Bayesian semi-parametric joint modeling of biomarker data with a latent changepoint: assessing the temporal performance of Enzyme-Linked Immunosorbent Assay (ELISA) testing for paratuberculosis

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

In this paper, we develop a class of semi-parametric statistical models that can be used for the important problem of analyzing longitudinal biomarker data with the purpose of quantifying their diagnostic capabilities, as a function of time from infection. We focus on the complicated problem where there is no gold standard assessment of the actual timing of infection/disease onset (our change point), which provides additional motivation for considering a second, binary test, in order to make it easier to estimate the change points for individuals that become diseased. An important additional feature of our model is its nonparametric part, which allows for distinct biomarker responses to the insult of infection/disease. In our case, the model allows for the possibility of an unknown number of clusters of individuals, each with distinct slopes corresponding to distinct biological reactions. Clusters with steeper slopes would correspond to individuals that could be diagnosed sooner than those with more gradual slopes.

1. INTRODUCTION

Expenditure on diagnostic testing in human and animal health in the United States is massive. Hanson et al. (2000) reported that, at about that time, approximately $30 billion was spent annually in the U.S. on diagnostic tests in human medicine, and they provided similar evidence that costs for testing of livestock and pet populations were substantial. Infectious diseases of animals that

*Corresponding author.

are zoonotic such as avian influenza, brucellosis, and tuberculosis are of particular concern to animal health authorities, and testing for these makes up a large part of US Department of Agriculture-Animal Plant Health Inspection Services-Veterinary Services disease control budget.

Other diseases such as Johne's disease (caused by the bacterium Mycobacterium avium subsp. paratuberculosis (Map)), are important because of a possible link with Crohn's disease in humans and because infection in dairy cows results in decreased milk production, weight loss, and premature culling. Because cows can transmit the infection to other cattle during the asymptomatic phase, early detection and subsequent removal of infected cows from the herd is important for controlling the spread of Johne's Disease. Losses in U.S. dairy cattle from Johne's disease have been projected to be $200 million to $250 million annually (Ott et al., 1999). However, antemortem diagnosis of Johne's disease is difficult especially in the subclinical phase of disease, which may last for years, because organism (fecal culture) and antibody detection tests (serology) have low to moderate sensitivity (30 to 60%) in this stage (Collins et al., 2005).

The Map data are from a binary organism detection test in feces and a continuous serum antibody (serology) test, both of which are administered repeatedly over time to each cow sampled from a herd. While our model and method are specifically designed and discussed in the context of Map, the development serves as a prototype for quantifying the performance of diagnostic procedures for other diseases, human and animal. We thus develop a general flexible statistical model for longitudinal joint diagnostic test outcome data. Statistical analysis of longitudinal test data appears to have had limited attention in the literature.

In the simplest case, biomarker values measured on a continuous scale require dichotomization at a selected cutoff value for classification of disease/infection status. The proportion of the time that the cutoff is exceeded when individuals are actually infected/diseased is the sensitivity of the test corresponding to that cutoff, and the proportion with values below the cutoff among individuals that are not

diseased is the corresponding specificity. It is well known that sensitivity of an ELISA test will be low for individuals that were recently infected since there would be insufficient time for the serologic response to achieve the cutoff. Thus the sensitivity of a test applied to a cross section of a general population must be a mixture of sensitivities, mixed according to an unknown and unknowable distribution of times since infection among infected animals at that point in time.

Since we have longitudinal data, however, it is possible to estimate the sensitivity of a diagnostic test as a function of time. We are unaware of work, other than our own (Norris and Johnson, 2009), that addresses this problem. Norris and Johnson (2009) developed a purely parametric model for analyzing the Map data. Key features of Norris and Johnson (2009) are: (i) the joint modeling of longitudinal binary and continuous diagnostic outcome data, (ii) the incorporation of latent disease status and timing of infection for infected individuals, which necessitates (iii) the incorporation or Reversible Jump Markov chain Monte Carlo (RJMCMC) techniques, (iv) modeling the biomarker responses after infection using a random slope effect centered on a fixed effect, (iv) the estimation of sensitivity as a function of time from infection, and (v) estimating receiver operating characteristic (ROC) curves based on the biomarker considered, in absence of a gold standard.

In the current paper, we extend the novelty of that work in the following ways: (a) we extend the parametric model to allow for clusters of individuals with distinct serologic responses to infection, e.g. different clusters have different slopes, which is accomplished by modeling slopes using a Dirichlet Process Mixture (DPM) (Escobar and West, 1995); (b) we obtain estimated ROC curves for increasing time after infection, and for different clusters, thus finding that some groups are more difficult to detect than others due to having more gradual slopes; (c) we consider joint diagnosis based on the combined binary and continuous response and compare with diagnosis based on the continuous biomarkers alone, which is accomplished by statistically comparing areas under the respective ROC curves (AUC); and (d) we obtain an estimate of the cumulative risk of infection through time. An exciting scientific finding for the Map data considered is that there are evidently two basic types of cows; one with a fairly steep serologic response to infection, and a smaller group, with a much smaller slope after infection. It would be interesting to now investigate the possible reasons that impede diagnosis for this second type. We finally remark that the extension to a semi-parametric model, and the methods for comparison of joint versus single outcome models are non trivial; our code is available from the first author upon request.

The remainder of the paper is structured as follows: we describe the data and discuss background references in Section 2, and our semi-parametric model is discussed in Section 3. Prior specification and Markov Chain Monte Carlo

(MCMC) methods for implementing the Bayesian paradigm are discussed in Section 4, and an Appendix. Simulation results are given in Section 5 and the Johne's disease data analysis is given in Section 6. Concluding remarks are given in Section 7.

2. BACKGROUND

The data consist of results of fecal culture and ELISA results (hereafter termed fecal tests and serology tests for brevity) for Map in 365 dairy cows. Attempts to screen the cows were made approximately every six months, but this regimen was not strictly followed. The time between screenings ranged between three months and two years. Reasons varied from not being able to locate the cow to contamination of the sample. One or both tests may be missing for a cow on an attempted screening date. The total number of longitudinal screenings varies considerably among cows, from a minimum of one to a maximum of 23. Since our interest is in longitudinal testing, we omitted cows with only a single test outcome from our analysis. The median number of screenings for the cows included in this analysis is six. The results of 2185 serum ELISA tests and 2288 fecal tests were included in our data set. Cows entered and left the study at different times corresponding to either joining the herd or being removed from it for various reasons, including disease events and low milk production.

Because Map infection is endemic in over 90% of U.S. dairy herds (Lombard et al., 2013), we assume that there are cows that have become infected at some time during the study, while others have remained uninfected. Since neither serology nor fecal culture are perfect, there is no certainty about when or whether an individual animal becomes infected with Map. So in addition to jointly modeling test outcomes, we will also model the latent infection status of each animal and, conditional on infection, the time of infection during the study. After infection, it is assumed that serology scores will eventually, after a lag, increase linearly to higher levels.

These data have been previously analyzed by Norris et al. (2009), using a purely parametric model. But there are biological reasons to suspect that the parametric assumption is not adequate for these data. In particular, the model proposed for the serology scores is piecewise linear with random, cow-specific slopes, which model the increase in serology scores after infection. In the parametric model, slopes are assumed to be draws from a log normal distribution. For biological reasons discussed in Section 3, it is expected that this distribution is actually multimodal or skewed. The implication is that there are subgroups of animals that have different serologic responses to infection, one with a gradual increase and another with a more pronounced increase in serologic values. We looked for empirical confirmation that the distribution of log-slopes might be non-normal. Using the results of the parametric model, we restricted out attention to the 92 cows classified as infected and having a serology reaction. For each cow in this subset, we obtained
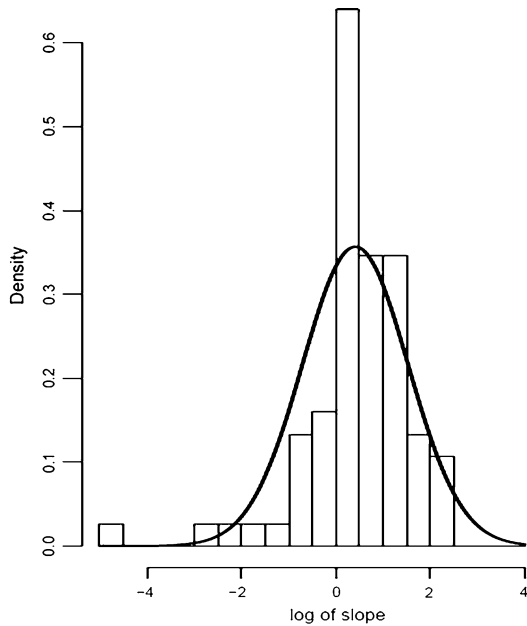
Figure 1. *Empirical Estimate of Log-slope Distribution.*

a least squares estimate of the slope of the post-lag serology trajectory. A histogram of these log-slopes is shown in Figure 1 along with a fitted normal. Compared to the fitted normal, the histogram shows a much higher than expected relative frequency for a bin near zero and also for the lowest bin (near -5 on the x-axis). The bin near -5 could represent a distinct cluster of cows having a weak serology response. In this application, insight into the distribution of random effects is, in and of itself, of scientific interest since we would like to know if the distribution is multimodal, is skewed, or has other non-normal features. If the distribution is multimodal, information about the location and portion of cows represented by each mode would be of interest to scientists.

There is some literature involving joint testing for infection that is related to ours. For example, Cook et al. (2000) developed methods for handling dependent binary tests observed longitudinally over a regular time grid and where disease state can alternate between diseased and not through time. Jones et al. (2011) developed a joint diagnostic test model for bovine digital dermatitis in dairy cows. Their model also allows that the agent for disease, once acquired, can disappear and then reappear. While their data are similar to ours in the sense that they observe binary (foot lesions, yes/no) and continuous biomarker values (serology score) through time, their parametric modeling of the relationships among disease, infection, and biomarkers is based on specific biology, and their goals are completely different. Wang et al. (2011) developed a parametric model for Johne's disease data that is similar to ours but taken over a shorter time frame (three years). Like us, they include a latent infection time in their model but also allow for covariate effects. However, they use a simpler model for longitudinal serology

results that only assumes a jump in serology score at the time of infection.

In the context of a clinical trial, Pauler and Laird (2000) modeled longitudinal biomarker data with the goal of detecting patients in noncompliance of prescribed treatment regimens. They modeled the latent switching status and time of switching, which resulted in a varying dimensional model and the subsequent need to use RJMCMC. In related work, Skates, Pauler and Jacobs (2001) analyzed case-control data with the goal of distinguishing cases of ovarian cancer which produce the antigen CA125 from those that do not. A substantial amount of data for controls was useful in determining the existence and location of changepoints for cases. The biomarker trajectory for cases that emit CA125 was modeled as a piecewise constant which changes to a linear function having positive slope following the changepoint. While improved diagnosis was the main goal of the research in these two papers, our goals are to characterize the accuracy of diagnostic tests through time, and to determine the utility of joint testing compared with serology testing only.

Finally, Li et al. (2010) developed a semi-parametric mixed model for longitudinal data where the nonparametric part consisted of modeling random effects with a flexible family of distributions, and also modeling a flexible functional time component. Our purposes again are quite distinct. We collect bivariate longitudinal diagnostic marker data with the goal of assessing and improving upon diagnostic accuracy with the expectation of substantial improvement in misclassification rates compared with sampling cross-sectional data as is the norm.

There are negative consequences of a poor distributional assumption for random effects. Verbeke and Lesaffre (1997) show that random effects are badly estimated if the distribution is assumed to be a single unmixed normal, while the true distribution is a mixture. Kleinman and Ibrahim (1998) have shown that the posterior distribution of model parameters can be affected by the choice of random effects distribution. They argue that the use of a nonparametric prior can potentially avoid inaccuracies resulting from an overly precise distributional assumption.

Many researchers have developed models and methods for incorporating more flexible random effects distributions into linear mixed models (LMM). In the Bayesian framework, West, Müller and Escobar (1994) developed flexible models for repeated measures, while Bush and MacEachern (1996) developed them for the randomized complete block design. Kleinman and Ibrahim (1998) specify a flexible random effects distribution using the Dirichlet Process prior of Ferguson (1973) in the general LMM framework. Frequentist contributions have been made by Madger and Zeger (1996), Tao, Palta, Yandell and Newton (1999), Verbeke and Lesaffre (1996), and Zhang and Davidian (2001).

3. MODEL FOR THE DATA

We propose a joint longitudinal model for fecal and serum (ELISA) results. For uninfected cows, we assume fecal and

serology results are independent since they have different biological bases and the fecal test is perfectly specific for practical purposes. For infected cows, the assumption of outcome independence has been accepted as reasonable by several authors, but we provide additional empirical evidence in Appendix I.

The proposed model accounts for the fact that bacteria can be shed in feces soon after infection whereas the production of detectable serum antibodies typically occurs later after infection. Consequently, we propose a joint model with a changepoint corresponding to time of infection, t^*. The probability of a positive fecal test changes at the time of infection, but the rise in serology score, which corresponds to an increase in antibody production, occurs some time later. Some of the research involving experimental infection of animals with the Map bacteria suggests this lag may be about one year after infection. However, since there is uncertainty about the lag for many screening tests, we formulate a general model in which the lag is an unknown parameter. Once the lag has elapsed, we model the increase in antibodies as linear. Since we are modeling no gold standard data, i.e. the infection status and time of infection are both unknown for the cows, we assume three latent states and construct an appropriate model for each. The latent states are: 1) no infection during the entire screening period; 2) infection, but insufficient time to mount an antibody reaction during screening period (since "lag" has not elapsed when screening ends); and 3) infection with antibody reaction within screening period (since "lag" elapses before the end of screening period). Note that latent states 2 and 3 depend on the "lag" parameter, which will be estimated along with all other parameters.

We use the latent variable, $k_i \in \{1, 2, 3\}$, to denote the latent disease state of cow i. We shall also use the following notation: $t_{ij} =$ time of the jth screening for the ith subject, $i = 1, 2, ..., n$ and $j = 1, 2, ..., m_i$; $(S_{ij}, F_{ij}) =$ the serology and fecal culture outcomes of the ith subject at time t_{ij}; $Se_F =$ sensitivity of fecal culture; $Sp_F =$ specificity of fecal culture; lag $=$ time interval between infection and serology reaction; $\Theta =$ vector of all model parameters; and $U =$ vector of all model latents. Also, we denote a normal distribution with mean, μ, and variance, σ^2, by $N(\mu, \tau = \frac{1}{\sigma^2})$ where τ is the precision. This parameterization facilitates MCMC calculations.

