Modeling bivariate longitudinal diagnostic outcome data in the absence of a gold standard

Michelle Norris, Wesley O. Johnson* and Ian A. Gardner

Diagnostic screening involves testing humans or animals for the presence of disease or infection. For some diseases, a perfect, "gold-standard" test does not exist or is too invasive or expensive to use. Hence, the goals of diagnostic testing may include: quantifying the performance of an imperfect test, diagnosing subjects, and estimating disease prevalence – possibly in the absence of a perfect reference test. To date, most work in this area has focused on cross-sectional data. We extend recent work by developing a model for bivariate longitudinal diagnostic outcomes in the no-gold standard case. We consider the situation where a continuous test and a binary test are repeatedly administered to each subject. For infected subjects, we assume the existence of a changepoint corresponding to time of infection and posit appropriate changes to the model thereafter. This results in a varying-dimensional parameter space since the true infection status of the subjects is unknown. We make inference using Bayesian Markov chain Monte Carlo methods, incorporating the Reversible Jump Markov chain Monte Carlo algorithm of Green for posterior simulation from a varying-dimensional parameter space. We test the model's performance on simulated data, and then analyze a data set based on Johne's disease in cattle.

Keywords and phrases: Changepoint model, Gibbs sampler, Johne's disease, Longitudinal data, Markov chain Monte Carlo, No-gold standard, Reversible Jump.

1. INTRODUCTION

Over the past few decades, considerable progress has been made in the area of diagnostic screening for disease. Methods now exist to handle no-gold standard data, relax assumptions about normality of diagnostic outcomes, and handle dependence among screening tests. Historically, work has focused on cross-sectional data, with important early work by Hui and Walter (1980) and Gastwirth (1987), which established the capability to make inferences about diagnostic test accuracy in the absence of a gold standard. Recent work explores models for longitudinal screening data. In this paper, we extend screening models to handle bivariate longitudinal screening data with one continuous and one binary response, in the absence of a gold-standard.

*Corresponding author.

Early work in this area focused on the construction of models for subjects who were *known* to have a disease of interest. In some studies, the date of onset of the disease was known, whereas in others the date of onset was only known to have occurred in some, possibly large, window of time. These studies involved a longitudinal collection of continuous outcomes. Typical models posit subject-specific trajectories varying about a population-average trajectory. Trajectories are often assumed to be piecewise linear with a subject-specific changepoint that is related to disease progression. Applications of this type of model to the study of HIV and dementia can be found in Lange, Carlin and Gelfand (1992) and Hall, Ying, Kuo and Lipton (2003), respectively.

Later work extended these longitudinal models to the case where disease status was unknown, either for all or a portion of the subjects. These models often incorporate latent disease indicator variables, and allow for different parametric forms for the subject-specific trajectories conditional on disease status. Examples of this type of model can be found in Pauler and Laird (2000) and Skates, Pauler and Jacobs (2001).

Less work has been done in the area of *joint* modeling of longitudinal diagnostic outcomes. One exception is the work of Cook, Ng, and Meade (2000) who consider the situation where multiple, possibly dependent binary tests are administered to each subject at each observation time. They consider the balanced case, propose modeling the underlying latent disease process using a hidden Markov model, and use maximum likelihood estimation. More recently, Engel, Backer and Buist (2009) proposed a Bayesian model for multiple binary tests applied longitudinally in the no-gold standard case. They employed Markov Chain Monte Carlo techniques to make inference.

Here, we develop a joint model for longitudinal diagnostic screening data. In particular, we jointly model continuous and binary outcomes over time with no gold standard. Our model accommodates irregular and missing data. This model will be useful for diagnosing infection in individual subjects, assessing the performance of diagnostic tests, understanding the progression of antibody response, and studying the temporal dynamics of infection within a population. An important feature of our model is that it allows us to estimate sensitivity of a dichotomized version of the continuous outcome as a function of time. Extensions to multiple tests are conceptually straightforward.

We work in the Bayesian framework and use Markov Chain Monte Carlo (MCMC) methods to make inferences. In addition, because (i) we work in the no-gold standard case where the true disease status of each subject is unknown and must be estimated along with the parameters and (ii) the number of parameters in the model varies, depending on the unknown true disease status, we incorporate the Reversible Jump MCMC (RJMCMC) algorithm of Green (1995). Green's algorithm is well-suited for Bayesian model selection problems which lead to varying-dimensional parameter spaces such as this one. We use RJMCMC in a manner similar to Pauler and Laird (2000) and Skates, Pauler and Jacobs (2001).

The motivating dataset for this paper consists of joint longitudinal screening data for Johne's disease (JD) in cattle. Johne's disease is a chronic bacterial infection caused by *Mycobacterium avium* subspp *paratuberculosis* (Map), which can cause diarrhea, weight loss and reduced milk production. However, these signs may not manifest themselves for months to years after infection, if at all. Early diagnosis in this asymptomatic phase is desirable since infected cows may pass the infection on to herdmates whether or not they are exhibiting signs of infection.

Our data consist of records from 10 dairy herds known to be infected with JD. Data were collected from 1984-2003, and tests were performed about every six months. There was, however, substantial deviation from this testing schedule with one-fourth of inter-test times below 4 months and one-fourth above 8.3 months. Herd size ranged from 50 to 160 milking cows (median = 60). Three hundred, sixty-five cows from this study were included in our analysis; the number of observations for each cow ranged from 2 to 23 with a median of 6. At each screening time, both fecal culture and serum ELISA tests were done, although for various reasons either of these may be missing on a given test date. Fecal culture test results were categorized as positive if at least one Map colony was visible on the culture medium and negative otherwise. ELISA tests measure the antibody concentration on a continuous scale through the Optical Density (OD). Plots for data from four cows are shown in Fig. 1.

The remainder of this paper is organized as follows: Section 2 gives background for the Johne's disease data, Section 3 describes the proposed model, Section 4 covers inference, Section 5 gives model extensions, Section 6 gives simulation results, Section 7 contains an analysis of the Johne's disease data, and Section 8 contains concluding remarks.

2. BACKGROUND

Figure 1 indicates that there is a clear difference in the behavior of the fecal and serology outcomes depending on whether the cow is infected or not. Additionally, the serology response takes 10–17 months to develop (Lepper et al.,
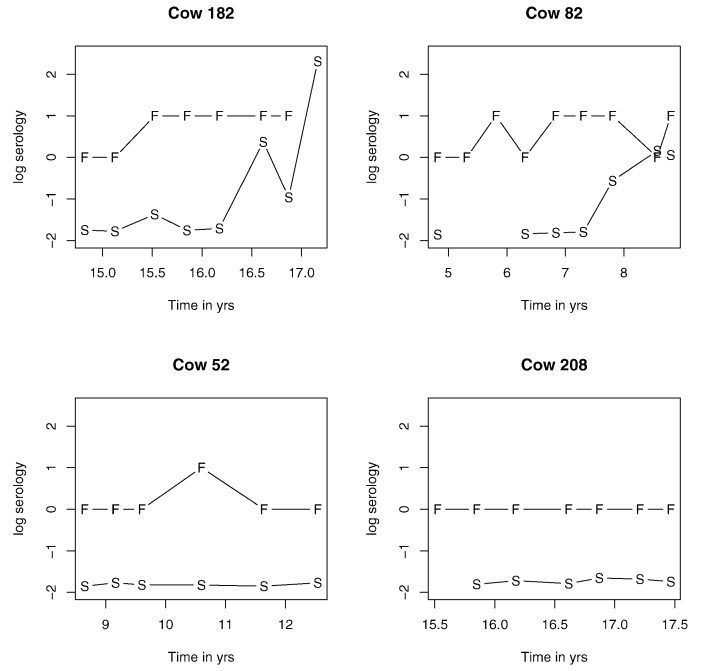

Figure 1. Plots of serology and fecal culture data by cow. Cows 82 and 182 are infected; cows 52 and 208 are uninfected. F = fecal culture (binary), S = log of serology score (OD).
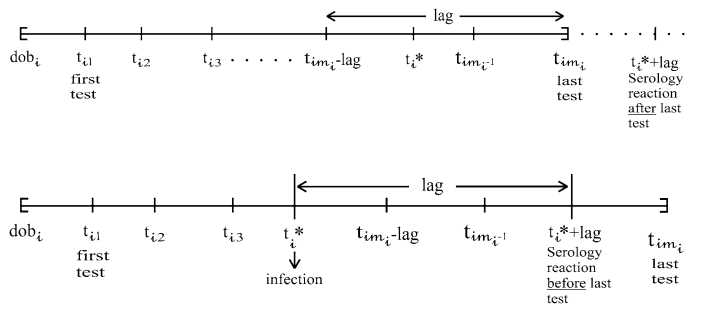

Figure 2. Upper graph is infection without serologic reaction; lower is infection with serologic reaction. $(t_{i1}, t_{i2}, \ldots, t_{im_i})$ is the vector of screening times for cow i. t_i^* is the time of infection and dob_i is the date of birth of the ith cow.

1989), a time period we refer to as the "lag." Due to the lag, a cow that becomes infected "close" to the last screening time will fail to exhibit a serology response during the screening window (see Fig. 2). Thus, considering the entire screening window, the behavior of a cow's data will depend on the latent infection state:

- "no infection" – no infection has occurred from the date of birth to the last screening, no changes in either fecal or serology behavior
- "infection without serologic reaction" – infection occurs prior to cow's last screening date but serology scores

do not exhibit a rise during screening window, change occurs in fecal behavior but not serology

- "infection with serologic reaction" – indicates the cow has been infected and had a serologic reaction within the screening window, change occurs in fecal and serology behavior

Figure 2 depicts the relationship between infection time, last screening time and lag for the "infection without serologic reaction" and "infection with serologic reaction" cases. We will define a different model for each of these latent states. Changepoints will be included to model changes in the fecal and serology behavior.

Technically, the length of time between infection and serologic response, or lag, will vary from subject to subject, but concentrate about some "average" value (Lassauzet, Johnson and Thurmond, 1989). Our model, however, will assume a common lag across subjects.

Further, we make some assumptions about independence of fecal and serology outcomes. For cows in the "no-infection" state, we assume the fecal culture and serology vectors are independent. This assumption seems reasonable since fecal and serologic tests have different biological bases that, presumably, only become correlated when the cow contracts JD. For cows in either of the two infection states, we assume the existence of an unknown cow-specific infection time, denoted t_i^*. We further assume the fecal and serology vectors are conditionally independent given this infection time. This assumption implies that the correlation between a cow's serology and fecal culture results at any particular screening time is due primarily to the progression of the infection. One would expect animals infected for a considerable length of time to tend to have jointly "positive" outcomes and animals not infected to tend to have jointly "negative" outcomes. This assumption may be too simplistic for situations where (i) it is thought that cows will have differing rates of overall response to infection and (ii) this cow-specific response rate affects both the serology and fecal outcomes. In such cases, the latent intensity model developed by Dendukuri and Joseph (2001), which correlates two concurrent tests on a subject through a random subject effect, might be more suitable. We present some empirical evidence for the conditional independence assumption for the JD data in Section 7.1.

