Interpretive summary: Johne's Disease Diagnostic Data. Wang

A new Bayesian statistical model was developed to evaluate the performance of milk ELISA and fecal culture tests for longitudinal Johne's Disease data in the absence of a gold standard diagnostic test. Data from a Danish longitudinal study from January 2000 to March 2003 were analyzed using the proposed method. Based on the Bayesian approach, the posterior probability distribution of the infection onset time could be obtained and used as a criterion for disease infection diagnosis. The posterior probability criterion was superior to the raw single reading ELISA and fecal culture tests for Johne's Disease diagnosis.

Bayesian analysis of longitudinal Johne's disease diagnostic data without a gold standard test

C. Wang*, B. W. Turnbull[†], S. S. Nielsen[‡], and Y. T. Gröhn[§]

[*] Department of Veterinary Diagnostic and Production Animal Medicine and Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A.

[†]Department of Statistical Science, Cornell University, Ithaca, NY 14853, U.S.A.

[‡]Department of Large Animal Sciences, Faculty of Life Sciences, University of Copenhagen, Frederiksberg, Denmark

[§]Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, U.S.A.

Corresponding Author: Chong Wang, Department of Veterinary Diagnostic and Production Animal Medicine and Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A., Ph: 515-294-3836, Fax: 515-294-1072, e-mail: chwang@iastate.edu

ABSTRACT

A Bayesian methodology was developed based on a latent change-point model to evaluate the performance of milk ELISA and fecal culture tests for longitudinal Johne's Disease diagnostic data. The situation where there is no perfect reference test was considered, i.e., no "gold standard". A change-point process with a Weibull survival hazard function was used to model the progression of the hidden disease status. The model adjusted for the fixed effects of covariate variables and random effects of subject on the diagnostic testing procedure. Markov chain Monte Carlo methods were used to compute the posterior estimates of the model parameters that provide the basis for inference concerning the accuracy of the diagnostic procedure. Based on the Bayesian approach, the posterior probability distribution of the change-point onset time can be obtained and used as a criterion for infection diagnosis. An application is presented to an analysis of ELISA and fecal culture test outcomes in the diagnostic testing of paratuberculosis (Johne's Disease) for a Danish longitudinal study from January 2000 to March 2003. The posterior probability criterion based on the Bayesian model with 4 repeated observations has an area under the receiver operating characteristic curve (AUC) of 0.984, and is superior to the raw ELISA (AUC = 0.911) and fecal culture (Sensitivity = 0.358 , specificity = 0.980) tests for Johne's Disease diagnosis.

Key words: Johne's disease, longitudinal, ROC curve, no Gold Standard

INTRODUCTION

This study was motivated by the need for accurate diagnostic tests of bovine paratuberculosis (Johne's disease), which is a major animal health problem (Rideout et al., 2003). Control of Johne's disease has been difficult due to lack of accurate diagnostic tests (Nielsen and Toft, 2008). Diagnostic tests for Johne's disease such as milk antibody ELISA and fecal culture (**FC**) are known to be imperfect. It is less straightforward to evaluate the performances of these existing imperfect tests when no gold standard (**GS**) tests are available for comparison.

For binary outcome test data, sensitivity (probability of a positive outcome in an infected individual) and specificity (probability of a negative outcome in a non-infected individual) are used to assess the test accuracy. To assess the accuracy of an ordinal or continuous-scaled diagnostic procedure, it is useful to consider the receiver operating characteristic (**ROC**) curve, which is a graph of pairs of sensitivity and 1-specificity values that result as the test's threshold value is varied. The area under the ROC curve (**AUC**) is a measure of the overall diagnostic ability of the test. Principles of ROC curve estimation using parametric and nonparametric methods are well reviewed in Pepe (2003) and Krzanowski and Hand (2009). Traditional diagnostic test evaluation assumes the existence of a GS reference test that has perfect sensitivity and specificity.

In the absence of a perfect reference test, Hui and Walter (1980) developed a maximum likelihood method to estimate the sensitivity and specificity of 2 imperfect binary tests. For ordinal or continuous outcome test data, construction of an ROC curve requires all possible pairs of sensitivity and 1-specificity values to be estimated. When no GS test but only an imperfect binary test was available for comparison, Nielsen et al. (2002a) estimated these pairs separately

using the method of Hui and Walter (1980) based on paratuberculosis data. Wang et al. (2007) developed a Bayesian methodology for nonparametric estimation of ROC curves in the same situation as that of Nielsen et al. (2002a), and assures the natural monotonicity property of the ROC curve. For the situation where the test values (or transformed test values) of both infected and non-infected individuals have continuous normal distributions, Choi et al. (2006) proposed a Bayesian method for estimating an ROC curve when there is no GS test. When covariate information is available, it can be used to increase the effectiveness of continuous markers in distinguishing between uninfected and infected subjects; Pepe (2003, Chapter 6) discussed covariate effects on tests and the need to identify them. Wang et al. (2006) proposed a 2-part statistical model to adjust the ROC curve estimates for covariates without GS tests.

All the above methods focus on cross-sectional studies, where the hidden infection statuses are usually assumed independent across observations. Besides cross-sectional studies, there have been longitudinal studies to understand the ability of the diagnostic tests to predict Johne's disease in dairy cows (Nielsen and Toft, 2006; Sweeney et al., 2006). The ROC analysis incorporating the time dimension was discussed in Pepe (2003, Chapter 9.2). Norris et al. (2009) proposed estimating ROC curves without a GS test for longitudinal studies by using a change-point model, in which the time to change-point was modeled conditional on unknown infection status. There is no statistical methodology currently available to adjust ROC curves for covariate effects without a GS test. Thus, the first objective was to estimate the ROC curve of a continuous-scaled diagnostic without a perfect reference test and adjust for covariate effects based on longitudinal data.