For the no infection state 1, we assume that each cow has its own baseline, so we model cow-specific baseline serology levels to be normally distributed random effects. We assume the fecal test results follow a Bernoulli distribution with probability of "success," defined as a positive test result, equal to $1 - Sp_F$. Thus, the model for cows in latent state 1, denoted $k_i = 1$, is:

$$S_{ij} \mid \Theta, U, k_i = 1 \sim \beta_{0i} + \epsilon_{ij}, \quad \perp$$
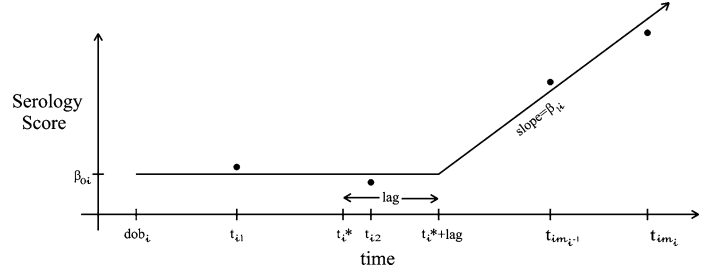$$F_{ij} \mid \Theta, U, k_i = 1 \sim \text{Bern}(1 - Sp_F)$$

Figure 2. Serology trajectory with data for cow with $k_i = 3$.

where $\beta_{0i} \overset{\perp}{\sim} N(\beta_0, \tau_{\beta_0})$, $\epsilon_{ij} \overset{\perp}{\sim} N(0, \tau_e)$, and $\beta_{0i} \perp \epsilon_{ij} \; \forall i, j$.

For latent state 2, infection without serology reaction, the probability of positive fecal test changes to Se_F at the time of infection, t_i^*. Further, we assume no change in the serology behavior since the "lag" time necessary to produce detectable antibodies has not elapsed. We also assume independence of serology and fecal tests conditional upon knowing the subject is in latent state 2. The model is:

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

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 $\beta_{0i} \perp \epsilon_{ij} \; \forall i, j$. The indicator function $I(x \in A)$ is 1 whenever x is an element of the set A and 0 otherwise.

The model for cows in latent state 3 is the same as latent state 2 for the fecal culture portion, but additionally assumes a linear increase in serology score at one lag past the infection time of the cow. A random, cow-specific slope for the post-lag serology trajectory is assumed in order to allow for differing rates of antibody production among infected cows. We define the function z^+ to equal z if $z > 0$ and 0 otherwise. The model is:

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

where β_{0i} and ϵ_{ij} are distributed as in the previous cases. We also assume that β_{1i}, β_{0i}, and ϵ_{ij} are pairwise independent for all i and j. Note that the term containing β_{1i} is zero until $t_{ij} = t_i^* +$ lag. Hence, the mean serology trajectory is a flat line until $t_i^* +$lag, then it increases linearly with slope β_{1i} as shown in Figure 2.

Because of the conditional independence of fecal and serology tests, given the latent state and infection time, none of the full conditionals are dependent on any type of data missingness. As a result, missing data do not present any new issues for making inferences. Specifically, once the latent state/infection times are known, the form of the models for the serology and fecal tests are determined, and there is

no longer dependency between these tests. The serology part of the model can handle any irregularity in the time interval between tests, including missing observations, since a linear relationship with time since infection is assumed. The component of the likelihood function arising from the serology portion of the model simply consists of a product of normal density functions with means dependent on the time since infection. Given the latent state/infection time, the fecal model is decoupled from the serology model and becomes a completely specified Bernoulli distribution. Hence, the portion of the likelihood representing the fecal data consists of a product of Bernoulli probability mass functions with parameters depending on state and time since infection. The time since infection is assumed to be a more basic variable which affects both the serology and fecal tests and accounts for the correlation between them. Making the models dependent on time since infection removes the need to include correlation structure relating the serology and fecal test.

We now consider the distribution of β_{1i}. It is reasonable to assume that many cows will have antibody production rates, i.e. slopes of the serology trajectory, that are normally distributed about some average rate. However, about 15% of cows with clinical disease and up to 70% of cows with subclinical disease will mount a weak antibody response that is below the recommended threshold for designation of a serology result as positive. These estimates are based on cross-sectional data, representing a range of times since infection (Collins et al., 2005). Hence, the distribution of slopes may be bimodal, or even multimodal, calling into question the traditional assumption that random effects are normally distributed. As a result, we model the distribution of slopes "nonparametrically," using a Dirichlet Process mixture of normals, to allow greater flexibility. See Ferguson (1973), for background on the Dirichlet Process and Escobar and West (1995) or Hanson et al. (2005), for background on the Dirichlet Process mixture (DPM). The DPM allows the number of terms in the mixture to be random and data-driven. Since biology also dictates that the antibodies must be *non-decreasing* after infection, we additionally constrain the slopes to be positive by modeling the log-slope as a DPM of normals as follows:

$$(1) \quad \log \beta_{1i} = \gamma_i \mid \mu_i, \tau_i \overset{\perp}{\sim} N(\mu_i, \tau_i) \quad \text{for } i : k_i = 3$$
$$(\mu_i, \tau_i) \mid G \overset{\perp}{\sim} G$$
$$G \mid \alpha, G_0 \sim DP(\alpha, G_0)$$

Heuristically, we write

$$\gamma_i \quad \mid \quad G \overset{\perp}{\sim} \int N(\cdot \mid \mu_i, \tau_i) G(\mathrm{d}\mu_i, \mathrm{d}\tau_i)$$
$$G \quad \mid \quad \alpha, G_0 \sim DP(\alpha, G_0)$$

Let (n_1, n_2, n_3) denote the numbers of cows that are in each of the three latent states. Then since G arising from the DP is almost surely discrete, there will be repeats among

the n_3 realized values of $\theta_i = (\mu_i, \tau_i)$ drawn from the random bivariate measure G. Letting $\{\phi_j : j = 1, 2, ..., k\}$ with $k \leq n_3$ represent the distinct θ_i, the group of subjects associated with each distinct ϕ_j will then form "clusters" of observations from the same component normal distribution. Hence, G is the mixing distribution that determines both the parameters for each component normal and the probability associated with it. While sampling the infinite-dimensional G as part of our MCMC scheme for sampling the joint posterior distribution might appear to be a daunting task, we actually marginalize over G, thus avoiding the issue.

4. PRIORS, NUMERICAL APPROXIMATION TO THE POSTERIOR, AND INFERENCE

We use the Bayesian framework to estimate model parameters and predict latent variables such as infection time and disease status. We use conditionally conjugate priors whenever possible. For convenience, we refer to parameters/random effects as either *cow-specific*, if they describe aspects of a particular cow, i.e. cow i's infection time, t_i^*, or *global* if they pertain to all cows, i.e. β_0, which is the population mean baseline serology score for the population of cows. We let q_i, $i = 1, 2, 3$ represent the probability of a cow being in latent infection states 1,2 and 3, respectively. The priors for global parameters are:

$$(q_1, q_2, q_3) \quad \sim \quad \text{Dirichlet}(\zeta_1, \zeta_2, \zeta_3)$$
$$se_F \quad \sim \quad \text{Beta}(a_{se_F}, b_{se_F})$$
$$sp_F \quad \sim \quad \text{Beta}(a_{sp_F}, b_{sp_F})$$
$$\text{lag} \quad \sim \quad U(\min_l, \max_l)$$
$$\beta_0 \quad \sim \quad N(\mu_{\beta_0}, b)$$
$$\tau_{\beta_0} \quad \sim \quad \Gamma(a_{\beta_0}, b_{\beta_0})$$
$$\tau_e \quad \sim \quad \Gamma(a_{\tau_e}, b_{\tau_e})$$
$$G \mid G_0, \alpha \quad \sim \quad DP(\alpha, G_0)$$

where $\min_l$ and $\max_l$ are reasonable lower and upper bounds on the lag.

To simplify computation, while maintaining flexibility, we use the normal/inverse-gamma conjugate to the normal sampling model for G_0 (Escobar and West, 1995). Hence, under $G_0(\cdot)$, we have $\tau_i \sim \Gamma(\frac{s}{2}, \frac{S}{2})$ and $\mu_i \mid \tau_i \sim N(m, \frac{\tau_i}{d})$. This choice for G_0 results in closed form solutions for constants and standard, recognizable forms for the distributions needed to carry out the MCMC algorithm described in Appendix IV. Other nonconjugate choices for G_0 are possible, but carry additional computational burden. MacEachern and Müller (1998) developed methods for posterior simulation in the nonconjugate case. We will discuss issues associated with choosing the hyperparameters s, S, m, and d as well as the DP parameter α in Sections 5 (Simulation) and 6 (Data Analysis).

Cow-specific infection times under state 2 are modeled as

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

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

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

Our interest is in the joint posterior probability density function (pdf) of the parameters and latents given the data. The posterior distribution is shown in Appendix III. Since it is unrecognizable, we use MCMC methods to simulate draws from it. Inferences for all parameters and functionals of interest are obtained from this sample. In order to avoid sampling the infinite-dimensional parameter G, we marginalize over the DP using the Polya Urn scheme (Escobar and West, 1995). Thus, we need only sample the finite-dimensional vector, $\theta_{i_1}, \theta_{i_2}, ..., \theta_{i_{n_3}}$, of parameters of the normal distribution associated with each latent state 3 subject.

Our sampling scheme is based on a Gibbs sampler that 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 (RJM-CMC) steps used to sample the posterior, are detailed in Appendix IV. The MCMC scheme was implemented by writing code in the R language (R Core Development Team, 2011). The R functions for MCMC simulation from the posterior for either the parametric or semi-parametric versions of this model are available from the first author. Simulation was executed on a Pentium(R) Dual-Core E5700 CPU running at 3.00 GHz with 4 GB of RAM. It took approximately eight hours to run simulations for 40,000 MC iterations for the Johne's disease data. This dataset contained 365 subjects, and approximately 2000 parameters and latents were simulated at each iteration of the chain.

Reversible jump MCMC is a tool for sampling varying-dimensional 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 since these sub-models have varying-dimensional vectors of terms (for model 1, 2, and 3, these vectors are $(\beta_{0i}), (\beta_{0i}, t_i^*)$, and $(\beta_{0i}, \beta_{1i}, t_i^*, \mu_i, \tau_i)$, respectively), the number of terms sampled at each iteration may change. Hence, we require a mechanism for sampling varying-dimensional spaces like RJMCMC. Note that RJM-CMC allows us to perform model selection at the individual cow level and estimate model parameters and latents simultaneously. Hence, our inferences and predictions will reflect the uncertainty associated with model selection.

Gibbs samplers for DPM models can be slow to converge because the chain can enter states where many of the "data," or slopes in our case, are assigned to the same cluster and get "stuck" there. The reason for this stickiness has to do with the discreteness of the DP and the fact that the weight on any current parameter value is proportional to the number of "data" points currently associated with it. To obtain improved mixing and quicker convergence, we incorporate the MacEachern and Müller (1998) algorithm that is designed to alleviate the problem of "sticky clusters," by augmenting the Gibbs sampler with additional steps to update the cluster parameters, ϕ_j. Other techniques for alleviating this problem can be found in Neal (2000) and in Jain and Neal (2004).

We are now set to make inferences. Standard inferences will be about the values of the various parameters and latents, including the sensitivity and specificity of the fecal culture test, the lag time from infection until serologic response and the proportions of animals in the three states. We will estimate the distribution of slopes for animals in state 3, and we will also estimate the cumulative proportions of infected animals through time. In addition, we will focus on assessing the performance of the serologic outcome as a marker for infection. In practice, populations will be screened with such biomarkers alone, without the help of a second test outcome, like fecal culture. Here, fecal culture testing serves a dual purpose (i) to help in the assessment of serology as a marker for infection, and (ii) to allow us to assess the performance of the joint testing versus using only serology.

The assessment of a continuous marker for infection begins with a discussion of sensitivity and specificity. In standard settings, a sample of outcomes is taken on animals at a particular time when it is unknown when or if animals in the cross-sectional sample were infected. A cutoff value, say c, is set, and if a serologic outcome exceeds this value, the corresponding animal is termed positive, and, otherwise, negative. The sensitivity of the test corresponding to cutoff c, $Se(c)$, is the probability that the serology outcome for an infected animal would exceed c, and the specificity, $Sp(c)$, is the probability that an uninfected animal would have a serology outcome that did not exceed c. A plot, over all possible cutoffs c, of the values of the false positive proportion $(1 - Sp(c))$ versus the true positive proportion, $Se(c)$, is called the receiver operating characteristic curve (ROC). Curves that have area under the curve (AUC) near one indicate that it is possible to find a cutoff that will achieve very high sensitivity and specificity, while if AUC is near 0.5, the biomarker is no better than a coin toss at deciding if an animal is infected or not. Letting D denote the event that the animal is infected, and $\bar{D}$ denote the event that it is not, the traditional method of calculating the AUC is to use the result that $AUC = Pr(S_D > S_{\bar{D}})$, where S_D and $S_{\bar{D}}$ are assumed to have been independently taken from the sampling distributions of serology scores from the D and $\bar{D}$ populations, respectively (Pepe, 2003).

Here, sensitivity will be a function of time since infection. Define the time related sensitivity as $Se_t(c) \equiv P(S(t) > c \mid D, \Theta, U)$, where $S(t)$ is a serology score that will be observed t units of time after t^* and where $t > lag$. So $t = 0$ corresponds to time t^* for the particular animal. Animals that

have only been recently infected, with $t < lag$, have little chance of detection; $Se_t(c)$ should be nondecreasing in t. Animals associated with a smaller (random) slope, β_1, should have smaller sensitivities than those with larger slopes, for a fixed $t > lag$. The specificity will of course not depend on t. We are thus able to obtain ROC curves that depend on t.

