**Biostatistics Advance Access published July 23, 2009**

*Biostatistics* (2009), **0**, 0, *pp.* 1–10
doi:10.1093/biostatistics/kxp026

# Bayesian inference for within-herd prevalence of *Leptospira interrogans* serovar Hardjo using bulk milk antibody testing

FRASER I. LEWIS*, GEORGE J. GUNN

*Epidemiology Research Unit, Scottish Agricultural College, King's Buildings, West Mains Road, Edinburgh EH9 3JG, UK*
fraser.lewis@sac.ac.uk

IAIN J. MCKENDRICK

*Biomathematics & Statistics Scotland, King's Buildings, West Mains Road, Edinburgh EH9 3JZ, UK*

FIONA M. MURRAY

*Veterinary Services, Scottish Agricultural College, King's Buildings, West Mains Road, Edinburgh EH9 3JG, UK*

SUMMARY

Leptospirosis is the most widespread zoonosis throughout the world and human mortality from severe disease forms is high even when optimal treatment is provided. Leptospirosis is also one of the most common causes of reproductive losses in cattle worldwide and is associated with significant economic costs to the dairy farming industry. Herds are tested for exposure to the causal organism either through serum testing of individual animals or through testing bulk milk samples. Using serum results from a commonly used enzyme-linked immunosorbent assay (ELISA) test for *Leptospira interrogans* serovar Hardjo (*L. hardjo*) on samples from 979 animals across 12 Scottish dairy herds and the corresponding bulk milk results, we develop a model that predicts the mean proportion of exposed animals in a herd conditional on the bulk milk test result. The data are analyzed through use of a Bayesian latent variable generalized linear mixed model to provide estimates of the true (but unobserved) level of exposure to the causal organism in each herd in addition to estimates of the accuracy of the serum ELISA. We estimate 95% confidence intervals for the accuracy of the serum ELISA of (0.688, 0.987) and (0.975, 0.998) for test sensitivity and specificity, respectively. Using a percentage positivity cutoff in bulk milk of at most 41% ensures that there is at least a 97.5% probability of less than 5% of the herd being exposed to *L. hardjo*. Our analyses provide strong statistical evidence in support of the validity of interpreting bulk milk samples as a proxy for individual animal serum testing. The combination of validity and cost-effectiveness of bulk milk testing has the potential to reduce the risk of human exposure to leptospirosis in addition to offering significant economic benefits to the dairy industry.

*Keywords*: Bayesian; Latent class analysis; Leptospirosis.

*To whom correspondence should be addressed.

## 1. INTRODUCTION

Leptospirosis is the most widespread zoonosis throughout the world (Meites *and others*, 2004) and carries with it implications for both human and animal health. Human mortality from severe disease forms, Weils disease and severe pulmonary hemorrhage syndrome, is high with mortality rates in excess of 10% and 50%, respectively, even when optimal treatment is provided (McBride *and others*, 2005). In regard to animal health, leptospirosis is one of the most common pathogen-related infections responsible for reproductive losses in cattle worldwide (Grooms, 2006) and is associated with significant economic costs to the dairy farming industry (Bennett and Ijpelaar, 2005). Studies from Ireland (Leonard *and others*, 2004) found that infection with *Leptospira interrogans* serovar Hardjo (*L. hardjo*), the most prominent strain found in Europe, was present in some 79% of 347 dairy herds sampled. Screening programs are commonplace across Europe and the United States and use a range of laboratory testing kits for detecting the presence of leptospira antibodies in either serum or milk.

Leptospirosis can be effectively controlled by annual vaccination. The risk of zoonosis is a main motivating factor behind full herd vaccination and has resulted in significant decreases in occupationally acquired infection (Thornley *and others*, 2002). Herds are tested for the presence of *L. hardjo* antibodies by either (i) sampling from a bulk milk tank, comprising a joint contribution of milk from multiple animals in a herd, or else (ii) serum samples taken from individual animals. The ability to predict herd prevalence through bulk milk sampling is economically very attractive to dairy producers compared with the significant veterinary and laboratory costs involved with individually testing all animals in a herd. Scottish Agricultural College (SAC) veterinary services use the Ceditest *L. hardjo* enzyme-linked immunosorbent assay (ELISA) kit for diagnostic testing of exposure to *L. hardjo* in milk and serum. The kit protocol provides cutoffs, which classify the presence of *L. hardjo* antibodies in either sera or milk as negative, inconclusive, or positive. The recommendation following an inconclusive bulk milk result is that follow-up serum testing of all individual animals may be appropriate.