3. THE MODEL

In this section, we give notation and formally define a model for joint longitudinal diagnostic outcomes. Actually, we define three models, one corresponding to each latent infection state. Since we make inference in the Bayesian framework, we also specify priors. Also, the number of parameters varies across infection states so we specify three sets of priors, one for each latent state.

3.1 Models by latent infection state

We propose the following hierarchical models, which allow cow-specific serology trajectories to vary around a population-average serology trajectory. The serology trajectory is assumed to be "flat" when a cow is not having a serologic reaction and increasing with a cow-specific positive slope starting from a "changepoint" that occurs one lag after infection. We assume actual serology measurements are subject to random error about the cow-specific trajectory and model them as independent and identically distributed (iid) normal.

In reality, serology scores tend to be skewed, so they were log transformed to make the normality assumption reasonable. Additionally, serology scores are expected to increase once a reaction starts so we constrain the slope of the serology trajectory after the changepoint to be positive by modeling the log-slope as normal.

In order to specify the models, we need the notion of sensitivity and specificity of a binary diagnostic test. The sensitivity is the probability that a diseased subject tests positive whereas the specificity is the probability a non-diseased subject tests negative. Perfect binary tests will have sensitivity and specificity both equal to 1.

We now define the following quantities: t_{ij} = time of the jth screening for the ith subject, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m_i$; (S_{ij}, F_{ij}) = the serology and fecal culture outcomes of the ith subject at time t_{ij}; k_i = infection status indicator, $k_i \in \{1, 2, 3\}$; (q_1, q_2, q_3) = probability $k_i = 1, 2, 3$, respectively; Se_F = sensitivity of fecal culture; Sp_F = specificity of fecal culture; lag = time interval between infection and serology reaction; Θ = vector of all model parameters; and U = vector of all model latents. The parameter and latent vectors will be given explicitly following the model definition. Also, we parameterize normal distributions with the mean and precision, where precision is the reciprocal of variance and is often denoted by τ with a subscript. This parameterization facilitates the MCMC calculations.

We first consider the model for a cow in the no infection state, which we denote as $k_i = 1$. We use the general notation $X \sim \text{Bernoulli}(p)$ to denote a random variable X that assumes the value 1 with probability p and the value 0 with probability $1 - p$. For the ith subject at time t_{ij},

$$
\begin{aligned}
(1) \qquad & S_{ij}|\Theta, U, k_i = 1 \sim \beta_{0i} + \epsilon_{ij}, \quad \perp \\
& F_{ij}|\Theta, U, k_i = 1 \sim \text{Bernoulli}(1 - Sp_F)
\end{aligned}
$$

where $\beta_{0i} \overset{\perp}{\sim} N(\beta_0, \tau_{\beta_0})$, $\epsilon_{ij} \overset{\perp}{\sim} N(0, \tau_e)$, and $\beta_{0i} \perp \epsilon_{ij} \quad \forall i, j$. We use the notation $X_i \overset{\perp}{\sim} F$ to indicate that the random variables, X_i, are independent and identically distributed according to F. Here, no change in the behavior of the data is anticipated. Hence, the serology trajectory is flat and no changepoints are included in the model. Next
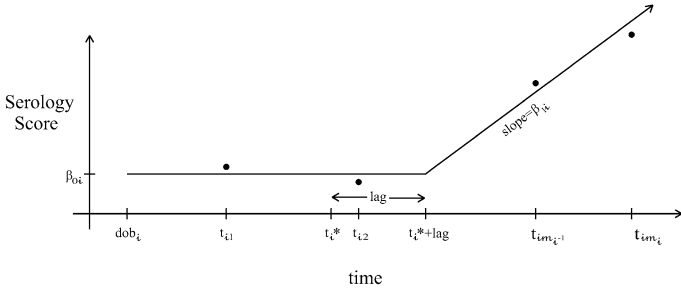
Figure 3. Serology trajectory with data for cow with $k_i = 3$.

consider $k_i = 2$ (infection without serology reaction). We have

$$S_{ij}|\Theta, U, k_i = 2 \sim \beta_{0i} + \epsilon_{ij}, \quad \perp$$
$$F_{ij}|\Theta, U, k_i = 2 \sim \text{Bernoulli}(\pi_{ij})$$

where $\pi_{ij} = I(t_{ij} \geq t_i^*)Se_F + I(t_{ij} < t_i^*)(1 - Sp_F)$, $\beta_{0i} \overset{\perp}{\sim} N(\beta_0, \tau_{\beta_0})$, $\epsilon_{ij} \overset{\perp}{\sim} N(0, \tau_e)$, and $\beta_{0i} \perp \epsilon_{ij} \quad \forall i, j$. The only difference from the $k_i = 1$ model is that the probability of a positive fecal culture changes from $1 - Sp_F$ to Se_F at the changepoint t_i^*. We continue to assume that S_{ij} and F_{ij} are conditionally independent given all parameters and latents. Finally, with $k_i = 3$ (infection with serology reaction),

$$S_{ij}|\Theta, U, k_i = 3 \sim \beta_{0i} + \beta_{1i}(t_{ij} - t_i^* - \text{lag})^+ + \epsilon_{ij}, \quad \perp$$
$$F_{ij}|\Theta, U, k_i = 3 \sim \text{Bernoulli}(\pi_{ij})$$

where β_{0i} and ϵ_{ij} are distributed as in the previous cases, $\log \beta_{1i} = \gamma_i \overset{\perp}{\sim} N(\mu_\gamma, \tau_\gamma)$ and the notation $z^+ = z$ if $z > 0$, 0 otherwise. We also assume that β_{1i}, β_{0i} and ϵ_{ij} are pairwise independent for all i and j. Also, S_{ij} and F_{ij} are conditionally independent given all parameters and latents. Note that the term containing β_{1i} is zero until $t_{ij} = t_i^* + \text{lag}$, at which point we will see a "kink" in the trajectory. More specifically, the serology trajectory will be flat until $t_i^* + \text{lag}$, at which point it will begin to increase linearly. Figure 3 shows a graph of the serology trajectory for a cow with $k_i = 3$.

In any of the latent infection states, we can think of β_{0i} as cow i's baseline, or disease-free, serology score (without measurement error) so it is reasonable to model these as iid from a common distribution across all three latent infection states.

3.2 Prior specification

We consider parameters as either "global" or "cow-specific." Global parameters are parameters such as τ_e that apply across all cow-level models. "Cow-specific" parameters are specific to an individual cow's model and, consequently, are subscripted with i. One example is cow i's infection time, t_i^*. Since we work in a Bayesian framework, we now specify

priors. Because the parameter space changes from model to model, we have slightly different prior specifications for each sub-model. For instance, the infection with serology reaction state, $k_i = 3$, requires that $t_i^* \leq t_{im_i} - \text{lag}$, whereas the infection with no serology reaction requires $t_{im_i} - \text{lag} < t_i^* \leq t_{im_i}$. Hence, the priors on t_i^* need to be specified separately for these two models.

Letting $q_l = P(k_i = l)$ for $l = 1, 2, 3$, the priors for global parameters are

$$(q_1, q_2, q_3) \sim \text{Dirichlet}(\zeta_1, \zeta_2, \zeta_3)$$
$$Se_F \sim \text{Beta}(a_{se_F}, b_{se_F})$$
$$Sp_F \sim \text{Beta}(a_{sp_F}, b_{sp_F})$$
$$\text{lag} \sim U(\min_l, \max_l)$$
$$\beta_0 \sim N(\mu_{\beta_0}, b)$$
$$\tau_{\beta_0} \sim \Gamma(a_{\beta_0}, b_{\beta_0})$$
$$\mu_\gamma \sim N(a_{\mu_\gamma}, b_{\mu_\gamma})$$
$$\tau_\gamma \sim \Gamma(a_{\tau_\gamma}, b_{\tau_\gamma})$$
$$\tau_e \sim \Gamma(a_{\tau_e}, b_{\tau_e})$$

where $\min_l$ and $\max_l$ are reasonable lower and upper bounds on the lag. Infection times under infection state 2 are modeled as

$$(2) \qquad t_i^*|k_i = 2, \text{lag} \sim U[t_{im_i} - \text{lag}, \ t_{im_i}]$$

Under latent infection state 3, the prior for infection times is

$$(3) \qquad t_i^*|k_i = 3, \text{lag} \sim U[\text{dob}_i, \ t_{im_i} - \text{lag}]$$

In situations where a disease is well-studied and a good deal is known about the diagnostic tests, we can strengthen our analysis by incorporating this information into the model through informative priors. The construction of these priors involve eliciting information from an expert. We explain how this would be accomplished for the fecal culture sensitivity.

To elicit a prior for Se_F, we would obtain the expert's most likely, or modal, value and $(\alpha \times 100)$th percentile for Se_F. α is typically taken to be 0.95 but may be any value that is easy for the expert to think about. Now we can easily obtain the parameters of the beta prior, (a_{se_F}, b_{se_F}), conforming to the elicited mode and 95th percentile using software called "BetaBuster" available at http://www.epi.ucdavis.edu/diagnostictests/.

4. INFERENCE

We work in a Bayesian framework so our interest is in the joint posterior probability density function (pdf) of the parameters and latents given the data. In this section, we derive this posterior distribution, which we approximate by sampling from it using Markov Chain Monte

Carlo (MCMC) techniques. Inference for all parameters and functionals of interest are obtained from this sample. The sampling scheme is based on a Gibbs sampler which incorporates Metropolis, slice and reversible jump steps (Green, 1995). Details of the MCMC sampling scheme, including the full conditionals necessary for the Gibbs sampler, the methods for sampling these full conditionals, and the Reversible Jump Markov Chain Monte Carlo (RJMCMC) steps used to sample the posterior, are detailed in the Appendix.

Reversible jump MCMC is a tool for sampling varying-dimensional parameter spaces. We require it because we work in the no gold-standard case where the latent infection state is unknown. Consequently, the latent infection states will be sampled in our Markov chain and will potentially change from iteration to iteration. Since the latent infection state defines the sub-model for each cow and these sub-models have varying-dimensional vectors of parameters and latents (for model 1, 2, and 3, these vectors are (β_{0i}), (β_{0i}, t_i^*), and $(\beta_{0i}, \beta_{1i}, t_i^*)$, respectively), the number of parameters we sample at each iteration may change. Hence, we require a mechanism for sampling varying-dimensional parameter spaces like RJMCMC. Note that RJMCMC allows us to perform model selection at the individual cow level and estimate model parameters simultaneously. Hence, our inferences and predictions will reflect the uncertainty associated with model selection.

