10-2015

Disease Diagnosis From Immunoassays With Plate to Plate Variability: A Hierarchical Bayesian Approach

Oliver Entine
University of Pennsylvania

Dylan S. Small
University of Pennsylvania

Shane T. Jensen
University of Pennsylvania

Gerardo Sanchez Garcia

Milagros Bastos Mazuelos

See next page for additional authors

Disease Diagnosis From Immunoassays With Plate to Plate Variability: A Hierarchical Bayesian Approach

Abstract

The standard methods of diagnosing disease based on antibody microtiter plates are quite crude. Few methods create a rigorous underlying model for the antibody levels of populations consisting of a mixture of positive and negative subjects, and fewer make full use of the entirety of the available data for diagnoses. In this paper, we propose a Bayesian hierarchical model that provides a systematic way of pooling data across different plates, and accounts for the subtle sources of variations that occur in the optical densities of typical microtiter data. In addition to our Bayesian method having good frequentist properties, we find that our method outperforms one of the standard crude approaches (the "3SD Rule") under reasonable assumptions, and provides more accurate disease diagnoses in terms of both sensitivity and specificity.

Disciplines
Biostatistics

Author(s)
Oliver Entine, Dylan S. Small, Shane T. Jensen, Gerardo Sanchez Garcia, Milagros Bastos Mazuelos, Manuela R. Verastegui Pimentel, and Michael Z. Levy

Disease Diagnosis from Immunoassays with Plate to Plate Variability: A Hierarchical Bayesian Approach

Oliver Entine   Dylan S. Small   Shane T. Jensen

Gerardo Sánchez García   Milagros Bastos Mazuelos

Manuela Verastegui Pimentel   Michael Z. Levy

December 17, 2013

Abstract

The standard methods of diagnosing disease based on antibody microtiter plates are quite crude. Few methods create a rigorous underlying model for the antibody levels of populations consisting of a mixture of positive and negative subjects, and fewer make full use of the entirety of the available data for diagnoses. In this paper, we propose a Bayesian hierarchical model that provides a systematic way of pooling data across different plates, and accounts for the subtle sources of variations that occur in the optical densities of typical microtiter data. In addition to our Bayesian method having good frequentist properties, we find that our method outperforms one of the standard crude approaches (the "3SD Rule") under reasonable assumptions, and provides more accurate disease diagnoses in terms of both sensitivity and specificity.

1   Introduction

The diagnosis of a large number of infectious diseases relies on the detection of antibodies in the sera or saliva of patients. The presence of a high amount of antibodies specific

1

to a disease agent is a strong indicator of the presence of the disease, as these antibodies are evidence that the body is trying to fight off an infection, or has fought off an infection in the recent past. Though these antibody levels are hard to measure directly, one can measure the optical density of a sample through enzyme linked immunosorbent assays (ELISA). Comparing these optical densities with those calibrated to antibody concentrations of known samples (i.e. controls) allows one to estimate the antibody level of those unknown samples.

ELISAs are typically run on a series of microtiter plates, each containing 96 wells. On each plate, the majority of wells contain samples from the population of interest, some of whom may be infected with the disease and others who are not. A portion of the wells on each plate are filled with negative controls (samples from patients known to be negative), while another portion of the wells contains samples from positive controls.

When analyzing data from microtiter plates, the standard procedures used to classify samples as positive or negative are surprisingly crude. One such example is what we informally call "The 3SD Rule." For each microtiter plate, the negative control samples are used to compute an empirical average and standard deviation that is taken to represent the population of uninfected individuals. For each plate, a cutoff value is assigned as three SDs above the mean of these negative controls. All of the unknown samples on that plate are classified as positive if their measurements exceed this cutoff, and are diagnosed as negative otherwise (Irion et al., 2002). This is an example of an *unpooled procedure*, as each plate has its own cutoff, determined only by measurements that are on that specific plate, and makes no use of the data from any of the other plates in the dataset. Even more sophisticated methods, which utilize serial dilution curves to better capture nonlinear relationships between optical densities and antibody concentrations, are unpooled procedures as each plate's dilution curve is derived only from measurements on that plate (Higgins et al., 1998; Gelman et al., 2004).

A more sophisticated method for diagnosis is by directly modeling the antibody levels of a population through a two-component mixture model, where each component corresponds to the negative and positive sub-populations, with an unknown fraction of people belonging to each group. Two-component mixture models have been used in previous

bioassay analyses (Moulton et al., 2002).

Our goal in this paper is to make more accurate diagnoses by making use of the data across all the plates simultaneously. If all of the plates used for diagnostics were identical, one could fit a single two-component mixture model across all wells and plates. However, even under highly controlled conditions, the plates are likely to have small, but noticeable differences that systematically skew the antibody levels of positive and negative subjects upward or downward. A second option would be to fit a separate mixture model for each plate. Such analysis would have limited power due to a relatively small sample size on each plate. Rather, we propose a compromise between these two extremes. Namely, we develop a hierarchical Bayesian approach, in which we consider the effects of the plates to come from a common distribution. This allows us to have distinct models for each plate, but also lets us "share" information across the plates regarding expected variation of within-plate antibody levels.

2   An Example of Microtiter Plate Data

To motivate the incorporation of plate effects into our model, we first look at a series of microtiter plates from a lab in Peru. Sera tested on these plates were previously found to react relatively weakly to a commercially-available ELISA test for Chagas disease (Verani et al., 2009; Levy et al., 2009). Researchers are therefore developing an ELISA test which derives antigens from a local strain of T. cruzi. In our exploratory data analysis, we denote the optical measurements of the Chagas data by Y_{ij}, where i indexes plates, and j indexes wells within plates. We focus for now on differences in the logarithm of Y_{ij} across plates i. In the first set of boxplots (Figure 1, top), we plot $log(Y_{ij})$ for all of our non-control subjects. There are substantial differences in the average response across the 18 plates. However, this is not necessarily due to plate effects because some plates will likely have more positive subjects than others (even though we do not have definitive positive/negative diagnostic results for these subjects), and the difference between density levels of these two groups is expected to be large (even on the log scale). However, when we look at the plot of the optical levels for the same 7 negative controls on each