Our main goal was to assess the performance of serologic testing as a function of time since infection. However, it is also of interest to quantify the effect of testing based on serology alone and serology with fecal testing. We consider diagnosis based on both fecal culture, F, and serology, S. Fecal culture takes on values $F = 1$ or $F = 0$. We compare with diagnosis based on S alone. Recall that D is the event of being in either state 2 or state 3, and define $\pi = P(D) = q_2 + q_3$. For our study, the distribution of serologic scores for infected animals depends on the time, t, since infection. Ignoring t for the moment, and assuming, given latents and parameters, diagnosis based on test outcomes, (F, S), would be accomplished by calculating the odds, after applying Bayes' Theorem

$$
\begin{aligned}
&O(D \mid F, S, \Theta, U, \beta_{0*}, \beta_{1*}) \\
&\equiv \frac{P(D \mid F, S, \Theta, U, \beta_{0*}, \beta_{1*})}{1 - P(D \mid F, S, \Theta, U, \beta_{0*}, \beta_{1*})} \\
&= \frac{\pi Se^F (1 - Se)^{1-F} f(S \mid \Theta, U, D, \beta_{0*}, \beta_{1*})}{(1 - \pi)(1 - Sp)^F Sp^{1-F} f(S \mid \Theta, U, \bar{D}, \beta_{0*})} \\
&= \frac{Se^F (1 - Se)^{1-F}}{(1 - Sp)^F Sp^{1-F}} O(D \mid S, \Theta, U, \beta_{0*}, \beta_{1*}).
\end{aligned}
$$

where (β_{0*}, β_{1*}) are random intercept and slope effects corresponding to the particular S (not in the data). The corresponding probability based on (F, S) is

(4)
$$
\begin{aligned}
&P(D \mid F, S, \Theta, U, \beta_{0*}, \beta_{1*}) = \\
&e^{\log[O(D|F,S,\Theta,U,\beta_{0*},\beta_{1*})]} / [1 + e^{\log[O(D|F,S,\Theta,U,\beta_{0*},\beta_{1*})]}]
\end{aligned}
$$

with a similar expression if only S is observed. We then obtain Bayes estimates of these probabilities by taking their posterior expectations. We define these to be $g(F, S)$ and $g(S)$ respectively. With unknown parameters and latents, we allocate animals with outcome (F, S) as D if $g(F, S) > c$ or if $g(S) > c$, respectively.

The density for S conditional on D above depends on the time since infection, t. In our examples, we select various values of t, and numerically approximate the posterior mean of (4) by using the Gibbs sampler (Appendix IV) to first obtain $\{(\Theta^j, U^j) : j = 1, 2, ..., B\}$, and then $(\beta_{0*}^j, \theta_*^j) \mid (\Theta^j, U^j)$ followed by $\beta_{1*}^j \mid \theta_*^j$. Thus

(5) $\quad g(F, S) \doteq \sum_{j=1}^{B} P(D \mid F, S, \Theta^j, U^j, \beta_{0*}^j, \beta_{1*}^j)/B$

Bayesian estimates of the corresponding AUCs can be calculated as the predictive probabilities, $\widehat{AUC}_{FS} \equiv$ $P(g(F_D, S_D) > g(F_{\bar{D}}, S_{\bar{D}}) \mid data)$ and $\widehat{AUC}_S \equiv P(g(S_D) > g(S_{\bar{D}}) \mid data)$, where (F_D, S_D) is sampled from the predictive distribution of $(F, S) \mid D, data$, and $(F_{\bar{D}}, S_{\bar{D}})$ is sampled from the predictive distribution of $(F, S) \mid \bar{D}, data$, independently. At each iterate of the Gibbs sampler, we sample $(F_D^j, S_D^j) \mid D, (\Theta^j, U^j, \beta_{0*}^j, \beta_{1*}^j)$, and $(F_{\bar{D}}^j, S_{\bar{D}}^j) \mid \bar{D}, (\Theta^j, U^j, \beta_{0*}^j)$, where the intercept is sampled, then $\theta_*^j \mid \theta_1^j, ..., \theta_{n_3}^j, data$ is sampled using a Polya Urn scheme (see Escobar and West, 1995, Section 2) that is similar to but distinct from the Polya Urn scheme for $\{\theta_i : i = 1, ..., n_3\}$ given in Appendix IV, followed by sampling β_{1*} from the log normal distribution for slopes with given θ_*^j. In this way, we have a Monte Carlo sample of B pairs from each of the predictive distributions of interest, and we thus numerically approximate the AUC based on (F, S) as

$$
\widehat{AUC}_{FS} \doteq \sum_{j=1}^{B} I\{g(F_D^j, S_D^j) > g(F_{\bar{D}}^j, S_{\bar{D}}^j)\}/B
$$

where we use the approximation in (5) to approximate the terms inside the indicator function. We obtain a similar approximation to $\widehat{AUC}_S$.

We are only able to obtain the point estimates $\widehat{AUC}_{FS}$ and $\widehat{AUC}_S$ defined above. In more standard situations, the AUC can be expressed as a function of the parameters, so that the AUC itself is a parameter, and then is it usually straightforward to obtain probability intervals for it by obtaining its induced posterior distribution. Here, it is not as simple, but we can define

$$
\begin{aligned}
AUC_{F,S}(\Theta, \tilde{U}) = \\
Pr\{g(F_D, S_D \mid \Theta, \tilde{U}) > g(F_{\bar{D}}, S_{\bar{D}} \mid \Theta, \tilde{U}) \mid \Theta, \tilde{U}\}
\end{aligned}
$$

where (F_D, S_D) and $(F_{\bar{D}}, S_{\bar{D}})$ are taken from the predictive distributions described above, only conditional on $(\Theta, \tilde{U}) \equiv (\Theta, U, \beta_{0*}, \beta_{1*})$. So this is the AUC that corresponds to using a procedure for deciding on disease status, only where the procedure for deciding depends on unknown $(\Theta, \tilde{U})$. We can nonetheless numerically approximate the posterior distribution of $AUC_{FS}(\Theta, \tilde{U})$ and similarly for $AUC_S(\Theta, \tilde{U})$. This is accomplished by taking further iterations from the above predictive distributions at each iteration within the Gibbs sampler. These iterates are then used to numerically approximate $AUC_{FS}(\Theta^j, \tilde{U}^j)$ and $AUC_S(\Theta^j, \tilde{U}^j)$, for $j = 1, ..., B$. We can then numerically approximate $Pr\{AUC_{FS}(\Theta, \tilde{U}) > AUC_S(\Theta, \tilde{U}) \mid data\}$. We thus obtain an additional measure of how much improvement there is by using both F and S rather than S alone.

5. SIMULATION

In this section, we describe the results of two simulations. For the first simulated dataset, we provide a description of

the simulated data (SD), discuss prior specification, indicate how mildly informative hyperparameters for the Dirichlet Process can be constructed, and report the model's performance. We close this section with a brief summary of the second dataset and the model's performance on it. Parameter estimates were quite good for both simulated datasets. Both datasets were designed to mimic Johne's disease data so that we could test our model on simulated data.

For the first dataset, fecal and serology data were simulated for 350 subjects — 102 in latent disease state 1, 75 in state 2, and 173 in state 3. Data were unbalanced but complete, and testing occurred every six months. The number of observations per subject had minimum = 5, median = 10.5 and maximum = 21. Log-slopes of state 3 subjects were drawn from a 50–50 mix of a $N(0.4, \sigma = 0.1)$ and a $N(0.8, \sigma = 0.2)$. The other parameters were set to values similar to those estimated for the Johne's disease data.

5.1 Prior specification

Diffuse proper priors were used for parameters other than those involved in the DP. Precisions were modeled as $\Gamma(0.001, 0.001)$, a $N(0, 0.0001)$ was used for β_0 and parameters having a closed interval for support were given uniform priors. Note that in the parameterization used here, the $\mathrm{Gamma}(a, b)$ distribution has mean a/b. The vector of probabilities (q_1, q_2, q_3), is modeled with a $\mathrm{Dirichlet}(1,1,1)$. For the DP, we need to specify the precision parameter, α, and a centering measure, G_0. "Large" values of α represent strong belief that the centering measure is the correct one. However, they also give rise to a large number of clusters. Likewise, "small" values of α result in a few clusters. Here, "large" and "small" are relative to the number of subjects in latent class 3. Antoniak (1974) derives the following result relating the expected number of clusters, C, to the sample size, n, and α.

$$C = \sum_{i=1}^{n} \frac{\alpha}{\alpha + i - 1}$$

Given that the simulated data contain 173 subjects in latent state 3, it seems reasonable to identify 2–3 clusters. Using Antoniak's formula, we find that $\alpha = 0.3$ corresponds to 2.6 expected clusters so we set α to 0.3. With actual data, it is unlikley that the number of subjects in state 3 would be known. However, a rough approximation to the number of subjects in state 3 could be obtained by first running a parametric version of this model (Norris et al., 2009). Under the parametric model, we assume $\log \beta_{1i} = \gamma_i \sim N(\mu_\gamma, \tau_\gamma)$.

The selection of the hyperparameters of the centering distribution, G_0, warrants discussion. Recall that G_0 is the normal-inverse gamma conjugate prior where $\tau_i \sim \Gamma(\frac{s}{2}, \frac{S}{2})$ and $\mu_i | \tau_i \sim N(m, \frac{\tau_i}{d})$. So, we need to choose $s, S, m,$ and d. On the scale of the simulated data, a diffuse prior could be attained by setting $m = 0$, $d = 10$, and choosing $S = s = 2$. Unfortunately, the Gibbs sampler mixed poorly with this prior. In particular, the estimated distribution of slopes appeared to get stuck in states where one cluster contained the vast majority of the data (99% or more). However, we obtained excellent results with a mildly informative choice for G_0, formulated as described in Appendix II. Our choice of G_0 is based on an initial parametric analysis of the data, where we used information from the posterior for $(\mu_\gamma, \tau_\gamma)$ to create these mildly informative choices for G_0 in the semi-parametric analysis. Consequently, our current analysis is empirical Bayesian, which is quite common in this area. Escobar and West (1995) and Kleinman and Ibrahim (1998) also use informative priors in similar models involving DPs.

We note that it is possible to put priors on $\alpha, S, s, d,$ and m (Escobar and West, 1995 and DeIorio et al., 2009), making their values more data-driven. However, Escobar and West (1995) indicate that the information content in the data for estimating d may be small, so even if d is modeled with a prior distribution, typically it should be mildly informative. We would then need to include additional steps in the Gibbs sampler to sample the full conditionals of $\alpha, S, s, d,$ and m. However, if we choose these priors wisely, the associated full conditionals assume nice forms and are not difficult to sample.

5.2 Inference for simulated data (SD) 1

Two chains with dispersed initial values quickly converged to the same stationary distribution. Discarding the first 5000 iterations of each chain, we obtain the parameter estimates shown in Table 1 by combining the remaining 35,000 iterations from both chains. We note that the point estimates, taken as the posterior median here since a few of the posteriors are skewed, are quite close to the true values. Additionally, all 95% probability intervals (taken as the 2.5 and 97.5 percentiles of each posterior distribution) contain the true parameter value.

Next, we consider estimation of the nonparametric slope distribution. Samples from the posterior slope distribution are shown in Figure 3. Samples from the posterior slope distribution under the alternate scenario with $\alpha = 0.1$ are shown in the right panel of Figure 3, and we will elaborate on this comparison shortly. First, we discuss inference under $\alpha = 0.3$. Most posterior samples of the slope distribution capture the bimodality of the true distribution and account for the difference in standard deviations of the component normals.

Although the shape of the distribution of slopes appears to be estimated reasonably well, we find that the number of components in the DP mixture of normals tends to be overestimated. The true distribution of slopes is a mixture of two normals, but the highest posterior probability is assigned to a mixture of three normals as shown in Table 2. However, when we examine the posterior probabilities for the number of modes in the distribution of slopes shown in Table 3, two modes has the highest posterior probability. We note two possible reasons that the number of component normals can differ from the number of modes in a distribu-

Table 1. Parameter Estimates for SD 1

Parameter	Truth	Post. Median	95% Probability Interval	
			Lower	Upper
β_0	0.080	0.079	0.060	0.099
σ_{β_0}	0.022	0.031	0.020	0.048
σ_e	0.224	0.224	0.217	0.233
se_F	0.65	0.62	0.59	0.66
sp_F	0.97	0.966	0.955	0.976
lag	1.00	1.13	0.91	1.35
q_1	0.29	0.28	0.22	0.34
q_2	0.21	0.23	0.18	0.30
q_3	0.49	0.49	0.43	0.54
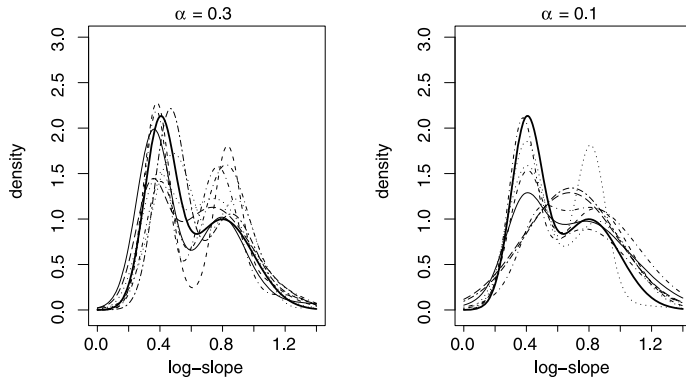
Figure 3. Posterior Samples of Log-slope Distribution for SD 1, for $\alpha = 0.3$ and $\alpha = 0.1$. Thick line = true distribution of log-slope.

Table 2. Comparison of Posterior Probabilities of Number of Clusters for SD 1

	Number of Clusters					
	1	2	3	4	5	6 or more
$\alpha = 0.3$	0.01	0.17	0.34	0.28	0.14	0.06
$\alpha = 0.1$	0.15	0.44	0.30	0.09	0.02	0.00

Table 3. Comparison of Posterior Probabilities of Number of Modes for SD 1

	Number of Modes			
	1	2	3	4
$\alpha = 0.3$	0.114	0.832	0.054	0.001
$\alpha = 0.1$	0.326	0.658	0.016	0.000

tion: (i) some components have an associated probability, or "weight," that is very small and don't show up as a distinct "bump" or (ii) some components may overlap so much that they become unimodal. For the simulation data and the Johne's disease data in the next section, the number of modes is smaller than the number of component normals due to the first reason.

To assess how sensitive inferences are to the choice of α, we reran the model with all priors the same except that α was set to 0.1, which corresponds to an expected number of clusters of 1.56. We found that the estimated number of component normals is sensitive to the choice of the precision parameter, α, in the DP prior. With $\alpha = 0.1$, a mixture of normals with two components has the highest posterior probability (compared to three components with $\alpha = 0.3$) as shown in Table 2. Posterior probabilities for the number of modes is shown in Table 3, and the highest posterior probability is still assigned to two modes. While inference about the number of component normals is sensitive to α, the overall shape of the distribution of log-slopes is similar under both choices of α, as shown in Figure 3. The distribution with $\alpha = 0.1$ appears to fit the true distribution more closely on iterations where the slope distribution

contains two modes. However, it assigns more probability to unimodal distributions. It seems that both choices of alpha provide reasonable inference about the distribution of slope distributions. Consequently, as long as interest lies in the shape of the distribution of the log-slopes, there is some flexibility in the choice of α.