4.1 Posterior distribution

We now specify the augmented data likelihood of this model and then use it to derive the posterior distribution. To economize on notation, we set $data = \{(S_{ij}, F_{ij}), t_{ij} \quad \forall i, j\}$. Note that, if a log transform is used for the serology data, S_{ij} actually represents the log of the serology score. The latents, denoted by U, include $\{\beta_{0i} : i = 1, 2, \ldots, n\} \bigcup \{\beta_{1i} : i$ such that $k_i = 3\} \bigcup \{k_i : i = 1, 2, \ldots, n\}$, while the parameters, denoted by Θ include $q_1, q_2, q_3, \beta_0, \tau_{\beta_0}, \tau_e, se_F, sp_F,$ lag, μ_γ, τ_γ and $\{t_i^* : i = 1, 2, \ldots, n\}$. We have

(4)

$$L(\Theta|U, \text{data}) = f(\text{data}|U, \Theta) \cdot f(U|\Theta) \propto$$

$$\left\{ \prod_{i:k_i=1} \left[q_1 \prod_{j=1}^{m_i} \tau_e^{\frac{1}{2}} \exp\{-\frac{\tau_e}{2}(S_{ij} - \beta_{0i})^2\} \right. \right.$$

$$\left. \times Sp_F^{1-F_{ij}}(1 - Sp_F)^{F_{ij}} \right] \tau_{\beta_0}^{\frac{1}{2}} \exp\{-\frac{\tau_{\beta_0}}{2}(\beta_{0i} - \beta_0)^2\} \right\}$$

$$\times \left\{ \prod_{i:k_i=2} \left[q_2 \prod_{j=1}^{m_i} \tau_e^{\frac{1}{2}} \exp\{-\frac{\tau_e}{2}(S_{ij} - \beta_{0i})^2\} \right] \right.$$

$$\times \prod_{j:t_{ij}<t_i^*} \left[Sp_F^{1-F_{ij}}(1 - Sp_F)^{F_{ij}} \right]$$

$$\times \prod_{j:t_{ij}\geq t_i^*} \left[Se_F^{F_{ij}}(1 - Se_F)^{1-F_{ij}} \right]$$

$$\left. \times \tau_{\beta_0}^{\frac{1}{2}} \exp\{-\frac{\tau_{\beta_0}}{2}(\beta_{0i} - \beta_0)^2\} \right\}$$

$$\times \left\{ \prod_{i:k_i=3} \left[q_3 \prod_{j=1}^{m_i} \tau_e^{\frac{1}{2}} \exp\{-\frac{\tau_e}{2}(S_{ij} - \beta_{0i} \right. \right.$$

$$\left. - \beta_{1i}(t_{ij} - t_i^* - \text{lag})^+)^2\} \right] \prod_{j:t_{ij}<t_i^*} \left[Sp_F^{1-F_{ij}}(1 - Sp_F)^{F_{ij}} \right]$$

$$\times \prod_{j:t_{ij}\geq t_i^*} \left[Se_F^{F_{ij}}(1 - Se_F)^{1-F_{ij}} \right] \times \tau_{\beta_0}^{\frac{1}{2}}$$

$$\left. \times \exp\{-\frac{\tau_{\beta_0}}{2}(\beta_{0i} - \beta_0)^2\} \frac{\tau_\gamma^{\frac{1}{2}}}{\beta_{1i}} \exp\{-\frac{\tau_\gamma}{2}(\log\beta_{1i} - \mu_\gamma)^2\} \right\}$$

Lines 2–3 represent the augmented data likelihood for subjects having no infection ($k_i = 1$), the next four lines correspond to subjects having infection without serology reaction ($k_i = 2$), and the final four lines correspond to subjects having infection with serology reaction ($k_i = 3$).

Thus, the joint "posterior" pdf is proportional to the product of the augmented data likelihood times the prior:

(5)

$$p(\Theta, U|\text{data}) \propto$$
$$L(\Theta|U, \text{data}) \cdot \pi(q_1, q_2, q_3, \beta_0, \tau_{\beta_0}, \tau_e, se_F, sp_F, \text{lag}, \mu_\gamma, \tau_\gamma)$$
$$\times \prod_{i:k_i=2} \pi_2(t_i^*) \cdot \prod_{i:k_i=3} \pi_{3i}(t_i^*)$$

where $\pi_2(t_i^*) = I(t_{im_i} - \text{lag} < t_i^* \leq t_{im_i})/\text{lag}$, the prior for t_i^* under model 2 and $\pi_{3i}(t_i^*) = I(\text{dob}_i \leq t_i^* \leq t_{im_i} - \text{lag})/(t_{im_i} - \text{lag} - \text{dob}_i)$, the prior for t_i^* under model 3.

The MCMC sampling scheme described in the Appendix converges because the posterior distribution is invariant under each step in the Gibbs sampler and the chain is irreducible and aperiodic under the entire procedure (Tierney, 1994).

5. MODEL EXTENSIONS

This section describes how the basic model can be extended to include a quadratic serology trajectory or covariates related to (q_1, q_2, q_3), the probabilities of being in each latent disease state.

5.1 Quadratic serology trajectory

Modeling the serology trajectory with an increasing linear trend makes sense for a chronic, slowly progressing disease like Johne's disease. However, such a model may be too restrictive for other diseases or even for Johne's if data are collected over a longer period of time. This is because the serology response will eventually level off and may even decline after a period of leveling. Modeling the serology trajectory as a quadratic would address such behavior. We can accomplish this by modifying the current model and posterior sampling scheme.

We need to be careful to constrain the quadratic so that it models serology behavior in a manner that makes biological sense. For example, to model an increase in antibodies

followed by a leveling and decrease, we need the quadratic function to be concave down with a maximum occurring after $t_i^* + \text{lag}$. A serology trajectory that is concave upward would mean that antibody production declines after infection (+lag), which is biological nonsense.

A formulation that addresses these biological constraints has $\beta_{1i} > 0$ and $\beta_{2i} < 0$ in the following model 3 serology component

$$S_{ij} | \Theta, U, k_i = 3 \sim \beta_{0i} + \beta_{1i}(t_{ij} - t_i^* - \text{lag})^+ \\ + \beta_{2i}((t_{ij} - t_i^* - \text{lag})^+)^2 + \epsilon_{ij}$$

where

$$\left(\begin{array}{c} \log \beta_{1i} \\ \log(-\beta_{2i}) \end{array} \right) = \left(\begin{array}{c} \gamma_i, \\ \nu_i \end{array} \right) \sim N\left(\left(\begin{array}{c} \gamma \\ \nu \end{array} \right), \Delta \right),$$

$\beta_{0i} \overset{\perp}{\sim} N(\beta_0, \tau_{\beta_0})$, and $\epsilon_{ij} \overset{\perp}{\sim} N(0, \tau_e)$. We assume $(\beta_{1i}, \beta_{2i}) \perp \beta_{0i} \perp \epsilon_{ij} \quad \forall i, j$.

If more complicated trajectories were anticipated, other, more flexible, forms such as penalized splines could be used (cf. Lang and Brezger, 2004). Additional flexibility could be incorporated by modeling both the trajectory and the distribution of cow-specific intercepts nonparametrically as described in Li, Lin and Müller (2009).

5.2 Latent disease state covariates

If covariates are thought to be related to the probability of being in latent states 1, 2 and 3, they can be incorporated into the model using two "nested" binary regressions. One obvious covariate might be age or some other proxy for length of time exposed to the infectious agent. It would seem reasonable that longer exposure would increase the chance of infection and the chance of progressing to latent state 3. The vector of probabilities of being in latent states 1, 2 and 3, (q_1, q_2, q_3), would be replaced by the set $\{(q_{1i}, q_{2i}, q_{3i}) : i = 1, 2, \ldots, n\}$, which contains a vector of probabilities for each cow. Letting X_i be the vector of covariates for cow i, we use a probit model to related them to (q_{1i}, q_{2i}, q_{3i}). Specifically, we set $Pr(k_i = 1) = q_{1i} = \Phi(X_i^T \omega)$. We model the probability of being in latent state 2, conditional on *not* being in latent state 1, by defining $\tilde{q}_{2i} = \frac{q_{2i}}{1 - q_{1i}}$ and setting $\tilde{q}_{2i} = \Phi(X_i^T \nu)$. We then have that $q_{2i} = (1 - q_{1i})\tilde{q}_{2i}$ and $q_{3i} = 1 - q_{1i} - q_{2i}$. The computational approach taken by Albert and Chib (1993) would simplify the Monte Carlo sampling. Of course, any cumulative distribution function could be used as the link function in the binary regression. This approach generalizes the one taken by Branscum et al. (2008). Finally, suitable priors for the regression coefficients can be formulated as in Bedrick et al. (1996), together with modifications to the MCMC scheme.

6. SIMULATION STUDY

We investigated the model's performance using a simulation study. The simulation consisted of data for 100 subjects with 37, 14, and 49 subjects in latent disease states 1, 2,

Table 1. Means and 95% probability intervals for simulated data

Para-meter	Truth	Post. Mean	Post. Median	Lower Limit	Upper Limit
β_0	0.08	0.075	0.075	0.059	0.091
τ_{β_0}	2000	1323	1116	421	3417
τ_e	20	20.7	20.7	19.0	22.5
μ_γ	0.40	0.44	0.44	0.37	0.51
τ_γ	25	28.8	27.6	14.5	49.7
Se_F	0.60	0.59	0.59	0.54	0.64
Sp_F	0.99	0.985	0.986	0.970	0.996
lag	1.00	0.93	0.93	0.86	0.99
q_1	0.37	0.38	0.38	0.28	0.49
q_2	0.14	0.13	0.13	0.05	0.22
q_3	0.49	0.49	0.49	0.39	0.59

Table 2. Classification accuracy for simulated data

		Classification		
		1	2	3
	1	37	0	0
True State	2	2	10	2
	3	1	2	46

and 3, respectively. The simulated data were unbalanced but complete. Screening occurred semi-annually. The number of observations per subject has median = 11, minimum = 4, and maximum = 31. We generated data using parameter values similar in magnitude to those estimated for the Johne's disease data and used diffuse proper priors. The R language (R Development Core Team, 2005) was used for simulating and analyzing data.

History plots indicated that two chains with dispersed initial values converged to the same distribution quickly, i.e. after a few thousand iterations. We used a burn-in of 5,000 iterations and based inference on the subsequent 45,000 iterations from both chains. The posterior means, medians and 95% probability intervals are shown in Table 1 along with the actual values used to simulate the data.

Trajectories fit to the serology data using the posterior means of β_{0i}, β_{1i} and t_i^* were quite accurate due to the large number of repeated observations on each cow. Observed values are shown along with fitted trajectories for 3 infected subjects in Fig. 4.

Using a 0-1 loss function, subjects were classified into the latent state with the highest posterior probability. Classification was quite accurate, with only 7 out of 100 subjects placed incorrectly as summarized in Table 2. Alternately, if we are only interested in which animals are infected or not infected, we can collapse latent states 2 and 3 into a single "infected" state and obtain an overall sensitivity and specificity for the joint model of 60/63 and 37/37, respectively.