Another objective of longitudinal studies is to improve the diagnosis of infection onset (change-point) by using repeated observations. A latent class model with the change-point

estimated by Bayesian methods is appealing because the posterior distribution can provide a direct answer to the question: what is the probability that the change-point has occurred? Lange et al. (1992), Gulyas (1998), and Slate and Turnbull (2000) utilized Bayesian change-point models to analyze longitudinal disease diagnostic data. The methods of Cronin (1995), Gulyas (1998) and Slate and Turnbull (2000) all lead to a dynamic index based on the posterior probability that change-point has occurred by the current time in an individual subject. All results in this paragraph were based on the presence of a GS test.

In this paper, a parametric model was proposed with change-points to estimate the ROC curve of a diagnostic test for longitudinal data.

MATERIALS AND METHODS

Statistical Model Structure

A statistical methodology was proposed for assessing the accuracy of the continuous-scaled test (**Test 1**) by comparing it with the imperfect binary reference test (**Test 2**) for longitudinal data. In the application in the Johne's disease study, as discussed in the Introduction, Test 1 is the ELISA and Test 2 the FC test. The covariate effects were corrected so as to avoid a biased estimate of test accuracy. The covariates for one individual may vary over time.

Parametric Models for Test Outcomes Given Infection Status. Suppose there is a sample of n individuals. As a general and realistic setting, the 2 tests, Test 1 and Test 2, are not necessarily taken always at the same times. So suppose for the ith individual ($i = 1, \ldots, n$), there are l_i time points $t_{i,1}, t_{i,2}, \ldots, t_{i,li}$ where at least 1 of the 2 diagnostic tests are taken. Suppose the Test 1 scores are observed at times $\{ t_{i,j}; j \in M_i \}$ and the reference Test 2 scores are observed at times $\{ t_{i,j}; j \in M_i^* \}$, where M_i and M_i^* are the sets of indices for Test 1 and Test 2, respectively. For the ith individual and jth observation, let $D_{i,j}$ (= 0, negative; or 1, positive)

denote the true unknown infection status, $T_{i,j}$ denote the value of Test 1, measured on a continuous scale, possibly transformed, and let $R_{i,j}$ (= 0, negative; or 1, positive) denote the diagnostic value of Reference Test 2, which is measured on a binary scale. We assume the 2 tests are conditionally independent, i.e., $T_{i,j}$ and $R_{i,j}$ are independent given infection status $D_{i,j}$.

For the *j*th observation of the *i*th individual, we have *K* measured covariates $Z_{i,j,1}, \ldots, Z_{i,j,K}$, which could affect the Test 1 scores given the infection status. To model the effects of the covariates on the Test 1 scores given infection status, we use a linear mixed effect model for the Test 1 scores:

$$T_{i,j} = \beta_0 + \beta_D D_{i,j} + \beta_1 Z_{i,j,1} + \ldots + \beta_K Z_{i,j,K} + \gamma_i + \varepsilon_{i,j}. \qquad [1]$$

Here $\beta_0, \beta_D, \beta_1, \ldots, \beta_K$ are the unknown regression coefficients for the mean of $T_{i,j}$, γ_i denotes the random effect of the *i*th cow on $T_{i,j}$, and $\varepsilon_{i,j}$ denotes the observation level random variation. In the analysis, we assume γ_i and $\varepsilon_{i,j}$ to be normally distributed with variances σ_g^2 and σ^2, respectively. This is not essential to the method; however, other distributional assumptions can be made.

For the imperfect binary reference Test 2, the natural misclassification model for the diagnostic values is $\{R_{i,j}: i = 1, \ldots, n;\ j \in M_i^*\}$:

$$(R_{i,j} \mid D_{i,j} = 1) \sim Bin(1, 1 - \alpha_1)$$

$$(R_{i,j} \mid D_{i,j} = 0) \sim Bin(1, \alpha_0),$$

where α_0 and α_1 are, respectively, the false positive probability and the false negative probability for Test 2. These parameters are unknown and will be estimated along with the parameters of the Test 1 model.

Latent Class Model to Detect Health Status Change-points. A change-point model is proposed for the progression of the infection status $\{D_{i,j}: i = 1, \ldots, n; j = 1, \ldots, l_i\}$. Let O_i be

the infection onset time for the *i*th cow, i.e., the time when the infection status changes from uninfected to infected. The latent infection state $D_{i,j}$ is determined by the visit time $t_{i,j}$ and change-point onset times O_i by

$$D_{i,j} = \begin{cases} 0 & \text{if} \quad t_{i,j} < O_i \\ 1 & \text{if} \quad t_{i,j} \geq O_i \end{cases}.$$

Of course the latent infection onset time O_i is not observed. A Weibull survival distribution model is proposed for O_i

$$O_i \sim Weibull(\rho, \lambda),$$

where ρ is the shape parameter and λ is the scale parameter. Here no covariate effects on the infection process model are considered for the analysis of Johne's disease. This is because the covariate milk yield may vary depending on the disease process other than affecting disease status, whereas the covariate age is equivalent to time.

Parametric Model for response covariates. To study the relationship between covariates and the infection process, 2 types of covariates were considered: the explanatory covariates (e.g., age) which could affect the hidden infection process and other variables; and the response covariates (e.g., milk yield), which can not affect the hidden infection process, but may vary depending on the infection status and the other covariates. Assume there are K_1 explanatory covariates, and denote them by $Z^X_{i,j,1}, \ldots, Z^X_{i,j,K_1}$; and assume there are K_2 response covariates, and denote them by $Z^Y_{i,j,1}, \ldots, Z^Y_{i,j,K_2}$.