In order to better understand how the performance of the serology marker changes over time, we can construct estimated ROC curves, with one curve for each value of time past infection, over a grid of time values. Since the posterior probability for two modes is the highest, we graph two sets of ROC curves to compare how the serology marker will perform on a hypothetical group with an ad hoc approximation to the smaller mode, termed the "low serology reaction" group, and another hypothetical group with ad hoc approximation to the larger mode, termed the "high serology reaction" group. We approximate the mean and standard deviation of each group using only the pdfs with two modes in Figure 3. The lower mode appears to have a mean of approximately 0.4 and a standard deviation of about 0.075 (using the fact that the range of a normal distribution is about four standard deviations and the range of the lower bump is about 0.3, so standard deviation is about 0.3/4=0.075). The higher mode appears to center on 0.9 and have a standard deviation of about 0.15. For a fixed time past infection and group, we generate an ROC curve by varying the cutoff value and plotting the ordered pairs $(1\text{-}Sp(c), Se(c))$. We obtain a MC approximation to $Se(c)$ at time t as follows:

- At iteration k of the Markov chain and for $w = 1, 2, 3, ..., l$, we simulate l serology scores corresponding to time t past infection by sampling l sets of cow-specific parameters, $\beta_{0w}^{(k)}, \beta_{1w}^{(k)}, \epsilon_w^{(k)}$ and calculating $\beta_{0w}^{(k)} + \beta_{1w}^{(k)}(t - lag^{(k)}) + \epsilon_w^{(k)}$
- Next, we determine what proportion of serology scores for time t are above the cutoff, i.e. we calculate

$$Se^{(k)}(c) = \frac{1}{l} \sum_{w=1}^{l} I(\beta_{0w}^{(k)} + \beta_{1w}^{(k)}(t - lag^{(k)}) + \epsilon_w^{(k)} > c)$$
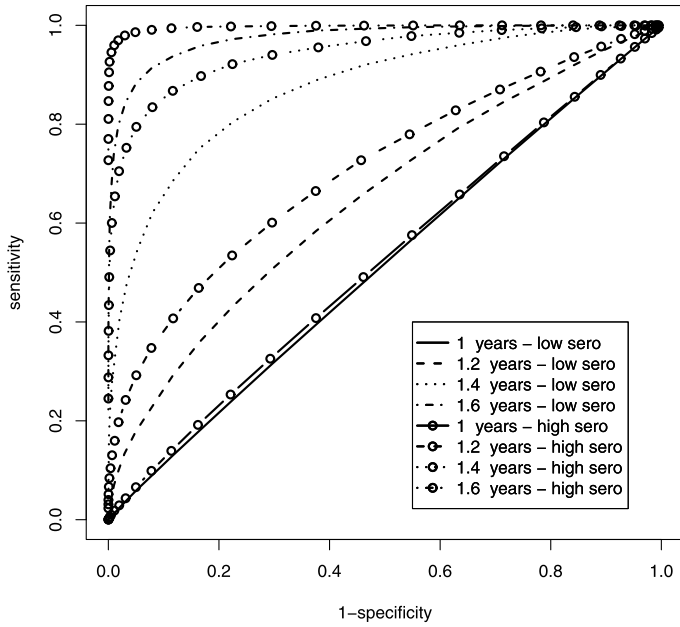
Figure 4. Estimated ROC curves for SD 1 at selected values of time past infection according to hypothetical low and high serology groups.
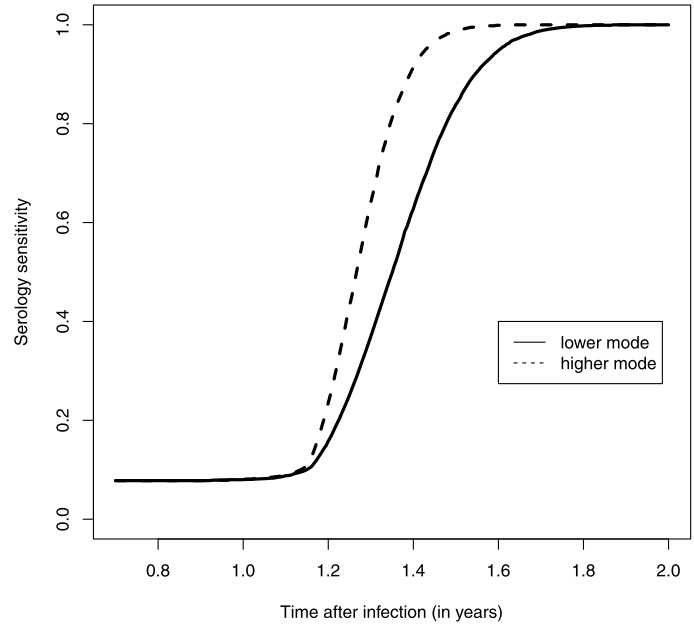

Figure 5. Estimated sensitivity as a function of time past infection for hypothetical low and high serology reaction groups for SD 1.

- Averaging $Se^{(k)}(c)$ over all iterations of the MC gives the approximation to $Se(c)$ at time t. In practice, we thin the MC to obtain these estimates because of the substantial compute time involved.

The slopes $\beta_{1w}^{(k)}$ are drawn from the lower mode, i.e. a normal with mean 0.4 and standard deviation 0.075 to obtain MCMC approximations to $Se(c)$ for this group. A similar computation is employed for $Sp(c)$. We repeat this procedure for the high serology group. The estimated ROC curves are shown in Figure 4. The two groups are clearly distinguished in terms of the biomarker's ability to discriminate infected from non-infected animals.

We can also study the manner in which the sensitivity of the serology test will increase over time by fixing a cutoff value for the serology test (scores above the cutoff value are interpreted as a positive test result) and graphing the serology sensitivity against time past infection. Graphs of estimated sensitivity as a function of time for the hypothetical low and high serology groups are shown in Figure 5.

Since one of our primary goals is diagnosis, we now examine the classification accuracy of the semi-parametric model. Each subject was classified based on their predictive probability vector for the three states; the state with the highest probability was the classification state. Classification was quite accurate, with only 34 out of 350 subjects, or 10%, misclassified as summarized in Table 4. Alternately, if we are only interested in which subjects are infected or not infected, we can collapse latent states 2 and 3 into a single "infected" state and obtain overall estimates of sensitivity

Table 4. Estimated classification accuracy for SD 1

		Classification		
		1	2	3
	1	92	9	1
True State	2	13	61	1
	3	1	9	163

and specificity for the joint model of 0.94 and 0.90, respectively. Since these allocations were performed using all the data, these numbers are likely somewhat optimistic.

5.3 Simulated dataset 2

The second simulated data (SD 2) consisted of fecal and serology test results for 300 subjects – 19% in latent state 1, 12% in state 2 and 69% in state 3. Data were unbalanced but complete, and testing occurred every six months. The number of observations per subject had minimum = 5, median = 11 and maximum = 17. Log-slopes of state 3 subjects were drawn from a mixture of two normals consisting of a $N(0.4, \sigma = 0.1)$ with probability 0.8 and a $N(0.7, \sigma = 0.2)$ with probability 0.2. The resulting distribution is unimodal and right-skewed. The true values of other parameters along with the estimated values and 95% probability intervals are shown in Table 5. Note that all 95% probability intervals contain the true parameter value.

Inference regarding the distribution of log-slopes was quite accurate. Figure 6 shows six systematically sampled iterates from log-slope distributions from the full MCMC

Table 5. *Parameter Estimates for SD 2*

| Parameter | Truth | Post. Median | 95% Probability Interval | |
			Lower	Upper
β_0	0.80	0.79	0.77	0.82
σ_{β_0}	0.032	0.032	0.019	0.047
σ_e	0.224	0.225	0.219	0.234
se_F	0.75	0.72	0.69	0.76
sp_F	0.92	0.93	0.91	0.94
lag	0.70	0.79	0.61	1.03
q_1	0.19	0.15	0.10	0.20
q_2	0.12	0.18	0.12	0.24
q_3	0.69	0.67	0.61	0.73

Table 6. *Posterior Probability of Number of Clusters for SD 2*

Number of Clusters	1	2	3	4	5 or more
Posterior Prob	0.00	0.22	0.35	0.25	0.17

sample, along with the true log-slope distribution. The iterates from the posterior tend to be unimodal and right-skewed like the true log-slope distribution.

Table 6 gives the posterior distribution for the number of clusters in the log-slope distribution while Table 7 gives the posterior distribution for number of modes. Note that the model correctly assigns highest posterior probability to one mode. Although the number of clusters having highest posterior probability is 3, when the true log-slope distribution has 2 clusters, we still obtain a good approximation to the true shape of the log-slope distribution. As shown in Figure 6, where a circle is plotted for the mean of each component distribution with area proportional to probability of the corresponding component, even when the number of clusters is 3, the resulting shape of the log-slope distribution is typically still unimodal and close to the true distribution of log-slopes. If there are three or four clusters, there is often a cluster with negligible probability or two clusters with means that are very close to each other.

6. ANALYSIS OF JOHNE'S DISEASE DATA

We now analyze the Johne's disease data. We first consider the specification of a relatively vague prior and follow up with inference under the semi-parametric model.

6.1 Prior specification

Except for the DP, all priors are vague. $\Gamma(0.001, 0.001)$ priors are used for all precision parameters and $N(0, 0.0001)$ priors are used for parameters whose support is the real line. Parameters whose support is a bounded interval are given uniform priors. As in the preceding simulation, we need to choose α, the precision parameter of the DP, as well as the hyperparameters of G_0, the centering distribution of the DP, i.e. $s, S, m,$ and d.
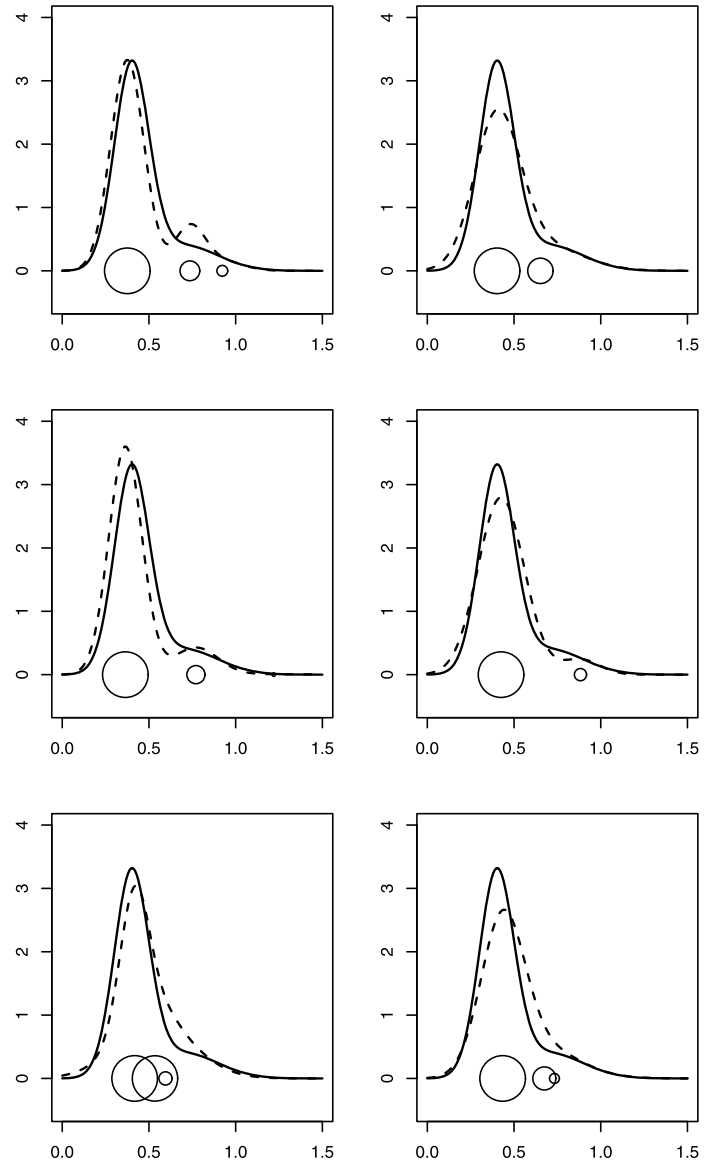

Figure 6. *Posterior Samples of Log-slope Distribution for SD 2. Solid line = true distribution of log-slope, dashed line = sample from posterior distribution. A circle is located at the mean of each component normal with size proportional to component weight.*

Since we don't have good independent information on which to base our prior for the distribution of log-slopes, we take the empirical Bayes approach discussed in Appendix II, as was done in the SD 1 in Section 5. We first consider $\tau_i \sim \Gamma(\frac{s}{2}, \frac{S}{2})$. Under the parametric model, $\log \beta_{1i} = \gamma_i \sim N(\mu_\gamma, \tau_\gamma)$. The posterior medians of μ_γ and τ_γ are 0.2184 and 1.0493, respectively. We set the mode of the prior $\tau_i \sim \Gamma(\frac{s}{2}, \frac{S}{2})$ equal to the posterior median under the parametric model, 1.0493. To obtain a upper bound on τ_i, i.e. a high value of the precision, we assume, a "narrow" component distribution in the DP mixture of normals has 16 times

Table 7. Posterior Probability of Number of Modes for SD 2

Number of Modes	1	2	3
Posterior Prob	0.59	0.38	0.03

Table 8. Parameter Estimates for Johne's Disease Data

| Parameter	Post. Mean	95% Probability Interval	
		Lower	Upper
β_0	-1.741	-1.761	-1.721
σ_{β_0}	0.067	0.052	0.087
σ_e	0.134	0.126	0.153
se_F	0.57	0.52	0.63
sp_F	0.976	0.955	0.990
q_1	0.48	0.41	0.55
q_2	0.25	0.19	0.32
q_3	0.26	0.22	0.32
lag	1.60	1.32	1.85

Figure 7. Posterior iterates of Log-slope Distribution, with posterior mean in bold; Johne's Data.

Table 9. Posterior Probability of Number of Clusters for Johne's Data

Number of Clusters	1	2	3	4	5 or more
Posterior Prob	0.12	0.27	0.30	0.19	0.11

the modal precision. Thus, the 97.5th percentile is set to $16 * \hat{\tau}_\gamma = 16 * 1.0493 = 16.79$. Using a grid search in R, we find $s = 2.5278$ and $S = 0.5030$ yield a gamma distribution with the desired mode and 97.5th percentile. We center the distribution of $\mu_i | \tau_i \sim N(m, \frac{\tau_i}{d})$ on the posterior median of μ_γ, so $m = 0.2184$. We also set $d = 10$ for the reasons discussed in Appendix II.

