \tau_{2b}) \tag{6}$$

Therefore, the status dynamics can be completely described by $p(S_1^+)$, τ_1 and τ_2.

2.1.4 Incorporation of information on risk factors for new infection

The probability of new infection is not the same across herds. For example, herds that introduce a lot of animals or are in areas where infection prevalence is high could be at increased risk of new infection (Qi *et al.*, 2019). Furthermore, the association between a given risk factor and the probability of new infection could be CP dependent. For example, the probability of introducing infection through animal introductions will depend on the infection prevalence in the population from which animals are introduced. As a consequence, estimates for these associations (as presented in the literature) could provide an indication about their order of magnitude, but their precision may be limited. On the other hand, the CPs which are of interest in this work usually generate large amounts of testing data which could be used to estimate the strengths of association between risk factors and new infections within a given CP. The variables that are associated with the probability of new infection could increase the sensitivity and timeliness of detection.

When risk factors for new infection are available, the model incorporates this information by modelling τ_1 as a function of these risk factors through logistic regression, instead of the prior distribution for τ_1.

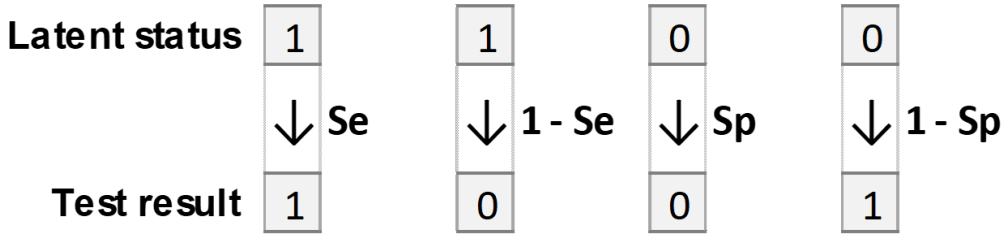

Figure 3: Relation of the model latent status to test result. Sensitivity is the probability of a positive test result in a status positive herd. Specificity is the probability of a negative test result in a status negative herd.

$$logit(\tau_{1ht}) = X_{ht}\theta \tag{7}$$

where X_{ht} is a matrix of predictors for herd h at time t and θ is a vector of coefficients. Normal priors are used for the coefficients of the logistic regression.

$$\theta_i \sim Normal(\mu_i, \sigma_i) \tag{8}$$

2.1.5 Test characteristics

The model allows the inclusion of several test types but for the sake of clarity, we show the model principles for only one test type. These principles can be extended to several tests by specifying prior distributions for all tests.

Tests are modelled as imperfect measures of the latent status (Figure 3). Test sensitivity is the probability of a positive test result given a positive latent status ($Se = p(T^+|S^+)$, refers to true positives) and test specificity is the probability of a negative test result given a negative latent status ($Sp = p(T^-|S^-)$, refers to true negatives).

Test result at time t is modelled as a Bernoulli event with probability $p(T_t^+)$ of being positive.

$$T^+ \sim Bernoulli(p(T_t^+)) \tag{9}$$

The relation between the probability of testing positive, the probability of a positive status, test sensitivity and test specificity is the following:

$$p(T_t^+) = S_t^+ Se + (1 - S_t^+)(1 - Sp) \tag{10}$$

Information or hypotheses regarding test characteristics are incorporated in the model as priors modelled by Beta distributions:

$$Se \sim Beta(Se_a, Se_b) \tag{11}$$

$$Sp \sim Beta(Sp_a, Sp_b) \tag{12}$$

It is important to note that the prior distributions used for sensitivity and specificity will determine what the latent status is. As an example, we consider the detection of BVDV infection with a test that detects BVDV specific antibodies in bulk tank milk. BVDV infection is associated with a long lasting antibody production. There can be cows that are seropositive long after the last PI animal has left the herd. In this situation, using a value of 1 for specificity will define the latent status as any herd with antibody positive cows. However, the herd-level specificity of the test, defined as the probability of a negative test result in a herd with no PI animals, is lower than the animal-level specificity defined as the probability of a negative test result in a sample from an non-PI animal. The specificity of interest, i.e. the detection of farms with PI animals, will depend on the proportion of antibody positive lactating dairy herds that are in farms with PI animals. In turn, this will depend on many factors that are CP dependent such as the prevalence of infection or the proportion of farms that use vaccination against the BVDV. With antibody testing alone, it is therefore difficult to define accurate prior distributions for sensitivity and specificity for the detection of farms with PI animals.

However, it is possible to align the meaning of the latent status with the status of interest. In most CPs, positive routine tests will be followed by confirmatory testing. The objective of routine testing is to detect any potentially infected herd. The tests used for routine testing should be sensitive. The objective of confirmatory testing is to identify truly infected herds among herds positive in routine testing. The testing procedure used for confirmatory testing should be both specific and sensitive. With our model, if these conditions are met and if prior distributions that reflect these hypotheses are used, the posterior distributions for the characteristics of both testing phases

should be more accurate. A useful property of HMMs is that accounting for the status dynamics makes the results of tests performed on different months in the same herd conditionally independent, because the conditional time dependence between statuses is modelled with the dynamics part of the model. For example, if a herd tests positive during routine testing, it will have a higher than average prior probability of infection in subsequent confirmatory testing. As a further consequence of this, the posterior distribution for the specificity of routine testing will depend on the proportion of herds that are confirmed positive in confirmatory testing.

2.1.6 Prediction of a probability of infection

In explaining how predictions are performed we use the following notation: $\tilde{y}$ is the predicted value for y, $\hat{\beta}$ is the estimated value for β. The equation $\tilde{y} = \hat{\beta}.x$ means that the predicted value for y is equal to x (data) times the estimated value for β.

The model predicts herd-level probabilities of infection on the last month in the data mimicking regular re-evaluation as new data come in. If there is no test result available on this month, the predicted probability of being status positive (called $p(\tilde{S}_t^{+*})$) is the predicted status on the previous month times $\tilde{\tau}_{1t}$ if the herd was predicted status negative or times $\hat{\tau}_2$ if the herd was predicted status positive (Table 1)[3]. This can be written as:

$$p(\tilde{S}_t^{+*}) = p(\tilde{S}_t^{+}|\hat{S}_{t-1}^{+}, \tilde{\tau}_{1t}, \hat{\tau}_2) = (1 - \hat{S}_{t-1}^{+}).\tilde{\tau}_{1t} + \hat{S}_{t-1}^{+}.\hat{\tau}_2 \tag{13}$$

where:

$$\tilde{\tau}_{1t} = logit^{-1}(X_t\hat{\theta}) \tag{14}$$

If a test result was available, the prediction must combine information from the test as well as previous information. The way to estimate this predicted probability from $p(\tilde{S}_t^{+*})$ and test results can be derived from Table 1. The predicted probability of being status positive can be computed as:

$$\begin{aligned} p(\tilde{S}_t^{+}|T_t^{+}, \tilde{S}_t^{+*}) = \quad & T_t^{+}.\frac{Se.p(\tilde{S}_t^{+*})}{Se.p(\tilde{S}_t^{+*})+(1-Sp)(1-p(\tilde{S}_t^{+*}))} + \\ & (1 - T_t^{+}).\frac{(1-Se).(1-p(\tilde{S}_t^{+*}))}{(1-Se).(1-p(\tilde{S}_t^{+*}))+Sp.(1-p(\tilde{S}_t^{+*}))} \end{aligned} \tag{15}$$

[3]Here $\tilde{\tau}_{1t}$ is *predicted* from herd-month specific risk factors while $\hat{\tau}_2$ is the same for all herds and *estimated* from historical data.

Table 1: Probability of test result by herd status. Cells on the first row are test positive herds with true positives on the left-hand side and false positives on the right-hand side. Cells on the second row are test negative herds with false negatives on the left-hand side and true negatives on the right-hand side.