The distribution of the response covariate $Z^Y_{i,j,k}$ ($k = 1, \ldots, K_2$) was modeled by

$$Z^Y_{i,j,k} = \mu_{k,0} + \mu_{k,D} D_{i,j} + \mu_{k,1} Z^X_{i,j,1} + \ldots + \mu_{k,K_1} Z^X_{i,j,K_1} + \gamma^Y_{i,k} + \varepsilon^Y_{i,j,k}, \qquad [2]$$

where $\mu_{k,0}$, $\mu_{k,D}$, $\mu_{k,1}, \ldots, \mu_{k,K_1}$ are the unknown regression coefficients for the mean of $Z^Y_{i,j,k}$, $\gamma^Y_{i,k}$ is the random effect of the *i*th cow on the *k*th response covariate, and $\varepsilon^Y_{i,j,k}$ is the observation level random variation. Here $\gamma^Y_{i,k}$ and $\varepsilon^Y_{i,j,k}$ are assumed to have normal distributions with variances τ_g^2 and τ^2, respectively. Again, these distributional assumptions are not essential to the method and other distributions can be used if applicable.

The Bayesian Approach and Markov Chain Monte Carlo Method

The parameters were estimated in the model by using Bayesian methods. Independent prior distributions were used for the model parameters. Specifically, it is assumed normal N(0, 10000) for each of $\beta_0, \beta_D, \beta_1, \ldots, \beta_K, \mu_{k,0}, \mu_{k,D}, \mu_{k,1}, \ldots, \mu_{k,K_1}$; Gamma(0.01, 0.01) for σ_g^{-2}, σ^{-2}, τ_g^{-2}, τ^{-2}, ρ, and λ; and Beta(0.5, 0.5) for α_0 and α_1. These priors are used as noninformative priors. Instead, informative priors could be used if previous knowledge is available for any of the parameters.

Based on the prior distributions and the likelihood of the observed data $\{T_{i,j}: i = 1, \ldots, n; j \in M_i\}$, $\{R_{i,j}: i = 1, \ldots, n; j \in M_i^*\}$, the posterior distribution of the parameters together with change-point times $\{O_i : i = 1, \ldots, n\}$ can be simulated by using Markov chain Monte Carlo (**MCMC**) methods (Robert and Casella, 2004). The simulation of posterior distributions was implemented by calling WinBUGS (Lunn et al., 2000) from R statistical software (R Development Core Team, 2010) using the R2WinBUGS (Sturtz et al., 2005) package.

Application to a Longitudinal Study of Johne's Disease

The method was applied to a longitudinal study that was performed to describe the probability of bacterial shedding of Johne's disease in feces and the antibody response as a

function of age. The sample population consisted of all cows present in 8 Danish dairy herds at any given time point in the study period, from January 2000 to March 2003.

All herds were infected with Johne's disease. During the period, milk samples were obtained from all lactating cows in the herds 11 times per year via the Danish milk recording system. Cows which were not lactating did not contribute milk samples on a given sampling date. Four times per year, fecal samples were collected from all cows in the herds, both lactating and non-lactating. In the study period of approximately 3 yr, repeated sampling of milk (23,265 samples) and feces (8,816 samples) was performed. A total of 1,997 Danish dairy cows provided material. The milk samples were analyzed for antibodies specific to *Mycobacterium avium* subsp. *paratuberculosis* (**MAP**) using an in-house ELISA, and the fecal samples were analyzed for MAP by FC and dichotomized as positive or negative. The in-house ELISA was an *M. phlei* absorbed ELISA based on *M. a. avium*. The FC method was based on decontamination and centrifugation followed by incubation on Herrold's Egg Yolk Medium for 12 wk. Positive cultures from FC were confirmed as MAP using IS*900* PCR. Both methods are described in detail in Nielsen and Toft (2006).

Data Modification. There were rapid increases and decreases of ELISA values at the beginning and end of lactation. To avoid unnecessary complexity of the modeling, the observations with DIM < 15 or DIM > 305 were excluded from the data analysis. After modification and before analysis, a log-transformation was applied to account for the skewness of the reading distribution and because the variance of the corrected optical density (**OD**) values increases with increasing mean values.

After data modification, 1,766 cows with both ELISA and FC tests were included in the analysis. A total of 18,966 ELISA test outcomes and 6,712 fecal culture test outcomes taken at

25,356 observation times (days) were used in the analysis. Among these, there were 18,644 observation times with only ELISA outcomes, 6,390 observation times with only fecal culture outcomes, and 322 observation times when both tests were performed.

Trajectories were plotted of log-transformed ELISA OD score vs. age for 20 randomly selected cows within age interval 3 to 5 yr in Figure 1. Ten of these cows had at least 1 FC positive outcome in the age interval, and the other 10 cows never had FC positive outcome in the age interval.

Johne's Disease Model Structure. Here the covariates milk yield (kg) and age (yr) were included in this analysis. Both milk yield and age were related to the ELISA test scores, as has been reported by Nielsen et al. (2002b), van Schaik et al. (2003) and Nielsen and Toft (2006). Thus, the effects of the covariates milk yield and age were fit in the regression model as explanatory variables for ELISA test score. The relationships between Johne's disease and these variables have been studied by Kudahl et al. (2004). As age is related to time, the effect of age has been accounted for in the hidden longitudinal model for infection process. It has been reported that infected cows are associated with decreased milk yield values (Stabel, 1998). Because the covariate milk yield not likely affects the infection process, it is used as a response variable in the model. It is known that age affects milk yield of dairy cows. The overall model structure is shown in Figure 2.

Posterior Density Simulation. The MCMC method produced the posterior density of each parameter of interest. Successive values of each of the parameters were generated by simulation and the steady-state distribution was the posterior for that parameter (Robert and Casella, 2004). WinBUGS was used to implement this procedure. The model has been run for 10,000 iterations and the samples from 5,001 to 10,000 iterations were used to provide

inferences of the posterior distributions. Convergence has been checked using CODA (Convergence Diagnosis and Output Analysis) for R software (Plummer et al., 2006). The source code is available upon request from the authors.