For serology tests, classification of subjects as diseased or non-diseased is typically based on a dichotomized score; sub-
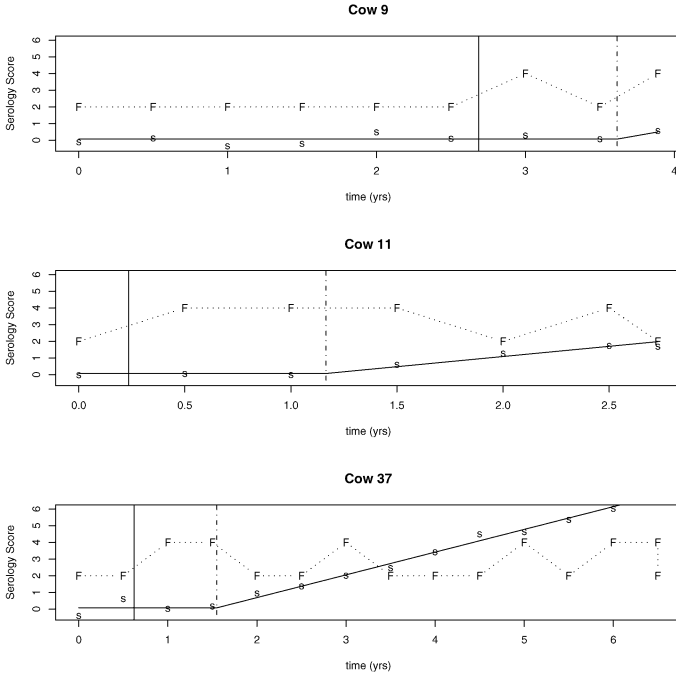
Cow 9

Cow 11

Cow 37

Figure 4. Plots of fitted serology trajectory and data by cow. All cows shown are infected. F = fecal culture, 2 if negative and 4 if positive, S = observed serology score. Solid vertical black line is estimated infection time, dash-dot vertical line is estimated serology reaction time.

jects with serology scores above a specified cutoff, denoted c, are classified as diseased while those below are classified as non-diseased. This dichotomization permits the calculation of the sensitivity and specificity of the serology test. Because the concentration of antibodies in an infected subject will rise over time (once the lag has elapsed), the serology sensitivity will be a function of time. Additionally, since subjects have their own baseline serology level and slope, sensitivity at a fixed time after infection will vary among subjects. The overall sensitivity at a fixed time after infection, t, can thus be estimated by "averaging" over the relevant posterior distributions using the following integral with $S_i(t)$ representing subject i's serology score t units of time *after* infection.

$$\hat{se}_c(t)$$
$$= Pr(S_i(t) > c|\text{data})$$
$$= E[Pr(S_i(t) > c|\text{data}, \Theta)]$$
$$= E[E\{I(S_i(t) > c)|\text{data}, \Theta\}]$$
$$= \int I(S_i(t) > c)f(U|\Theta)f(\Theta|\text{data})\mathrm{d}U\mathrm{d}\Theta$$
$$= \int I(\beta_{0i} + \beta_{1i}(t - \text{lag})^+ + \epsilon_{ij} > c)f(\beta_{0i}|\beta_0, \tau_{\beta_0})$$
$$\cdot f(\beta_{1i}|\mu_\gamma, \tau_\gamma)f(\epsilon_{ij}|\tau_e)f(\beta_0, \tau_{\beta_0}, \mu_\gamma, \tau_\gamma, \tau_e, \text{lag}|\text{data})$$
$$\cdot \mathrm{d}(\beta_{0i}, \beta_{1i}, \epsilon_{ij}, \beta_0, \tau_{\beta_0}, \mu_\gamma, \tau_\gamma, \tau_e, \text{lag})$$
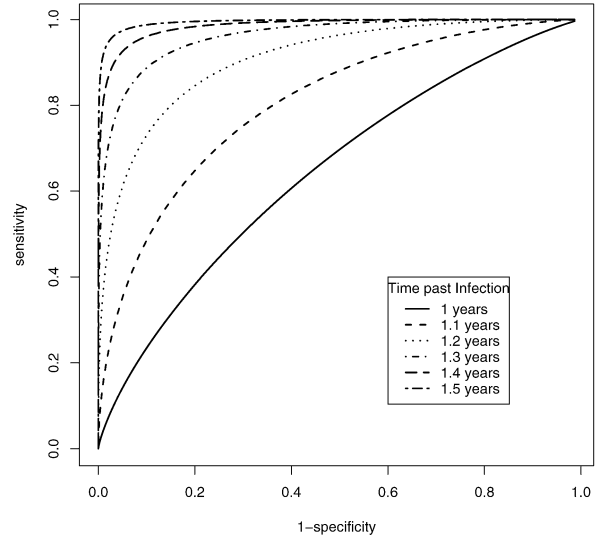

Figure 5. ROC curves for selected values of time past infection.

Since this integral is difficult to evaluate analytically, we obtain a Monte Carlo approximation from our chain as follows. For $z = 1, 2, \ldots, MC$, we sample $\beta_0^{(z)}$, $\tau_{\beta_0}^{(z)}$, $\mu_\gamma^{(z)}$, $\tau_\gamma^{(z)}$, and $\tau_e^{(z)}$. For each fixed value of z and for $l = 1, 2, \ldots, w$, we sample $\beta_{0i}^{(l,z)} \sim N(\beta_0^{(z)}, \tau_{\beta_0}^{(z)})$, $\log \beta_{1i}^{(l,z)} \sim N(\mu_\gamma^{(z)}, \tau_\gamma^{(z)})$, and $\epsilon_{ij}^{(l,z)} \sim N(0, \tau_e^{(z)})$. We estimate the sensitivity at time t after infection and cutoff, c, as

$$\hat{se}_c(t) \doteq$$
$$\frac{1}{MC}\sum_{z=1}^{MC}\{\frac{1}{w}\sum_{l=1}^{w}I(\beta_{0i}^{(l,z)} + \beta_{1i}^{(l,z)}(t - \text{lag}^{(z)})^+$$
$$+ \epsilon_{ij}^{(l,z)} > c)\}$$

Similar reasoning can be used to obtain a Monte Carlo estimator of the specificity corresponding to cutoff c, $\hat{sp}_c$. Then, if we fix t and vary the cutoff, c, we obtain a set of points, $(1 - \hat{sp}_c, \hat{se}_c(t))$. The curve generated by these points is the estimated receiver-operating characteristic (ROC) curve at time past infection t. By varying t over a grid of values, we obtain a family of ROC curves as shown in Fig. 5. Such a graph can give insight into how the performance of the serology test changes with time past infection.

7. ANALYSIS OF JOHNE'S DISEASE DATA

We now model the Johne's disease data for 365 cows described in the introduction. While the data were collected from ten herds, since we fit random effects for each cow in the data, we did not believe it necessary to include random effects for herds. We comment on this again at the end of this section.

Table 3. Parameter estimates for Johne's disease data

| Parameter | Post. Mean | 95% Probability Interval | |
		Lower	Upper
β_0	-1.745	-1.754	-1.737
σ_{β_0}	0.052	0.045	0.063
τ_e	64.1	59.7	68.7
μ_γ	0.09	-0.24	0.38
τ_γ	0.67	0.41	1.03
Se_F	0.62	0.57	0.67
Sp_F	0.985	0.973	0.993
q_1	0.46	0.39	0.53
q_2	0.25	0.18	0.31
q_3	0.29	0.24	0.35

We elicited priors for the sensitivity and specificity of the fecal culture from the third author (IG). Based on his extensive experience with the fecal test for Johne's disease, he believed that the most likely value for the sensitivity was 0.60 (Collins et al., 2006). In addition, he was 95 percent sure the sensitivity was above 0.40. This resulted in a Beta(10.9,7.6) prior for sensitivity. For fecal culture specificity, the prior was centered on 0.999 and had 5th percentile equal to 0.99, yielding a Beta(364.9,1.36) prior. The prior on μ_γ, the mean of the log slopes, was $N(0, \text{precision} = 0.1)$, which is rather diffuse, putting about 68% of the prior probability for the median of the untransformed slopes between $e^{-3} \approx 0.04$ and $e^3 \approx 24$. All other priors were likewise diffuse.

Unfortunately, we experienced poor mixing in both simulations and data analysis. This is due to the high correlation among the lag, latent disease indicators and infection times. Consequently, we used a fixed lag based on the work of Lepper et al. (1989). Their study of experimentally infected calves indicated that positive serum antibody ELISA responses occur 10–17 months after infection. Thus, we fixed the lag at 1 year and performed a sensitivity analysis to see if inference differed substantially for lags of 10 or 17 months. In the future, we hope to remedy this mixing problem as the lag is an important and often unknown parameter that diagnosticians would be interested in estimating, especially for new diseases and serology tests. We mention an extension of this model which posits cow-specific lags in lieu of a common lag for all cows. In addition to being more realistic biologically, such a model has the potential to solve this mixing problem because it reduces the correlation among the lag and other parameters.

A log transformation was applied to serology scores to make the normality assumption reasonable. We then ran two chains with different initial values for 50,000 iterations. They converged to the same posterior distribution after a few thousand iterations and mixed well. Discarding the first 5,000 iterations of each chain, we obtained the parameter estimates shown in Table 3 from the last 45,000 iterations of both chains.
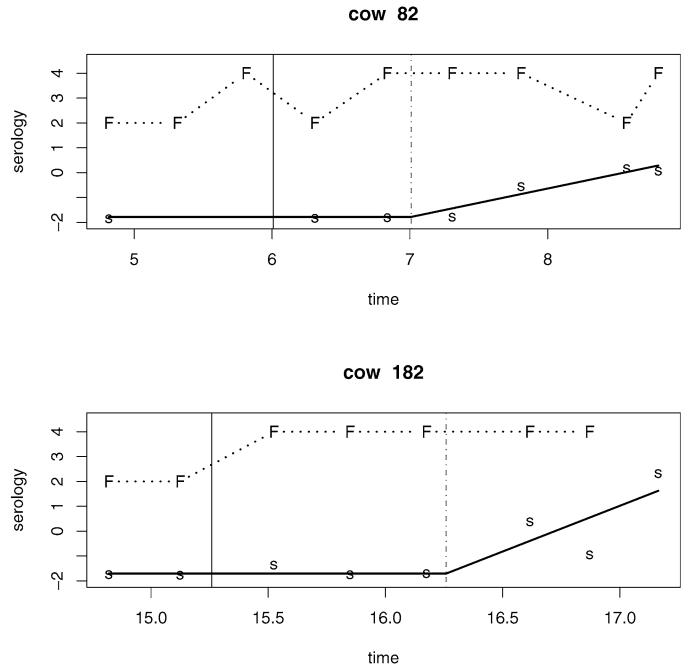
Figure 6. Fitted trajectories with data. F = fecal culture, 2 if negative and 4 if positive; S = transformed serology score. Solid vertical line is estimated infection time, dash-dot vertical line is estimated serology reaction time.