		$Herdstatus_t$	
		+	-
$Test_t$	+	$Se.p(S_t^+)$	$(1-Sp)(1-p(S_t^+))$
	-	$(1-Se).p(S_t^+)$	$Sp.(1-p(S_t^+))$

where $T_t^+ = 1$ when the test at time t is positive, $T_t^+ = 0$ when it is negative

2.2 Application of the model to a control programme for BVDV infection in cattle

2.2.1 Data

The model was evaluated on data collected for the surveillance of BVDV infection in cattle in Loire-Atlantique, France. Data were available from 1687 dairy herds between the beginning of 2010 and the end of 2016. Under the programme, each herd was tested twice a year with a bulk tank milk antibody ELISA test. For each campaign of testing, tests were performed for all the herds over a few weeks. Data on the number of cattle introduced into each herd with the associated date of introduction were also available. For the model evaluation, test data from the beginning of 2014 to the end of 2016 were used. Risk factor data collected between 2010 and 2016 were available to model (possibly lagged) associations between risk factors and latent status.

2.2.2 Test results

Test results were reported as optical density ratios (ODR). In the Loire-Atlantique CP, these ODRs are discretised into 3 categories using threshold

values of 35 and 60. ODR values below 35 are associated with low antibody levels and ODR values above 60 are associated with high antibody levels. Decision regarding which herds require further testing for the identification and removal of PI animals is complex and involves the combination of test categories on 3 consecutive tests, spanning a year.

In this work, the ODR values were discretised in order to convert them into either seropositive (antibodies detected) or seronegative (no antibodies detected) outcomes. The choice of the threshold to apply for the discretisation was based on the ODR distribution, which was clearly bimodal. For this purpose, the ODR distribution was modelled as a mixture of 2 normal distributions using the R mixdist package (Macdonald & Du, 2018). Assuming that one of the distributions was associated with seronegativity and the other one with seropositivity, the threshold that discriminated best between the 2 distributions was selected.

2.2.3 Selection of risk factors

A difficulty in the evaluation of putative risk factors was that Bayesian models usually take time to run, especially with large datasets as used here. It was therefore not possible to perform this selection with our Bayesian model. To circumvent this problem, logistic models as implemented in the R glm function (R Core Team, 2019) were used[4]. The outcome of these models was seroconversion defined as a binary event, and covariates of interest were risk factors for becoming status positive as defined through the τ_1 variable. All herds with 2 consecutive test results whose first result was negative (ODR below the chosen threshold) were capable of seroconverting. Of these herds, the ones that had a positive result (ODR above the chosen threshold) on the second test were considered as having seroconverted. The time of event (seroconversion or not) was considered the mid-point between the 2 tests.

Two types of risk factors of new infection were evaluated: infection through cattle introductions and infection through neighbourhood contacts (Qi *et al.*, 2019). Cattle introduction variables were constructed from the number of animals introduced into a herd on a given date. In addition to the raw number of animals introduced, the natural logarithm of the number of animals (+1 because ln(0) is not defined) was also evaluated. This was to allow a decreasing effect of each animal as the number of animals introduced increased. Regard-

[4]The functions used to perform this evaluation are included in the STOCfree package.

ing the neighbourhood risk, the test result data were used. For each testing campaign, the municipality-level prevalence of test positives (excluding the herd of interest) was calculated, and is subsequently termed 'local prevalence'. It was anticipated that when local seroprevalence would increase, the probability of new infection in the herd of interest would increase as well.

For all candidate variables, a potential problem was delayed detection, which relates to the fact that a risk factor recorded at one point in time may be detected through testing much later, even if the test is sensitive. For example, if a trojan cow (a non-PI female carrying a PI calf) is introduced into a herd, the lactating herd will only seroconvert when the PI calf is born and has had contact with the lactating herd. Therefore, for each candidate variable, the data were aggregated between the beginning of an interval (labelled lag1, in months from the outcome measurement) and the end of this interval (labelled lag2, in months from the outcome measurement). Models with all possible combinations of time aggregation between lag1 and lag2 were run, with lag1 set to 0 and lag2 set to 24 months. The best variables and time aggregation interval were selected based on low AIC value, biological plausibility and suitability for the Bayesian model.

2.2.4 Bayesian models

Four different Bayesian models were considered. For all models, historical data were used for parameter estimation and the probability of infection on the last month in the dataset was predicted.

Model 1 - Perfect routine test: in order to evaluate the monthly dynamics of seropositivity and seronegativity, the Bayesian model was run without any risk factor and assuming that both test sensitivity and test specificity were close to 1. The prior distributions for sensitivity and specificity were $Se \sim Beta(10000, 1)$ (percentiles: 5 = 1, 50 = 1, 95 = 1) and $Sp \sim Beta(10000, 1)$. Regarding infection dynamics, prior distributions were also specified for the prevalence of status positives (also test positives in this scenario) on the first testing time $p(S_1^+) \sim Beta(1, 1)$ (uniform on 0-1), the probability of becoming status positive $\tau_1 \sim Beta(1.5, 10)$ (percentiles: 5 = 0.017, 50 = 0.109, 95 = 0.317), and the probability of remaining status positive $\tau_2 \sim Beta(10, 1.5)$ (percentiles: 5 = 0.683, 50 = 0.891, 95 = 0.983).

Model 2 - Perfect routine test and risk factors: in order to quantify the association between risk factors and the probability of becoming status positive if the test were close to perfect, the Bayesian model was run with the risk factors identified as associated with seroconversion on the previous step and using the same priors for sensitivity, specificity and τ_2 as in Model 1 ($Se \sim Beta(10000, 1)$, $Sp \sim Beta(10000, 1)$, $\tau_2 \sim Beta(10, 1.5)$). The priors for risk factors were specified as normal distributions on the logit scale. The prior for the intercept was $\theta_1 \sim \mathcal{N}(-3, 1)$ (on the probability scale - percentiles: 5 = 0.01, 50 = 0.047, 95 = 0.205). This represented the prior probability of a new infection in a herd purchasing no animal and with a local seroprevalence of 0. The priors for the other model coefficients were centred on 0 with a standard deviation of 2. On the logit scale, values of -4 (2 standard deviations in this case) correspond to probabilities close to 0 (logit(-4) = (0.018) and values of 4 to probabilities that are close to 1 (logit(4) = (0.982).

Model 3 - Imperfect routine test and risk factors: the objective of this model was to incorporate the uncertainty associated with test results in both parameter estimation and in the prediction of the probabilities of infection. The priors for test sensitivity and specificity were selected based on the ODR distributions for seronegatives and seropositives identified by the mixture model. The following prior distributions were used: $Se \sim Beta(5000, 260)$ (percentiles: 5 = 0.946, 50 = 0.951, 95 = 0.955) and $Sp \sim Beta(5000, 260)$. For the associations between risk factors and the probability of new infection, the same prior distributions as in Model 2 were used.

Model 4 - Imperfect routine test, confirmatory testing and risk factors: the objective of this model was to assess the impact of confirmatory testing. The same prior distributions as in scenario 3 were used. In this case however, every time a positive test result was recorded, a new confirmatory test was randomly generated in the following month so that 85% of these tests were positive and 15% were negative. The confirmatory test was assumed to have both a sensitivity and a specificity close to 1.