RESULTS AND DISCUSSION

Descriptive Findings

For ELISA test scores, the parameter estimates and corresponding 95% credible intervals (**CI**) are in Table 1. The effects of covariates can be considered statistically significant at level 0.05 if the corresponding credible intervals do not include zero. There was a significant difference in baseline mean ELISA scores between uninfected and infected cows. Milk yield had a significant negative effect on the ELISA test outcome, whereas age had a significant positive effect.

The level of milk yield is capable of affecting the OD in terms of dilution, which explains the negative signs of the estimated coefficients of this effect. Another possible explanation is that cows with production of high levels of antibodies often have a reduced milk yield as demonstrated in Kudahl et al. (2004). The positive effect of age is natural as older cows usually have higher amounts of antibody than younger cows. The signs of the effects of milk yield and age in this longitudinal study support the cross-sectional study in Wang et al. (2006). Notice that the values of these coefficients are of the same order of magnitude as well. The difference is that both milk yield and age were significant in this longitudinal study, whereas only milk yield was significant in the cross sectional study in Wang et al. (2006).

The ROC curve for the ELISA test was calculated based on the coefficient estimates and is in Figure 3. The AUC was 0.911 with 95% CI (0.903, 0.920). The estimates (95% CI) of the

specificity ($1 - \alpha_0$) and sensitivity ($1 - \alpha_1$) for the FC test were 0.980 (0.976, 0.984) and 0.358 (0.327, 0.391), respectively. These accuracy estimates of the FC test support those that are published in the literature (Nielsen and Toft, 2008).

For the parameters in the Weibull distribution model of infection onset, the scale parameter λ was 9.621 with 95% CI (8.919, 10.520), and the estimate of the shape parameter ρ was 1.805 with 95% CI (1.632, 1.965). Weibull distributions with $\rho > 1$ have a failure rate that increases with time. This indicates that the failure rate for older cows is higher than for younger ones, as expected.

In Norris et al. (2009), an increase in slope of ELISA score was proposed after infection onset instead of a jump in mean as in this paper. A generalization of our model was considered by accommodating a possible change in the slope of the mean. This was done by adding an additional covariate, $(t_{ij} - O_i)^+$ (i.e., time past infection onset), into the mean (Formula 1) of the ELISA test score. The estimated regression coefficient for this term provided no evidence of an increased slope in ELISA OD values and the effect was negligible.

Posterior Probability as a New Diagnostic Rule

In this section, use of the posterior probability that change-point has occurred by the current time was proposed as a dynamic index for infection diagnosis in an individual subject. Because the case definition is latent, the diagnosis reflected the mutual condition that exists between presence of MAP, MAP-specific antibodies and any covariates included.

Use of the Bayesian Model to Diagnose a New Cow. There is interest in diagnosing a new cow, cow $n + 1$, for Johne's disease. We repeatedly collect milk ELISA samples and fecal culture samples at a series of times to find the cow's infection onset time O_{n+1}. Based on the Bayesian model and substituting the parameter values as estimated from the Danish data set,

O_{n+1} remains the only unknown value for cow $n + 1$. We use the observed ELISA and fecal culture test outcomes of cow $n + 1$ to update its posterior distribution of infection onset, leading to an estimate of O_{n+1}. Note this posterior distribution of $\{O_{n+1}\}$ changes as more and more observations are taken from this cow.

Figure 4 shows the posterior distributions of $\{O_{n+1}\}$ of 2 "simulated" cows at different observation times. The ELISA test outcomes, FC outcomes, and milk yield values are all simulated from their estimated distributions from analysis of the Danish longitudinal data. The first row shows a cow that never had an infection onset. Although the posterior probability may increase to some moderate value (third column) when some relatively high ELISA score are observed by chance, it will decrease after observation of the relatively low ELISA score again. The second row shows a cow that has an infection onset at 4.4 yr of age. The posterior probability of infection onset becomes large after the true onset time, and keeps increasing toward 1 afterwards, as consistently relatively high ELISA scores are observed.

Performance of Posterior Probability Test. To investigate how well the proposed posterior probability test performs compared with the original ELISA test and FC test, a moderate size simulation study was undertaken.

Two thousand cows mimicking the analyzed Johne's disease data set were simulated. For each cow, observation values were simulated every 3 mo from age 4 to age 4.75 yr. For each observation time, the true infection status, milk yield, ELISA test outcomes, and FC test outcomes were simulated based on the values of parameters estimated. These values were fit into a Bayesian model and the posterior probabilities were calculated. The performances of the posterior probability test based on 1 to 4 sample collections together with those of the original ELISA test are in Figure 5.

The performance of the posterior probability test improves over time, as more and more data become available. The AUC estimates (with 95% CI) for the ROC curves are 0.935 (0.921, 0.950), 0.964 (0.953, 0.974), 0.976 (0.968, 0.984), and 0.984 (0.979, 0.989) for diagnoses with one, two, three and four observations, respectively. Furthermore, even with a single observation, the posterior probability test performs better than the ELISA test. This is due to the use of an informative prior and covariates in the Bayesian model.

CONCLUSIONS

A method was proposed to estimate the ROC curve of a continuous-scaled diagnostic test without a perfect reference test for longitudinal data. A change-point model was used to estimate the infection onset time of the hidden infection process, based on the observed test outcomes of a continuous-scaled test and a binary test. Both the continuous-scaled test and the binary test were evaluated by comparing them with the estimated latent infection status of the model. The Weibull survival model for the latent change-point time could be generalized to accommodate covariate effects by linking a linear combination of covariates to the survival hazard function when applicable. An application of the method to ELISA and FC test outcomes for Johne's disease shows that the effects of milk yield and age were significant for the ELISA test outcome. Finally, the posterior probability tests behave better than naive single reading tests, because more information is considered than if the test were based on single readings.