Cow-specific fitted trajectories can be obtained using the posterior means of β_{0i}, β_{1i} and t_i^*. Fitted trajectories for two cows known to be infected at necropsy are plotted in Fig. 6.

Estimated ROC curves over a grid of time past infection values are shown in Fig. 7.

7.1 Empirical investigation of the conditional independence assumption

We now consider the empirical evidence for the assumption that serology and fecal outcomes on the same cow are conditionally independent given all parameters and latents. We first consider the case for uninfected cows. Actually, we consider the broader set of test outcomes corresponding to any time when a cow was uninfected. Because it is easier to investigate the dependence between two univariate random variables (as opposed to two random vectors), we remove the dependency between repeated serology measurements on the same cow by considering the predicted residual, $\hat{\epsilon}_{ij} = S_{ij} - \hat{\beta}_{0i}$. Then, the ordered pairs $(\hat{\epsilon}_{ij}, F_{ij})$ are independent for all i and j. We can now focus on the relationship between the serology residuals and fecal outcomes. Since we work in the no-gold standard case, we use the posterior mean to estimate t_i^* so we can identify data corresponding to pre-infection test times. Based on 1,443 pairs of data, we calculate a Pearson correlation coefficient of 0.0073 (asymptotic P-value = 0.78), which is quite strong support
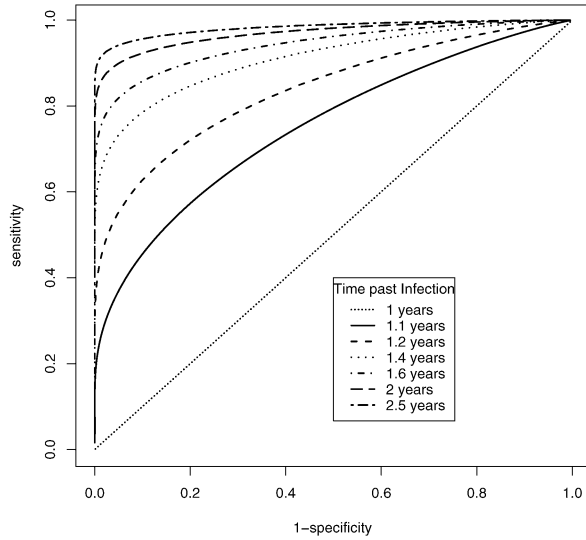
Figure 7. ROC curves for selected values of time past infection.

Table 4. Posterior means of global parameters for different values of lag time since infection

	Lag in months		
	10	12	17
β_0	-1.746	-1.745	-1.743
σ_{β_0}	0.052	0.052	0.057
τ_e	64.4	64.1	62.4
μ_γ	-0.01	0.09	0.26
τ_γ	0.59	0.67	0.91
Se_F	0.64	0.62	0.58
Sp_F	0.984	0.985	0.986
q_1	0.47	0.46	0.46
q_2	0.23	0.25	0.27
q_3	0.30	0.29	0.27

for the assumption of conditional independence when the test outcomes are from uninfected cows.

Next, we consider the possibility of correlation between concurrent serology and fecal outcomes for infected cows at times *after* infection. We again consider the predicted residual and fecal outcome pairs, i.e. $(\hat{\epsilon}_{ij} = S_{ij} - \hat{\beta}_{0i} - \hat{\beta}_{1i}(t_{ij} - \hat{t}_i^* - \text{lag})^+, F_{ij})$. There are 342 such pairs having a Pearson correlation coefficient of -0.10 (asymptotic P-value $= 0.07$). The 95% confidence interval estimate of the correlation coefficient is $(-0.200, 0.01)$. Hence, the evidence against conditional independence is not particularly strong, and, even if the serology residuals and fecal outcomes are correlated, the correlation is rather weak.

7.2 Sensitivity analysis

We performed a sensitivity analysis to determine the effect of assuming that lag is 1 year by analyzing the data assuming that lag is 10 months and 17 months, the lower and upper bounds given by Lepper et al. Table 4 compares the posterior means of the parameters for the different lags. The differences are not substantial in view of the 95% probability intervals in Table 3.

One possible explanation for the decline in estimated fecal culture sensitivity as lag increases has to do with the fact that it is easier to detect the time of serology reaction, $t_i^* + \text{lag}$, than the time of infection, t_i^* when τ_e is large and the fecal culture sensitivity is low. In this case, there is a lot more certainty in the estimation of the time of serology reaction because the linear trend is easy to detect and extrapolate whereas there is large uncertainty in the infection time because false negatives are quite likely to occur. A larger value for the fixed lag, therefore, will not change the estimated

serology reaction time much, but will shift the estimated infection time toward lower values, resulting in the inclusion of earlier screening times in the infection-prior-to-serology-reaction window. If the lag is overestimated, the subject will be incorrectly classified as "infected" at some of these early screening times. Moreover, at screening times when animals are actually uninfected, the probability of a positive test is $1 - Sp_F$, which is typically quite low. Thus, the estimated sensitivity will be diminished by negative screening outcomes corresponding to screening times at which the subject is incorrectly considered to be infected.

For 15 of the 365 cows, classification by latent disease status varied, depending on the assumed value of lag. However, only 6 cows actually changed between infected (status 2 or 3) and noninfected status. The remaining 9 switched from state 2 (infection without serology reaction) at a 17-month lag to latent state 3 (infection with serology reaction) at a 10-month lag. The switch between states 2 and 3 makes sense for a cow whose infection time has a high posterior probability of being within 10 to 17 months of the last screening and who has few/no observations in this time frame. Then, an "infection without serology reaction" model under a 17-month lag would fit nearly as well as "an infection with serology reaction" model under a 10-month lag, provided the slope in the latter model is close to zero. Removal of these near-zero slopes under the 17-month lag explains why μ_γ, the mean of the log-slopes, increases as the lag increases.

To assess sensitivity to priors, the analysis was rerun with the following diffuse priors: $Se_F \sim \text{Uniform}[0, 1]$, $Sp_F \sim \text{Uniform}[0, 1]$, and $\mu_\gamma \sim N(0, \text{precision} = 0.01)$. As shown in Table 5, these changes had a negligible effect on the estimates. The biggest differences were for the estimates of Sp_F and μ_γ. The point estimate of Sp_F declined by 0.01 and the interval widened slightly while the point and interval estimates of μ_γ all shifted to slightly higher values. The robustness of our analysis to prior assumptions is due to the large overall sample size and high probabilities across all three latent infection states. With a smaller overall sample size or low probability for some of the latent states, the

Table 5. Parameter estimates for Johne's disease data: diffuse priors

| Parameter | Post. Mean | 95% Probability Interval | |
		Lower	Upper
β_0	-1.745	-1.753	-1.737
σ_{β_0}	0.052	0.045	0.062
τ_e	64.3	60.0	68.7
μ_γ	0.16	-0.13	0.44
τ_γ	0.77	0.48	1.15
Se_F	0.63	0.58	0.68
Sp_F	0.97	0.95	0.99
q_1	0.49	0.41	0.56
q_2	0.23	0.17	0.30
q_3	0.28	0.23	0.34

priors could have a substantial impact on parameter estimation. As for classification by latent disease status, six cows changed from "infected" states (2 or 3) under informative priors to the noninfected state under diffuse priors. These cows had at least one positive fecal test and tended to have posterior probability spread quite evenly across two or three states rather than having very high posterior probability in one latent state. Thus, it seems that the informative prior for fecal culture specificity was enough to tip the balance for these six cows in favor of classification in an infected latent state.

Finally, we investigated our assumption that random herd effects were not necessary. After running our analysis, we calculated the average random effect among cows within each herd and compared averages across herds. We found minuscule differences in these averages, giving further confidence that our selected model is appropriate.

8. CONCLUDING REMARKS

We have developed a Bayesian model for jointly distributed longitudinal diagnostic test outcome data in the absence of a gold standard. The method was shown to work well using simulated data and it was illustrated using real data with expert elicited prior information. Sensitivity analysis showed results to be minimally dependent on the prior for the sample sizes considered.

Diagnosticians have long depended on cross-sectional data to obtain information about the sensitivity of tests based on dichotomized outcomes. This results in obtaining estimated sensitivity across units whose times from infection are generally quite variable. The actual sensitivity being estimated is a weighted average of sensitivities of the form $\int se_c(t)w(t)dt$, where $w(t)$ is the density of infected individuals in the population sampled that have been infected for t units of time. Our work provides, to our knowledge, a first attempt to provide a method of ascertaining how the sensitivity might increase with time since infection.

Of course, we made several assumptions in our model development. Although we were able to estimate the lag in our simulation, we had to assume the lag was known based on previous information in the data analysis. We also assumed a very regular distribution for the slopes associated with infected units when in fact there may be clusters of units that have smaller slopes while other clusters have larger slopes. In biological terms, some animals/persons may be more reactive than others to infection. We assumed that binary test outcome, fecal culture, had a constant sensitivity through time after infection. There may be other strictly binary tests for which the sensitivity increases with time as is the case for the ELISA. In fact, fecal cultures are often given as scores ranging from $0, 1, \ldots, M$; the larger the score, the greater the indication of infection. This creates even more possibilities for modeling. While we did not address these issues in this first attempt at handling this fundamentally difficult problem, we are currently attempting to do just that.

APPENDIX A. POSTERIOR SIMULATION

We now detail the Markov Chain Monte Carlo Methods used to simulate from the posterior distribution. We use a Gibbs sampler (Gelfand and Smith, 1990; Geman and Geman, 1984) that incorporates reversible jump steps. Most of the full conditionals have recognizable forms and can be sampled directly; we use Metropolis (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, 1953; Hastings, 1970) and slice sampling (Neal, 1997) for the rest. Robert and Casella (2004) catalog these and other popular MCMC techniques in their text. A tutorial on reversible jump MCMC (RJMCMC) can be found in Waagepetersen and Sorensen (2001).

Using the notations $\chi^{(k_i)}$ to represent the vector of cow-specific parameters associated with cow i under model k_i and $[\omega|\text{else}]$ to represent the distribution of the parameter or vector of parameters, ω, given all other parameters and the data, one sweep of the Gibbs sampler includes the following steps:

1. For each i, the joint full conditional, $[(k_i, \chi^{(k_i)})|\text{else}]$, is sampled. This step updates the model indicator and all corresponding parameters for cow i. We use RJMCMC for this step since changes to the model indicator result in changes to the parameter space. The reversible jump algorithm requires a proposal for k_i plus proposals for the parameters corresponding to model k_i, i.e. the $\chi^{(k_i)}$; the entire lot is then accepted or rejected. Consequently, the sub-model parameters, $\chi^{(k_i)}$, for subject i are also updated if a between-model move is accepted in this step.