Using bulk milk samples collected from 12 unvaccinated farms and corresponding serum samples from all animals contributing to each of the bulk milk tanks, our goal was to develop a robust statistical model, which predicts the proportion of animals exposed to *L. hardjo* in a herd conditional on the concentration of *L. hardjo* antibodies present in the herd's bulk milk sample. For a number of years, SAC veterinary services have provided such model-based predictions as part of their diagnostic testing service. This value-added interpretation has proved popular and appears to be of significant value to dairy farmer clients. The modeling subsequently discussed is an attempt to improve the approach currently used by removing the crucial, yet unsupported, assumption that the serum ELISA test currently used is sufficiently accurate to be considered error free in its predicted classifications. Introducing even a small probability of error into the model of this diagnostic test could greatly affect our confidence in its predictions. We are unaware of any peer-reviewed work that supports high accuracy of the Ceditest *L. hardjo* kit and it is not validated as a gold standard test by the OIE—the World Organisation for Animal Health—which validates and certifies all animal diagnostic tests used in the European Union.

Through our analyses, we aim to provide improved guidance on the interpretation of bulk milk test results with a view to avoiding unnecessary follow-up serum testing of individual animals. Additional by-products of this work are estimates of the sensitivity and specificity of the Ceditest *L. hardjo* serum ELISA kit, which are of general interest as this is a commonly used ELISA test.

## 2. DATA AND METHODS

Bulk milk samples were collected from 34 dairy herds from distinct farms across Scotland and screened using the Ceditest indirect ELISA for *Leptospira interrogans* serovar Hardjo antibodies. The ELISA uses antigens to capture and quantify the amount of target antibody present in a serum or milk sample. The test

results in a color reaction measured in terms of optical density values by an ELISA reader. Optical densities provide a numerical quantification of the amount of *L. hardjo* antibody present in the sample being tested. The final numerical output is a standardized percentage positivity (PP) relative to a fixed reference sample (PP = optical density of sample being tested/optical density of reference sample). To ensure robustness, each test result is required to meet extensive validation criteria set out in the manufacturer guidelines in addition to SAC internal standard operating procedures for laboratory quality assurance. The recommended Ceditest interpretation for PP from sera is PP$<$ 20%—negative for *L. hardjo* specific antibodies, 20% $\leqslant PP \leqslant$ 45%—inconclusive, and $PP >$ 45%—positive. The Ceditest interpretation for PP from bulk milk is $PP <$ 40%—negative for *L. hardjo* specific antibodies, 40% $\leqslant PP \leqslant$ 60%—inconclusive, and $PP >$ 60% positive.

Our initial study of 34 herds collected only bulk milk samples, as opposed to matched samples from the bulk milk tank and all individual animals contributing to this tank. From these 34 herds, a subset of herds were selected for follow-up whole herd serum testing. The empirical distribution of bulk milk PP values across all 34 herds was stratified into 3 blocks, and from within each of these blocks, farms were recruited on the basis of practical considerations such as geographical location and perceived willingness of the farmer to take part in the study. A total subset of 12 herds were recruited for matched bulk milk and whole herd serum testing. Section A in the supplementary material available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org) compares the empirical bulk milk PP distribution from all 34 herds in the initial study with the follow-up subset of 12 herds, with the latter appearing representative of the former.