ACKNOWLEDGMENTS

This study was funded in part by grant USDA-CSREES/2004-35605-14243, JDIP: Johne's Disease Integrated Program in Research, Education, and Extension and in part by grant R01 CA66218 from the National Institutes of Health.

REFERENCES

Choi, Y., W. O. Johnson, I. A. Gardner, and M. T. Collins. 2006. Bayesian inference for receiver operating characteristic curves in the absence of a gold standard. J. Agric. Biol. Envir. Stat., 11:210 – 229.

Cronin, K. A. 1995. Detection of changepoints in longitudinal data. PhD thesis, Cornell University.

Gulyas, S. W. 1998. Latent disease changepoint models for longitudinal biomarkers. PhD thesis, Cornell University.

Hui, S. L., and S. D. Walter. 1980. Estimating the error rates of diagnostic tests. Biometrics 36:167–171.

Krzanowski, W. J., and D. J. Hand. 2009. ROC curves for continuous data, Boca Raton, Chapman & Hall/CRC.

Kudahl, A., S. S. Nielsen, and J. T. Sørensen. 2004. Relationship between antibodies against *Mycobacterium avium* subsp. *paratuberculosis* in milk and shape of lactation curves. Prev. Vet. Med. 62:119-134.

Lange, N., B. P. Carlin, and A. E. Gelfand. 1992. Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers (with discussion). J. Am. Stat. Assoc. 87:615–632.

Lunn, D., A. Thomas, N. Best, and D. Spiegelhalter. 2000. Winbugs – a Bayesian modelling framework: concepts, structure, and extensibility. Statist. Comput. 10:325–337.

Nielsen, S. S., C. Gronbak, J. F. Agger, and H. Houe. 2002a. Maximum likelihood estimation of sensitivity and specificity of ELISAs and faecal culture for diagnosis of paratuberculosis. Prev. Vet. Med. 53:191–204.

Nielsen, S. S., C. Enevoldsen, and Y. T. Gröhn. 2002b. The *Mycobacterium avium* subsp. *paratuberculosis* ELISA response by parity and stage of lactation. Prev. Vet. Med. 54:1–10.

Nielsen, S. S., and N. Toft. 2006. Age-specific characteristics of ELISA and fecal culture for purpose-specific testing for paratuberculosis. J. Dairy Sci. 89:569-579.

Nielsen, S. S., and N. Toft. 2008. Ante mortem diagnosis of paratuberculosis: a review of accuracies of ELISA, interferon-gamma assay and faecal culture techniques. Vet. Microbiol. 129:217–235.

Norris, M., W. O. Johnson, and I. A. Gardner. 2009. Modeling bivariate longitudinal diagnostic outcome data in the absence of a gold standard. Stat Interface. 2: 171–185.

Pepe, M. 2003. The statistical evaluation of medical tests for classification and prediction, New York: Oxford University Press.

Plummer, M., N. Best, K. Cowles and K. Vines. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC, R News, 6(1): 7—11.

R Development Core Team. 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rideout, B. A., S. Brown, W. C. Davis, J. M. Gay, R. A. Giannella, M. E. Hines, W. D. Hueston, L. J. Hutchinson, and T. Rouse. 2003. The diagnosis and control of Johne's disease, Washington DC: National Academy Press.

Robert, C. P., and G. Casella. 2004. Monte Carlo Statistical Methods, 2nd ed., New York: Springer.

Slate, E. H., and B. W. Turnbull. 2000. Statistical models for longitudinal biomarkers of disease onset. Stat. Med. 19:617–637.

Stabel, J. R. 1998. Johne's disease: A hidden threat. J. Dairy Sci. 81:283–288.

Sturtz, S., U. Ligges, and A. Gelman. 2005. R2WinBUGS: A package for running WinBUGS from R. J. Stat. Software 12: 1–16.

Sweeney, R. W., R. H. Whitlock, S. McAdams, and T. Fyock. 2006. Longitudinal study of ELISA seroreactivity to *Mycobacterium avium* subsp. *paratuberculosis* in infected cattle and culture-negative herd mates. J. Vet. Diagn. Invest. 18:2–6.

van Schaik, G., S. M. Stehman, Y. H. Schukken, C. R. Rossiter, and S. J. Shin. 2003. Pooled fecal culture sampling for Mycobacterium avium subsp. paratuberculosis at different herd sizes and prevalence. J. Vet. Diagn. Invest. 15:233-241.

Wang, C., B. W. Turnbull, Y. T. Gröhn, and S. S. Nielsen. 2006. Estimating receiver operating characteristic curves with covariates when there is no perfect reference test for diagnosis of Johnes disease. J. Dairy Sci. 89:3038–3046.

Wang, C., B. W. Turnbull, Y. T. Gröhn, and S. S. Nielsen. 2007. Nonparametric estimation of ROC curves based on Bayesian models when the true disease state is unknown. J. Agric. Biol. Envir. Stat. 12:128–146.

Table 1. Parameter estimates and corresponding 95% credible intervals (CI) for log-transformed ELISA test score.

Parameters		Posterior Mean	95% CI
Healthy Baseline Mean	β_0	-0.837	(-0.870 , -0.805)
Diseased Baseline Mean	$\beta_0 + \beta_D$	-0.160	(-0.199 , -0.124)
Milk Yield (kg)	β_1	-0.004	(-0.005 , -0.003)
Age (yr)	β_2	0.118	(0.111 , 0.125)
Cow Level Variance	σ_g^2	0.066	(0.059 , 0.073)
Observation Level Variance	σ^2	0.063	(0.061 , 0.064)

CAPTIONS FOR FIGURES 1—4

Figure 1. Line plot of log-transformed ELISA optical density score vs. age for 20 randomly selected cows within age interval 3-5 yr. Bold lines represent cows with at least 1 Fecal Culture positive outcome in the age interval, thin lines represent cows with no Fecal Culture positive outcome in the age interval. The first ELISA optical density score after the first Fecal Culture positive outcome is indicated by '□' for each Fecal Culture positive cow.