For each model, 4 chains were run in parallel. The first 5 000 MCMC iterations were discarded (burn-in). The model was run for 5 000 more iterations of which 1 in 20 was stored for analysis. This yielded 1 000 draws

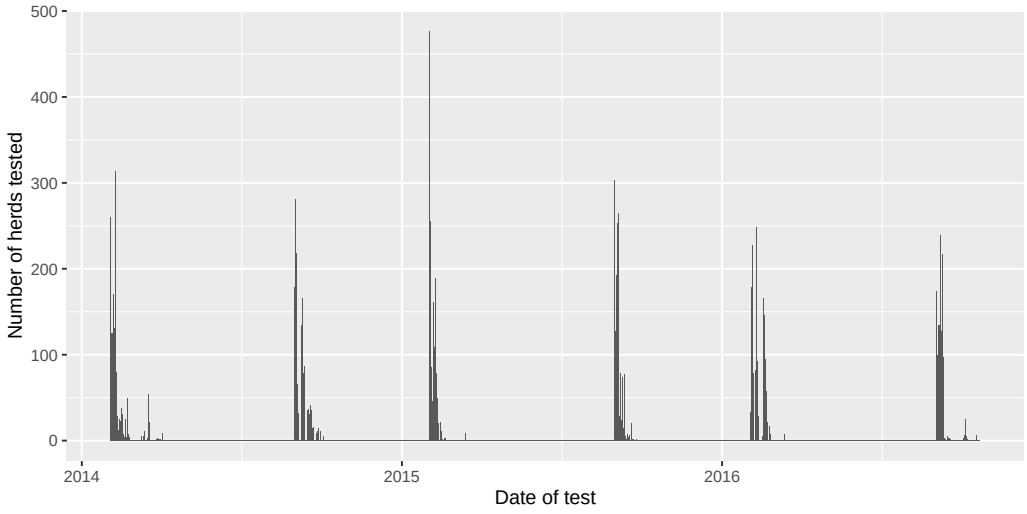

Figure 4: Distribution of the test dates between 2014 and 2017 in 1687 herds from Loire-Atlantique, France.

from the posterior distribution of each parameter. Convergence was assessed visually using traceplots. Each distribution was summarised with its median and 95% credibility interval.

3 Results

3.1 Test results

There were 9725 available test results from 1687 herds. Most herds were tested in February and September (See Figure 4). Two normal distributions were fit to the ODR data using the R mixdist package (Figure 5). The distribution for seronegatives had a mean and standard deviation of 7.1 and 16.3 respectively. The distribution for seropositives had a mean and standard deviation of 57 and 13 respectively. There were 58.6% and 41.4% of observations in the seronegative and seropositive distributions respectively. ODR values above 35 (21% of ODR values) were categorised as test positive and ODR values below 35 were categorised as test negative. The sensitivity and the specificity of the threshold value of 35 for the classification of test results with respect to seropositivity were estimated using the fitted distributions as the gold standard. These estimated sensitivity and specificity were 0.956

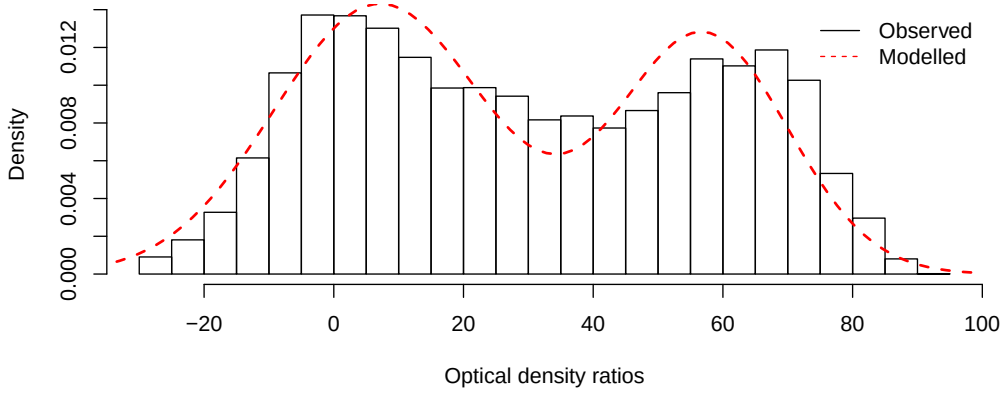

Figure 5: Distribution of the observed optical density ratios (histogramme) and fitted mixture of normal distributions (red dashed curves) for the bulk tank milk test results used in the analyses.

and 0.955 respectively. In the Bayesian models in which the latent status was seropositivity, the prior distributions for sensitivity and specificity were centred on these values.

3.2 Selection of risk factors

Risk factors related to animal introductions and seroprevalence were evaluated with logistic models. The model outcome was a seroconversion event. A first step of the analysis was, for each variable, to identify the time interval that was the most predictive of an observed seroconversion. Figure 6 presents the AIC values associated with each possible interval for the variables ln(Number of animals introduced + 1) and local seroprevalence.

For the animal introduction variables, for the same time interval, the AICs of the models of the untransformed number of animals were higher than the ones for the log transformed values (not shown). It can also be noted that considering longer intervals (further away from the diagonal) was usually better than considering short intervals (close to the diagonal). It may be that some herds never buy any animal while, on average, herds that buy once have already done it in the past. In this case, it is possible that

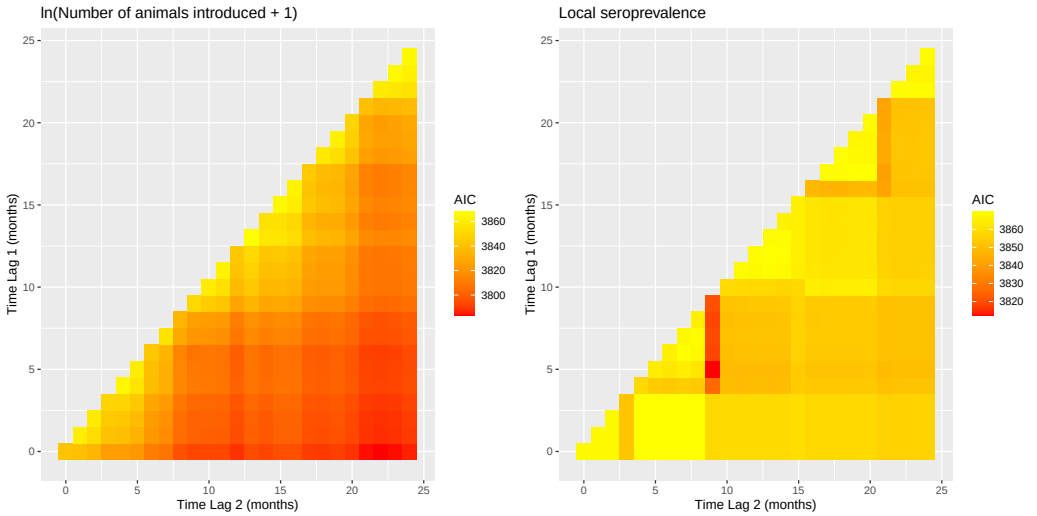

Figure 6: AIC values associated with logistic models of the association between 2 variables and the probability of seroconversion between 2 tests. The variable evaluated on the left-hand side panel is the sum of the log(number of animals introduced + 1) between lag1 and lag2. The variable evaluated on the right-hand side panel is the max of the local seroprevalence between lag1 and lag2.

the infection was introduced several times, while it is not possible to know which animal introduction was associated with herd seroconversion. This could explain the apparent cumulative effect of the number of introductions. The cells that are close to the diagonal are associated with short intervals. Considering one month intervals, the probability of infection was highest for introductions made 8 months from the month of seroconversion.

Local seroprevalence was evaluated from data collected in 2 different testing campaigns per year, as shown in Figure 4. For this reason, in the investigation of lagged relationships between local seroprevalence and the probability of seroconversion, the maximum local seroprevalence was computed, and not the sum as for the number of animals introduced. The strength of association between local seroprevalence and herd seroconversion was greatest for local seroprevalence 9 months prior to herd seroconversion.