3

Log Measurements of Unknowns across Plates

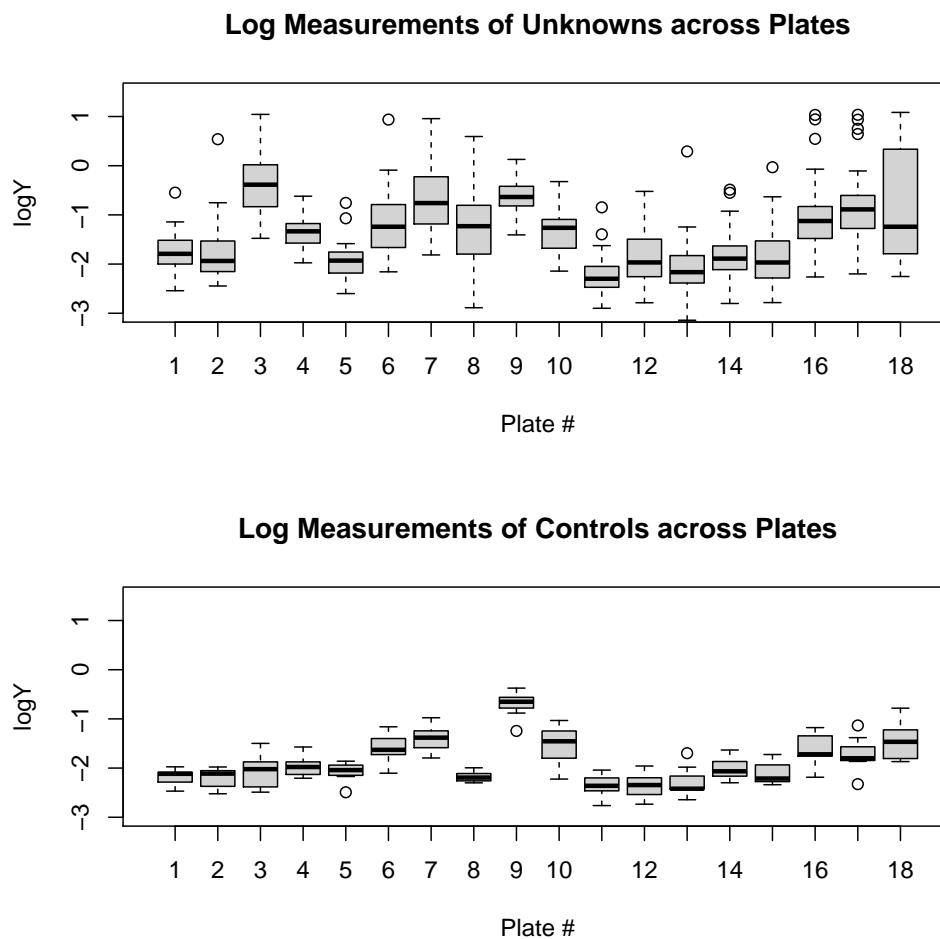**Log Measurements of Controls across Plates**

Figure 1: Boxplots of Logarithms of Optical Densities for Patients (above) and Negative Controls (below).

plate, the disparity in the readings from plate-to-plate is also stark. The bottom plot of Figure 1 suggests that the plate effects are substantial.

3 The Model

We consider a model in which the optical densities of the population come from a mixture of log-normal distributions (one component for people infected with the disease, one component for those not infected), but scaled by a factor corresponding to the contami-

nation of a particular plate. Again, as introduced before, let Y_{ij} be the optical density for an observation in the i-th plate and j-th well. β_i is the multiplicative effect from plate i upon the optical densities in its corresponding wells. α is the mixture plate parameter that indicates the fraction of subjects in the population who are infected. We augment our data with the latent variables I_{ij} where $I_{ij} = 1$ if Y_{ij} comes from an infected subject, and 0 otherwise. That is, I_{ij} comes from a Bernoulli with parameter α.

Our model for the unknown data is $Y_{ij} = \beta_i(X_{ij}^{neg})^{1-I_{ij}}(X_{ij}^{pos})^{I_{ij}}$, where

$$log(X_{ij}^{neg}) \sim N(\mu_{neg}, \sigma_{neg}^2)$$

$$log(X_{ij}^{pos}) \sim N(\mu_{pos}, \sigma_{pos}^2)$$

On the log scale, our model can be reexpressed as follows:

$$log(Y_{ij}) = log(\beta_i) + (1 - I_{ij})(log(X_{ij}^{neg})) + I_{ij}(log(X_{ij}^{pos}))$$

If in conjunction to these unlabeled observations, we include some labeled data, namely the optical densities of a predetermined number of negative control wells, one can in theory obtain much better estimates for the parameters of this mixture model. In the particular dataset we will examine in Section 5, the negative control subjects are simply a team of scientists in the lab who are assumed to be free of Chagas disease. Each plate has seven wells, each corresponding to one of these negative controls. Thus we observe variability both between controls, as indicated by the variation across these seven optical densities on a given plate, and also variability within the control subjects, measured by the variation of readings of each individual across the plates. To fully reflect the nature of these replications, our model should be able to capture both of these forms of variations. We let Y_{ij}^c be the optical density for the i-th control subject on the j-th plate (or the j-th replication).

For the negative control data, we have $Y_{ij}^c = \beta_i(X_{ij}^c)$, and to distinguish the two types of variation, we use the following construction.

$$log(X_{ij}^c) \sim N(\tau_j, \kappa^2)$$

$$\tau_j \sim N(\mu^{neg}, \lambda^2)$$

5

Here, κ^2 represents the plate-to-plate variation of a control's replicated samples, and λ^2 captures the variation between labeled control subjects. The advantage of this formulation is that the unconditional distribution of $log(X^c_{ij})$ is itself normal, with mean μ_{neg} and variance $\kappa^2 + \lambda^2$, where we naturally set $\sigma^2_{neg} = \kappa^2 + \lambda^2$.

Next, since we would like the plate effects to be strictly positive and centered around 1 (corresponding to "neutral" plate effect), we let β_i come from a lognormal distribution, and $log(\beta_i) \sim N(-\frac{\nu^2}{2}, \nu^2)$. (Since the mean of a log-normal distribution is given by $e^{\mu + \frac{\sigma^2}{2}}$, this parametrization ensures that the mean of β_i is 1). Another reasonable approach would be to have $log(\beta_i) \sim N(0, \nu^2)$, thereby forcing the *median* of β_i to be 1. However, we prefer our current approach as we find it easier to interpret on scales of means rather than medians.

Finally, we impose the following uninformative prior distributions on our parameters:

$$\mu_{neg}, \mu_{pos} \sim Unif(-\infty, \infty)$$

$$\sigma^2_{pos} \sim Gamma(0, 0)$$

$$\lambda^2, \kappa^2 \sim Unif(0, \infty)$$

$$\alpha \sim Unif(0, 1)$$

$$\frac{1}{\nu^2} \sim Gamma(5, 5)$$

.