Based on the estimated distribution of log-slopes from the parametric analysis shown in Figure 1, it appears that the log-slope distribution might be a mixture of three component normals having means at $x = -5, 0$ and 1. Consequently, we select the precision parameter of the DP, α, to allow for 2–3 clusters. Assuming that about 92 cows will be in latent state 3 (based on the parametric analysis) and using Antoniak's formula we find that $\alpha = 0.3$ gives 2.4 expected clusters.

6.2 Inference

Two chains with dispersed initial values quickly converged to the same stationary distribution. Discarding the first 10,000 iterations of each chain, we obtain the parameter estimates shown in Table 8 by combining the remaining 30,000 iterations from both chains. Classifying the 365 subjects into latent disease states yields 197 in state 1, 76 in state 2, and 92 in state 3.

Several iterates corresponding to systematic samples from the posterior slope distribution, and the posterior mean corresponding to these iterates are shown in Figure 7. We note that some of the iterates from the distribution of log-slopes are bimodal with global maximum near zero and a smaller mode near -2. Since $e^{-2} \approx 0.14$, it is possible that the lower mode represents cows who become infected and either fail to have a global antibody response or have a very low-level antibody response.

We now consider inference about the number of component normal distributions in our distribution of log-slopes.
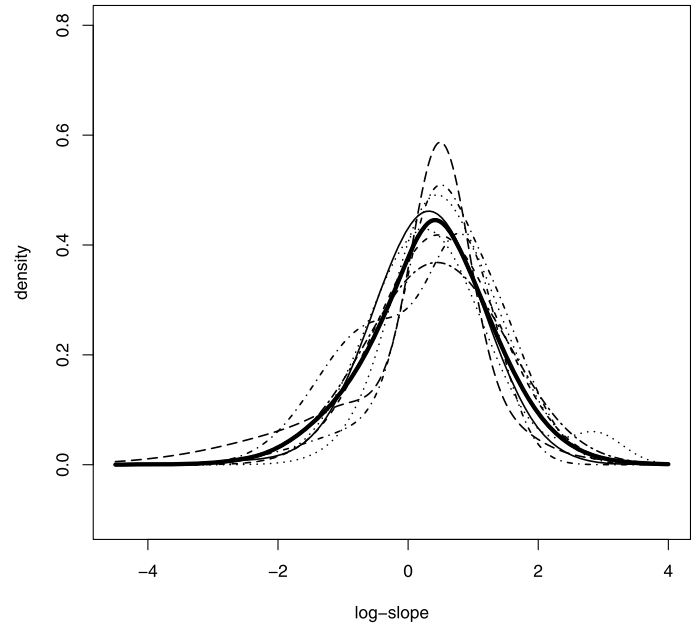
Table 9 displays the posterior probabilities corresponding to various numbers of clusters, or normal components, in the distribution of log-slopes. A normal mixture with three components has the maximum posterior probability. However, Table 10 indicates that the most probable number of modes is one, with two modes having second highest probability. Note that the fact that this analysis indicates that one mode is most likely does not mean that modeling slopes nonparametrically is unnecessary. Not only do nonparametric models allow for multiple modes, they also allow the distribution of slopes to be skewed, to have thicker or thinner tails than a normal, or to depart from a normal distribution in a myriad of ways. The fact that 3 clusters is most probable suggests that the log-slopes of the serology trajectories are insufficiently modeled with a single normal distribution and have features which are better modeled using a mixture of 3 normals. Indeed, the empirical data in Figure 1 do suggest skewness, among other departures from normality. It is possible to mix three component normals together in such a way as to create a unimodal distribution (for example, if their modes are close together).

For the Johne's disease data, the ROC curves at selected times past infection for estimated high and low serology reaction groups are displayed in Figure 8. By analyzing the posterior iterates of the log-slope distribution shown in Fig-

Table 10. Posterior Probability of Number of Modes for Johne's data

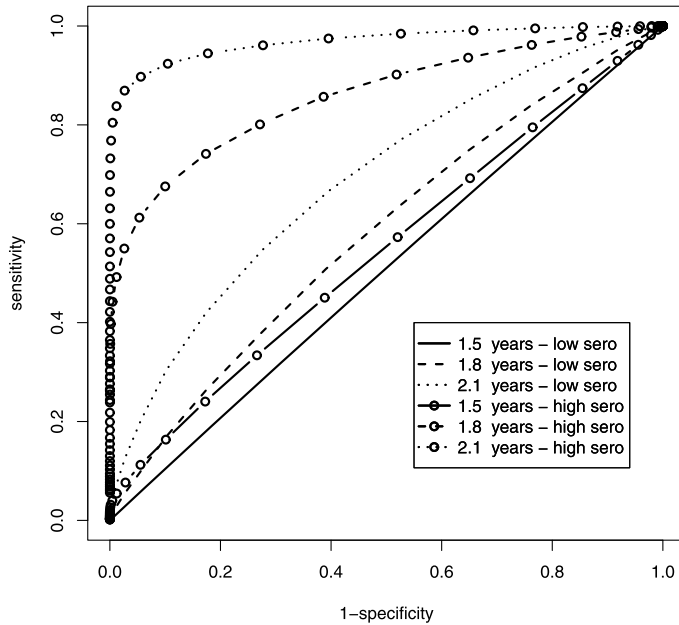| Number of Modes	1	2	3	4
Posterior Prob	0.62	0.30	0.07	0.00

Figure 8. Estimated ROC curves for Johne's disease data for hypothetical groups at selected values of time past infection.

Figure 9. Estimated sensitivity as a function of time for hypothetical high and low serology groups; cutoff $= -1.29$.

Table 11. Comparison of estimated AUCs for serology with fecal and serology alone

Years past infection	Sero. with fecal AUC_{FS}	Sero. alone AUC_S	Posterior Prob $AUC_{FS} > AUC_S$
1.7	0.868	0.718	1.000
1.9	0.948	0.904	0.994
2.1	0.976	0.944	0.974
2.3	0.981	0.962	0.942
2.7	0.986	0.977	0.870

ure 7, we obtained rough estimates of the mean and standard deviation of the high and low clusters. Many of the iterates suggest the low cluster is centered around -1.6 with a standard deviation of about 0.4 and the higher cluster is centered at about 0.6 with standard deviation of 0.9. We recognize that these estimates are ad hoc. However, obtaining estimates of the mean and standard deviation of these hypothetical high and low clusters is not a well-defined problem since the number of clusters will vary from iteration to iteration of the Markov chain, and components of the mixture arrive and disappear across iterations, so defining groups/clusters of cows is a non-trivial task. The bottom line issue is the lack of identifiability of the model for estimating components of the mixture. Although our solution to this difficulty is ad hoc, this method aids in giving an idea of the difference in how the serology test performs between the two hypothetical groups. The same method was used to obtain the graph of serology as a function of time past infection for the high and low clusters shown in Figure 9. We note there is a huge difference in the performance of the serology marker between these two hypothetical groups. Even at three years past infection, the serology test still has an estimated sensitivity less than 0.20 for the low serology reaction group, whereas the sensitivity for the high serology group is 1. An awareness of these types of limitations of
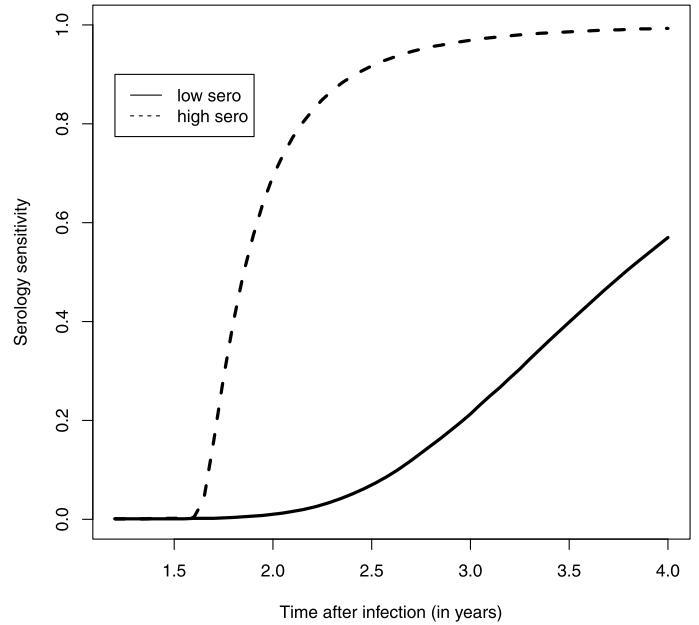
serology tests can aid scientists in devising better screening regimens.

In order to compare the performance of serology with fecal culture against serology alone, we estimate the AUCs over a grid of time points as shown in Table 11. We note that serology with fecal outperforms serology alone at all the times past infection considered. However, the difference in performance is largest at 1.7 years after infection, just after the lag has elapsed. The difference in performance becomes negligible by 2.7 years after infection. This analysis is without taking hypothetical groups into account. More work needs to be done to develop a systematic method of defining these groups.

Finally, we consider how this model could be used to understand the temporal dynamics of disease progression within a herd. For each animal predicted by the model to be in an infected state (state 2 or 3), we can obtain the animal's predicted infection time. We can then graph the cumulative number of animals infected as a function of time as shown in Figure 10 for the Johne's disease data. Unfor-
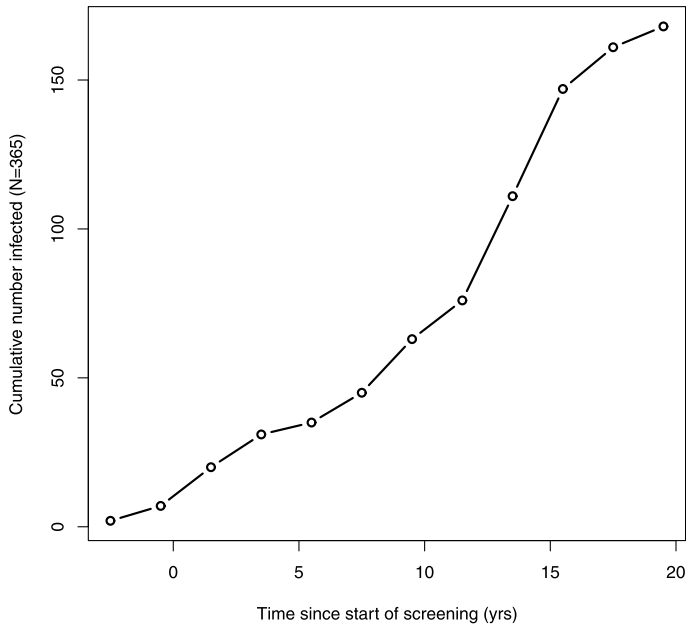
Figure 10. Estimated cumulative number of animals infected over time.

tunately, herd membership was not known with certainty for all animals in the Johne's disease study so the graph was produced assuming all animals were in the same herd. Hence, the graph serves an an illustration of the model's potential, and one should be cautious about drawing any biological conclusions from it.

7. CONCLUSIONS

Detection of chronic bacterial diseases, such as Johne's disease in cattle, is especially challenging in the subclinical phase because available diagnostic tests are of low to moderate sensitivity. Simultaneous use of multiple, conditionally-independent tests can increase the overall sensitivity and accuracy (as measured by the area under the ROC) of the joint test result. When longitudinal test results are available and there is no perfect reference standard, repeated test results may be difficult to interpret and ensure high predictive probabilities of disease or absence of disease. We developed a flexible semi-parametric approach for longitudinal test results for Johne's disease based on one binary test (fecal culture) and one continuous-scaled test (serum ELISA scores) using a Bayesian latent class model. The model allows for subpopulations of animals that were not identifiable assuming a parametric (log normal) model for ELISA scores. Our novel model has potential utility because differences in rates of serum antibody response to infection (high or low) in subpopulations of animals might be useful predictors of progression to clinical disease, or the effect of diseases on milk production or on reproductive success which often occur during the subclinical phase. In addition, the model can

provide insight into the temporal dynamics of infection with the causative bacterium in cattle herds thereby facilitating management decisions based on test results.

A major finding of our analysis is that there are distinct groups of animals where one group has a fairly quick serologic reaction while the other has a very slow one. If there are herds with large numbers of slow reactors, the overall sensitivity across all animals in such a herd could be quite low as it would involve some sort of average of estimates like those presented in Figure 9, for example. A second, not surprising, conclusion is that diagnosis based on both fecal culture and serology can result in much improved diagnosis over serology testing alone, at additional cost.

The statistical methodology developed for this practical veterinary problem, or simple variations of it, will apply to a host of other diagnostic testing situations. There are many animal infections that would involve binary and continuous diagnostic testing and thus our methodology would in theory apply to all of them. There are of course details that will vary, for example, tests might be dependent conditional on disease status, in which case the dependence could be modeled and our method adapted accordingly. In human medicine, there is usually follow-up when a person is suspected of having some form of infection or disease. For example, it is common to use a PSA test and a digital examination for prostate cancer. Follow-up can involve invasive and expensive biopsy. In HIV testing, a relatively inexpensive ELISA test is followed up with a more expensive Western Blot test (virtually a gold standard), when it is positive, and not when it is negative. We are currently working on adapting our methodology to handle the partial gold standard case, which can involve selection bias that must be taken into account. Our method is easily adapted to situations where a gold standard test is always applied. Finally, we are also working on more complex models for the serologic/biomarker response, which allow for the possibility of a sigmoid function, for example, where individuals have a continuous increase in response followed by a leveling off. More complex functions could also be studied.

ACKNOWLEDGEMENT

APPENDIX I. EVIDENCE FOR CONDITIONAL INDEPENDENCE ASSUMPTION

In order to determine if the assumption that serology and fecal results are independent conditional on knowing disease state and all other parameters and latents seems reasonable

for the Johne's disease data, we first fit the model to obtain estimates of all parameters and predictions of all latents. Then, for each observed serology score, we calculated the residual $e_{ij} = S_{ij} - [\hat{\beta}_{0i} + \hat{\beta}_{1i}(t_{ij} - \hat{t}_i^* - \hat{\text{lag}})^+]$. Note the residual simplifies to $e_{ij} = S_{ij} - \hat{\beta}_{0i}$ for all serology scores of uninfected animals and for serology scores of infected animals taken at any time before infection time plus lag. We calculated the Pearson correlation coefficient for the pairs (e_{ij}, F_{ij}) for four disjoint groups of screening results as follows: 1) all screenings for animals who were never infected during the study, 2) screenings for infected animals that were taken before the estimated infection time, 3) screenings for infected animals that were taken from time of infection to time of infection plus lag, and 4) screenings taken for infected animals that were taken after the lag had elapsed. The screening results were assigned to the four groups based on estimated infection times and latent states from the fitted model. We calculated correlations within the four groups to ensure that aggregation did not wash out any potential correlations between serology and fecal results. The correlation coefficients for the four groups, respectively, are: 0.009 ($P = 0.758$, $n = 1287$), -0.11 ($P = 0.05$, $n = 346$), 0.14 ($P = 0.02$, $n = 260$) and -0.003 ($P = 0.968$, $n = 204$). The conditional independence assumption is questionable for the infected animals prior to infection and during the lag between infection and serology reaction ($r = -0.11$ and 0.14). However, we note that if dependence between serology and fecal tests exists during these phases, it is not of practical importance.