A final multivariable logistic model with an animal introduction variable and a local seroprevalence variable was constructed. In the choice of the time intervals to include in this model, the following elements were consid-

Table 2: Results of the final logistic model of the probability of seroconversion between consecutive tests.

	lag1	lag2	Estimate	p-value
Intercept	-	-	-1.96	7.99e-306
ln(Number animals introduced +1)	8	8	0.38	5.70e-10
local seroprevalence	9	9	4.59	3.39e-13

ered. First, the Bayesian model runs with a monthly time step. Aggregating data over several months would result in including the same variable several times. Secondly, historical data may sometimes be limited. Having the smallest possible value for the end of the interval could be preferable. For this reason the variables considered for the final model were the natural logarithm of the number of animals introduced 8 months prior to the month of seroconversion as well as the local seroprevalence 9 months prior to the month of seroconversion. The results of this model are presented in Table 2. All variables were highly significant. The model intercept was the probability of seroconversion in a herd introducing no animals and with local seroprevalence of 0 in each of the time intervals considered. The probability of seroconversion between 2 tests corresponding to this scenario was of 0.124. Buying 1, 10 or 100 animals increased this estimated probability to 0.171, 0.866 and 1 respectively. Buying no animals and observing a seroprevalence of 0.2 (proportion of seropositives in the dataset) was associated with a probability of seroconversion of 0.261.

3.3 Bayesian models

Running each of the 4 models for the 1687 herds with 3 years of data took on average 7 hours per model. In models 2 to 4, the candidate covariates were the natural logarithm of the number of animals introduced 8 months before status evaluation/prediction as well as the local seroprevalence 9 months prior. The 95% credibility interval for the estimated coefficient associated with local seroprevalence included 0. This variable was therefore removed from the models and only cattle introductions were considered.

Table 3: Median (2.5%, 97.5%) of the parameter posterior distributions used in the 4 Bayesian models evaluated. Model 1: Perfect routine test; Model 2: Perfect routine test and risk factors; Model 3: Imperfect routine test and risk factors; Model 4: Imperfect routine test, confirmatory testing and risk factors.

Parameter	Model 1	Model 2	Model 3	Model 4
Se BTM ODR	1 (0.999, 1)	1 (1, 1)	0.948 (0.942, 0.953)	0.949 (0.944, 0.955)
Se confirmatory	-	-	-	0.976 (0.973, 0.98)
Sp BTM ODR	1 (0.999, 1)	1 (0.999, 1)	0.932 (0.926, 0.938)	0.971 (0.964, 0.978)
Sp confirmatory	-	-	-	1 (1, 1)
τ_1	0.029 (0.027, 0.032)	-	-	-
τ_2	0.965 (0.962, 0.967)	0.964 (0.961, 0.967)	0.994 (0.993, 0.996)	0.974 (0.97, 0.977)
θ_1 (Intercept)	-	-3.631 (-3.718, -3.545)	-4.803 (-4.985, -4.646)	-3.825 (-3.94, -3.711)
θ_2	-	0.589 (0.482, 0.684)	0.682 (0.522, 0.813)	0.665 (0.547, 0.776)

3.3.1 Model parameters

Figure 7 and Table 3 show the distributions of model parameters for the 4 models. Figure 8 shows the predicted probability of becoming status positive as a function of the number of animals introduced 9 months before status evaluation.

In Models 1 and 2, the prior distributions put on sensitivity and specificity were very close to 1. With these models, the latent status corresponded to the test result. In effect, they modelled the monthly probabilities of transition between BTM test negative and BTM test positive. In this case, the median (percentile 2.5 - percentile 97.5) probability of becoming status positive between consecutive months was 0.029 (0.027 - 0.032). This represents a probability of becoming status positive over a 12 month period of 0.298 (0.280 - 0.323). For status positive herds, the monthly probability of remaining positive was of 0.965 which represents a probability of still being status positive 12 months later of 0.652 (0.628-0.669). In model 2, a risk factor was incorporated into the estimation. The model intercept was much lower than the estimate from the logistic model estimated in the variable selection step. This was due to the different time steps considered (1 month vs. half a year). On the other hand, the estimate for the log number of animals introduced was higher.

In model 3, the prior distributions for test sensitivity and specificity were centred on 0.95 based on the mixture of 2 normal distributions for seroneg-

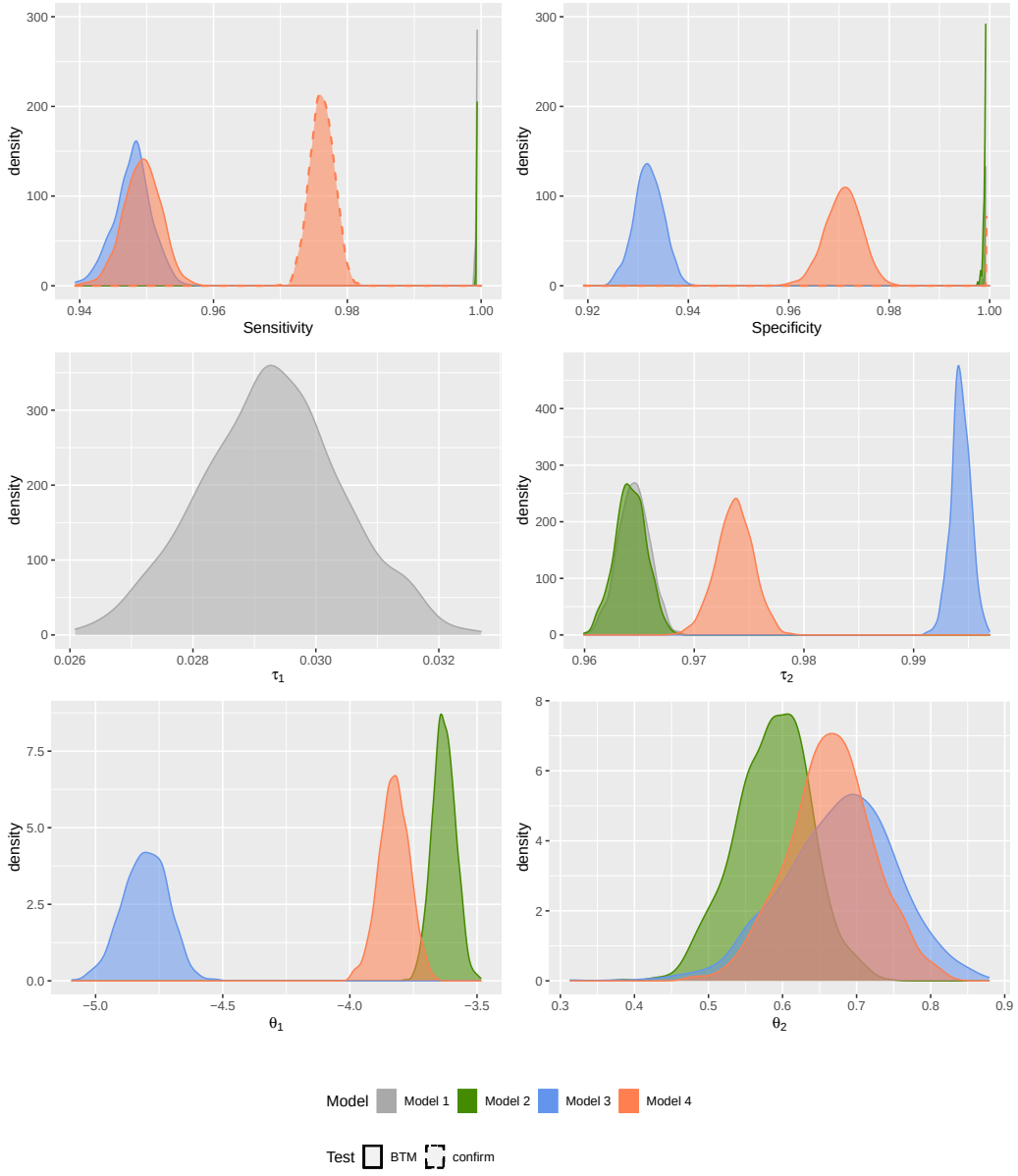

Figure 7: Parameters posterior distributions for the 4 Bayesian models. Model 1: Perfect routine test; Model 2: Perfect routine test and risk factors; Model 3: Imperfect routine test and risk factors; Model 4: Imperfect routine test, confirmatory testing and risk factors. Sensitivities and specificities close to 1 are not shown to facilitate reading. The dashed lines correspond to the distributions of the confirmatory tests. Parameters related to status dynamics are τ_1 (probability of becoming status positive between consecutive months) and τ_2 (probability of remaining status positive). τ_1 was only estimated for the model without risk factors (model 1). The parameters for the association between risk factors and the probability of becoming status positive are θ_1 and θ_2. θ_1 is the intercept of the logistic model and θ_2 is the coefficient associated with the log of the number of animals introduced 8 months before status evaluation/prediction.