To approximate the posterior distribution our unknown parameters, we would like to simulate draws from the posterior distribution

$$P(\mu_{neg}, \mu_{pos}, \sigma^2_{neg}, \sigma^2_{pos}, \alpha, \beta, \nu^2 | \mathbf{Y}).$$

To simplify our notation, let us define $y_{ij} = log(Y_{ij})$, $y^c_{ij} = log(Y^c_{ij})$ $b_i = log(\beta_i)$, and let $\theta = (\mu_{neg}, \mu_{pos}, \sigma^2_{pos}, \kappa^2, \lambda^2)$, so that θ represents the set of parameters with improper flat prior distributions. Taking advantage of the conditional independence of $\mathbf{y}$ and $\mathbf{y^c}$, we write out the joint posterior as:

$$P(\mathbf{I}, \vec{b}, \vec{\tau}, \nu^2, \alpha, \theta | \mathbf{Y}) \propto$$

$$P(\mathbf{y}|\mathbf{I},\vec{b},\theta)P(\mathbf{y^c}|\mathbf{I},\vec{b},\vec{\tau},\theta)P(\mathbf{I}|\alpha)P(\vec{b}|\nu^2)P(\vec{\tau}|\theta)P(\nu^2)P(\alpha,\theta)$$

The components of this posterior distribution are:

$$P(\mathbf{y}|\mathbf{I},\mathbf{b},\theta) \propto \prod_{i,j}\left(\frac{1}{\sigma_{neg}}e^{\frac{-(y_{ij}-b_i-\mu_{neg})^2}{2\sigma_{neg}^2}}\right)^{1-I_{ij}}\left(\frac{1}{\sigma_{pos}}e^{\frac{-(y_{ij}-b_i-\mu_{pos})^2}{2\sigma_{pos}^2}}\right)^{I_{ij}}$$

$$P(\mathbf{y^c}|\mathbf{I},\mathbf{b},\vec{\tau},\theta) \propto \prod_{i,j}\left(\frac{1}{\kappa}e^{\frac{-(y_{ij}^c-b_i-\tau_j)^2}{2\kappa^2}}\right)$$

$$P(\mathbf{I}|\alpha) \propto \alpha^{\sum_{i,j}I_{ij}}(1-\alpha)^{\sum_{i,j}(1-I_{ij})}$$

$$P(\mathbf{b}|\nu^2) \propto \prod_{i}\frac{1}{\nu}e^{\frac{-(b_i+\frac{\nu^2}{2})^2}{2\nu^2}}$$

$$P(\vec{\tau}|\theta) \propto \prod_{j}\frac{1}{\lambda}e^{\frac{-(\tau_j-\mu_{neg})^2}{2\lambda^2}}$$

$$P(\nu^2) \propto \frac{1}{(\nu)^6}e^{\frac{-5}{\nu^2}}$$

$$P(\alpha,\theta) \propto 1$$

where $\mathbf{Y}=(\mathbf{y},\mathbf{y^c})$. We obtain samples from the posterior distribution via a Markov Chain Monte Carlo (MCMC) scheme, consisting of Gibbs Sampling and Metropolis-Hastings steps (Hastings, 1970; Gelfand and Smith, 1990; Casella and George, 1992).

For each iteration of our MCMC implementation, we obtain a vector of parameters sampled from the joint posterior distribution. Combined with the optical density readings, we can calculate the posterior probability that a particular sample comes from the infected group. We then average these posterior probabilities (on a well-by-well basis) over all the iterations within that chain. Our classification rule is based on cutoff values for these posterior probabilities. If a sample's average posterior probability is less than the predetermined cutoff, we classify the sample as negative. If the probability of being positive exceeds the cutoff, we classify that sample as positive. We then compute usual measure of sensitivity and specificity of our procedure by determining how many of our classified negatives/positives match up with the actual negative/positive state of the sample (since the data is simulated, we definitively know the true diagnosis of each

subject). A higher probability cutoff will naturally lead to fewer true positives (and hence a lower sensitivity), and also lead to more true negatives (and thus higher specificity).

To choose this probability cutoff, a practitioner would choose a cutoff that best reflected their utility with respect to the balance between false positives and false negatives. A higher cut-off would reduce false positives (increased specificity) at the expense of more false negatives (decreased sensitivity). A lower cut-off would give preference to sensitivity over specificity. Reassuringly, our simulation results reviewed below suggest that our method has generally improved sensitivity and specificity over the 3SD rule regardless of the cutoff that is used.

4 Details of MCMC Implementation

Due to the conjugacy of our prior distributions with the joint-likelihood function, many of conditional posterior distributions come from standard distributions, and we can sample them directly.

Sampling the Well Indicators

The positive/negative labels for each well come from Bernoulli distributions.

$$I_{ij}|y_{ij}, b_i, \nu^2, \alpha, \theta \sim Bin(1, p_{ij})$$

$$p_{ij} = \frac{\alpha \phi(z_{ij}^{pos})}{\alpha \phi(z_{ij}^{pos}) + (1 - \alpha)\phi(z_{ij}^{neg})}$$

where ϕ is the pdf for a standard normal, $z_{ij}^{pos} = \frac{y_{ij} - b_i - \mu_{pos}}{\sigma_{pos}}$, and $z_{ij}^{neg} = \frac{y_{ij} - b_i - \mu_{neg}}{\sigma_{neg}}$.

Sampling the Normal-Normal Mixture Parameters

Letting n_C be the number of controls on each plate, $n_0 = \sum(1 - I_{ij})$ and $n_1 = \sum I_{ij}$, the conditional posterior distribution for the disease prevalence α is a beta distribution:

$$\alpha|\mathbf{y}, \mathbf{I}, \vec{b}, \nu^2, \theta \sim Beta(1 + n_1, 1 + n_0)$$

8

Letting $d_{ij} = y_{ij} - b_i$, the conditional densities of the two means of the mixture model are the following normal distributions.

$$\mu_{neg}|\mathbf{y}, \mathbf{I}, \vec{b}, \vec{\tau}, \nu^2, \alpha, \mu_{pos}, \sigma_{pos}^2, \kappa^2, \lambda^2 \sim N\left(\frac{\frac{\sum d_{ij}\mathbf{1}_{\{I_{ij}=0\}}}{\sigma_{neg}^2} + \frac{\sum \tau_j}{\lambda^2}}{\frac{n_0}{\sigma_{neg}^2} + \frac{n_C}{\lambda^2}}, \frac{1}{\frac{n_0}{\sigma_{neg}^2} + \frac{n_C}{\lambda^2}}\right)$$

$$\mu_{pos}|\mathbf{y}, \mathbf{I}, \vec{b}, \vec{\tau}, \nu^2, \alpha, \mu_{neg}, \sigma_{pos}^2, \kappa^2, \lambda^2 \sim N(\frac{\sum d_{ij}\mathbf{1}_{\{I_{ij}=1\}}}{n_1}, \frac{\sigma_{pos}^2}{n_1})$$

Since the inverse-gamma prior distribution for σ_{pos}^2 is conjugate to the normal likelihood, its posterior distribution is also inverse-gamma. Letting $SS_{pos} = \sum(z_{ij}^{pos} - \mu_{pos})^2$, the conditional posterior distribution of the variance of the positive group is given by:

$$\sigma_{pos}^2|\mathbf{y}, \mathbf{I}, \vec{b}, \vec{\tau}, \nu^2, \alpha, \mu_{neg}, \mu_{pos}, \kappa^2, \lambda^2 \sim InvGamma(\frac{n_1}{2}, \frac{SS_{pos}}{2})$$