Figure 2. Model structure diagram of relationships among disease status, ELISA score, Fecal Culture (FC) score, age and milk yield.

Figure 3. Estimated receiver operating characteristic (ROC) curve of ELISA measurement. Performance of Fecal Culture test is indicated by '□'.

Figure 4. Posterior distributions of $\{O_{n+1}\}$ of 2 "simulated" cows estimated by kernel smoothing method. '- -' indicates the current time of observation, '*' indicates the true onset time. Cow 1 never has a disease onset, whereas Cow 2 has a disease onset at 4.4 yr of age.

Figure 5. Performances of the original single reading ELISA test (- -), and the posterior probability test (–). Performance of Fecal Culture test is indicated by '□'.

Figure 1.

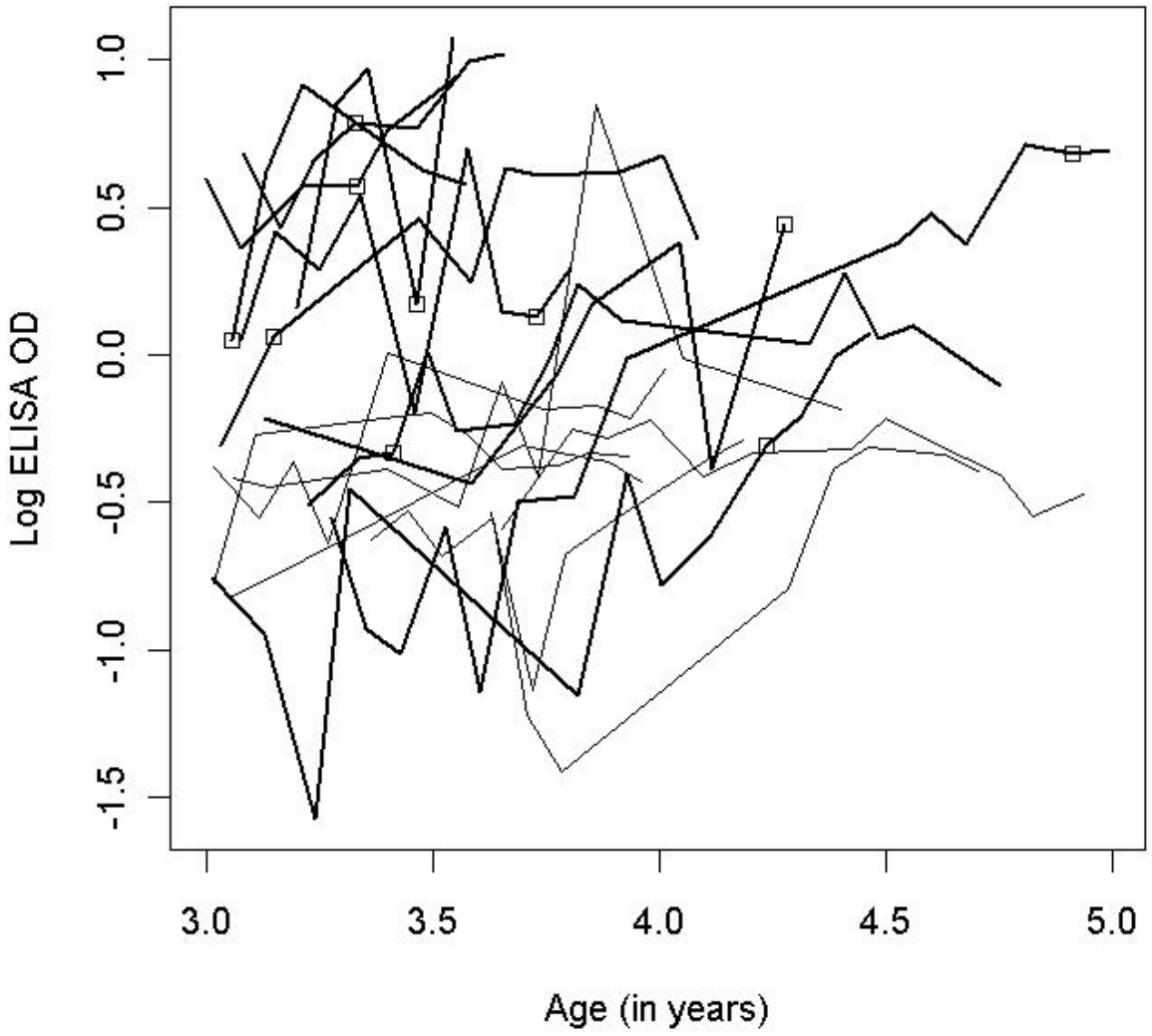


Figure 2.

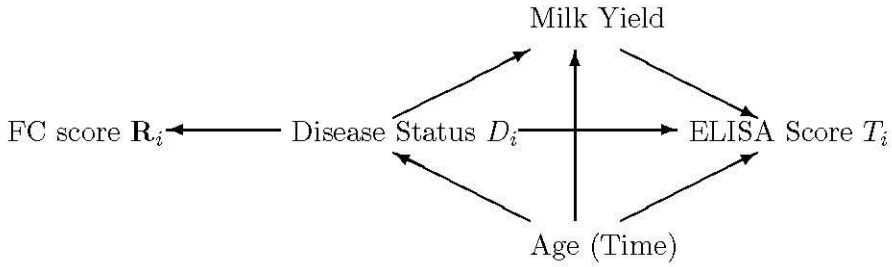

Milk Yield
FC score $\mathbf{R}_i$
Disease Status D_i
ELISA Score T_i
Age (Time)

Figure 3.