atives and seropositives that described best the BTM ODR data (see Section 3.1). With this model, the latent status corresponded to seropositivity. This assumption allowed the effect of having an imperfect test on the estimation of the different model parameters to be investigated. In this scenario, the posterior distribution for sensitivity was close to the prior, but the posterior for the specificity was slightly lower. On the other hand, the distribution for τ_2 was higher than when the test was considered perfect. This implies that the model identified some test positives as false positives, but that the ones that retained a positive status remained positive for longer. Compared to Model 2, the probability of becoming status positive was lower in herds buying no animals (model intercept), and tended to increase more rapidly with the number of animals introduced (θ_2), although for 100 animals introduced, the probability of becoming status positive was still lower than with the other models (Figure 8). Because of the imperfect sensitivity of routine testing, some herds that were seronegative at a test while seropositive at the previous or following tests were classified as false negative by the model and thereby were not included in the estimation of τ_1, which may have decreased the estimated strength of association between cattle introduction and new infection. However, the estimates produced by this should be more accurate.

In model 4, confirmatory testing was added, with a testing procedure assumed to have perfect sensitivity and specificity for the detection of farms with infected animals. This resulted in several differences with model 3, which illustrate the interplay between data and prior information. The added confirmatory negative results often contradicted the data because, they were generally followed by a positive routine test. This had the following consequences. The posterior distribution for the sensitivity of confirmatory testing was lower than its prior distribution, indicating that herds negative to confirmatory testing were classified as false negatives more often than suggested by the priors. The fact that the estimated value for the specificity of BTM testing was higher than in Model 3 shows that herds positive to routine testing were considered to be true positives slightly more often. The fact that the estimated value for τ_2 was lower than in Model 3 shows that status positive herds tended to clear infection more quickly, which allowed a more rapid status change between routine and confirmatory testing. Because Model 4 resulted in more frequent changes in status, the coefficients for the association between cattle introduction and new infections (Figure 8) were closer between Model 4 and Model 2 than between Model 4 and Model 3.

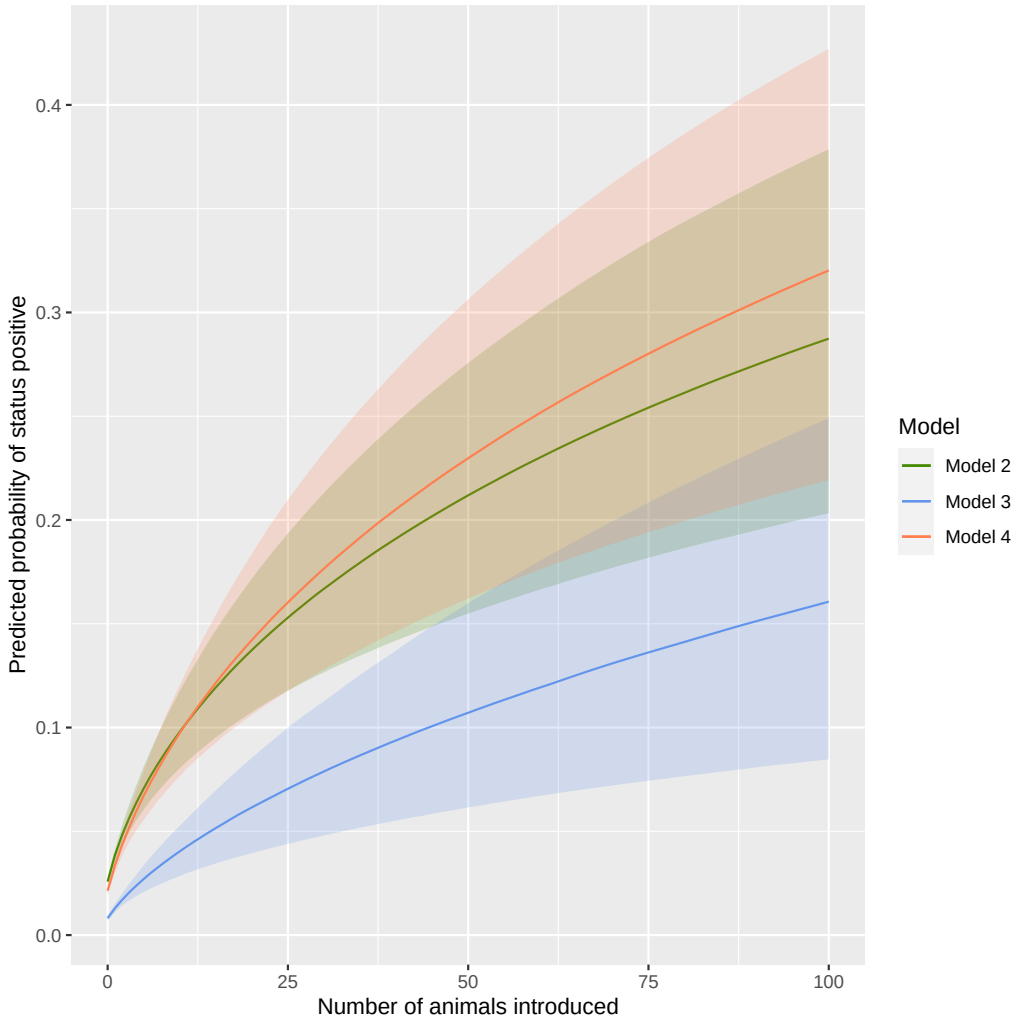

Figure 8: Predicted probability of new infection as a function of the number of animals introduced 8 months before the month of interest for the Bayesian models 2 to 4. Model 2: Perfect routine test and risk factors; Model 3: Imperfect routine test and risk factors; Model 4: Imperfect routine test, confirmatory testing and risk factors. The lines represent the median predicted values. The shaded areas represent the 95% credibility intervals.

3.3.2 Predicted probabilities of infection

Figure 9 shows the distributions of herd-level probabilities of infection predicted by the 4 Bayesian models. These probability distributions are bimodal for all models. The left-hand side corresponds to herds that were predicted status negative on the month before the month of prediction. These are associated to becoming status positive, i.e. τ_1. The right-hand side of the distributions corresponds to herds that were predicted status positive on the month before the month of prediction. These are associated to remaining status positive, i.e. τ_2. For models 3 and 4, which incorporate both risk factors and test uncertainty, the modes are closer to 0 and 1 than for the other 2 models. For Model 4, there is a third mode between 0.4 and 0.5. This mode was associated with confirmatory testing.

Figure 10 shows the distributions of the predicted probability of being status positive for 4 herds. It can be seen that herds that were consistently negative (positive) to the test had extremely low (high) probabilities of being status positive. Accounting for the number of animals introduced increased the probability of infection in the herds that were test negative.