Sampling the Plate Effect Parameters

Next, we sample the (log) plate effect parameters, b. If we let $z_{ij}^{\beta} = y_{ij} - \mu_{ij}$, then $z_{ij}^{\beta} \sim N(b_i, \sigma_{ij}^2)$ where μ_{ij} and σ_{ij}^2 correspond to the positive/negative assignment of observation y_{ij}, y_{ij}^c. This implies that the corresponding conditional posterior distribution for b_i is:

$$b_i|y_{ij}, I_{ij}, \nu^2, \alpha, \theta \sim N(m_i, s_i^2) \text{ where}$$

$$m_i = \frac{\sum_j \frac{z_{ij}^{\beta}}{\sigma_{ij}^2} + \frac{1}{2}}{\sum_j \frac{1}{\sigma_{ij}^2} + \frac{1}{\tau^2}}$$

$$s_i^2 = \frac{1}{\sum_j \frac{1}{\sigma_{ij}^2} + \frac{1}{\tau^2}}$$

Sampling the variance ν^2 of the log of the plate effects b is trickier. Since the distribution of b is constrained so that β has mean 1, the modeled distribution of b_i has the form $b_i \sim N(-\frac{\nu^2}{2}, \nu^2)$, which is no longer conjugate to the $InvGamma(5, 5)$ prior distribution imposed on ν^2. Thus, the posterior conditional distribution $\nu^2|y_{ij}, I_{ij}, b_i, \alpha, \theta$ is sampled indirectly using a Metropolis-Hastings step.

Sampling the Means of the Negative Controls, $\vec{\tau}$

Letting n_P be the total number of plates (and hence number of replications for each negative control subject), and $d_{ij}^c = y_{ij}^c - b_i$, the posterior conditional distributions for each τ_j are normal.

$$\tau_j | \mathbf{y}, \mathbf{I}, \vec{b}, \nu^2, \alpha, \theta \sim N \left(\frac{\frac{\mu_{neg}}{\lambda^2} + \frac{\sum_i d_{ij}^c}{\kappa^2}}{\frac{1}{\lambda^2} + \frac{n_P}{\kappa^2}}, \frac{1}{\frac{1}{\lambda^2} + \frac{n_P}{\kappa^2}} \right)$$

Sampling the Two Variance Components of σ_{neg}^2

Finally, we need to sample the two components of σ_{neg}^2, κ^2 and λ^2. The conditional posterior distribution for the negative control parameters (conditioned on μ_{neg} and b), where n_C is the number of control subjects, and n_P is number of plates, is now:

$$P(\tau, \kappa^2, \lambda^2 | \mathbf{y^c}, \mathbf{y^{neg}}) \propto P(\tau | \lambda^2) P(\mathbf{y^c} | \tau, \kappa^2) P(\mathbf{y^{neg}} | \lambda^2, \kappa^2)$$

$$\propto \frac{1}{(\lambda^2)^{\frac{n_C}{2}}} e^{\left(\frac{-\sum_i (\tau_j - \mu_{neg})^2}{2\lambda^2} \right)} \frac{1}{(\kappa^2)^{\frac{n_C * n_P}{2}}} e^{\left(\frac{-\sum_{i,j} (y_{ij}^c - b_i - \tau_j)^2}{2\kappa^2} \right)} \frac{1}{(\lambda^2 + \kappa^2)^{\frac{n0}{2}}} e^{\left(\frac{-\sum_{i,j} (y_{ij}^{neg} - b_i - \mu_{neg})^2}{2(\kappa^2 + \lambda^2)} \right)}$$

Since the resulting conditional distributions for λ^2 and κ^2 are non-standard, we incorporate another Metropolis step to sample from λ^2 and then κ^2 (Alternatively, one could also sample these two parameters jointly (Hastings, 1970)).

Implementation

The entire MCMC procedure above can be implemented in R using the *ImmunoassayMixture()* function, whose output is a matrix containing the iterations for all the key parameters in the model. This function also allows one to monitor the chains carefully to ensure that the sampler is converging to non-degenerate parameters. In particular, when the two components of the mixture model have too much overlap, the MCMC can lead to parameter estimates that are poorly identified (Lindsay and Roeder, 1993). *ImmunoassayMixture()* and the corresponding documentation file can be found in the supplemental materials.

5 Simulation Study

In Bayesian models based on noninformative priors, we want the model to produce results that are consistent with classical frequentist interpretations (Hobert et al., 2011; Morris, 1983; Rubin, 1984). To assess the frequentist performance of our methodology, we will consider the following:

1. The overall coverage rate of our posterior distributions. Do our 95% Bayesian posterior intervals in fact cover their respective "true" parameter values 95% of the time (sample-to-sample variation)?

2. Do the posterior means of our parameter estimates accurately capture the true parameter values, or are they biased upward or downward?

3. Does adding a substantial number of negative controls greatly increase the accuracy of our procedures?

In testing this, we first generated 100 distinct simulated datasets, setting "true values" for α, μ_{neg}, μ_{pos}, κ^2, λ^2, σ^2_{pos}, and ν^2. These values were chosen to approximately reflect the typical optical measurements in our data. Using these parameters, we then sampled values for b and τ. For simplicity, we assumed there were 20 plates, and 40 unlabeled wells on each plate, and that there were 10 negative controls replicated on each plate. For each dataset, we ran our MCMC sampler for 6500 iterations (the first 2000 of which were discarded, after checking that our sampler has converged to the correct joint distribution) and recorded the 95% posterior intervals for each of these parameters, as well as the posterior means. After running this procedure for all 100 datasets, we examined each parameter and observed both the fraction of times that the parameter's posterior interval contained its true value, and also the average squared deviation between the posterior mean and the true value.

Table 1 shows our results for simulated datasets with 20 plates, and 10 negative controls per plate. For the parameters of interest, we observed seemingly good coverage rates. Namely, the fraction of datasets that resulted in intervals that contained each of the

11

"true" values for these parameters ranged between 90 and 98%, close to the desired 95% coverage.

Another check as to the accuracy of our procedure is to see how the posterior means differ from the true underlying parameter. Here, we define bias as the simple mean of $(\hat{\theta} - \theta)$ across the 100 simulations, where $\hat{\theta}$ are the posterior means of the respective parameters. Here, the biases of these parameters are all positive. However, the magnitudes of these biases are generally pretty small compared to the true values, and thus should not be of too much concern.

Though the root-mean-squared deviation of a posterior mean and a fixed parameter is a bit of an awkward metric, it does represent the overall accuracy of this procedure. These RMSE values seem quite low. Most encouraging are the small errors in estimating κ^2, λ^2, and σ_{neg}^2, as this is the most subtle part of our model, and most complicated to implement. The only possible concern is the relatively large bias and RMSE for ν^2. This probably just reflects the uncertainty of measuring an effect where we only have 20 plates from which to infer the between-plate variation.