A single sample of blood was taken from each individual animal and a single milk sample was collected from each herd's bulk milk tank. The total volume of each serum/bulk milk sample collected was more than that required to fill a single well in the ELISA plate. This was to enable laboratory quality control measures where each serum/bulk milk sample is assayed in duplicate to ensure the ELISA kit is functioning properly. The bulk milk PP values reported in Table 1 are the means across both replicates as is typical laboratory practice. All 12 bulk milk samples met all quality control criteria. A small number of tests on sera from individual animals showed unusually large variation between replicates and were

Table 1. *Observed number of animals in each herd, which tested negative or nonnegative for Leptospira interrogans serovar Hardjo using Ceditest serum ELISA test, and the bulk milk PP from each herd using Ceditest milk ELISA test. Serum PP cutoff criteria were as per test manufacturer guidelines: 0* $\leqslant$ *negative* $<$ *20,* 20 $\leqslant$ *inconclusive* $\leqslant$ 45*, and positive* $>$ 45*. Due to the very small number of animals testing inconclusive (3% of total), this category was combined with the positive category*

| Farm | Bulk milk PP | No. negative cows | No. inconclusive/positive cows |
|---|---|---|---|
| 1 | 14.28 | 200 | 2 |
| 2 | 17.78 | 51 | 0 |
| 3 | 20.35 | 125 | 1 |
| 4 | 34.12 | 47 | 1 |
| 5 | 45.50 | 52 | 1 |
| 6 | 73.99 | 64 | 5 |
| 7 | 80.68 | 107 | 9 |
| 8 | 109.60 | 19 | 27 |
| 9 | 115.08 | 21 | 56 |
| 10 | 121.36 | 35 | 32 |
| 11 | 122.83 | 28 | 55 |
| 12 | 144.27 | 5 | 36 |
| | Total | 754 | 225 |

discarded (there was no evidence to reject an assumption of independence of discarded samples from farms or laboratory batches).

The Ceditest *L. hardjo* ELISA for sera provides 3 categories of diagnosis; however, due to the paucity of inconclusive responses, we present analyses where inconclusive and positive responses are collapsed into a single combined class; in total, only 3% of animals tested inconclusive compared with 77% negative and 20% positive. There is a good practical rationale for collapsing the inconclusive and positive categories. From the perspective of the dairy farmer either a sufficiently high level of disease is present in the herd for action to be required or else the herd is disease free, that is, the level of disease is estimated to be sufficiently low to be ignored on both economic and welfare grounds in which case no action is required. Therefore, a pragmatic and conservative approach given the available data is to combine inconclusive and positive responses into a single nonnegative class.

Using Bayesian inference, we develop a generalized linear model with a single binary response denoting the presence (absence) of exposure to disease, as indicated by a positive test for the presence of antibodies in each individual animal based on serum test results, with bulk milk PP as an explanatory covariate. We adopt a Bayesian approach as this provides a robust and numerically tractable way of fitting our model to data given the presence of latent variables (the unobserved diagnostic test error and herd prevalence, as discussed later). Bayesian methods are not the only available methodology, however, they have been shown in some cases to provide more stable estimates than alternative techniques such as the expectation maximization algorithm (Dempster *and others*, 1977) in the estimation of diagnostic test accuracy when the true disease state is unknown (Enoe *and others*, 2000). The adoption of Bayesian methods is also increasingly common in this area (Branscum *and others*, 2005).

Our model is intended as an aid in supporting disease management on dairy farms and as such we require that it be as robust as possible in its predictions. Hence, we do not assume *a priori* that the serum ELISA is a gold standard test, and we allow for the possibility that the data may exhibit clustering and hence model overdispersion. A random effect term at farm level is considered as a means of dealing with overdispersion. Our general model is defined in (2.1)–(2.4)

$$Y_i \sim \text{Bin}(n_i, q_i), \tag{2.1}$$

$$q_i = \text{S}p_i + (1 - \text{C})(1 - p_i), \tag{2.2}$$

$$f(p_i) = \theta + \beta \text{bulkmilk}_i + \phi_i, \tag{2.3}$$

$$\text{and} \qquad \phi_i \sim N(0, \sigma_f^2) \quad \text{for } i = 1, \ldots, 12 \tag{2.4}$$

where $Y_i$ denotes the number of animals that tested positive for the presence of antibodies on farm $i$, $q_i$ is the probability for each animal of it independently testing positive from a total of $n_i$ animals, $\phi_i$ is a farm-level random effect, $f(\cdot)$ denotes the link function, $f(p) = \log\{p/(1-p)\}$ or $f(p) = \log[-\log\{1-p\}]$.