2. Given the current k_i's, the full conditionals of the cow-specific parameters, $\chi^{(k_i)}$ are sampled as follows: for $k_i = 1$, sample full conditional of β_{0i}; for $k_i = 2$, sample full conditionals of β_{0i} and t_i^*; and for $k_i = 3$, sample full conditionals of β_{0i}, β_{1i} and t_i^*. Note that

if the model has accepted a change to $(k_i, \chi^{(k_i)})$ in step 1, this step will represent a "second" updating of the cow-specific parameters, $\chi^{(k_i)}$. As discussed later, this "double-sampling" does not impair convergence of the chain. It does, however, improve mixing since, without it, the cow-specific parameters will only be updated when a between model move is accepted in step 1.

3. Given model indicators and cow-specific parameters, the global parameters are updated.

All full conditionals, except the ones for $(k_i, \chi^{(k_i)})$, condition on k_i and, hence, are fixed-dimensional. Before describing how RJMCMC is used to sample the full conditional of $(k_i, \chi^{(k_i)})$, we give the fixed-dimensional full conditionals.

Letting n_i represent the number of subjects in model i for $i = 1, 2, 3$, the full conditionals for the global parameters are:

(6)

$$(q_1, q_2, q_3)|\text{else} \sim \text{Dirichlet}(n_1 + \zeta_1, n_2 + \zeta_2, n_3 + \zeta_3)$$

$$\beta_0|\text{else} \sim N\left(\frac{n\tau_{\beta_0}}{n\tau_{\beta_0} + b}\bar{\beta}_0 + \frac{b}{n\tau_{\beta_0} + b}\mu_{\beta_0}, n\tau_{\beta_0} + b\right)$$

$$\text{where } \bar{\beta}_0 = \frac{\sum_i \beta_{0i}}{n}$$

$$\tau_{\beta_0}|\text{else} \sim \Gamma\left(\frac{n}{2} + a_{\beta_0}, \frac{1}{2}\sum_i (\beta_{0i} - \beta_0)^2 + b_{\beta 0}\right)$$

$$\tau_e|\text{else} \sim \Gamma\Big(\frac{1}{2}\sum_i m_i + a_{\tau_e}, \frac{1}{2}\Big[\sum_{i:k_i \in \{1,2\}}\sum_j (S_{ij} - \beta_{0i})^2$$
$$+ \sum_{i:k_i=3}\sum_j (S_{ij} - \beta_{0i} - \beta_{1i}(t_{ij} - t_i^* - \text{lag})^+)^2\Big]$$
$$+ b_{\tau_e}\Big)$$

$$sp_F|\text{else} \sim \text{Beta}\Big(\sum_{i:k_i=1}\sum_j (1 - F_{ij})$$
$$+ \sum_{i:k_i \in \{2,3\}}\sum_{j:t_{ij}<t_i^*} (1 - F_{ij}) + a_{spF},$$
$$\sum_{i:k_i=1}\sum_j F_{ij} + \sum_{i:k_i \in \{2,3\}}\sum_{j:t_{ij}<t_i^*} F_{ij} + b_{spF}\Big)$$

$$se_F|\text{else} \sim \text{Beta}\Big(\sum_{i:k_i \in \{2,3\}}\sum_{j:t_{ij}\geq t_i^*} F_{ij} + a_{seF},$$
$$\sum_{i:k_i \in \{2,3\}}\sum_{j:t_{ij}\geq t_i^*} (1 - F_{ij}) + b_{seF}\Big)$$

$$\text{lag}|\text{else} \propto \prod_{i:k_i=3}\Big[\prod_{j=1}^{m_i}\exp\{-\frac{\tau_e}{2}(S_{ij} - \beta_{0i} -$$
$$\beta_{1i}(t_{ij} - t_i^* - \text{lag})^+)^2\}\Big] \cdot \frac{1}{\text{lag}^{n_2}}$$
$$\cdot \Big[\prod_{i:k_i=3}\frac{1}{t_{im_i} - \text{lag} - \text{dob}_i}\Big]$$
$$\cdot I(\min_l^* < \text{lag} < \max_l^*), \quad \text{where}$$
$$\min_l^* = \max\{\{t_{im_i} - t_i^*, \forall i \ni k_i = 2\}, \min_l\}$$
$$\max_l^* = \min\{\{t_{im_i} - t_i^*, \forall i \ni k_i = 3\}, \max_l\}$$

$$\mu_\gamma|\text{else} \sim N\left(\frac{\tau_\gamma n_3}{\delta} \cdot \frac{\sum_{i:k_i=3}\log\beta_{1i}}{n_3} + \frac{b_{\mu_\gamma}}{\delta}a_{\mu_\gamma}, \delta\right)$$
$$\text{where } \delta = \tau_\gamma n_3 + b_{\mu_\gamma}$$

$$\tau_\gamma|\text{else} \sim \Gamma\left(\frac{n_3}{2} + a_{\tau_\gamma}, \frac{1}{2}\sum_{i:k_i=3} (\log\beta_{1i} - \mu_\gamma)^2 + b_{\tau_\gamma}\right)$$

The full conditional for lag is not recognizable. In fact, it is not even continuous. In addition, the permissible values of lag are constrained by the current latent state assignments, k_i, and the corresponding t_i^*'s. In particular, the requirement $t_{im_i} - \text{lag} < t_i^* \leq t_{im_i}$ for all current state 2 cows yields the constraint $\max_{i:k_i=2}\{t_{im_i} - t_i^*\} < \text{lag}$. Similarly, the requirement that $\text{dob}_i < t_i^* \leq t_{im_i} - \text{lag}$ for cows with $k_i = 3$ yields the other constraint, $\text{lag} \leq \min_{i:k_i=3}\{t_{im_i} - t_i^*\}$. Note that these constraints ensure the sampled value of lag does not change any of the current k_i's. A slice sampler is used to sample the full conditional of lag.

In one sweep or iteration of the Gibbs sampler, we sample the global parameters using the full conditionals just described. We also sample the cow-specific coefficients/latents for each cow given the current state of that cow, k_i. As previously mentioned, the full conditionals that are sampled for the ith cow in the current iteration depend on k_i. We thus have, if $k_i = 1$ in the current iteration, which corresponds to an animal that is not infected over the course of the study,

$$\beta_{0i}|\text{else}, k_i = 1 \sim N(\hat{\beta}_{0i}, \tau_e m_i + \tau_{\beta_0}),$$

where $\hat{\beta}_{0i} = \frac{\tau_e m_i}{\tau_e m_i + \tau_{\beta_0}}\bar{S}_i + \frac{\tau_{\beta_0}}{\tau_e m_i + \tau_{\beta_0}}\beta_0$ and $\bar{S}_i = \frac{1}{m_i}\sum_j S_{ij}$. If $k_i = 2$, the current classification is infection without serology reaction so we sample the full conditionals of β_{0i} and t_i^*. The full conditional for β_{0i} is the same as in the $k_i = 1$ case because in either case there is no serology reaction, just a "flat" baseline serology trajectory. For t_i^*, we have the following piecewise constant function

$$p(t_i^*|k_i = 2, \text{else}) \propto$$
$$\prod_{j:t_{ij}<t_i^*}\Big[Sp_F^{1-F_{ij}}(1 - Sp_F)^{F_{ij}}\Big] \prod_{j:t_{ij}\geq t_i^*}\Big[Se_F^{F_{ij}}(1 - Se_F)^{1-F_{ij}}\Big]$$
$$\times I(t_{im_i} - \text{lag} < t_i^* < t_{im_i})$$

Finally, if $k_i = 3$, the infection with serology reaction case, we sample β_{0i}, β_{1i} and t_i^*. The full conditional for β_{0i} is the same as for the $k_i = 1$ case except we replace $\bar{S}_i$ with $\bar{\bar{S}}_i = \frac{1}{m_i}\sum_j\{S_{ij} - \beta_{1i}(t_{ij} - t_i^* - \text{lag})^+\}$. Because β_{1i} and t_i^* are highly correlated, we sample them jointly using the following full conditional

(7)

$$p(\beta_{1i}, t_i^*|k_i = 3, \text{else}) \propto$$
$$\exp\{-\frac{\tau_e}{2}\sum_{j=1}^{m_i}(S_{ij} - \beta_{0i} - \beta_{1i}(t_{ij} - t_i^* - \text{lag})^+)^2\}$$

$$\times \prod_{j:t_{ij}<t_i^*}\left[Sp_F^{1-F_{ij}}(1-Sp_F)^{F_{ij}}\right]\prod_{j:t_{ij}\geq t_i^*}\left[Se_F^{F_{ij}}(1-Se_F)^{1-F_{ij}}\right]$$

$$\times \frac{1}{\beta_{1i}}\exp\{-\frac{\tau_\gamma}{2}(\log(\beta_{1i})-\mu_\gamma)^2\}\cdot I(\mathrm{dob}_i<t_i^*<t_{im_i}-\mathrm{lag})$$

This joint full conditional does not have a recognizable form so we use a random walk Metropolis sampler. Since the support is $\mathbb{R}^+\times(\min_l,\max_l)$ and we wish to use a bivariate normal proposal distribution having support $\mathbb{R}^2$, we make the following transformation:

$$\left(\begin{array}{c}\log\tilde{\beta}_{1i}\\\mathrm{logit}(\frac{\tilde{t}_i^*-\mathrm{dob}_i}{t_{im_i}-\mathrm{lag}-\mathrm{dob}_i})\end{array}\right)=\left(\begin{array}{c}\tilde{\gamma}\\\tilde{\omega}\end{array}\right)$$

$$\sim N\left(\left(\begin{array}{c}\log\beta_{1i}^{(t)}\\\mathrm{logit}(\frac{t_i^{*(t)}-\mathrm{dob}_i}{t_{im_i}-\mathrm{lag}-\mathrm{dob}_i})\end{array}\right),\tilde{\Sigma}\right)$$

Thus, we generate $\tilde{\gamma}$ and $\tilde{\omega}$ from the indicated bivariate normal. We then transform back to the original variables using $\tilde{\beta}_{1i}=\exp(\tilde{\gamma})$ and $\tilde{t}_i^*=(t_{im_i}-\mathrm{lag}-\mathrm{dob}_i)\mathrm{expit}(\tilde{\omega})+\mathrm{dob}_i$.