Parameter	True Value	Coverage Rate	RMSE	Bias
μ_{neg}	-1.88	0.94	0.198	0.0332
μ_{pos}	0.20	0.95	0.219	0.0246
σ_{neg}^2	0.27	0.96	0.017	0.0027
σ_{pos}^2	0.36	0.94	0.102	0.0259
α	0.125	0.94	0.017	0.0052
λ^2	0.17	0.93	0.020	0.0012
κ^2	0.10	0.91	0.010	0.0015
ν^2	0.49	0.97	0.253	0.1285

Table 1: Simulation Results for 20 plates, 10 Negative Controls per plate

To get a sense as to the effect of number of plates in the accuracy of our results, we performed another simulation with just 5 plates instead of 20 (each plate still consisted of 10 negative controls and 40 unknown subjects). The coverage rates for these parameters are

only slightly lower than they were above, except for ν^2, which has dropped to 87% (Table 2). However, we begin to see drastic differences in the RMSE's and biases, which are now an order of magnitude larger than they were when we used more plates. This suggests that our Bayesian procedure is quite sensitive to the number of plates we incorporate into our model.

Parameter	True Value	Coverage Rate	RMSE	Bias
μ_{neg}	-1.88	0.90	0.661	0.3518
μ_{pos}	0.20	0.92	0.656	0.2888
σ^2_{neg}	0.27	0.94	0.037	0.0109
σ^2_{pos}	0.36	0.95	0.237	0.0747
α	0.125	0.96	0.037	0.0164
λ^2	0.17	0.97	0.036	0.0027
κ^2	0.10	0.96	0.022	0.0082
ν^2	0.49	0.87	1.122	0.7318

Table 2: Simulation Results for 5 plates, 10 Negative Controls per plate

Lastly, we assess the performance of our model when we reduce the number of controls on each plate. Here, we use 20 plates, but now use just 3 controls (replicated on each plate) instead of 10. Compared with the 10-control simulation, our biases and RMSE's seem more extreme. Our coverage rates though are pretty good, except for μ_{neg}, whose coverage rate drops to 89% (Table 3). This could simply be a consequence of the fact that when we have fewer controls, we have fewer negative data points, so our estimates for the parameters of the subgroup of negative patients would likely be less precise.

We also evaluated our classification of positive vs. negative samples via a simulation study. For the first analysis, we look carefully at the model with 20 plates, but only 3 negative controls. In Figure 2, we plot the sensitivity and specificity of our criterion as a function of the posterior probability cutoff. For our procedure, the sensitivity remains fairly high for cutoff probabilities as high as 0.8 (with sensitivity of 0.8), before dropping off for more stringent tests. The specificity for our procedure is extremely high for prob-

Parameter	True Value	Coverage Rate	RMSE	"Bias"
μ_{neg}	-1.88	0.89	0.4605	0.2187
μ_{pos}	0.20	0.94	0.4817	0.1787
σ^2_{neg}	0.27	0.98	0.0245	-0.0006
σ^2_{pos}	0.36	0.96	0.1486	0.0362
α	0.125	0.96	0.0283	0.0079
λ^2	0.17	0.95	0.0412	-0.0193
κ^2	0.10	0.91	0.0361	0.0187
ν^2	0.49	0.91	0.7417	0.4301

Table 3: Simulation Results for 20 Plates, 3 Negative Controls per Plate

ability cutoffs as low as 0.1. Thus, nearly all true negative samples have tiny posterior probabilities of being in the positive group. These extremely high sensitivity and specificity characteristics for our procedure are a consequence of the fact that the lognormal mixture distribution for the optical densities leads to fairly high separation between positive and negative values.
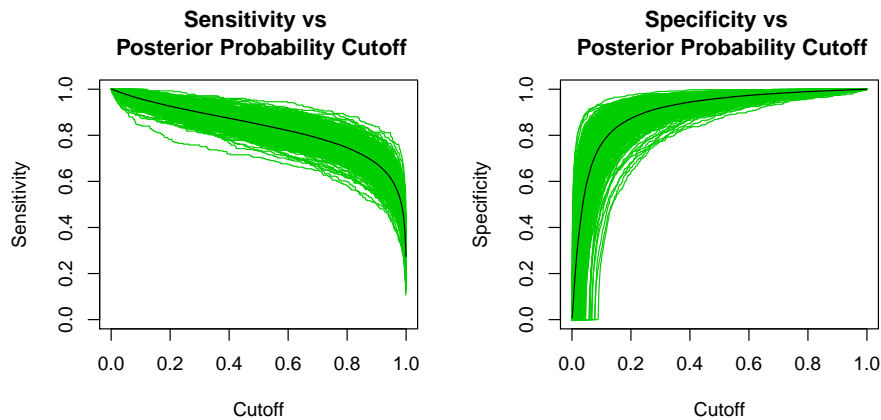

Figure 2: Sensitivity and Specificity for 20 Plates, 3 Negative Controls. Green curves represent individual simulations, the black line reflects the average rates for all 1000 simulations.

Next, we look at the Receiver Operating Characteristic (ROC) for our method, and compare it to the classification rates using the 3 SD Rule (Figure 3) (Opsteegh et al., 2010; Greiner et al., 2000). Though this type of analysis was originally developed as a frequentist technique (Kurkjian et al., 2005), it extends nicely to the Bayesian framework (Choi et al., 2006; Wang et al., 2007), and has been used previously in Bayesian analysis of ELISA tests (Limmathurotsakul et al., 2011; Nielsen et al., 2002). Our method clearly outperforms the 3 SD rule, as the blue X (denoting the average ROC values of the 3 SD rule across 1000 datasets) and the majority of the red circles (corresponding to individual simulations) lie to the lower right of the black line (our procedure). The 3 SD rule results in roughly a 0.94 and 0.89 sensitivity and specificity respectively. Using our method, if we required 0.94 sensitivity, we could improve specificity to 0.98. If instead we require a specificity of 0.89, we could improve our specificity to nearly 0.99. Thus, one could say that when using just 3 controls, our method dominates the industry standard in terms of classification rates.
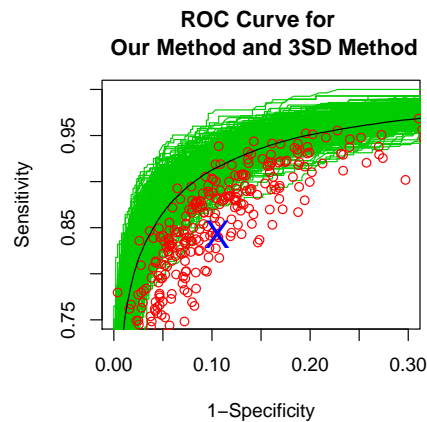

Figure 3: Plot of Sensitivity vs. 1-Specificity for the two methods (3 controls). Green curves represent simulated posterior ROC curves under our method. Black line reflects the average ROC curve across all 1000 simulations. Red dots correspond to the sensitivity/specificities of the individual simulations under the 3SD rule. The blue "X" is the average sens./spec. of the 3SD rule under all 1000 simulations.