APPENDIX II. DETAILS FOR SELECTION OF MILDLY INFORMATIVE G_0

Formulation of the DP as a mildly informative prior could be based on data from past studies or information elicited from an expert. Since neither of these sources of information are available for simulated data, we take an empirical Bayes approach and inform the choice of G_0 based on a parametric run of the model. We take the posterior medians of μ_γ and τ_γ as their point estimates. For the simulated data, $\hat{\mu}_\gamma = 0.65$ and $\hat{\tau}_\gamma = 12.3$. Since $\tau_i \sim \Gamma(\frac{s}{2}, \frac{S}{2})$, we choose s and S so that the mode of this distribution equals the precision for the parametric (one normal component) model. Thus, $\frac{\frac{s}{2}}{\frac{S}{2}-1} = 12.3$. We also choose s and S so that the 97.5 percentile corresponds to a "large" value of τ_i. Since a "large" precision corresponds to a small standard deviation, we choose a "large" value of τ_i to be the precision corresponding to a component distribution (in the DP mixture of normals) having one-fourth the standard deviation obtained under the parametric model. One-fourth the standard deviation corresponds to 16 times the precision, thus $P_{0.975} = 16 * 12.3 = 196.8$. Using a grid search in R, we find that the gamma with mode 12.3 and 97.5 percentile 196.8 has parameters $s = 2.52$ and $S = 0.04$.

We center the prior $\mu_i | \tau_i \sim N(m, \frac{\tau_i}{d})$ on $m = \hat{\mu}_\gamma = 0.65$ from the parametric model. Next, we choose d, a parame-

ter which affects the number of modes in the distribution of the log-slopes. "Large" values of d result in priors which place higher probability on distributions with a large number of modes (Escobar and West, 1995). Recall that the prior parameter, α, influences the number of clusters, which is different from the number of modes since clusters that are close together may appear as a single mode. At a minimum, we wanted to ensure that d was chosen so that, when $\tau_i = \hat{\tau}_\gamma$, the prior for μ_i has more dispersion than the distribution of the log-slopes, γ_i, under the parametric model, i.e. $\frac{1}{\text{Var}(\mu_i | \tau_i)} = \frac{\tau_i}{d} = \frac{\hat{\tau}_\gamma}{d} < \hat{\tau}_\gamma$. Thus, d must be at least 1 since lower precision means more dispersion. We ran simulations with $d = 2.25, 10, 50$ and 200. All values of d produced good approximations to the true distribution of log-slopes except $d = 2.25$, which tended to produce unimodal distributions. We chose $d = 10$ since too large a value of d can inhibit the mixing of the chain by yielding draws for new cluster means that are ridiculously small or large. Consequently, when τ_i equals its mode, the standard deviation of the distribution of μ_i is roughly 3.3 times standard deviation of the distribution of the log-slopes under the parametric model.

APPENDIX III. LIKELIHOOD AND MARGINALIZED JOINT POSTERIOR DISTRIBUTION

We derive the likelihood and joint posterior, and subsequently marginalize the joint posterior over the DP. To economize on notation, we set $data = \{(S_{ij}, F_{ij}), t_{ij} \; \forall i, j\}$. 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 $\{\beta_{0i} : i = 1, 2, ..., n\} \bigcup \{\beta_{1i} : i \text{ such that } k_i = 3\} \bigcup \{k_i : i = 1, 2, ..., n\}$, while the parameters, denoted by $\Theta = \{q_1, q_2, q_3, \beta_0, \tau_{\beta_0}, \tau_e, se_F, sp_F, \text{lag}, \{t_i^* : i = 1, 2, ...n\}, \{(\mu_i, \tau_i) \; \forall i \text{ such that } k_i = 3\}, G\}$. We have

(6) $\quad L(\Theta | U, data)$
$$= f(data | U, \Theta) \cdot f(U | \Theta)$$
$$\propto \left\{ \prod_{i:k_i=1} \left[q_1 \prod_{j=1}^{m_i} \tau_e^{\frac{1}{2}} \exp\{-\frac{\tau_e}{2}(S_{ij} - \beta_{0i})^2\} \right. \right.$$
$$\left. \times Sp_F^{1-F_{ij}}(1 - Sp_F)^{F_{ij}} \right]$$
$$\left. \times \tau_{\beta_0}^{\frac{1}{2}} \exp\{-\frac{\tau_{\beta_0}}{2}(\beta_{0i} - \beta_0)^2\} \right\}$$
$$\times \left\{ \prod_{i:k_i=2} \left[q_2 \prod_{j=1}^{m_i} \tau_e^{\frac{1}{2}} \exp\{-\frac{\tau_e}{2}(S_{ij} - \beta_{0i})^2\} \right] \right.$$
$$\times \prod_{j:t_{ij}<t_i^*} \left[Sp_F^{1-F_{ij}}(1 - Sp_F)^{F_{ij}} \right]$$
$$\times \prod_{j:t_{ij}\geq t_i^*} \left[Se_F^{F_{ij}}(1 - Se_F)^{1-F_{ij}} \right]$$
$$\left. \times \tau_{\beta_0}^{\frac{1}{2}} \exp\{-\frac{\tau_{\beta_0}}{2}(\beta_{0i} - \beta_0)^2\} \right\}$$

$$\times \left\{ \prod_{i:k_i=3} \left[q_3 \prod_{j=1}^{m_i} \tau_e^{\frac{1}{2}} \exp\{-\frac{\tau_e}{2}\tilde{S}_{ij}^2\} \right] \right.$$

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

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

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

$$\left. \times \frac{\tau_i^{\frac{1}{2}}}{\beta_{1i}} \exp\{-\frac{\tau_i}{2}(\log(\beta_{1i})-\mu_i)^2\} \right\}$$

where

$$\tilde{S}_{ij} = S_{ij} - \beta_{0i} - \beta_{1i}(t_{ij}-t_i^*-\text{lag})^+$$

Lines 3–5 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 five lines correspond to subjects having infection with serology reaction ($k_i = 3$).

Letting $\pi(\cdot)$ represent the joint prior on its arguments, and $\mathcal{P}$ represent the Dirichlet Process prior on the random measure G, and with a slight misuse of notation, we obtain the following posterior distribution for the semi-parametric model.

$$p(\Theta, U|data) \propto$$
$$L(\Theta|U, data) \cdot \pi(q_1, q_2, q_3, \beta_0, \tau_{\beta_0}, \tau_e, se_F, sp_F, \text{lag})$$
$$\times \prod_{i:k_i=2} \pi_2(t_i^*) \cdot \left[\prod_{i:k_i=3} \pi_{3i}(t_i^*) \cdot G(d\theta_i) \right] \cdot \mathcal{P}(dG)$$

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.

In order to avoid sampling the infinite-dimensional parameter G, we marginalize over the DP using the Polya Urn scheme (Ferguson, 1973). We let $i_1, i_2, ..., i_{n_3}$ be the indices for subjects in state 3.

$$(7) \int p(\Theta, U|data)\mathcal{P}(dG) \propto$$
$$L(\Theta|U, data) \cdot \pi(q_1, q_2, q_3, \beta_0, \tau_{\beta_0}, \tau_e, se_F, sp_F, \text{lag})$$
$$\times \prod_{i:k_i=2} \pi_2(t_i^*) \cdot \prod_{i:k_i=3} \pi_{3i}(t_i^*)$$
$$\times \prod_{r=1}^{n_3} \frac{\alpha G_0(d\theta_{i_r}) + \sum_{w=1}^{r-1} \delta_{(\theta_{i_w})}(d\theta_{i_r})}{\alpha + r - 1}$$

We note that the marginalized joint posterior in Equation 7 no longer includes the random measure G. However, we now need to sample $\theta_{i_1}, \theta_{i_2}, ..., \theta_{i_{n_3}}$, the parameters of the normal distribution associated with each latent state 3 subject.

APPENDIX IV. MCMC SIMULATION OF POSTERIOR

We use Markov Chain Monte Carlo methods to approximate the corresponding posterior distribution. More specifically, we use a Gibbs sampler that incorporates reversible jump steps. Most of the full conditionals have recognizable forms and can be sampled directly; we use Metropolis and slice sampling for the rest.

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 $\beta_{0i}, \beta_{1i}, t_i^*, \mu_i$ and τ_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.

4. DPM models can be slow to converge. To speed convergence, we implement the "sticky clusters" algorithm of MacEachern and Müller (1998) which amounts to augmenting the parameter space with the distinct cluster parameters $\{\phi_j : j = 1, 2, ..., k\}$ and sampling their full conditionals. More detail will follow in Appendix IV.2.

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 $i = 1, 2, 3$ and $\overrightarrow{q} = (q_1, q_2, q_3)$, the full conditionals for

the global parameters are:

(8)

$$\vec{q}|\text{else} \sim \text{Dirichlet}(n_1 + \zeta_1, n_2 + \zeta_2, n_3 + \zeta_3)$$

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

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

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

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

$$+ \sum_{i:k_i=3}\sum_j (S_{ij} - \beta_{0i} - \beta_{1i}(t_{ij} - t_i^* - \text{lag})^+)^2\Big]$$

$$+ b_{\tau_e})$$

$$sp_F|\text{else} \sim \text{Beta}(\sum_{i:k_i=1}\sum_j (1 - F_{ij}) +$$

$$\sum_{i:k_i \in \{2,3\}}\sum_{j:t_{ij}<t_i^*} (1 - F_{ij}) + a_{spF},$$

$$\sum_{i:k_i=1}\sum_j F_{ij} + \sum_{i:k_i \in \{2,3\}}\sum_{j:t_{ij}<t_i^*} F_{ij} + b_{spF})$$

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

$$\sum_{i:k_i \in \{2,3\}}\sum_{j:t_{ij}\geq t_i^*} (1 - F_{ij}) + b_{seF})$$

$$\text{lag}|\text{else} \propto \prod_{i:k_i=3}\Big[\prod_{j=1}^{m_i} \exp\{-\frac{\tau_e}{2}\tilde{S}_{ij}^2\}\Big] \cdot \frac{1}{\text{lag}^{n_2}}$$

$$\cdot \Big[\prod_{i:k_i=3} \frac{1}{t_{im_i} - \text{lag} - \text{dob}_i}\Big] \cdot I(\min_l^* < \text{lag} < \max_l^*)$$

$$\text{where } \tilde{S}_{ij} = S_{ij} - \beta_{0i} - \beta_{1i}(t_{ij} - t_i^* - \text{lag})^+$$

$$\text{where } \min_l^* = \max\{\{t_{im_i} - t_i^*, \forall i \ni k_i = 2\}, \min_l\}$$

$$\text{where } \max_l^* = \min\{\{t_{im_i} - t_i^*, \forall i \ni k_i = 3\}, \max_l\}$$

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

$$p(t_i^*|k_i = 2, \text{else}) \propto$$

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

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

$$\times I(t_{im_i} - \text{lag} < t_i^* < t_{im_i})$$

Finally, if $k_i = 3$, the infection with serology reaction case, we sample $\beta_{0i}, \beta_{1i}, t_i^*, \mu_i$ and τ_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

$$p(\beta_{1i}, t_i^*|k_i = 3, \text{else}) \propto$$

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

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

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

$$\times \frac{1}{\beta_{1i}}\exp\{-\frac{\tau_i}{2}(\log(\beta_{1i}) - \mu_i)^2\}$$

$$\cdot I(\text{dob}_i < t_i^* < t_{im_i} - \text{lag})$$

This joint full conditional does not have a recognizable form so we use a random walk Metropolis sampler as described in Norris, Johnson and Gardner (2009). The full conditional of (μ_i, τ_i) is derived in the next section.

IV.1 Sampling full conditional of (μ_i, τ_i)

In this section, we derive the full conditional of $\theta_i = (\mu_i, \tau_i)$ using the argument of Escobar and West (1995) but with added detail. We misuse notation slightly by allowing θ_i for $i = 1, 2, ..., n_3$ to represent the collection of θ's associated with model 3 subjects, realizing that subjects in model

3 on any given iteration of the Markov Chain are unlikely to have indices in consecutive order. Letting $\theta^{(i)} = \{\theta_j : j \neq i\}$ and $\Theta^{(i)}$ = the parameter vector Θ with θ_i removed, the full conditional of θ_i is derived as follows.

$$
\begin{aligned}
&f(\theta_i \mid k_i = 3, data, U, \Theta^{(i)}) \\
&= f(\theta_i \mid \beta_{1i}, \theta^{(i)}) \\
&\propto \sqrt{\tau_i}\exp\{-\frac{\tau_i}{2}(\log\beta_{1i} - \mu_i)^2\} \\
&\times \frac{\alpha G_0(\mathrm{d}\theta_i) + \sum_{w\neq i, w=1}^{n_3} \delta_{\theta_w}(\mathrm{d}\theta_i)}{\alpha + n_3 - 1}
\end{aligned}
$$

Completing the square and some algebra give the following result.

$$
\begin{aligned}
&f(\theta_i \mid \beta_{1i}, \theta^{(i)}) \propto \\
&C\tau_i^{\frac{s-1}{2}}\exp\{-\frac{\tau_i}{2}[S + \frac{1}{d+1}(\log\beta_{1i} - m)^2]\} \\
&\times \tau_i^{\frac{1}{2}}\exp\{-\frac{\tau_i}{2}(1 + \frac{1}{d})(\mu_i - \hat{\mu}_i)^2\} \\
&+ \sqrt{\frac{\tau_i}{2\pi}}\exp\{-\frac{\tau_i}{2}(\log\beta_{1i} - \mu_i)^2\} \sum_{w\neq i, w=1}^{n_3} \delta_{\theta_w}(\mathrm{d}\theta_i) \\
&\text{where } C = \frac{\alpha}{\sqrt{2\pi d}} \cdot \frac{(\frac{S}{2})^{\frac{s}{2}}}{\Gamma(\frac{s}{2})} \text{ and } \hat{\mu}_i = \frac{d\log\beta_{1i} + m}{d+1}
\end{aligned}
$$

The first term is a constant times the kernel of a $\Gamma(\frac{s+1}{2}, \frac{1}{2}[S + \frac{1}{d+1}(\log\beta_{1i} - m)^2])$ times a $N(\hat{\mu}_i, (1 + \frac{1}{d})\tau_i)$ kernel. The second term is a discrete distribution with mass proportional to $q_w \equiv \sqrt{\frac{\tau_w}{2\pi}}\exp\{-\frac{\tau_w}{2}(\log\beta_{1i} - \mu_w)^2\}$ on $\theta_w = (\mu_w, \tau_w)$, for $w \neq i, w = 1, 2, ..., n_3$. So the full conditional is a mixture of a continuous and a discrete bivariate distribution.