We follow the "no gold standard" parameterization set out by Joseph *and others* (1995), hence

$$\begin{aligned} q &= \Pr(\text{test positive}) \\ &= \Pr(\text{test positive}|\text{animal positive}) \times \Pr(\text{animal positive}) \\ &\quad + \Pr(\text{test positive}|\text{animal negative}) \times \Pr(\text{animal negative}) \\ &= Sp + (1 - C)(1 - p) \end{aligned}$$

where $p$ denotes the true within-herd prevalence of disease, $S$ is the test sensitivity, and $C$ the specificity. These are the latent variables in our model which, given sufficient degrees of freedom, are estimated indirectly from the data. We refer to $p$ as denoting true within-herd prevalence distinguishing it from $q$, the

probability that an animal will test positive. We use noninformative priors for all parameters, specifically: $\beta \sim N(0, 1000)$, $\theta \sim N(0, 1000)$, $S \sim U(0, 1)$, and $C \sim U(0, 1)$, where $N(\mu, \sigma^2)$ denotes a Gaussian density with mean $\mu$ and variance $\sigma^2$; and $U(a, b)$ is a uniform density on the interval $(a, b)$. The standard deviation of the farm-level random effect, $\sigma_f$, was given a prior of $U(0, 100)$. Posterior distributions for all model parameters were estimated using an implementation of the slice sampler due to Neal (2003) written in C using the GNU scientific library (Galassi *and others*, 2006). This code was validated against a range of models for which the posterior distributions could be calculated analytically or were known from existing published studies. Many chains were run using different initial seeds and the typical burn-in period appeared to be very small (several thousand iterations) for all models examined. All results presented are based on output from chains that were run for a considerably large number of iterations, with output from multiple chains combined so that all parameter estimates were based on samples with effective sample sizes (for definition, see Gelman *and others*, 2004) of at least 10 000. Our justification for this is 2-fold: (i) there exists very high correlation between the intercept and gradient parameters, $\theta$ and $\beta$, respectively, which can result in slow traversal of the parameter space; and (ii) due to the nature of our models, there is the potential for complex mixing, specifically oscillation between multiple stationary distributions, as discussed below.

## 3. Results

Our analysis has 2 main objectives (i) assess the predictability of within-herd prevalence from bulk milk PP and (ii) estimation of the accuracy of the Ceditest serum ELISA. It is important to note that these 2 objectives are not independent: the model developed in (i) also estimates the accuracy of the serum ELISA. Hence, our subsequent estimates of test sensitivity and specificity are conditional on the assumption that our choice of model is a good fit to the observed data. For this reason, we examine a number of competing models to identify an appropriate optimal model given the data available.

### 3.1 *Model selection*

A series of models of increasing complexity were fitted to the data to assess the statistical support for the use of bulk milk PP as a predictor of mean herd prevalence. Models with logistic and complementary log–log link functions, with and without overdispersion, were all examined. Model complexity was increased systematically from the null model (comprising only a constant term without overdispersion) up to the most general model defined in (2.1)–(2.4). As is typical in Bayesian model comparison, we used Bayes factors (Gelman *and others*, 2004), specifically a comparison of log marginal likelihoods, as the goodness of fit criterion when comparing the various models.

A potential complication in estimating the parameters in our various models is the presence of multiple solutions. The existence of multiple solutions can easily be explained by analogy with the Hui–Walter model (Hui and Walter, 1980), a standard model for estimating disease prevalence in the absence of a gold standard test. The Hui–Walter model has 2 optimal solutions in a maximum likelihood sense. If the set of parameters $\{p, S, C\}$ represents a solution, then $\{1 - p, 1 - C, 1 - S\}$ is also a solution. Assessing convergence and estimation of stationary distributions in respect of the Hui–Walter model are discussed by Toft *and others* (2007).

Parameter estimation can become problematic if multiple solutions exist, as the sampler may jump between solutions requiring then the disentanglement of the posterior distributions for each respective solution. Despite extensive tuning of the slice sampler parameters (see Neal, 2003) and running extremely long chains of up to $2 \times 10^7$ iterations, we were unable to force such jumping to occur when sampling from any of our models. Further details can be found in section B in the supplementary material available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org).