We perform a similar analysis for a series of microtiter plates with 7 controls, instead of 3, to more mimic our real optical density data. In Figure 4,the sensitivity and specificity curves are more ideal including the additional 4 controls. The same probability cutoffs lead to slightly higher sensitivity and specificity rates than before.
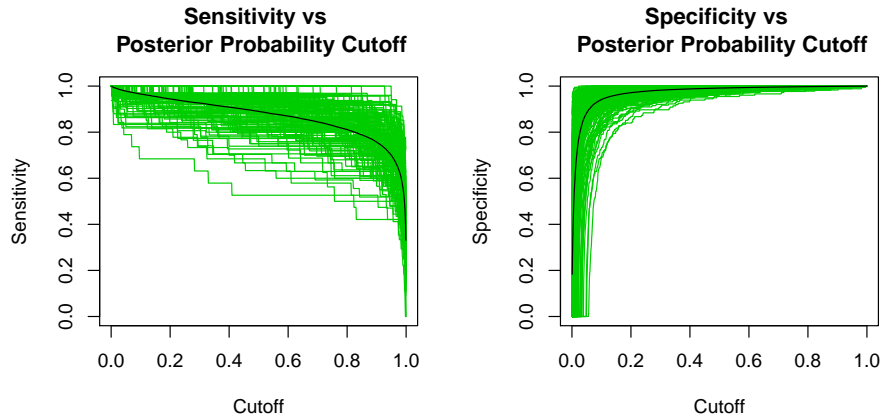

Figure 4: Sensitivity and Specificity for 20 Plates, 7 Negative Controls. Green curves represent individual simulations, the black line reflects the average rates for all 1000 simulations.

Looking at the ROC curves (Figure 5), we see that the 3SD rule performs much better, as the red dots hover much closer to the black line. This makes sense, as the 3 standard deviation cutoff for each plate is now computed using 7 controls, as opposed to just 3. However, while better than before, the 3SD classification procedure is still dominated by our method.

6 Analysis of Chagas Disease Data

Applying our model to the 18 microtiter plates introduced in Section 2, we obtain the posterior estimates for our parameters given in Table 4.

Next, we examine the magnitude of the estimated plate effects, as determined by the median $\log(\beta)$ (Figure 6). Given how we constructed the model for the plate effects, it is not surprising that the 18 parameters are centered around 0. The magnitudes of these
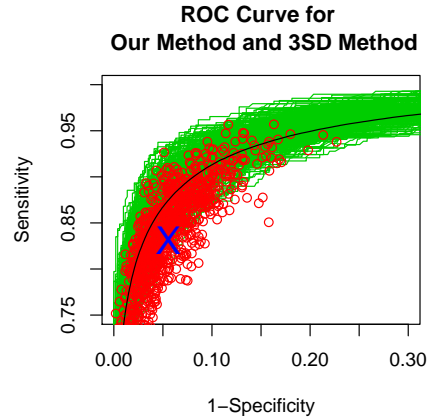
**ROC Curve for
Our Method and 3SD Method**

Figure 5: Plot of Sensitivity vs. 1-Specificity for the two methods (7 controls). Green curves represent simulated posterior ROC curves under our method. Black line reflects the average ROC curve across all 1000 simulations. Red dots correspond to the sensitivity/specificities of the individual simulations under the 3SD rule. The blue "X" is the average sens./spec. of the 3SD rule under all 1000 simulations.

effects is surprisingly strong. These values suggest that if we take a sample from one plate, and transfer it over to another, the log of the measured optical density can easily increase by 1.0 units. This is consistent with the parameter values above, where ν^2 is noticeably larger than σ^2_{neg} and σ^2_{pos}, suggesting that a larger amount of the variation in the optical densities can be attributed to the plate effect than to the heterogeneity of the actual test subjects. Also somewhat surprising is the nature in which the magnitudes of the plate effects cluster together, in particular for those around -0.5.

For this particular application, our method ends up being far more conservative in terms of classifying patients as having the disease. In Figure 7, we look at 4 representative plates, where we overlay our mixture density onto the histogram of the log of the measurements (using the posterior mean of our parameters to generate the density curves). The vertical lines represent the plates' respective cutoffs under the 3SD rule. For every one of our plates, the 3SD classification threshold is to the left of the mode of the positive population density in our mixture. Thus, for any posterior probability cutoff we choose in our method, our procedure will be far more conservative than the one based on

17

Parameter	2.5% Quantile	50% Quantile	97.5% Quantile
μ_{neg}	-1.614	-1.364	-0.989
μ_{pos}	0.019	0.602	0.979
σ^2_{neg}	0.189	0.229	0.261
σ^2_{pos}	0.026	0.064	0.378
α	0.027	0.042	0.086
λ^2	0.066	0.110	0.157
κ^2	0.089	0.116	0.151
ν^2	0.125	0.256	0.609

Table 4: Posterior Intervals for Parameters

the 3SD rule, for each of the plates in our data.

When we use these two procedures to compare the classifications, there are four possible regimes;

1. Observations classified as positive under both methods.

2. Observations classified as negative under both methods.

3. Observations classified as positive under the 3SD rule, but negative under our Bayesian model.

4. Observations classified as negative under the 3SD rule, but positive under our Bayesian model.

Due to the stark differences in the sensitivity of these two tests, there were no individual subjects that fell into this last regime (negative under 3SD, but positive under our model). Figure 8 shows how often observations fell into one of these three regimes as a function of the posterior probability cutoff in our method.