Using the change of variables technique, we obtain the induced proposal distribution on $(\tilde{\beta}_{1i},\tilde{t}_i^*)$:

$$q(\tilde{\beta}_{1i},\tilde{t}_i^*)\propto|\tilde{\Sigma}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\tilde{V}^T\tilde{\Sigma}^{-1}\tilde{V}\}\cdot|J|$$

$$\text{where }\tilde{V}=\left(\begin{array}{c}\log\tilde{\beta}_{1i}-\log\beta_{1i}^{(t)}\\\mathrm{logit}(\frac{\tilde{t}_i^*-\mathrm{dob}_i}{t_{im_i}-\mathrm{lag}-\mathrm{dob}_i})-\mathrm{logit}(\frac{t_i^{*(t)}-\mathrm{dob}_i}{t_{im_i}-\mathrm{lag}-\mathrm{dob}_i})\end{array}\right)$$

$$\text{and }|J|=\left|\begin{array}{cc}\frac{1}{\tilde{\beta}_{1i}}&0\\0&\frac{1}{\tilde{t}_i^*-\mathrm{dob}_i}+\frac{1}{t_{im_i}-\mathrm{lag}-\tilde{t}_i^*}\end{array}\right|$$

The proposal is accepted with probability $\min(1,\alpha)$ where

$$\alpha=\frac{p(\tilde{\beta}_{1i},\tilde{t}_i^*|\mathrm{else})}{p(\beta_{1i}^{(t)},t_i^{*(t)}|\mathrm{else})}\cdot\frac{q(\beta_{1i}^{(t)},t_i^{*(t)})}{q(\tilde{\beta}_{1i},\tilde{t}_i^*)}$$

$$=\frac{p(\tilde{\beta}_{1i},\tilde{t}_i^*|\mathrm{else})}{p(\beta_{1i}^{(t)},t_i^{*(t)}|\mathrm{else})}\cdot\frac{\tilde{\beta}_{1i}(\tilde{t}_i^*-\mathrm{dob}_i)(t_{im_i}-\mathrm{lag}-\tilde{t}_i^*)}{\beta_{1i}^{(t)}(t_i^{*(t)}-\mathrm{dob}_i)(t_{im_i}-\mathrm{lag}-t_i^{*(t)})}$$

The covariance matrix of the proposal distribution, $\tilde{\Sigma}$ can be initially taken as the identity matrix times a tuning parameter, σ_p. After a pilot run, the covariance matrix should be tuned. An empirical estimate of the covariance matrix from a pilot run of about 5,000 iterations resulted in good acceptance rates and mixing in the simulations and data analysis in this paper.

Reversible jump MCMC to sample model indicators

We now devise a reversible jump strategy to update $(k_i,\chi^{(k_i)})$, the vector of parameters corresponding to cow i. We embed the RJMCMC algorithm in a Gibbs sampler, using it to sample the joint full conditional of $(k_i,\chi^{(k_i)})$,

because these are the only full conditionals that are not fixed-dimensional. This approach was taken by Richardson and Green (1995).

We start by defining the allowed between-model moves. Next, for each move type, we define the mechanism for generating the vector of proposal parameters and calculate the acceptance probability.

We permit moves to "adjacent" models only. In practice, restricting moves to models that are "close" to the current model improves mixing since moves to "distant" models are unlikely to be accepted (Brooks, Giudici and Roberts, 2003). Since our model space (for a cow) only includes three parameter spaces, our intuition is that mixing would not be hampered significantly by including jumps between models 1 and 3. However, we still restrict moves to adjacent models to simplify computation, noting that the resulting chain remains irreducible since a cow may move between any two models one step at a time. Specifically, we propose between-model moves as follows: if a cow is in disease status 1 or 3, we propose a jump to model 2 with probability 1. If the disease status is 2, we propose a jump to model 1 or 3 with equal probability. Using the notation $\pi_{kk'}$ to denote the probability of proposing a move from state k to state k', we can summarize the between-model moves as follows: $\pi_{12}=\pi_{32}=1$ and $\pi_{21}=\pi_{23}=\frac{1}{2}$.

Moving between models 1 and 2

In order to construct a move from disease status 1 to 2, we need to define a bijection between the two associated parameter spaces, β_{0i} and (β_{0i},t_i^*). Since we can only define a bijection between two spaces if their dimensions match, we augment the model 1 parameter "vector," β_{0i}, with a generated auxiliary variable, u. We generate u using the prior for t_i^* under model 2 (see equation 2). Recall this prior is denoted by $\pi_2(\cdot)$. Letting $\beta_{0i}^{(t)}$ denote the current latent associated with cow i, we can now define a bijection

$$T_{12}(\beta_{0i}^{(t)},u)=(\tilde{\beta}_{0i},\tilde{t}_i^*),\quad\text{where }\tilde{\beta}_{0i}=\beta_{0i}^{(t)},\quad\tilde{t}_i^*=u$$

This bijection is actually just the identity map.

We now calculate the acceptance probability. Using $[Y|X]$ to denote the distribution of Y conditional on X, we have that the serology portion of the likelihood for cow i under model z, $z\in\{1,2\}$, can be represented $[S_i|\beta_{0i},\tau_e,k_i=z]$. Likewise, cow i's fecal culture contribution to the likelihood is $[F_i|sp_F,k_i=1]$ under model 1 and $[F_i|t_i^*,se_F,sp_F,k_i=2]$ under model 2. In the following calculations, we assume that parameters not directly involved in the bijection are set at their current values although we drop the usual superscript notation, i.e. τ_e actually represents $\tau_e^{(t)}$. Also, we let $\pi(\cdot)$ generically represent the prior on a parameter or the distribution of a random slope or intercept. The proposal to (i) move from model 1 to model 2 *and* (ii) accept the proposed parameters $(\tilde{\beta}_{0i},\tilde{t}_i^*)$ has acceptance probability $\min(\alpha_{12},1)$ where

α_{12}

$$= \frac{p(\tilde{\beta}_{0i}, \tilde{t}_i^* | k_i = 2, \text{ else})}{p(\beta_{0i}^{(t)} | k_i = 1, \text{ else})} \cdot \frac{\pi_{21}}{\pi_{12} \cdot \pi_2(u)} \cdot |J|$$

$$= \frac{[S_i | \tilde{\beta}_{0i}, \tau_e, k_i = 2] \cdot [F_i | \tilde{t}_i^*, se_F, sp_F, k_i = 2]}{[S_i | \beta_{0i}^{(t)}, \tau_e, k_i = 1] \cdot [F_i | sp_F, k_i = 1]} \cdot \frac{q_2 \pi(\tilde{\beta}_{0i}) \pi_2(\tilde{t}_i^*)}{q_1 \pi(\beta_{0i}^{(t)})}$$
$$\times \frac{\pi_{21}}{\pi_{12} \cdot \pi_2(u)} \cdot |J|$$

$$= \frac{[S_i | \beta_{0i}^{(t)}, \tau_e, k_i = 2] \cdot [F_i | u, se_F, sp_F, k_i = 2]}{[S_i | \beta_{0i}^{(t)}, \tau_e, k_i = 1] \cdot [F_i | sp_F, k_i = 1]} \cdot \frac{q_2 \pi(\beta_{0i}^{(t)}) \pi_2(u)}{q_1 \pi(\beta_{0i}^{(t)})}$$
$$\times \frac{\pi_{21}}{\pi_{12} \cdot \pi_2(u)} \cdot |J|$$

$$= \frac{[F_i | \tilde{t}_i^* = u, se_F, sp_F, k_i = 2]}{[F_i | sp_F, k_i = 1]} \cdot \frac{q_2}{q_1} \cdot \frac{\pi_{21}}{\pi_{12}} \cdot |J|$$

$$= \frac{sp_F^{\sum_{j:t_{ij} < \tilde{t}_i^*}(1 - F_{ij})} (1 - sp_F)^{\sum_{j:t_{ij} < \tilde{t}_i^*} F_{ij}} se_F^{\sum_{j:t_{ij} \geq \tilde{t}_i^*} F_{ij}}}{sp_F^{\sum_j (1 - F_{ij})} (1 - sp_F)^{\sum_j F_{ij}}}$$
$$\times \frac{(1 - se_F)^{\sum_{j:t_{ij} \geq \tilde{t}_i^*}(1 - F_{ij})}}{1} \cdot \frac{q_2}{q_1} \cdot \frac{\pi_{21}}{\pi_{12}} \cdot 1$$

We obtain line 3 by writing out the posterior in the first factor in line 2 as the likelihood times the prior. Lines 5 and 6 are obtained by substituting the proposed values into the numerator. Since the serology portion of the likelihood is the same under models 1 and 2, these factors cancel in line 7. Finally, the determinant of the Jacobian is 1 since

$$\frac{\partial T_{12}(\beta_{0i}^{(t)}, u)}{\partial(\beta_{0i}^{(t)}, u)} = \frac{\partial(\tilde{\beta}_{0i}, \tilde{t}_i^*)}{\partial(\beta_{0i}^{(t)}, u)} = \frac{\partial(\beta_{0i}^{(t)}, u)}{\partial(\beta_{0i}^{(t)}, u)} = I_2$$

where I_2 is the 2 by 2 identity matrix.

For the reverse move, from model 2 to 1, we transform $T_{21}(\beta_{0i}^{(t)}, t_i^{*(t)}) = (\beta_{0i}^{(t)}, t_i^{*(t)}) = (\tilde{\beta}_{0i}, \tilde{u})$ and the acceptance probability is $\min(\alpha_{12}^{-1}, 1)$ where α_{12} is calculated with $\tilde{t}_i^*$ replaced by $t_i^{*(t)}$.

Moving between models 2 and 3

In proposing a move from model 2 to model 3, we construct a more complex bijection in order to obtain proposals that have a reasonable chance of being accepted. We generate a proposal in the parameter space of model 3 by generating u_1, which will be the proposal of t_i^*, from the density proportional to the model 3 fecal culture portion of the likelihood. We draw u_2 from a $N(\mu_\gamma, \tau_\gamma)$ pdf, using the current values of μ_γ and τ_γ, and take e^{u_2} as the proposal for β_{1i}. These choices of auxiliary variables simplify the acceptance probability and appear to give reasonable mixing between latent infection states 2 and 3. Note that we define a bijection of dimension 4 even though we only add one cow-specific parameter, β_{1i}, to the vector of model 2 parameters, $(\beta_{0i}^{(t)}, t_i^{*(t)})$. This is because the support of t_i^* varies by model. Hence, it is simpler to generate a new t_i^* as a proposal for model 3 using the auxiliary variable u_1.

However, we still need the mapping to be a bijection so we map $t_i^{*(t)}$ to yet another auxiliary variable, v, generated using the prior $\pi_2(\cdot)$ given in equation 2 (this results in a sensible proposal for the reverse move $T_{32} = T_{23}^{-1}$ given in equation 9). Define $T_{23}(\beta_{0i}^{(t)}, t_i^{*(t)}, u_1, u_2) = (\tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^*, v)$, where u_1 has pdf $\psi(u_1) \propto sp_F^{\sum_{j:t_{ij} < u_1}(1 - F_{ij})}(1 - sp_F)^{\sum_{j:t_{ij} < u_1} F_{ij}} se_F^{\sum_{j:t_{ij} \geq u_1} F_{ij}}(1 - se_F)^{\sum_{j:t_{ij} \geq u_1}(1 - F_{ij})}$. $I(\text{dob}_i < u_1 < t_{im_i} - \text{lag})$ and $u_2 \sim N(\mu_\gamma, \tau_\gamma)$. We also have $\tilde{\beta}_{0i} = \beta_{0i}^{(t)}$, $\tilde{\beta}_{1i} = e^{u_2}$, $\tilde{t}_i^* = u_1$, and $v = t_i^{*(t)}$.