4 Discussion

This article describes a statistical framework for the prediction of an infection related status from longitudinal data generated by CPs against infectious diseases of farm animals. The statistical model developed estimates a herd level probability of being *latent status* positive on a specific month, based on input data that can vary in terms of the types of test used, frequency of testing and risk factor data. This is achieved by modelling the latent status with the same discrete time step, regardless of the frequency with which input data are available, and by modelling changes in the latent status between consecutive time steps. This model therefore fulfils one of our main objectives which was to be able to integrate heterogeneous information into the estimation. However, in order to be able to compare the output of this model run on data from different CPs, the definition of the latent status should be the same.

In this model, the latent status is mostly defined by the prior distributions put on the different model parameters. In setting the prior distributions there are two issues: setting the distribution's central value (mean, median

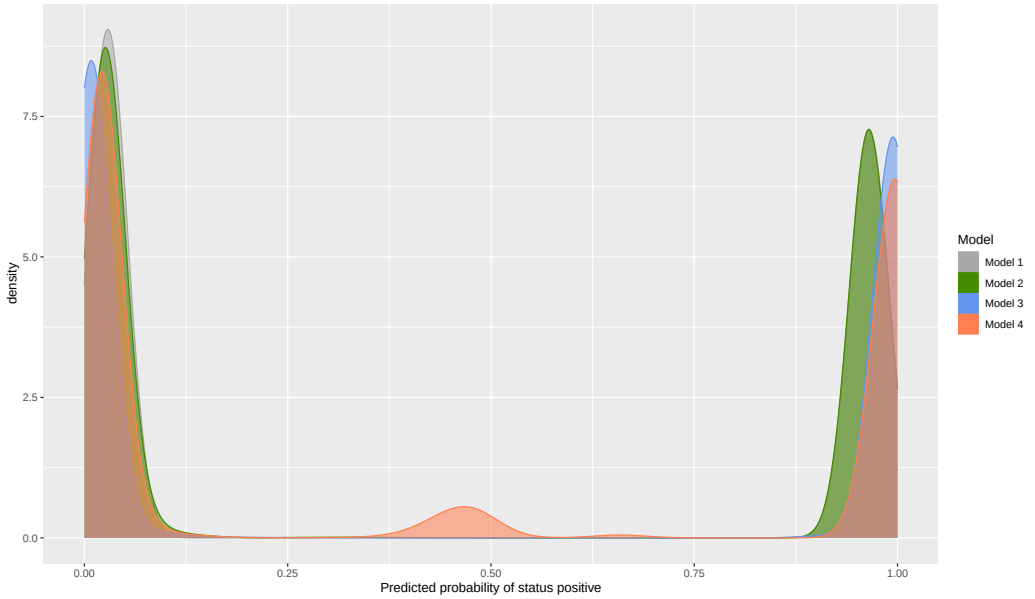

Figure 9: Distributions of the predicted probabilities of being status positive for all herds with the 4 Bayesian models evaluated. Model 1: Perfect routine test; Model 2: Perfect routine test and risk factors; Model 3: Imperfect routine test and risk factors; Model 4: Imperfect routine test, confirmatory testing and risk factors.

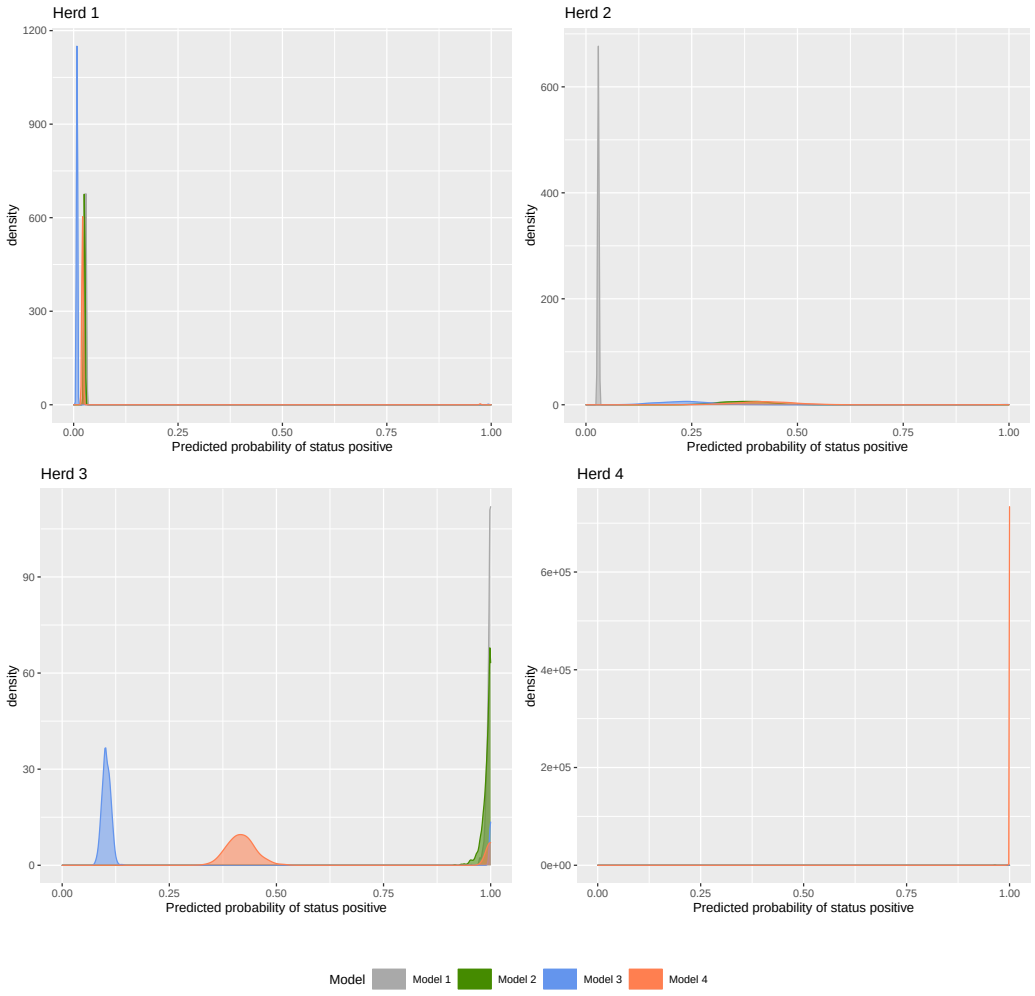

Figure 10: Distribution of predicted probabilities of being status positive on the month of prediction for 4 herds with the 4 models compared. Model 1: Perfect routine test; Model 2: Perfect routine test and risk factors; Model 3: Imperfect routine test and risk factors; Model 4: Imperfect routine test, confirmatory testing and risk factors. Herd 1 was test negative for 6 consecutive tests, introduced no animal. Herd 2 was test negative for 6 consecutive tests, introduced animals regularly (196 associated with the month of prediction). Herd 3 was test negative on the first 5 tests and test positive on the month of prediction, introduced animals regularly (3 introductions associated with the month of prediction). Herd 4 was test negative on the first 2 tests and test positive on the last 4 tests, introduced animals regularly.

638 ...) and setting the distribution width. Choosing the wrong central value,
639 i.e. the prior distribution does not include the true parameter value, can lead
640 to systematic error (bias) or absence of convergence. This problem will be
641 more important as prior distributions become narrower. Setting prior distri-
642 butions that are too wide can lead to a lack of convergence, when multiple
643 combinations of parameter values are compatible with the data. This was a
644 problem in initial modelling of the BVDV data (not shown). Putting narrow
645 prior distributions on test sensitivity and test specificity allowed the model
646 to converge. These narrow distributions imply very strong hypotheses on
647 test characteristics.