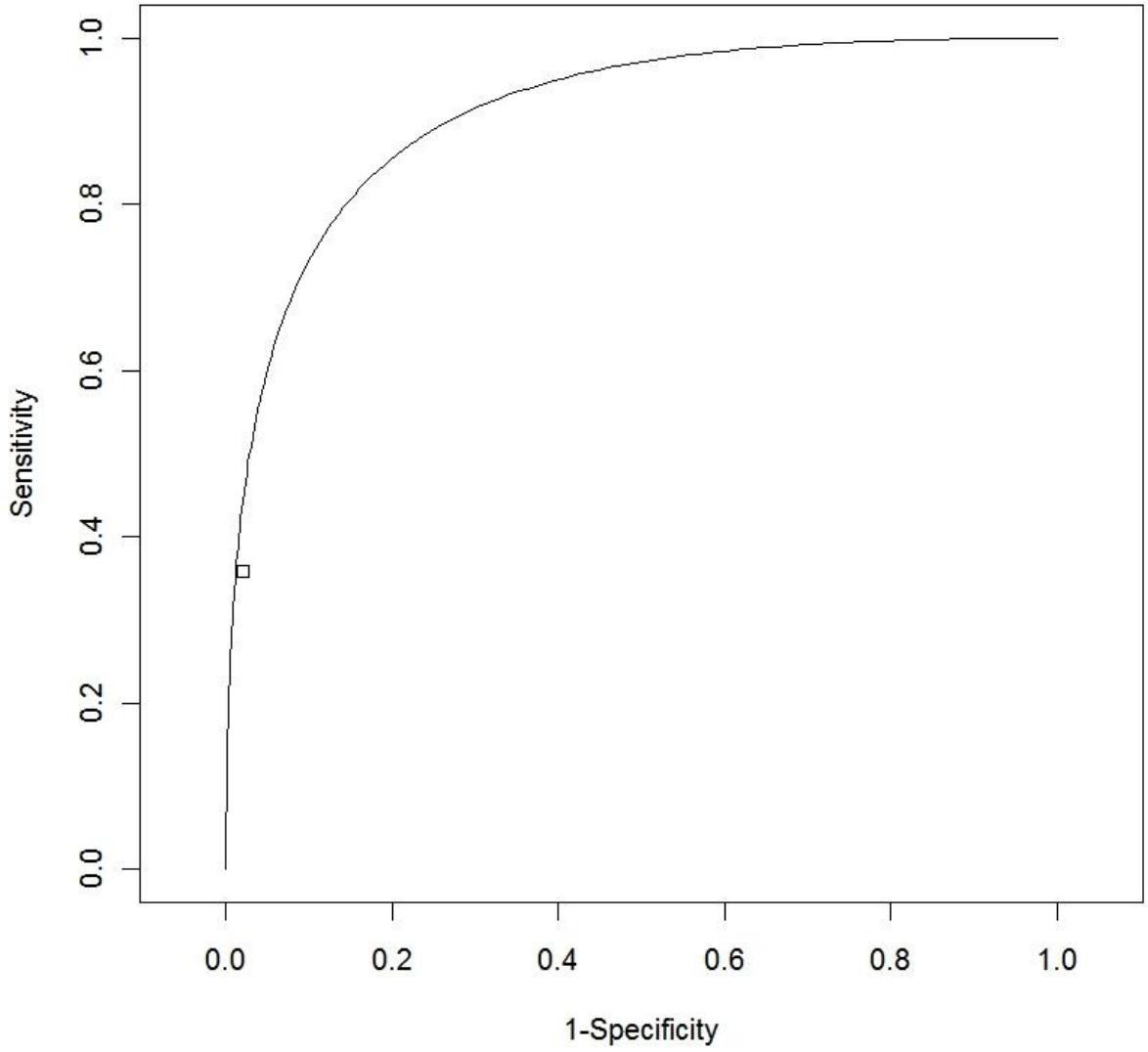


Figure 4.