In order to determine the probability of drawing from the continuous component, we need to integrate the first term in the above equation with respect to (μ_i, τ_i) to determine its normalizing constant, q_{0i}. This calculation is standard but tedious so we only show the result below:

$$
\begin{aligned}
q_{0i} &= \int C\tau_i^{\frac{s-1}{2}}\exp\{-\frac{\tau_i}{2}[S + \frac{1}{d+1}(\log\beta_{1i} - m)^2]\} \\
&\times \tau_i^{\frac{1}{2}}\exp\{-\frac{\tau_i}{2}(1 + \frac{1}{d})(\mu_i - \hat{\mu}_i)^2\}\,\mathrm{d}\mu_i\mathrm{d}\tau_i \\
&= \sqrt{\frac{1}{\pi}} \cdot \frac{\Gamma(\frac{s+1}{2})}{\Gamma(\frac{s}{2})}(sM)^{-\frac{1}{2}}[1 + \frac{(\log\beta_{1i} - m)^2}{sM}]^{-\frac{s+1}{2}}
\end{aligned}
$$

where $M = \frac{(1+d)S}{s}$.

Thus, (μ_i, τ_i) is a new draw from the continuous part of the full conditional with probability proportional to q_{0i} and equal to one of the remaining current values, (μ_w, τ_w), with probability proportional to q_w for $w = 1, 2, ..., i-1, i+1, ..., n_3$. This mixture distribution can be written more compactly as

$$
f(\theta_i \mid \text{else}) = \frac{q_{0i}G_i(\mathrm{d}\theta_i) + \sum_{w\neq i, w=1}^{n_3} q_w\delta_{\theta_w}(\mathrm{d}\theta_i)}{q_{0i} + \sum_{w\neq i, w=1}^{n_3} q_w}
$$

where $G_i(\mathrm{d}\theta_i)$ is the bivariate normal-gamma pdf with components $\tau_i \sim \Gamma(\frac{s+1}{2}, \frac{1}{2}[S + \frac{1}{d+1}(\log\beta_{1i} - m)^2])$ and $\mu_i \sim N(\hat{\mu}_i, (1 + \frac{1}{d})\tau_i)$. These results are also given in Escobar and West (1995) for the case where the data, y_i, are independent and identically distributed draws from a DPM. Replacing y_i by $\log\beta_{1i}$ in their results yields the results shown above.

IV.2 Sticky clusters algorithm

DPM models can be slow to converge because the chain can enter states where many of the "data," or slopes in our case, are assigned to the same cluster and get "stuck" there. The reason for this stickiness has to do with the discreteness of the DP and the fact that the weight on any current parameter value is proportional to the number of "data" points currently associated with it. Specifically, recall that $\{\phi_j : j = 1, 2, ..., k\}$ for $k \leq n_3$ represents the distinct values among the components of $\theta^{(i)}$. Letting l_j = the number of components of $\theta^{(i)}$ that equal ϕ_j, we can rewrite the full conditional of (μ_i, τ_i) as

$$
\begin{aligned}
&f(\theta_i \mid \beta_{1i}, \theta^{(i)}) \propto \\
&C\tau_i^{\frac{s-1}{2}}\exp\{-\frac{\tau_i}{2}[S + \frac{1}{d+1}(\log\beta_{1i} - m)^2]\} \\
&\times \tau_i^{\frac{1}{2}}\exp\{-\frac{\tau_i}{2}(1 + \frac{1}{d})(\mu_i - \hat{\mu}_i)^2\} + \sum_{j=1}^{k} l_jq_j\delta_{\phi_j}(\mathrm{d}\theta_i)
\end{aligned}
$$

The probability weight associated with an existing value, $\phi_j = (\mu_j^*, \tau_j^*)$, is $l_j \times \sqrt{\frac{\tau_j^*}{2\pi}} \cdot \exp\{-\frac{\tau_j^*}{2}(\log\beta_{1i} - \mu_j^*)^2\} = l_jq_j$. So values of ϕ_j associated with large clusters and, hence, large l_j are likely to be picked for updated values of θ_i. In order for the cluster parameter, ϕ_j, to update to a new value, all members of that cluster must update to the new value individually via their full conditionals. If the new value has few observations associated with it, the full conditionals will not put much probability mass on it. Hence, the chain gets "stuck" when a large number of observations become associated with the same ϕ_j. Even if other values of ϕ_j have higher posterior probability, the chain must move through a low-probability intermediate state of shifting observations away from large clusters at less-probable values of ϕ_j.

The MacEachern and Müller algorithm, which we refer to as the "sticky clusters" algorithm because it is designed to alleviate the problem of "sticky clusters," augments the Gibbs sampler with the ϕ_j's. Since the θ_i's are still in the chain, none of the previous full conditionals change. In other words, conditioning any of the previous full conditionals on both $\{\theta_i : i = 1, 2, ..., n_3\}$ and $\{\phi_j : j = 1, 2, ..., k\}$ is the same as conditioning on the θ_i's since the θ_i's actually contain all the information of the ϕ_j's plus the association of each observation with a cluster. We only need to add a step that samples the full conditionals of the ϕ_j. In effect, this step will update a cluster's parameters using the data for all observations currently associated with that cluster. So the θ_i's will be sampled according to the full conditionals developed in Appendix IV.1. With high probability, there

will be duplicate values among the sampled θ_i's, giving rise to clusters of slopes associated with each distinct value, ϕ_j. We then update the ϕ_j's according to their full conditionals, which we derive next. Finally, all θ_i's associated with cluster j are updated with the new ϕ_j.

We use the following notation in deriving the full conditional of ϕ_j: $\Gamma_j = \{i : \theta_i = \phi_j\}$, l_j is the number of elements in Γ_j and $\overline{\log \beta_{1i}} = \frac{1}{l_j}\sum_{i \in \Gamma_j} \log \beta_{1i}$. Then

$$
\begin{aligned}
f(\phi_j &\mid \phi^{(j)}, \text{else}) \\
&\propto \left[\prod_{i \in \Gamma_j} f(\log \beta_{1i} \mid \phi_j)\right] \cdot dG_0(\phi_j) \\
&\propto \left[\prod_{i \in \Gamma_j} \sqrt{\tau_j^*}\exp\{-\frac{\tau_j^*}{2}(\log\beta_{1i} - \mu_j^*)^2\}\right] \\
&\quad \times \tau_j^{*\frac{s}{2}-1}\exp\{-\frac{S}{2}\tau_j^*\} \times \sqrt{\tau_j^*}\exp\{-\frac{\tau_j^*}{2d}(\mu_j^* - m)^2\} \\
&= \tau_j^{*\frac{l_j+s}{2}-1}\exp\{-\frac{\tau_j^*}{2}\Big[\sum_{i\in\Gamma_j}(\log\beta_{1i} - \overline{\log\beta}_{1i})^2 + S \\
&\quad + \frac{l_j}{l_jd+1}(\overline{\log\beta}_{1i} - m)^2\Big]\} \\
&\quad \times \sqrt{\tau_j^*}\exp\{-\frac{\tau_j^*}{2}(l_j+\frac{1}{d})(\mu_j^* - \hat{\mu}_j^*)^2\}
\end{aligned}
$$

The second- and third-to-last line of this equation as a function of τ_j^* is the kernel of a gamma distribution. The last line as a function of μ_j^* has the form of a normal distribution. Hence, we sample the full conditional of $\phi_j = (\mu_j^*, \tau_j^*)$ by first sampling $\tau_j^* \mid \text{else} \sim \Gamma(\frac{l_j+s}{2}, \frac{1}{2}\{\sum_{i\in\Gamma_j}(\log\beta_{1i} - \overline{\log\beta}_{1i})^2 + S + \frac{l_j}{l_jd+1}(\overline{\log\beta}_{1i} - m)^2\})$. Then, conditional on the τ_j^* just sampled, sample $\mu_i^* \mid \tau_j^*, \text{else} \sim N(\hat{\mu}_j^*, (l_j+\frac{1}{d})\tau_j^*)$ where $\hat{\mu}_j^* = \frac{d\,l_j\overline{\log\beta}_{1i}+m}{d\,l_j+1}$.

IV.3 Reversible jump MCMC

We use a reversible jump (RJ) step to sample the full conditional $[k_i, \chi^{(k_i)}|\text{else}]$. Recall that $\chi^{(k_i)}$ is the cow-specific parameter vector for cow i under model k_i and that $\chi^{(k_i)}$ changes dimension depending on the value of k_i. Specifically, for k_i equal to 1, 2 and 3, we now have that $\chi^{(k_i)}$ is β_{0i}, (β_{0i}, t_i^*) and $(\beta_{0i}, \beta_{1i}, t_i^*, \mu_i, \tau_i)$, respectively.

Moving between models 1 and 2

In order to construct a move from disease status 1 to 2, we must define a bijection between the two associated parameter spaces, β_{0i} and (β_{0i}, t_i^*). In order to match the dimensions of these parameter vectors, we augment the model 1 parameter "vector", β_{0i}, with an auxiliary variable, u, which is randomly generated from 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

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

which is just the identity map.

We now calculate the acceptance probability. If we let $[Y|X]$ denote the distribution of Y conditional on X, then 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

$$
\begin{aligned}
\alpha_{12} &= \frac{p(\tilde{\beta}_{0i}, \tilde{t}_i^*|k_i = 2, \text{ else})}{p(\beta_{0i}^{(t)}|k_i = 1, \text{ else})} \cdot \frac{\pi_{21}}{\pi_{12} \cdot \pi_2(u)} \cdot |J| \\
&= \frac{[S_i|\tilde{\beta}_{0i}, \tau_e, k_i = 2] \cdot [F_i|\tilde{t}_i^*, se_F, sp_F, k_i = 2]}{[S_i|\beta_{0i}^{(t)}, \tau_e, k_i = 1] \cdot [F_i|sp_F, k_i = 1]} \\
&\quad \times \frac{q_2\pi(\tilde{\beta}_{0i})\pi_2(\tilde{t}_i^*)}{q_1\pi(\beta_{0i}^{(t)})} \cdot \frac{\pi_{21}}{\pi_{12} \cdot \pi_2(u)} \cdot |J| \\
&= \frac{sp_F^{\sum_{j:t_{ij}<\tilde{t}_i^*}(1-F_{ij})}(1-sp_F)^{\sum_{j:t_{ij}<\tilde{t}_i^*}F_{ij}}}{sp_F^{\sum_j(1-F_{ij})}(1-sp_F)^{\sum_j F_{ij}}} \\
&\quad \times (se_F^{\sum_{j:t_{ij}\geq\tilde{t}_i^*}F_{ij}}(1-se_F)^{\sum_{j:t_{ij}\geq\tilde{t}_i^*}(1-F_{ij})}) \\
&\quad \times \frac{q_2}{q_1} \cdot \frac{\pi_{21}}{\pi_{12}} \cdot 1
\end{aligned}
$$

We obtain lines 2 and 3 by writing out each posterior in the first factor in line 1 as the likelihood times the prior. Since $\tilde{\beta}_{0i} = \beta_{0i}^{(t)}$ and $\tilde{t}_i^* = u$, several factors in the numerator and denominator cancel. 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 1 and 3

A move from model 1 to model 3 requires that we construct a bijection from β_{0i} to $(\beta_{0i}, \beta_{1i}, t_i^*, \mu_i, \tau_i)$. We generate four auxiliary variables u_1, u_2, u_3 and u_4 and define the bijection from the model 1 to the model 3 parameter space as follows:

$$
T_{13}(\beta_{0i}^{(t)}, u_1, u_2, u_3, u_4) = (\tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^*, \tilde{\mu}_i, \tilde{\tau}_i)
$$

where $\tilde{\beta}_{0i} = \beta_{0i}^{(t)}$, $\tilde{t}_i^* = u_1$, $\tilde{\mu}_i = u_2$, $\tilde{\tau}_i = u_3$, and $\tilde{\beta}_{1i} = e^{u_4}$. We generate u_1 using a pdf which is proportional to the fecal culture portion of the likelihood under model 3. Thus, the proposal pdf for u_1 is: $\psi(u_1) \propto sp_F^{\sum_{j:t_{ij}<u_1}(1-F_{ij})}(1 - sp_F)^{\sum_{j:t_{ij}<u_1}F_{ij}} se_F^{\sum_{j:t_{ij}\geq u_1}F_{ij}}(1 - se_F)^{\sum_{j:t_{ij}\geq u_1}(1-F_{ij})} \cdot I(\text{dob}_i < u_1 < t_{im_i} - \text{lag})$. Additionally, (u_2, u_3), the proposal for $(\tilde{\mu}_i, \tilde{\tau}_i)$, is generated from a distribution which resembles the marginalized Dirichlet Process mixture model, conditional on the n_3 current values of $\{(\mu_{i_r}, \tau_{i_r}) : r = 1, 2, ..., n_3\}$ and has the following distribution

$$(9) \quad \psi^*(du_2, du_3 \mid (\mu_{i_r}, \tau_{i_r}), r = 1, 2, ..., n_3) = \frac{\alpha^* G^*(du_2, du_3) + \sum_{r=1}^{n_3} \delta_{(\mu_{i_r}, \tau_{i_r})}(du_2, du_3)}{\alpha^* + n_3}$$

Under $G^*(\cdot)$, we have $\tau \sim \Gamma(\alpha^*, \beta^*)$ and, independently, $\mu \sim N(\mu^*, \tau_{rj}^*)$. Note that the marginalized Dirichlet process mixture model, denoted $\psi(\cdot)$, assumes the same form as equation 9 except that $\alpha = \alpha^*$ and $G^* = G_0$ where, under $G_0(\cdot)$, we have $\tau_i \sim \Gamma(\frac{s}{2}, \frac{S}{2})$ and $\mu_i \mid \tau_i \sim N(m, \frac{\tau_i}{d})$. Using G_0 to generate proposals often results in poor proposals since the proposed precision may be quite small if a vague proposal such as $\Gamma(0.01, 0.01)$ is used. Since the proposed value of μ depends on a very small precision, very large μ's may be generated resulting in numerical overflow.