Under our Bayesian model, the estimated posterior probabilities for particular observations belonging to the positive group are highly polarized. Small optical densities lead to posterior probabilities that are often less than 0.001, while those samples with high

18

Distribution of Plate Effects for Chagas Disease Data
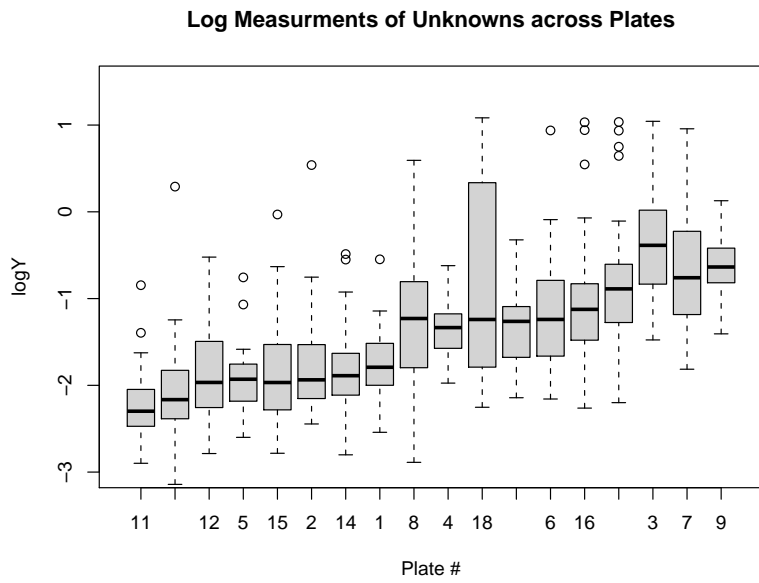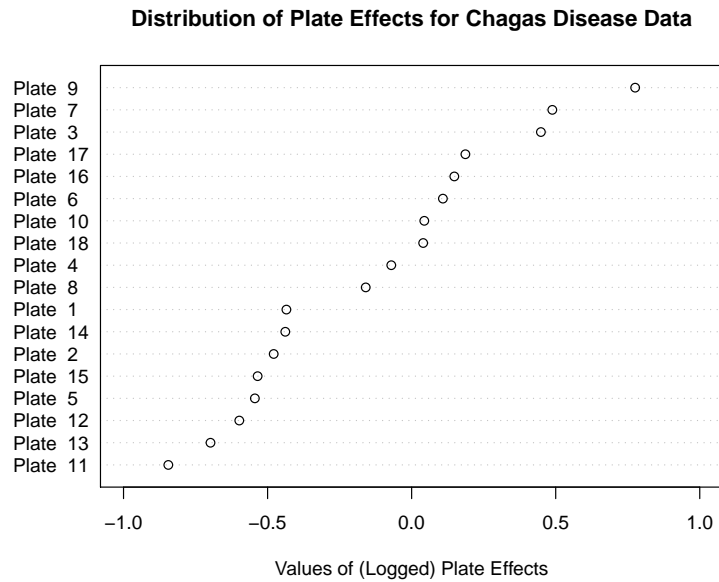
Log Measurments of Unknowns across Plates

Figure 6: (above) Dot plot of the magnitudes of the 18 estimated plate effects, sorted in increasing order. The lowest dashed line corresponds to the plate with the most negative effect, the highest dashed line to the most positive of the 18 plates, and so on. (below) Log measurements of unknowns. Same as Figure 1 but sorted in ascending order of plate effect.

antibody concentrations often have posterior probabilities higher than 0.99. Since there are very few observations whose probabilities are between 0.1 and 0.9, the curves for the
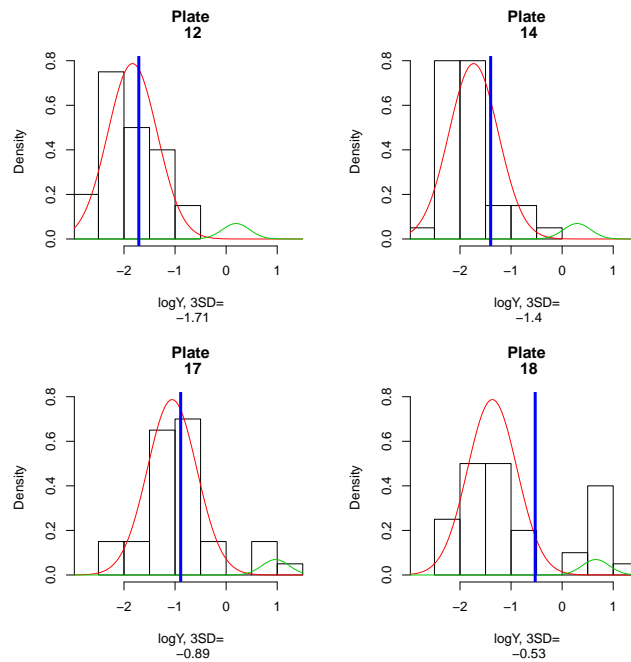
Figure 7: Histogram of the log of the optical densities, overlayed with a mixture density derived from our procedure. For each plate, the leftmost bell curve is the estimated Gaussian density curve for the negative subjects, the smaller rightmost bell curve is the density for the positives. The vertical line represents the (plate-specific) cutoff values derived from the 3SD rule ($\log(\bar{Y} + 3s_Y)$)
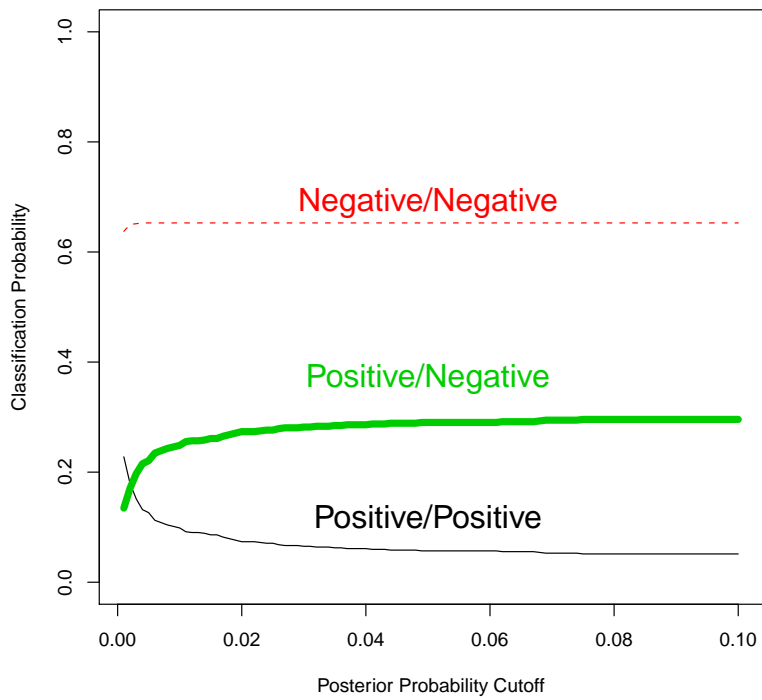
Figure 8: Fraction of patients who are diagnosed as positive under both methods, negative under both methods, and positive under the 3SD rule, and negative under our Bayesian model, as a function of posterior probability cutoff.

three regimes remain mostly flat in this region.

7 Discussion

Our goal has been to develop a robust model that allows one to perform more accurate diagnoses for diseases when testing samples on a series to series of different microtiter plates. Our classification method, based on computing posterior probabilities from a Bayesian mixture model, works quite well in simulation settings. In particular, it provides more plausible results than the conventional 3SD rule, and has improved sensitivity and specificity. Furthermore, the Bayesian approach has excellent frequentist properties. One can further improve the utility of this model by increasing the number of negative controls that are replicated on each plate.