648 The definition of prior distributions for test characteristics that reflect
649 the latent status of interest is challenging (Duncan *et al.*, 2016). This was
650 apparent in the application to infection by the BVDV we presented. For
651 the trade of animals from herds that are free from infection by the BVDV,
652 the latent status of interest was the *presence of at least one PI animal in*
653 *the herd*. The test data available to estimate the probability of this event
654 were measures of bulk tank milk antibody levels which were used to define
655 seropositivity as a binary event. Although milk antibody level is associated
656 with the herd prevalence of antibody positive cows (Beaudeau *et al.*, 2001),
657 seropositive cows can remain long after all the PIs have been removed from
658 a herd. Furthermore, vaccination induces an antibody response which may
659 result in vaccinated herds being positive to serological testing regardless of
660 PI animal presence (Raue *et al.*, 2011; Booth *et al.*, 2013). Therefore, the
661 specificity of BTM seropositivity, i.e. the probability for herds with no PI
662 animals to be test negative, is less than 1. More importantly, this specificity
663 depends on the context; i.e. on the CP. PI animals can be identified and
664 removed more or less quickly depending on the CP, the proportion of herds
665 vaccinating and the reasons for starting vaccination can differ between CPs.
666 Test sensitivity can also be imperfect. Continuing with the example of bulk
667 tank milk testing, contacts between PI animals present on the farm and the
668 lactating herd may be infrequent, which would decrease sensitivity. The
669 probability of contact between PI animals and the lactating herd depends
670 on how herds are organised, which could vary between CPs. Furthermore,
671 the contribution of each seropositive cow to the BTM decreases as herd size
672 increases which can result in differences in BTM test sensitivity associated
673 with different herd sizes between CPs.

674 The effects of using different prior distributions for test characteristics
675 on latent status definition, parameter estimation and probability prediction

were evaluated. In models 1 and 2, the dichotomised BTM antibody test results were modelled assuming perfect sensitivity and perfect specificity. With these assumptions, the latent status was the dichotomised test results. In Model 3, the BTM test was assumed to have both a sensitivity and a specificity concentrated around 95%, based on the normal distributions associated with seronegativity and seropositivity identified by a mixture model. The latent status in Model 3 can therefore be described as *seropositivity*. Because overall the probability of changing status was small, assuming an imperfect sensitivity lead to isolated negative test results in sequences of mostly positive test results to be considered false negatives, as shown by the increase in the estimated value for τ_2 between Model 2 and Model 3. This illustrates that in addition to test characteristics, status dynamics will determine the latent status within herds. Model 4 was constructed to evaluate the impact of incorporating confirmatory testing into the model. In CPs, herds that test positive are usually re-tested in order to rule out a false positive test, and to identify infected animals if needed. The testing procedure used in confirmatory testing usually has a high sensitivity and a higher specificity than routine testing in relation to the gold standard. When incorporated into the model, this high quality information, in conjunction with wider prior distributions on routine testing specificity, should allow the posterior distribution of the specificity of routine testing to be revised towards the gold standard. Indeed, if a confirmatory test comes back negative, then the corresponding latent status will become negative with high probability. Given the low probability of becoming status negative between consecutive months, the latent status on the month of routine testing has an increased probability of being negative, leading to a decrease in the specificity of routine testing. This could not be adequately demonstrated in Model 4, because simulating test results at random was often not consistent with patterns of test results in individual herds. However, this confirmed the importance of status dynamics in estimating the latent status.

Status dynamics contributed to the definition of the latent status in several ways. Negative test results interspersed with sequences of positive test results will be classified as latent status positive (i.e. as false negatives) more often as test sensitivity decreases and τ_2 increases. Positive test results interspersed with sequences of negative test results will be classified as latent status negative (i.e. as false positives) with increased frequency as test specificity and τ_1 each decrease. With a perfect test (sensitivity and specificity equal to 1), the model can learn the values of τ_1 and τ_2 from the data, and

the prior distributions put on these parameters can be uninformative. With decreasing values for test sensitivity and specificity, the information provided through the prior distributions put on τ_1 and τ_2 becomes increasingly important. The informative value of τ_1 and τ_2 will increase as the probability of transition between latent status negative and latent status positive decrease, i.e. when τ_1 is small and τ_2 is high.

When data on risk factors of new infection are available, the τ_1 parameter is modelled as a function of these risk factors using logistic regression. In such a case, prior distributions are put on the parameters of the logistic regression and not on the the τ_1 parameter. In the application that we presented, we used a prior distribution corresponding to a low probability of new infection in the reference category (intercept: herds which introduced no animals) and we centred the prior distribution for the association with cattle introductions on a hypothesis of no association (mean = 0 on the logit scale). This allowed the model to estimate the association between the risk factor and the latent status from historical data and to use the estimated association to predict probabilities of being latent status positive on the month of prediction. As expected, the prior distributions put on test characteristics had an impact on the parameter estimates. In Model 3, the model intercept was lower and the estimated association between becoming latent status positive and cattle introduction was higher than in the other models. The most likely explanation for this is that Model 3 allowed the highest level of discrepancy between dichotomised test result and latent status, while assuming a low probability of changing status between months. This resulted in negative test results in herds that were regularly positive to be classified as latent status positive (false negatives, associated with lower test sensitivity, see Table 3) thereby removing opportunities for new infections in herds that were regularly positive while also buying animals. This would imply that the estimated association from model 3 is more closely associated with new infections than estimates from the other models because herds that are regularly test positive have less weight in the estimation. It would also have been possible to base the prior distributions for the model coefficients on published literature. Unfortunately, estimates of the strengths of association between risk factors and the probability of new infection are not readily available from the published literature or are hard to compare between studies (van Roon *et al.*, 2020a). However, estimates from the literature could allow the prior distributions to be bounded within reasonable ranges.

Because the model takes a lot of time to run, the variables included in

the logistic regression were first identified with logistic models estimated by maximum likelihood. This confirmed the importance of animal introduction and neighbourhood contacts in new infections (Qi *et al.*, 2019). However, in the Bayesian models, the 95% credibility for the association between local seroprevalence and new infection included 0 and this variable was therefore not included. The reason for this was not elucidated in this work. Other risk factors such as herd size, participation in shows or markets, the practice of common grazing have shown a consistent association with the probability of new infection by the BVDV (van Roon *et al.*, 2020a). These variables were not included in our model because the corresponding data were not available. One advantage of our approach is the possibility to choose candidate risk factors to include in the prediction of infection based on the data available in a given CP. The associations between the selected putative risk factors and the probability of new infection can be estimated from these data.

Given the reasonably good performance of tests for the detection of BVDV infection, the main advantage of incorporating these risk factors was not to complement the test results on a month a test was performed, but rather to enhance the timeliness of detection. Risk factors that are associated with new infection will increase the predicted probability of infection regardless of the availability of a test result. Therefore, when testing is not frequent, infected herds could be detected more quickly if risk factors of infection are recorded frequently. If the available data on risk factors of new infection captured all the possible routes of new infection, it would be possible to perform tests more frequently in herds that have a higher probability of infection as predicted by our model. In other words, our model could be used for risk-based surveillance (Cameron, 2012).

In the CP from which the current data were used, herds are tested twice a year. This could lead to a long delay between the birth of PI calves and their detection through bulk tank milk testing. We addressed this problem of *delayed detection* by proposing a method for the investigation of lagged relationships between risk factor occurrence and new infections, and by including lagged risk factor occurrences in the prediction of the probability of infection. In our dataset, herds purchasing cattle were more likely to have seroconverted 8 months after the introduction. In the Bayesian model, cattle introduction was modelled as affecting the probability of becoming status positive 8 months after the introduction. It can be argued that infection is present but not detected during this period, as the expression *delayed detection* suggests, and that the probability of infection should increase as soon

as risk factor occurrence is recorded. Modelling this phenomenon would be possible by decreasing the test sensitivity for a period corresponding to the lag used in the current version of the model. This would imply that for this duration, any negative BTM test result would not provide any information about the true status regarding infection and that the herd would have an increased predicted probability of infection. This could be incorporated in future versions of the model.