Table 2. *Model selection using log marginal likelihood as the goodness of fit criteria. Including a gradient term (bulk milk PP coefficient) greatly improves the model fit. The inclusion of a random effect term is also strongly supported, however, the precise parameterization has little effect as does the choice of link function. Of the parameterizations explored, scaling the farm level random effect by the gradient parameter maximizes the marginal likelihood, as does the use of a logistic link function*

| Model | log (marginal likelihood) | |
|---|---|---|
| | logistic | cloglog |
| $\theta$ | −528.24 | −528.25 |
| $\theta + \beta\text{bulkmilk}_i$ | −281.05 | −282.15 |
| $\theta + \beta\text{bulkmilk}_i + \phi_i$ | −273.58 | −274.20 |
| $\theta + \beta(\text{bulkmilk}_i + \phi_i)$ | −272.98 | −273.11 |

Table 2 details the goodness of fit using log marginal likelihoods for the different models explored. We additionally include in our analyses, a variant of the parameterization of our general model where $f(p_i) = \theta + \beta(\text{bulkmilk}_i + \phi_i)$; we use the same noninformative priors as previously. These 2 parameterizations are mathematically equivalent; however, extensive simulations have shown that scaling the random effect term by the bulk milk regression coefficient, $\beta$, results in improved mixing when the sample size is small, giving parameter estimates with lower variances and increased goodness of fit. As the sample size increases, the 2 parameterizations give indistinguishable results as should be the case. Section C in the supplementary material available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org) contains a comparison of the mixing of each parameterization using simulated data with parameter estimates similar to those from the observed data.

Multiple chains were run for each model and the log likelihood values calculated every 1 000 steps, with the output from various chains pooled (after allowing sufficient burn-in) until the combined effective sample size was in excess of 10 000. To ensure robust estimation of the log marginal likelihood, we follow Congdon (2001) and divide the output into batches, calculate the harmonic mean in each batch, and then take the mean of these values. Up to 8 batches were tried along with the median rather than the mean over batches. Such variations had negligible impact on the resulting marginal likelihood values, giving confidence in the robustness of our estimates. The difference in log marginal likelihoods between the models with and without a bulk milk term is large. Congdon (2001), table 10.1, provides guidelines on the magnitude of differences required between Bayes factors to be notable, ranging from weak support denoting the smallest difference between log marginal likelihoods, through to very strong support denoting a difference in log marginal likelihoods of at least 5. The choice of link function in the various models has little effect (weak support). The inclusion of a random effect term to allow for overdispersion at the farm level has very strong support; however, the precise parameterization of this term has little effect (only weak support) on the overall model fit.

We have a number of models with comparable goodness of fit and it is therefore informative to compare the predictions in mean within-herd prevalence and diagnostic test accuracy between these various models. Section D in the supplementary material available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org) contains a detailed comparison. Of the parameterizations explored, scaling the farm-level random effect by the bulk milk regression coefficient maximizes the marginal likelihood, and from the comparisons between the alternative models and parameterizations, we choose as our optimal model $\log\{p/(1-p)\} = \theta + \beta(\text{bulkmilk}_i + \phi_i)$. Parameter estimates for this model, including test sensitivity and specificity, are detailed in Table 3. Finally, to investigate the robustness of our chosen model, we fitted it to jackknife samples (Efron and Tibshirani, 1993) from our data set of 12 farms (see section E in the supplementary material available at *Biostatistics* online http://www.biostatistics.oxfordjournals.org). The

Table 3. *Parameter estimates for the optimal bulk milk model (bulk milk term, overdispersion scaled to gradient term, and logistic link); 95% confidence intervals use the 2.5% and 97.5% quantiles of the posterior distribution estimated from Markov chain Monte Carlo output*

| Parameter | Median (95% confidence interval) |
|---|---|
| $\boldsymbol{\theta}$ | −9.629 (−21.213, −6.367) |
| $\boldsymbol{\beta}$ | 0.0912 (0.0593, 0.222) |
| $\boldsymbol{S}$ | 0.859 (0.688, 0.987) |
| $\boldsymbol{C}$ | 0.989 (0.975, 0.998) |
| $\boldsymbol{\sigma}_f$ | 9.756 (3.931,28.225) |