While our procedure is designed for measurements that arise from a two-component mixture model, simulations suggest that this approach is robust to certain deviations from this model. In particular, when we applied our approach to datasets that arose either from an all-negative population, or from a three-component normal mixture model (in which the distribution of the positive subgroup was itself bimodal), we obtained accurate frequentist coverage of the relevant parameters, and in the latter case, encouraging ROC characteristics. Results of these simulations can be found in the supplemental materials. It would be of future research interest to extend our model to allow for the distribution of the positive subgroup to itself be a mixture of several components.

The model makes the crucial assumption that each unknown sample is randomly allocated to one of the microtiter plates. If unknown samples are put into the plates in sequential order (that is, filling up plate 1 before adding samples to plate 2, and so on), then the β parameters might not be valid measures of the underlying plate-effect, but rather be a measure of the overall antibody levels for that particular batch of samples, which may all share some confounding factor that systematically affects their measured optical densities. For instance, the samples of a particular batch could all come from a community with a high prevalence of T. cruzi infection, or could all be prepared in the same laboratory. In the presence of such confounding factors, application of our model

may lead to too many false negatives on plates with patients from highly infected villages, and too many false positives for samples placed on plates that come from fairly healthy regions. Randomizing samples to plates has the benefit of enabling us to make unbiased estimates of the plate effects, which in turn, lets us pool information across plates to make better classifications using our model.

In future research, it would be worth considering different priors for the plate-effect variation parameter, as the prior distribution we have chosen for $\frac{1}{\nu^2}$ may be somewhat informative. Plausible alternatives include uniform or half-Cauchy priors on ν (Gelman, 2006). However, in our simulation studies, our uniform prior on $\frac{1}{\nu^2}$ led to good frequentist properties.

References

G. Casella and E. George. Explaining the Gibbs sampler,. *Amer. Statist.*, 46:167–174, 1992.

Y. Choi, W. Johnson, M. Collins, and I. Gardner. Bayesian inferences for receiver operating characteristic curves in the absence of a gold standard. *Journal of Agricultural, Biological, and Environmental Statistics*, 11:210–229, 2006. ISSN 1085-7117. URL `http://dx.doi.org/10.1198/108571106X110883`. 10.1198/108571106X110883.

Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. 85(410):398–409, June 1990. ISSN 0162-1459 (print), 1537-274X (electronic).

A. Gelman, G. L. Chew, and M. Shnaidman. Bayesian analysis of serial dilution assays. *Biometrics*, 60:407–417, Jun 2004.

Andrew Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1:1–19, 2006.

M. Greiner, D. Pfeiffer, and R. D. Smith. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev. Vet. Med.*, 45:23–41, May 2000.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, Apr 1970.

K. M. Higgins, M. Davidian, G. Chew, and H. Burge. The effect of serial dilution error on calibration inference in immunoassay. *Biometrics*, 54:19–32, Mar 1998.

J. P. Hobert, V. Roy, and C. Robert. Improving the Convergence Properties of the Data Augmentation Method with an Application to Bayesian Mixture Modeling. *Statist. Sci.*, 26(3):332–351, 2011.

A. Irion, H. P. Beck, and T. Smith. Assessment of positivity in immuno-assays with variability in background measurements: a new approach applied to the antibody response to Plasmodium falciparum MSP2. *J. Immunol. Methods*, 259:111–118, Jan 2002.

K. M. Kurkjian, L. E. Vaz, R. Haque, C. Cetre-Sossah, S. Akhter, S. Roy, F. Steurer, J. Amann, M. Ali, R. Chowdhury, Y. Wagatsuma, J. Williamson, S. Crawford, R. F. Breiman, J. H. Maguire, C. Bern, and W. E. Secor. Application of an improved method for the recombinant k 39 enzyme-linked immunosorbent assay to detect visceral leishmaniasis disease and infection in Bangladesh. *Clin. Diagn. Lab. Immunol.*, 12:1410–1415, Dec 2005.

M. Z. Levy, N. M. Bowman, V. Kawai, J. B. Plotkin, L. A. Waller, L. Cabrera, F. Steurer, A. E. Seitz, V. V. Pinedo-Cancino, J. G. Cornejo del Carpio, E. Cordova Benzaquen, F. E. McKenzie, J. H. Maguire, R. H. Gilman, and C. Bern. Spatial patterns in discordant diagnostic test results for Chagas disease: links to transmission hotspots. *Clin. Infect. Dis.*, 48(8):1104–1106, Apr 2009.

D. Limmathurotsakul, N. Chantratita, N. Teerawattanasook, K. Piriyagitpaiboon, A. Thanwisai, V. Wuthiekanun, N. P. Day, B. Cooper, and S. J. Peacock. Enzyme-linked immunosorbent assay for the diagnosis of melioidosis: better than we thought. *Clin. Infect. Dis.*, 52:1024–1028, Apr 2011.

B. G. Lindsay and K. Roeder. Uniqueness of estimation and identifiability in mixture models,. *Canad. J. Statist.*, 21:139–147, 1993.

C. Morris. Parametric empirical Bayes inference: theory and applications. *J. Amer. Statist. Assoc.*, 78(381):47–65, 1983. ISSN 0162-1459. With discussion.

L. H. Moulton, F. C. Curriero, and P. F. Barroso. Mixture models for quantitative HIV RNA data. *Stat Methods Med Res*, 11:317–325, Aug 2002.

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

M. Opsteegh, P. Teunis, M. Mensink, L. Zuchner, A. Titilincu, M. Langelaar, and J. van der Giessen. Evaluation of ELISA test characteristics and estimation of Toxoplasma gondii seroprevalence in Dutch sheep using mixture models. *Prev. Vet. Med.*, 96:232–240, Sep 2010.

D. B. Rubin. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.*, 12(4):1151–1172, 1984. ISSN 0090-5364. doi: 10.1214/aos/1176346785. URL http://dx.doi.org/10.1214/aos/1176346785.

J. R. Verani, A. Seitz, R. H. Gilman, C. LaFuente, G. Galdos-Cardenas, V. Kawai, E. de La-Fuente, L. Ferrufino, N. M. Bowman, V. Pinedo-Cancino, M. Z. Levy, F. Steurer, C. W. Todd, L. V. Kirchhoff, L. Cabrera, M. Verastegui, and C. Bern. Geographic variation in the sensitivity of recombinant antigen-based rapid tests for chronic Trypanosoma cruzi infection. *Am. J. Trop. Med. Hyg.*, 80(3):410–415, Mar 2009.

C. Wang, B. Turnbull, Y. Grhn, and S. Nielsen. Nonparametric estimation of roc curves based on bayesian models when the true disease state is unknown. *Journal of Agricultural, Biological, and Environmental Statistics*, 12:128–146, 2007. ISSN 1085-7117. URL http://dx.doi.org/10.1198/108571107X178095. 10.1198/108571107X178095.