There are several questions related to this modelling framework that would require further work. The model outputs are distributions of herd level probabilities of infection. Defining herds that are free from infection from these distributions will require decision rules to be developed based on distribution summaries (likely a percentile) and cut-off values. It would also be possible to model the probability of remaining infected between consecutive tests (τ_2) as a function of the control measures put in place in infected herds. Another area that requires further investigations is the evaluation of the modelling framework against a simulated gold standard to determine whether it provides an added value compared to simpler methods. The availability of the model code as a Github repository allows interested users to improve or suggest improvements to our modelling framework.

Acknowledgments

This work is part of the STOC free project that was awarded a grant by the European Food Safety Authority (EFSA) and was co-financed by public organizations in the countries participating in the study.

We thank Groupement de Défense Sanitaire de Loire-Atlantique (GDS-44) for providing the data.

Conflict of interest disclosure

The authors of this article declare that they have no financial conflict of interest with the content of this article.

References

Beaudeau, F., Belloc, C., Seegers, H., Assié, S., Pourquier, P., & Joly, A. 2001. Informative value of an indirect enzyme-linked immunosorbent assay (ELISA) for the detection of bovine viral diarrhoea virus (BVDV) antibodies in milk. *Journal of veterinary medicine. B, Infectious diseases and veterinary public health*, **48**(Nov.), 705–712.

Booth, R. E., Cranwell, M. P., & Brownlie, J. 2013. Monitoring the bulk milk antibody response to BVDV: the effects of vaccination and herd infection status. *The Veterinary record*, **172**(Apr.), 449.

Cameron, A. R. 2012. The consequences of risk-based surveillance: Developing output-based standards for surveillance to demonstrate freedom from disease. *Preventive veterinary medicine*, **105**(Aug.), 280–286.

Collins, J., & Huynh, M. 2014. Estimation of diagnostic test accuracy without full verification: a review of latent class methods. *Statistics in medicine*, **33**(Oct.), 4141–4169.

Duncan, A. J., Gunn, G. J., & Humphry, R. W. 2016. Difficulties arising from the variety of testing schemes used for bovine viral diarrhoea virus (BVDV). *The Veterinary record*, **178**(Mar.), 292.

Fernandes, L. G., Denwood, M. J., de Sousa Américo Batista Santos, C., Alves, C. J., Pituco, E. M., de Campos Nogueira Romaldini, A. H., De Stefano, E., Nielsen, S. S., & Santos de Azevedo, S. 2019. Bayesian estimation of herd-level prevalence and risk factors associated with BoHV-1 infection in cattle herds in the State of Paraíba, Brazil. *Preventive veterinary medicine*, **169**(Aug.), 104705.

Hui, S. L., & Walter, S. D. 1980. Estimating the error rates of diagnostic tests. *Biometrics*, **36**(Mar.), 167–171.

Johnson, W. O., Gardner, I. A., Metoyer, C. N., & Branscum, A. J. 2009. On the interpretation of test sensitivity in the two-test two-population problem: assumptions matter. *Preventive veterinary medicine*, **91**(Oct.), 116–121.

Le Strat, Y., & Carrat, F. 1999. Monitoring epidemiologic surveillance data using hidden Markov models. *Statistics in medicine*, **18**(Dec.), 3463–3478.

Macdonald, P., & Du, J. 2018. *mixdist: Finite Mixture Distribution Models*. R package version 0.5-5.

Martin, P. A. J., Cameron, A. R., & Greiner, M. 2007. Demonstrating freedom from disease using multiple complex data sources 1: a new methodology based on scenario trees. *Preventive veterinary medicine*, **79**(May), 71–97.

Norström, M., Jonsson, M. E., Åkerstedt, J., Whist, A. C., Kristoffersen, A. B., Sviland, S., Hopp, P., & Wahlström, H. 2014. Estimation of the probability of freedom from bovine virus diarrhoea virus in Norway using scenario tree modelling. *Preventive veterinary medicine*, **116**(Sept.), 37–46.

Nusinovici, S., Madouasse, A., Hoch, T., Guatteo, R., & Beaudeau, F. 2015. Evaluation of Two PCR Tests for *Coxiella burnetii* Detection in Dairy Cattle Farms Using Latent Class Analysis. *PloS one*, **10**, e0144608.

Plummer, M. 2017. *JAGS version 4.3.0 user manual*.

Qi, L., Beaunée, G., Arnoux, S., Dutta, B. L., Joly, A., Vergu, E., & Ezanno, P. 2019. Neighbourhood contacts and trade movements drive the regional spread of bovine viral diarrhoea virus (BVDV). *Veterinary research*, **50**(Apr.), 30.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**(2), 257–286.

Raue, R., Harmeyer, S. S., & Nanjiani, I. A. 2011. Antibody responses to inactivated vaccines and natural infection in cattle using bovine viral diarrhoea virus ELISA kits: assessment of potential to differentiate infected and vaccinated animals. *Veterinary journal (London, England : 1997)*, **187**(Mar.), 330–334.

Toft, N., Jørgensen, E., & Højsgaard, S. 2005. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. *Preventive veterinary medicine*, **68**(Apr.), 19–33.

van Roon, A., Mercat, M., van Schaik, G., Nielen, M., Graham, D., More, S., Guelbenzu, M., Fourichon, C., Madouasse, A., & Santman-Berends, I. 2020a. Quantification of risk factors for BVDV in cattle herds: a systematic search and meta-analysis. *Journal of dairy science (In press)*.

van Roon, A.M., Santman-Berends, I.M.G.A., Graham, D., More, S.J., Nielen, M., van Duijn, L., Mercat, M, Fourichon, C, Madouasse, A, Gethmann, J, Sauter-Louis, C, Frössling, J, Lindberg, A, Correia-Gomes, C, Gunn, G J, Henry, M K, & van Schaik, G. 2020b. A description and qualitative comparison of the elements of heterogeneous bovine viral diarrhea control programs that influence confidence of freedom. *Journal of dairy science*, Mar.

Whittington, R., Donat, K., Weber, M. F., Kelton, D., Nielsen, S. S., Eisenberg, S., Arrigoni, N., Juste, R., Sáez, J. L., Dhand, N., Santi, A., Michel, A., Barkema, H., Kralik, P., Kostoulas, P., Citer, L., Griffin, F., Barwell, R., Moreira, M. A. S., Slana, I., Koehler, H., Singh, S. V., Yoo, H. S., Chávez-Gris, G., Goodridge, A., Ocepek, M., Garrido, J., Stevenson, K., Collins, M., Alonso, B., Cirone, K., Paolicchi, F., Gavey, L., Rahman, M. T., de Marchin, E., Van Praet, W., Bauman, C., Fecteau, G., McKenna, S., Salgado, M., Fernández-Silva, J., Dziedzinska, R., Echeverría, G., Seppänen, J., Thibault, V., Fridriksdottir, V., Derakhshandeh, A., Haghkhah, M., Ruocco, L., Kawaji, S., Momotani, E., Heuer, C., Norton, S., Cadmus, S., Agdestein, A., Kampen, A., Szteyn, J., Frössling, J., Schwan, E., Caldow, G., Strain, S., Carter, M., Wells, S., Munyeme, M., Wolf, R., Gurung, R., Verdugo, C., Fourichon, C., Yamamoto, T., Thapaliya, S., Di Labio, E., Ekgatat, M., Gil, A., Alesandre, A. N., Piaggio, J., Suanes, A., & de Waard, J. H. 2019. Control of paratuberculosis: who, why and how. A review of 48 countries. *BMC veterinary research*, **15**(June), 198.

Zucchini, W., MacDonald, I. L., & Langrock, R. 2017. *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CRC.