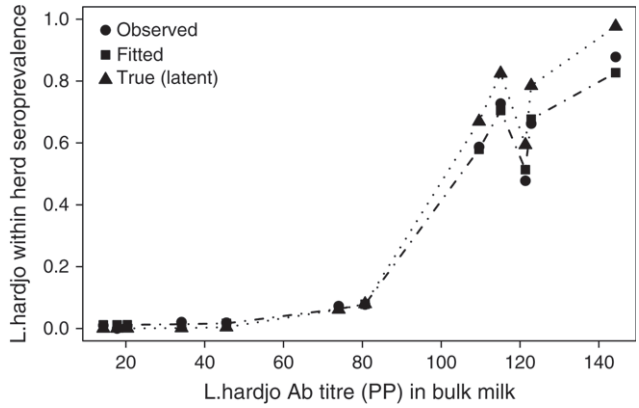


Fig. 1. Observed data and predicted values. A comparison of the observed proportion of animals testing positive in each farm, the predicted mean proportion of positive tests using the optimal model, and the predicted mean true prevalence of exposure to disease within each herd via a latent variable.

model appears relatively robust to the choice of farms with the exception of farm 12 when estimating test sensitivity $S$. Exclusion of farm 12 from the data has a substantial impact on the resulting estimate of $S$. This can be explained by the relative position of farm 12 in Figure 1. Generally, the higher the proportion of test positive animals in a farm the greater influence it will exert on estimates of $S$, which in this case is also combined with farm 12 having by far the largest bulk milk PP value compared with farms 8–11, which have lower but relatively uniform bulk milk PP values. However, farms 8–11 do exhibit substantial variance in prevalence, which affects the estimation of $S$.

### 3.2 *Prevalence prediction using bulk milk PP*

Figure 1 shows a comparison of the observed data (the proportion of positive tests and corresponding observed bulk milk PP), fitted values from our optimal model (the predicted mean proportion of positive tests conditional on bulk milk PP), and predictions of the mean prevalence of exposure to disease in the herd (the latent variable in our model denoting true exposure status). We find that as the bulk milk PP increases, mean prevalence in the herd also increases, as must intuitively be the case.

Of particular interest is the comparison between the predicted proportion of positive tests and mean prevalence in the context of a high specificity and moderate sensitivity. In herds with low prevalence, the proportion of positive tests was a good estimator of true prevalence, however, in herds with higher

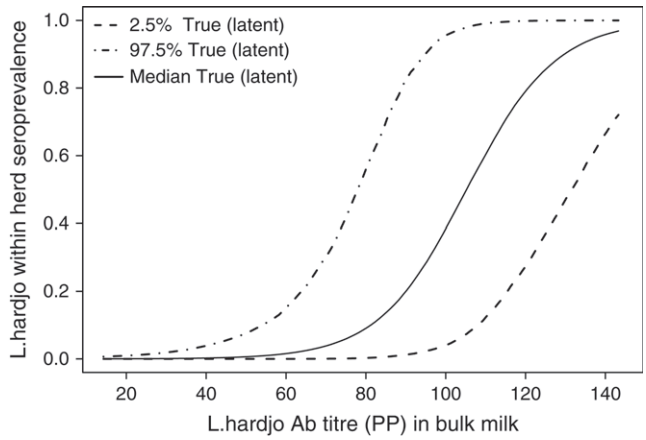


Fig. 2. Posterior distribution for mean prevalence in the optimal model.

prevalence using the serum ELISA test alone underestimates the true prevalence. This can be seen in Figure 1 by the difference between the proportion of positive tests and true prevalence of exposure in the five farms with highest prevalence. As the bulk milk PP increases, more antibodies are present in the milk tank and, by implication, more of the herd is likely to have been exposed to disease, therefore test accuracy in detecting exposed animals becomes relatively more important. Plotting the predicted proportion of positive tests against predicted mean prevalence (not illustrated) shows increasing divergence as bulk milk PP increases. This is to be expected given our estimates of test sensitivity (see Section 3.3).

Figure 2 shows the posterior distribution for the mean prevalence of exposure to disease predicted by our model. The Ceditest kit interpretation for bulk milk is that a PP of less than 40% is indicative of an unexposed herd. This is of particular interest as we find from our model that using a PP cutoff of 41% or less ensures that there is at least a 97.5% probability of less than 5% of the herd being exposed to *L*. hardjo.