To calculate the corresponding acceptance probability, we represent cow i's serology likelihood under model 3 by $[S_i | \beta_{0i}, \beta_{1i}, t_i^*, \tau_e, k_i = 3]$. Also, cow i's fecal culture contribution to the likelihood under model 3 is $[F_i | t_i^*, se_F, sp_F, k_i = 3]$. Using $\phi(\cdot)$ to represent the pdf of the standard normal distribution, we accept the proposed move from model 2 to 3 with probability $\min(\alpha_{23}, 1)$ where

(8)

$$\alpha_{23} = \frac{p(\tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^* | k_i = 3, \text{ else})}{p(\beta_{0i}^{(t)}, t_i^{*(t)} | k_i = 2, \text{ else})}$$
$$\times \frac{\pi_{32} \cdot \pi_2(v)}{\pi_{23} \cdot \psi(u_1) \cdot \sqrt{\tau_\gamma} \phi(\sqrt{\tau_\gamma}(u_2 - \mu_\gamma))} \cdot |J|$$

$$= \frac{[S_i | \tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^*, \tau_e, k_i = 3] \cdot [F_i | \tilde{t}_i^*, se_F, sp_F, k_i = 3]}{[S_i | \beta_{0i}^{(t)}, \tau_e, k_i = 2] \cdot [F_i | t_i^{*(t)}, se_F, sp_F, k_i = 2]}$$
$$\times \frac{q_3 \pi(\tilde{\beta}_{0i}) \cdot \frac{1}{\tilde{\beta}_{1i}} \sqrt{\tau_\gamma} \phi(\sqrt{\tau_\gamma}(\log \tilde{\beta}_{1i} - \mu_\gamma)) \cdot \pi_3(\tilde{t}_i^*)}{q_2 \pi(\beta_{0i}^{(t)}) \pi_2(t_i^{*(t)})}$$
$$\times \frac{\pi_{32} \cdot \pi_2(v)}{\pi_{23} \cdot \psi(u_1) \cdot \sqrt{\tau_\gamma} \phi(\sqrt{\tau_\gamma}(u_2 - \mu_\gamma))} \cdot |J|$$

$$= \frac{[S_i | \beta_{0i}^{(t)}, e^{u_2}, u_1, \tau_e, k_i = 3] \cdot [F_i | u_1, se_F, sp_F, k_i = 3]}{[S_i | \beta_{0i}^{(t)}, \tau_e, k_i = 2] \cdot [F_i | t_i^{*(t)}, se_F, sp_F, k_i = 2]}$$
$$\times \frac{q_3 \pi(\beta_{0i}^{(t)}) \cdot \frac{1}{e^{u_2}} \sqrt{\tau_\gamma} \phi(\sqrt{\tau_\gamma}(\log(e^{u_2}) - \mu_\gamma)) \cdot \pi_3(u_1)}{q_2 \pi(\beta_{0i}^{(t)}) \pi_2(t_i^{*(t)})}$$
$$\times \frac{\pi_{32} \cdot \pi_2(t_i^{*(t)})}{\pi_{23} \cdot \psi(u_1) \cdot \sqrt{\tau_\gamma} \phi(\sqrt{\tau_\gamma}(u_2 - \mu_\gamma))} \cdot |J|$$

$$= \frac{[S_i | \beta_{0i}^{(t)}, e^{u_2}, u_1, \tau_e, k_i = 3] \cdot [F_i | u_1, se_F, sp_F, k_i = 3]}{[S_i | \beta_{0i}^{(t)}, \tau_e, k_i = 2] \cdot [F_i | t_i^{*(t)}, se_F, sp_F, k_i = 2]}$$
$$\times \frac{q_3 \cdot \frac{1}{e^{u_2}} \cdot \pi_3(u_1)}{q_2} \cdot \frac{\pi_{32}}{\pi_{23} \cdot \psi(u_1)} \cdot |J|$$

Lines 3 and 4 of equation 8 are obtained by writing out the posterior given on the right-hand side of the equation in line 1 as likelihood (line 3) times prior (line 4). Lines 6–8 are obtained by substituting the proposed values defined by the bijection T_{23} as known functions of $\beta_{0i}^{(t)}, t_i^{*(t)}, u_1$, and u_2 in for $\tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^*$, and v. Since some of the proposals were chosen to be priors, we obtain some cancelation in lines 6–8, which gives the simpler expression in lines 9–10. The acceptance probability is evaluated using the current

values of all global parameters (se_F, τ_e, etc). Note that we need to include a proposal distribution in the numerator, i.e. $\pi_2(v)$, for this move since the reverse move from model 3 to model 2 would require the generation of a model 2 infection time from the prior $\pi_2(\cdot)$. The Jacobian is calculated as follows:

$$|J| = \left| \frac{\partial(\tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^*, v)}{\partial(\beta_{0i}^{(t)}, t_i^{*(t)}, u_1, u_2)} \right| = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{u_2} \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{vmatrix} = e^{u_2}$$

The proposal for a move from model 3 to model 2 is already defined as it is the inverse of the function T_{23}. Specifically, we have

$$(9) \qquad T_{32}(\beta_{0i}^{(t)}, \beta_{1i}^{(t)}, t_i^{*(t)}, v) = T_{23}^{-1}(\beta_{0i}^{(t)}, \beta_{1i}^{(t)}, t_i^{*(t)}, v)$$
$$= (\tilde{\beta}_{0i}, \tilde{t}_i^*, \tilde{u}_1, \tilde{u}_2)$$

where $\tilde{\beta}_{0i} = \beta_{0i}^{(t)}$, $\tilde{t}_i^* = v$, $\tilde{u}_1 = t_i^{*(t)}$ and $\tilde{u}_2 = \log \beta_{1i}^{(t)}$. The probability of accepting a proposed move from model 3 to model 2 is $\min(\alpha_{32}, 1)$ where $\alpha_{32} = \alpha_{23}^{-1}$. Since we are attempting a move from model 3 to model 2, α_{23} is calculated with the current values of the parameters for model 3 and the proposed values for model 2.

ACKNOWLEDGEMENTS

REFERENCES

ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88** 669–679. MR1224394

BEDRICK, E. J., CHRISTENSEN, R., and JOHNSON, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association* **91** 1450–1460. MR1439085

BRANSCUM, A. J., JOHNSON, W. O., HANSON, T. E., and GARDNER, I. A. (2008). Bayesian semiparametric ROC curve estimation and diagnosis. *Statistics in Medicine* **27** 2474–2496.

BROOKS, S. P., GIUDICI, P., and ROBERTS, G. O. (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society, Ser. B* **65** 3–55. MR1959092

COLLINS, M. T., GARDNER, I. A., GARRY, F. B., ROUSSEL, A. J., and WELLS, S. J. (2006). Consensus recommendations on diagnostic testing for the detection of paratuberculosis in cattle in the United States. *Journal of the American Veterinary Medical Association* **229** 1912–1919.

COOK, R. J., NG, E. T. M., and MEADE, M. O. (2000). Estimation of operating characteristics for dependent diagnostic tests based on latent Markov models. *Biometrics* **56** 1109–1117. MR1815590

DENDUKURI, N. and JOSEPH, L. (2001). Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* **57** 158–167. MR1833302

ENGEL, B., BACKER, J., and BUIST, W. (2009). Evaluation of the accuracy of diagnostic tests from repeated measurements without a gold standard. *Journal of Agricultural, Biological, and Environmental Statistics*, to appear.

GASTWIRTH, J. L. (1987). The statistical precision of medical screening procedures: application to polygraph and AIDS antibodies test data. *Statistical Science* **2** 213–238. MR0920139

GELFAND, A. and SMITH, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85** 398–409. MR1141740

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** 721–741.

GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. MR1380810

HALL, C. B., YING, J., KUO, L., and LIPTON, R. B. (2003). Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computational Statistics and Data Analysis* **42** 91–109. MR1963357

HASTINGS, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.

HUI, S. L. and WALTER, S. D. (1980). Estimating the error rates of diagnostic tests. *Biometrics* **36** 167–171.

LANG, S. and BREZGER, A. (2004). Bayesian P-Splines. *Journal of Computational and Graphical Statistics* **13** 183–212. MR2044877

LANGE, N., CARLIN, B. P., and GELFAND, A. E. (1992). Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-Cell numbers. *Journal of the American Statistical Association* **87** 615–626.

LASSAUZET, M.-L., JOHNSON, W. O., and THURMOND, M. (1989). Regression models for time-to-seroconversion following experimental bovine leukemia virus infection. *Statistics in Medicine* **8** 725–741.

LEPPER, A. W., WILKS, C. R., KOTIW, M., WHITEHEAD, J. T., and SWART, K. S. (1989.) Sequential bacteriological observations in relation to cell-mediated and humoral antibody responses of cattle infected with *Mycobacterium paratuberculosis* and maintained on normal or high iron intake. *Australian Veterinary Journal* **66** 50–55.

LI, Y., LIN, X., and MÜLLER, P. (2009). Bayesian inference in semiparametric mixed models for longitudinal data. *Biometrics*, to appear.

METROPOLIS, N., ROSENBLUTH, A., ROSENBLUTH, M., TELLER, A., and TELLER, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21** 1087–1092.

NEAL, R. M. (1997). Markov chain Monte Carlo methods based on "slicing" the density function. Technical report, Dept. of Computer Science, Univ. of Toronto.

PAULER, D. K. and LAIRD, N. M. (2000). A mixture model for longitudinal data with application to assessment of noncompliance. *Biometrics* **56** 464–472.

R DEVELOPMENT CORE TEAM (2005). R: A language and environment for statistical computing, reference index version 2.4.0. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

RICHARDSON, S. and GREEN, P. J. (1995). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Ser. B* **59** 731–792. MR1483213

ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York. MR2080278

SKATES, S. J., PAULER, D. K., and JACOBS, I. J. (2001). Screening based on the risk of cancer calculation from Bayesian hierarchical changepoint and mixture models of longitudinal markers. *Journal*

of the American Statistical Association **96** 429–439. MR1939346

TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics* **22** 1701–1786. MR1329166

WAAGEPETERSEN, R. and SORENSEN, D. (2001). A tutorial on reversible jump MCMC with a view toward applications in QTL-mapping. *International Statistical Review* **69** 49–61.

Michelle Norris
Department of Mathematics and Statistics
California State University
Sacramento, CA 95819, USA
E-mail address: norris@csus.edu

Wesley O. Johnson
Department of Statistics
University of California
Irvine, CA 92697, USA
E-mail address: wjohnson@ics.uci.edu

Ian A. Gardner
Department of Medicine and Epidemiology
University of California
Davis, CA 95616, USA
E-mail address: iagardner@ucdavis.edu