Review Report

21st September 2020

Madouasse et al 2020. A modelling framework for the prediction of the herd-level probability of infection from longitudinal data

**Overall comments:**

Difference in surveillance and disease control intensities makes the interpretation of herd status difficult, particularly for diseases such as BVD and Johne's that have potentially long latent periods. The authors used a Bayesian framework to integrate various sources of information to calculate the probability of herds having at least one BVD persistently infected (PI) animal. I appreciate the motivation behind this study because the ability to predict herd disease status in near real-time may provide useful information that guide relevant stakeholders to make decisions on trading and disease controls. Providing the code and R package is excellent and this will certainly facilitate the discussions on this topic within the community. Therefore, overall, I would very like to see this work to be shared widely and this kind of framework to be applied to many data held in industries and governments.

I have several questions that I would like the authors to address and/or clarify as follows.

*1. Prior for test accuracy*

I have a few comments on the prior for the test accuracy. Which years of testing results (9725 observations) were used to estimate the prior for the bulk milk ELISA sensitivity and specificity? I understand that Bayesian modelling was done for data between 2014 and 2016. Because prior should not be elicited from the same data that is analysed (by definition of prior), please clarify whether test accuracy prior was obtained from data before 2014.

As I initially read this manuscript, I wondered why the prior for the test sensitivity was extremely narrow (5th percentile 94.6 and 95th percentile 95.5). This question was answered by Line 640 onwards suggesting that there was non-identifiability issue when wider prior was used. Using a narrow prior itself is not an issue as long as this can be justified; however, non-identifiability indicates the lack of information contained in the data and this would not really justify the use of narrower prior. Having said that, I understand the parameter estimation was probably not the interest of this study. I would therefore suggest removing Line 640 onwards because I'm afraid that these sentences may give an impression to users who are less familiar with Bayesian analysis that prior can be adjusted flexibly so that convergence can be achieved. Instead, please state clearly that prior for test accuracies should be drawn independently from the data using e.g. literature, expert opinions, previous data, or diffuse priors.

*2. Logistic regression*

It is excellent that the logistic model dynamically predicted probabilities of the latent status changing from negative to positive, accounting for risk factors as well as uncertainties in the test outcome. It is a trivial comment; however, It seems that the latent status was defined as presence/absence of at least one PI animal (Line 206) but explanations about the logistic model often said it was predicting the probability of infection (e.g. Line 339). Please use consistent definitions for the latent status throughout the manuscript.

*3. Confirmatory test*

The manuscript mentioned that the results of confirmatory test were randomly generated once herds tested positive by the routine test. Was the data of confirmatory test (e.g. when, which herds were tested and their results) available? If so I wonder why this data was not explicitly integrated into the modelling framework i.e. use the observed positive/negative test results to estimate parameters rather than generating random results. Adding the observed second test results may have solved the non-identifiability issue. I may have missed something, so please clarify. Also please add information about the prior distribution used for the test accuracy of the confirmatory test (Line 466). Or did it assume 100% DSe and DSp (Line 567)?

*4. Model comparison*

It was not immediately clear to me what the motivation was behind for developing 4 different models. Was the intention to show that this framework can handle different information source and demonstrate how it can be done? Perhaps if a sentence could be added somewhere that clearly states the objective, it would make it easier to read the text.

**Minor comments:**

Line 139: For the estimation to work, the same tests should be used in different populations…

I assume this sentence talks about frequentist framework; these assumptions can be relaxed in various ways in Bayesian framework. Please clarify this in the text.

Line 165: …although not with longitudinal data from multiple epidemiological units such as herds.

I'm aware that Simon Spencer from Warwick has been working on this field and there are some applications of HMM for veterinary data (although precisely speaking they don't have many units). I would suggest adding some previous works e.g. (Touloupou et al., 2020).

Touloupou, P., Finkenstädt, B., Spencer, S.E.F., 2020. Scalable Bayesian Inference for Coupled Hidden Markov and Semi-Markov Models. Journal of Computational and Graphical Statistics 29, 238–249. https://doi.org/10.1080/10618600.2019.1654880

Line 246: The status on the first time step ($S_1$+)…

It was slightly confusing that S+ was used to represent the latent status itself (regardless being positive or negative), while it also sometimes represent 'positive' status only e.g. formula in Line 248, which is showing the probability of being status positive. Perhaps re-writing as p($S_1$+=+ve) or similar throughout the manuscript would avoid this confusion.

Line 308 'specificity of interest'

Did it mean sensitivity?

Line 712: dynamics contributed to the definition…

I would suggest using 'estimation' rather than 'definition'.

Line 762: 95% credibility [interval] for the association between local seroprevalence and new infection included 0 and this variable was therefore not included.

This is just a trivial comment and out of curiosity; I wonder the inclusion of this variable may have improved the estimation of other parameters? Did you record Bayes Factor with/without this variable, or credibility interval of this variable?