### 3.3 *Serum ELISA test accuracy*

Figure 3 shows estimates of the posterior densities for the serum ELISA specificity ($C$) and sensitivity ($S$) from our optimally fitting model (see Table 3 for 95% confidence intervals). The test is extremely good at correctly predicting unexposed animals; however, there is considerably more uncertainty regarding the correct classification of exposed animals. The uncertainty in our estimate of sensitivity could be due, at least in part, to the relatively small proportion of animals that tested nonnegative.

## 4. DISCUSSION

We have developed a method to predict within-herd prevalence of exposure to an important endemic and zoonotic pathogen using estimates of bulk milk PP. Commonly used ELISA kits were used for both milk and serum testing with observed data on 979 animals split across 12 bulk milk samples, with each sample collected from a distinct farm. A Bayesian latent variable generalized linear mixed model was used to estimate the accuracy of the serum ELISA test and provide a robust predictive model. Our goal was to provide a method able to evaluate the bulk milk interpretation provided by the test manufacturers guidelines on a "live" data set and provide additional value in terms of robust predictions of within-herd prevalence of exposure to disease.

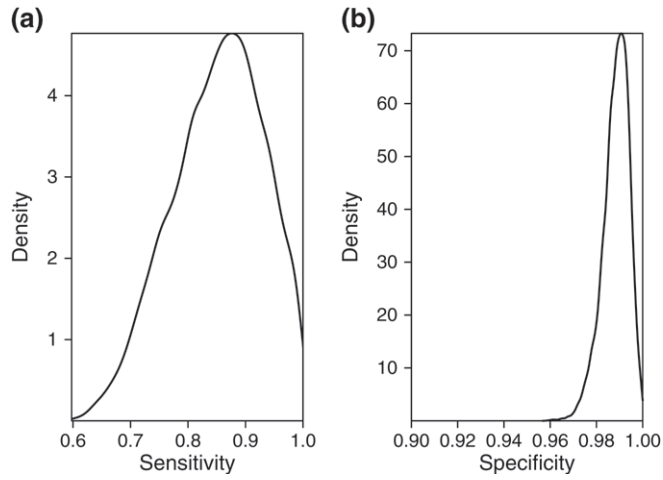


Fig. 3. Posterior distributions for serum ELISA sensitivity (a) and specificity (b). See Table 3 for summary statistics.

The test interpretation guidelines from Ceditest for their ELISA for *L. hardjo* in bulk milk state that a PP of less than 40% is indicative of an unexposed herd. We estimate with 97.5% probability that less than 5% of a herd is exposed if the PP is less than or equal to approximately 41%. The latter interpretation is an entirely reasonable and practical measure of an unexposed herd. The consistency between this interpretation and the interpretation provided with the ELISA kit, given that each was based on different data sources and different estimation methods, provides very strong evidence in support of bulk milk testing as a means of identifying herds, which have not been exposed to disease.

Our method provides value-added interpretation in the form of predictions for the mean prevalence of exposure to disease conditional on bulk milk PP. However, these predictions do suffer from a relatively high degree of uncertainty (see Figure 3), a significant contributor being the large observed variation in the proportion of positive tests between farms with relatively similar bulk milk PP values (see farms 9 and 10 in Table 1). Despite extensive checks, we were unable to identify satisfactorily reasons for these variations. This additional variance between farms necessitated the inclusion of sizeable random effects in our model and thus a loss in predictive precision.

We hope that our analyses will further support the use of bulk milk testing as an effective disease surveillance tool and encourage rigorous statistical validation of commonly used diagnostic tests.

## Acknowledgments

The authors thank 2 anonymous reviewers and the editor for constructive comments that have improved the article. *Conflict of Interest*: None declared.

## Funding

This work was supported by Schering–Plough who sponsored the data collection and veterinary fieldwork. Statistical analyses were undertaken by part of the Scottish Government funded Centre of Excellence in epidemiology, population health, and infectious disease control. Biomathematics and Statistics Scotland and the Scottish Agricultural College both receive financial support from the Scottish Government (RERAD). Funding to pay the Open Access publication charges for this article was provided by the Scottish Agricultural College.