Finally, u_4, the auxiliary variable related to the slope proposal by $e^{u_4} = \tilde{\beta}_{1i}$, is distributed $N(\tilde{\mu}_i, \tilde{\tau}_i)$ where $(\tilde{\mu}_i, \tilde{\tau}_i) = (u_2, u_3)$.

Letting ϕ represent the pdf of the standard normal distribution, the proposed move from disease state 1 to state 3 is accepted with probability $\min(1, \alpha_{13})$ where

α_{13}

$$= \frac{p(\tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^*, \tilde{\mu}_i, \tilde{\tau}_i \mid k_i = 3, \text{ else})}{p(\beta_{0i}^{(t)} \mid k_i = 1, \text{ else})}$$
$$\times \frac{\pi_{31}}{\pi_{13} \cdot \psi(u_1) \cdot \psi^*(u_2, u_3 \mid (\mu_{i_r}, \tau_{i_r}), r = 1, 2, ..., n_3)}$$
$$\times \frac{1}{\sqrt{u_3}\phi(\sqrt{u_3}(u_4 - u_2))} \cdot |J|$$
$$= \frac{[S_i|\tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^*, \tau_e, k_i = 3] \cdot [F_i|\tilde{t}_i^*, se_F, sp_F, k_i = 3]}{[S_i|\beta_{0i}^{(t)}, \tau_e, k_i = 1] \cdot [F_i|sp_F, k_i = 1]}$$
$$\times \frac{q_3 \pi(\tilde{\beta}_{0i}) \cdot \psi(\tilde{\mu}_i, \tilde{\tau}_i \mid (\mu_{i_r}, \tau_{i_r}), r = 1, 2, ..., n_3)\cdot}{q_1 \pi(\beta_{0i}^{(t)})}$$
$$\times \frac{1}{\tilde{\beta}_{1i}}\sqrt{\tilde{\tau}_i}\phi(\sqrt{\tilde{\tau}_i}(\log \tilde{\beta}_{1i} - \tilde{\mu}_i)) \cdot \pi_3(\tilde{t}_i^*)$$
$$\times \frac{\pi_{31}}{\pi_{13} \cdot \psi(u_1) \cdot \psi^*(u_2, u_3 \mid (\mu_{i_r}, \tau_{i_r}), r = 1, 2, ..., n_3)}$$
$$\times \frac{1}{\sqrt{u_3}\phi(\sqrt{u_3}(u_4 - u_2))} \cdot |e^{u_4}|$$

$$= \frac{[S_i|\tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^*, \tau_e, k_i = 3] \cdot [F_i|\tilde{t}_i^*, se_F, sp_F, k_i = 3]}{[S_i|\beta_{0i}^{(t)}, \tau_e, k_i = 1] \cdot [F_i|sp_F, k_i = 1]}$$
$$\times \frac{q_3 \cdot \pi_3(u_1) \cdot \pi_{31}\psi(u_2, u_3 \mid (\mu_{i_r}, \tau_{i_r}), r = 1, ..., n_3)}{q_1 \cdot \psi(u_1) \cdot \pi_{13} \cdot \psi^*(u_2, u_3 \mid (\mu_{i_r}, \tau_{i_r}), r = 1, 2, ..., n_3)}$$

We obtain the last line by substitution using the following equalities $\tilde{\beta}_{0i} = \beta_{0i}^{(t)}$, $\tilde{t}_i^* = u_1$, $\tilde{\mu}_i = u_2$, $\tilde{\tau}_i = u_3$, and $\tilde{\beta}_{1i} = e^{u_4}$ and canceling like factors.

Moving between models 2 and 3

A move from model 2 to model 3 now requires a bijection between the model 2 and model 3 parameter spaces. The model 2 parameters are (β_{0i}, t_i^*), and the model 3 parameters are $(\beta_{0i}, \beta_{1i}, t_i^*, \mu_i, \tau_i)$. Since the support of t_i^* is different under these two models, we do not map the previous model 2 value of t_i^* to the proposed model 3 value of t_i^*. Instead, we map the t_i^* from model 2 to a new auxiliary variable, v, in the state 3 augmented parameter space.

Borrowing from the auxiliary variables constructed for moves from state 1 to state 3, we formally define the bijection for a move from model 2 to 3 as follows:

$$T_{23}(\beta_{0i}^{(t)}, t_i^{*(t)}, u_1, u_2, u_3, u_4) = (\tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^*, v, \tilde{\mu}_i, \tilde{\tau}_i)$$
$$\text{where } \tilde{\beta}_{0i} = \beta_{0i}^{(t)}, \quad \tilde{\beta}_{1i} = e^{u_4}, \quad \tilde{t}_i^* = u_1, \quad v = t_i^{*(t)}$$
$$\text{and } (\tilde{\mu}_i, \tilde{\tau}_i) = (u_2, u_3)$$

The generation of all auxiliary variables except v is as defined in the last section for moves between models 1 and 3. v is generated from the prior, $\pi_2(\cdot)$ for t_i^* under model 2 as shown in Equation 2.

We accept the move to model 3 along with the proposed model 3 parameters with probability $\min(1, \alpha_{23})$. We now calculate α_{23}.

α_{23}

$$= \frac{p(\tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^*, \tilde{\mu}_i, \tilde{\tau}_i \mid k_i = 3, \text{ else})}{p(\beta_{0i}^{(t)}, t_i^{*(t)} \mid k_i = 2, \text{ else})}$$
$$\times \frac{\pi_{32} \cdot \pi_2(v)}{\pi_{23} \cdot \psi(u_1) \cdot \psi^*(u_2, u_3 \mid (\mu_{i_r}, \tau_{i_r}), r = 1, 2, ..., n_3)}$$
$$\times \frac{1}{\sqrt{u_3}\phi(\sqrt{u_3}(u_4 - u_2))} \cdot |J|$$
$$= \frac{[S_i|\tilde{\beta}_{0i}, \tilde{\beta}_{1i}, \tilde{t}_i^*, \tau_e, k_i = 3] \cdot [F_i|\tilde{t}_i^*, se_F, sp_F, k_i = 3]}{[S_i|\beta_{0i}^{(t)}, \tau_e, k_i = 2] \cdot [F_i|t_i^{*(t)}, se_F, sp_F, k_i = 2]}$$
$$\times \frac{q_3 \pi(\tilde{\beta}_{0i}) \cdot \psi(\tilde{\mu}_i, \tilde{\tau}_i \mid (\mu_{i_r}, \tau_{i_r}), r = 1, 2, ..., n_3)\cdot}{q_2 \pi(\beta_{0i}^{(t)})\pi_2(t_i^{*(t)})}$$
$$\times \frac{1}{\tilde{\beta}_{1i}}\sqrt{\tilde{\tau}_i}\phi(\sqrt{\tilde{\tau}_i}(\log \tilde{\beta}_{1i} - \tilde{\mu}_i)) \cdot \pi_3(\tilde{t}_i^*)$$
$$\times \frac{\pi_{32} \cdot \pi_2(v)}{\pi_{23} \cdot \psi(u_1) \cdot \psi^*(u_2, u_3 \mid (\mu_{i_r}, \tau_{i_r}), r = 1, 2, ..., n_3)}$$
$$\times \frac{1}{\sqrt{u_3}\phi(\sqrt{u_3}(u_4 - u_2))} \cdot |J|$$

$$
\begin{aligned}
= \ & \frac{[S_i|\beta_{0i}^{(t)}, e^{u_4}, u_1, \tau_e, k_i = 3] \cdot [F_i|u_1, se_F, sp_F, k_i = 3]}{[S_i|\beta_{0i}^{(t)}, \tau_e, k_i = 2] \cdot [F_i|t_i^{*(t)}, se_F, sp_F, k_i = 2]} \\
& \times \frac{q_3 \cdot \pi_3(u_1)}{q_2} \cdot \frac{\pi_{32}}{\pi_{23} \cdot \psi(u_1)} \\
& \times \frac{\psi(u_2, u_3 \mid (\mu_{i_r}, \tau_{i_r}), r = 1, 2, ..., n_3)}{\psi^*(u_2, u_3 \mid (\mu_{i_r}, \tau_{i_r}), r = 1, 2, ..., n_3)}
\end{aligned}
$$

It is now a simple matter to define the proposal for a move from model 3 to model 2 as T_{23}^{-1}; additionally, this proposal is accepted with probability $\min(1, \alpha_{32})$ where $\alpha_{32} = \alpha_{23}^{-1}$.

REFERENCES

ANTONIAK, C. E. (1974) Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *The Annals of Statistics* **2**, 1152–1174. MR0365969

BUSH, C. A. and MACEACHERN, S. N. (1996) A Semi-parametric Bayesian Model for Randomised Block Designs. *Biometrika* **83**, 275–285.

COLLINS, M. T., WELLS, S. J., PETRINI, K. R., COLLINS, J. E., SCHULTZ, R. D. and WHITLOCK, R. H. (2005). Evaluation of Five Antibody Detection Tests for Diagnosis of Bovine Paratuberculosis. *Clinical and Diagnostic Laboratory Immunology* **12**, 685–692.

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

DEIORIO, M., JOHNSON, W., MÜLLER, P. and ROSNER, G. (2009) Bayesian Nonparametric Non-Proportional Hazards Survival Modeling. *Biometrics* **65**, 762–771. MR2649849

ESCOBAR, M. D. and WEST, M. (1995) Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association* **90**, 577–588. MR1340510

FERGUSON, T. S. (1973) A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics* **1**, 209–230. MR0350949

GREEN, P. J. (1995) Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* **82**, 711–732. MR1380810

HANSON, T. E., JOHNSON, W. O. and GARDNER, I. A. (2000) Log-linear and Logistic Modeling of Dependence Among Serologic Tests for Paratuberculosis, Toxoplasmosis and Brucellosis. *Preventive Veterinary Medicine* **45**, 123–137.

HANSON, T., BRANSCUM, A. and JOHNSON, W. O. (2005) Bayesian Nonparametric Modeling and Data Analysis: An Introduction. In Bayesian Thinking, Modeling and Computation (*Handbook of Statistics*, volume 25), 245–278. Edited by D. K. Day and C. R. Rao. Amsterdam: Elsevier. MR2490529

JAIN, S. and NEAL, R. M. (2004) A Split-merge Markov Chain Monte Carlo Procedure for the Dirichlet Process Mixture Model. *Journal of Computational and Graphical Statistics*, **13**, 158–182. MR2044876

JONES, G., JOHNSON, W. O., VINK, D. and FRENCH, N. (2012). A Framework for the Joint Modeling of Longitudinal Diagnostic Outcome Data and Latent Infection Status: Application to Investigating the Temporal Relationship Between Infection and Disease. *Biometrics* **68**, 371–9. MR2959603

KLEINMAN, K. P. and IBRAHIM, J. G. (1998) A Semi-parametric Bayesian Approach to the Random Effects Model. *Biometrics* **54**, 921–938.

LI, Y., LIN, X. and MÜLLER, P. (2010) Bayesian Inference in Semiparametric Mixed Models for Longitudinal Data. *Biometrics* **66**, 70–78. MR2756692

LOMBARD, J. E., GARDNER, I. A., JAFARZADEH, S. R., FOSSLER, C. P., HARRIS, B., CAPSEL, R. T., WAGNER, B. A. and JOHNSON, W. O. (2013) Herd-level Prevalence of Mycobacterium Avium Subsp. Paratuberculosis Infection in United States Dairy Herds in 2007. *Preventive Veterinary Medicine* **108(2–3)**, 234–238.

MACEACHERN, S. N. and MÜLLER, P. (1998) Estimating Mixture of Dirichlet Process Models. *Journal of Computational and Graphical Statistics* **7**, 223–238.

MADGER, L. S. and ZEGER, S. L. (1996) A Smooth Nonparametric Estimate of a Mixing Distribution Using Mixtures of Gaussians. *Journal of the American Statistical Association* **91**, 1141–1151. MR1424614

NEAL, R. M. (2000) Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics* **9**, 249–265. MR1823804

NORRIS, M., JOHNSON, W. O. and GARDNER, I. A. (2009) Modeling Bivariate Longitudinal Diagnostic Outcome Data in the Absence of a Gold Standard. *Statistics and Its Interface* **2**, 171–185. MR2516068

OTT, S. J., WELLS, S. J. and WAGNER, B. A. (1999). Herd Level Economic Losses Associated with Johne's Disease on US Dairy Operations. *Preventive Veterinary Medicine* **40**, 179–192.

PAULER, D. K. and LAIRD, N. M. (2000) A Mixture Model for Longitudinal Data with Application to Assessment of Noncompliance. *Biometrics* **56**, 464–472.

PEPE, M. (2003). The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press. MR2260483

R DEVELOPMENT CORE TEAM (2011). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

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

TAO, H., PALTA, M., YANDELL, B. S. and NEWTON, M. A. (1999) An Estimation Method for the Semi-parametric Mixed Effects Model. *Biometrics* **55**, 102–110. MR1705675

VERBEKE, G. and LESAFFRE, E. (1996) A Linear Mixed-effects Model with Heterogeneity in the Random-effects Population. *Journal of the American Statistical Association* **91**, 217–221.

VERBEKE, G. and LESAFFRE, E. (1997) The Effect of Misspecifying the Random Effects Distribution in Linear Mixed Models for Longitudinal Data. *Computational Statistics and Data Analysis* **23**, 541–556. MR1437679

WANG, C., TURNBULL, B. W., NIELSEN, S. S. and GRÖHN, Y. T. (2011) Bayesian Analysis of Longitudinal Johne's Disease Diagnostic Data Without a Gold Standard. *Journal of Dairy Science* **94**, 2320–2328.

WEST, M., MÜLLER, P. and ESCOBAR, M. D. (1994) Hierarchical Priors and Mixture Models, with Applications in Regression and Density Estimation. In *Aspects of Uncertainty: A Tribute to D. V. Lindley*, A. F. M. Smith and P. R. Freeman (eds). London: Wiley. MR1309702

ZHANG, D. and DAVIDIAN, M. (2001) Linear Mixed Models with Flexible Distributions of Random Effects for Longitudinal Data. *Biometrics* **57**, 795–802. MR1859815

Michelle Norris
Department of Mathematics and Statistics
California State University, Sacramento
6000 J Street
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 Health Management
Atlantic Veterinary College
University of Prince Edward Island
Charlottetown, Prince Edward Island
Canada
C1A 4P3
E-mail address: iagardner@upei.ca