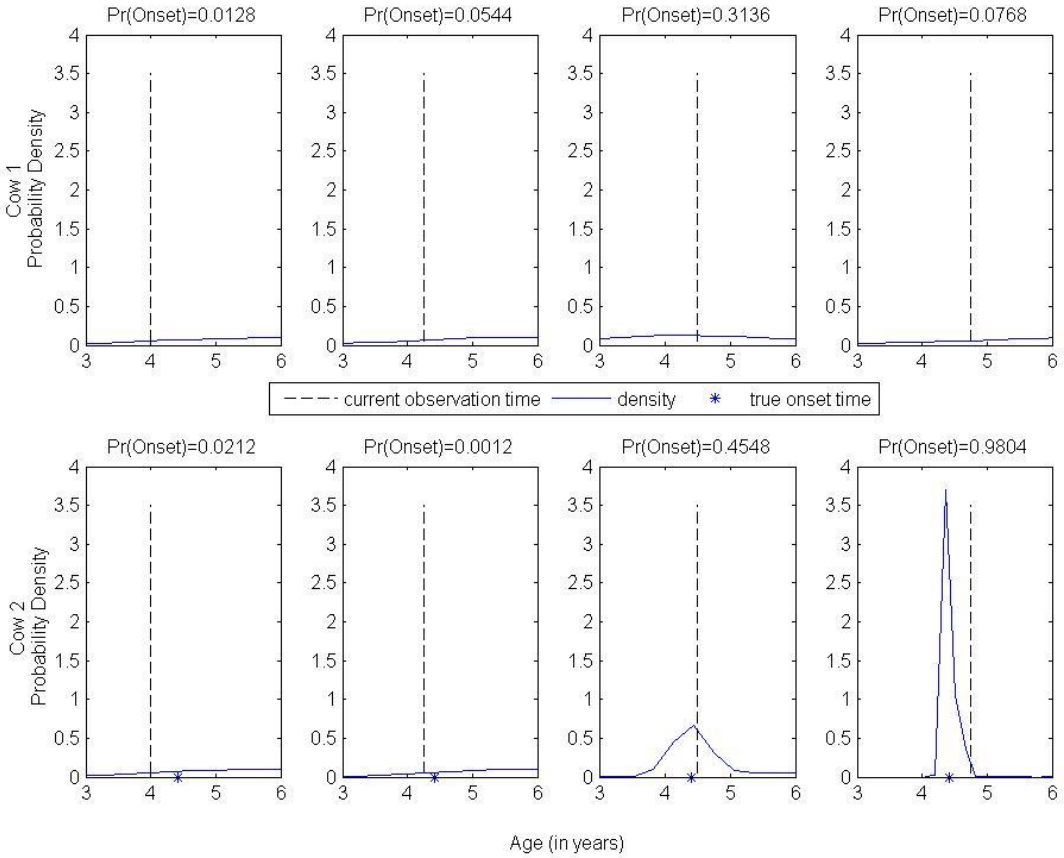
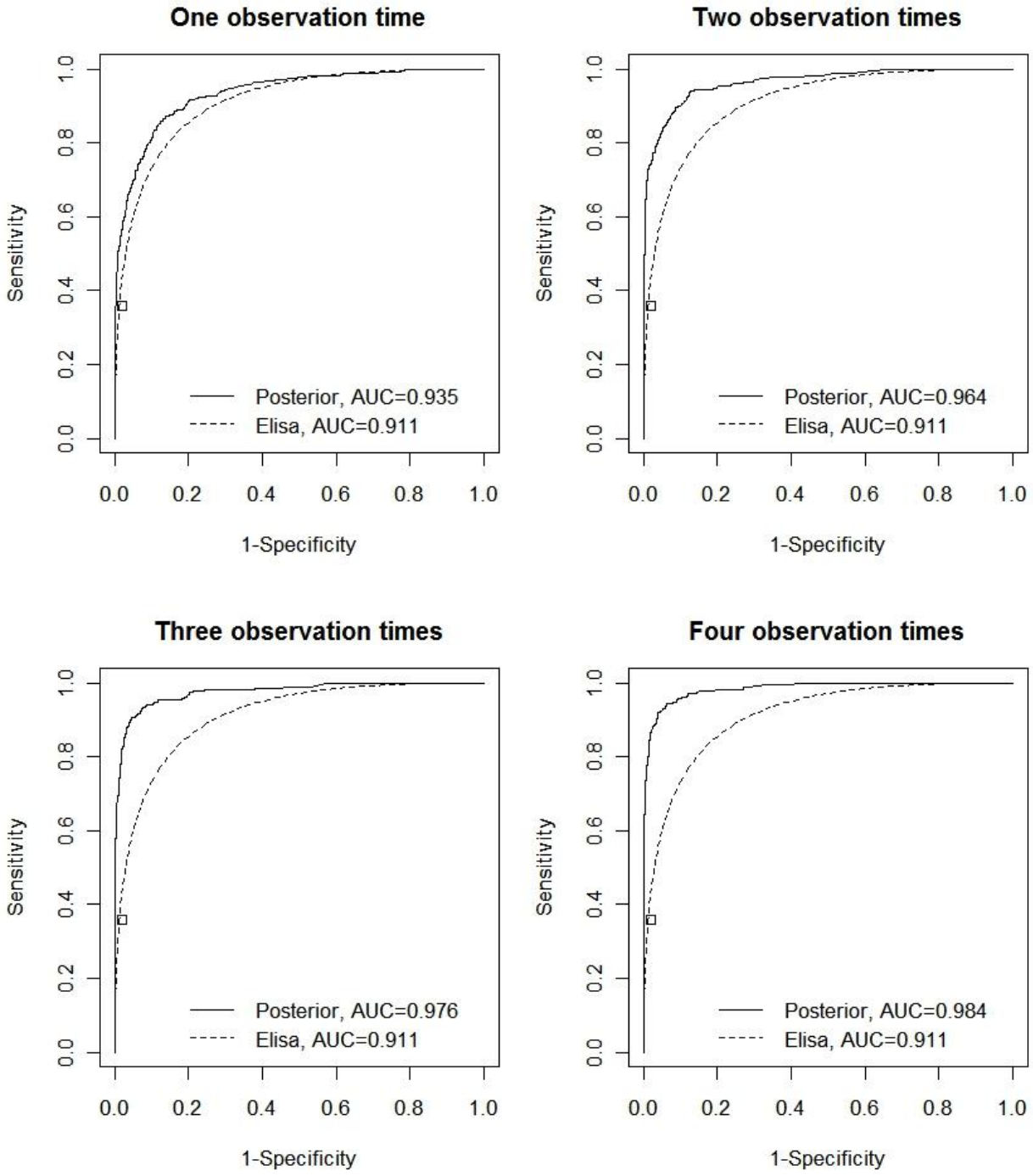

Pr(Onset)=0.0128
Pr(Onset)=0.0544
Pr(Onset)=0.3136
Pr(Onset)=0.0768
Cow 1
Probability Density
current observation time
density
true onset time
Pr(Onset)=0.0212
Pr(Onset)=0.0012
Pr(Onset)=0.4548
Pr(Onset)=0.9804
Cow 2
Probability Density
Age (in years)

Figure 5.


One observation time
Sensitivity
1-Specificity
Posterior, AUC=0.935
Elisa, AUC=0.911
Two observation times
Posterior, AUC=0.964
Elisa, AUC=0.911
Three observation times
Posterior, AUC=0.976
Elisa, AUC=0.911
Four observation times
Posterior, AUC=0.984
Elisa, AUC=0.911