BIOMETRICS 57, 158–167
March 2001

Bayesian Approaches to Modeling the Conditional Dependence Between Multiple Diagnostic Tests

Nandini Dendukuri* **and Lawrence Joseph**

Department of Epidemiology and Biostatistics, McGill University,
1020 Pine Avenue West, Montreal, Quebec H3A 1A2, Canada
**email:* nandini.dendukuri@smhc.qc.ca

SUMMARY. Many analyses of results from multiple diagnostic tests assume the tests are statistically independent conditional on the true disease status of the subject. This assumption may be violated in practice, especially in situations where none of the tests is a perfectly accurate gold standard. Classical inference for models accounting for the conditional dependence between tests requires that results from at least four different tests be used in order to obtain an identifiable solution, but it is not always feasible to have results from this many tests. We use a Bayesian approach to draw inferences about the disease prevalence and test properties while adjusting for the possibility of conditional dependence between tests, particularly when we have only two tests. We propose both fixed and random effects models. Since with fewer than four tests the problem is nonidentifiable, the posterior distributions are strongly dependent on the prior information about the test properties and the disease prevalence, even with large sample sizes. If the degree of correlation between the tests is known *a priori* with high precision, then our methods adjust for the dependence between the tests. Otherwise, our methods provide adjusted inferences that incorporate all of the uncertainty inherent in the problem, typically resulting in wider interval estimates. We illustrate our methods using data from a study on the prevalence of *Strongyloides* infection among Cambodian refugees to Canada.

KEY WORDS: Bayesian analysis; Binary data; Correlation; Diagnostic tests; Gold standard; Latent class model; Markov chain Monte Carlo; Random effects model; Sensitivity; Specificity.

1. Introduction

Disease diagnosis is often based on information obtained from multiple diagnostic tests, none of which is a gold standard providing perfect sensitivity and specificity. In such a situation, two or more of the diagnostic tests may be conditionally dependent due to a factor other than the disease status, arising, e.g., from a common biological phenomenon on which two tests are based. Though it is generally assumed that the accuracy of a test remains constant across all subjects to whom it is applied, this may not be the case in practice. For example, in a stool examination for a parasitic disease, a positive test means that the parasite of interest was directly observed under a microscope. In a severely diseased case, there is a larger concentration of parasites, making it easier to detect and resulting in a more sensitive test. If the stool examination is used in combination with another test whose performance is affected by the severity of disease, then a dependence would be induced between the two tests. In order to simplify the modeling and statistical analysis of such data, it is often assumed that the results from different tests are independent of each other conditional on the true disease status. However, several authors have demonstrated that it is important to account for this dependence while analyzing the results from diagnostic tests in order to obtain unbiased estimates of the prevalence of disease and accuracy of the tests (Fryback, 1978; Vacek, 1985; Brenner, 1996; Torrance-Rynard and Walter, 1997).

Frequentist approaches that address this problem require results from at least four different tests in order to have sufficient degrees of freedom to estimate each of the parameters of interest uniquely (Walter and Irwig, 1988; Espeland and Handelman, 1989; Qu, Tan, and Kutner, 1996). In practice, it is not always possible to have results from four different tests, particularly when tests are expensive, time consuming, or invasive. For example, in a study conducted to estimate the prevalence of *Strongyloides* infection among a group of Cambodian refugees to Canada (Joseph, Gyorkos, and Coupal, 1995), only two non-gold-standard tests—a serology test and a stool examination—were available (Table 1). The results from the stool examination suggest that the prevalence is 40/162 ($\approx$ 25%), but the serology test suggests a prevalence of 125/162 ($\approx$ 77%). Clearly, in order to draw any inferences about the prevalence, the sensitivities and specificities of the two tests have to be estimated. Since we have only 3 d.f. but at least five parameters to estimate (the sensitivity and specificity of each test and the prevalence of *Strongyloides*), this problem is nonidentifiable. Therefore, the results in Table 1 could arise from an infinite number of possible situations, such as (1) prevalence = 40/162, sensitivity (stool) = specificity (stool) = 100%, sensitivity (serology) = 38/40, specificity (serology) = 35/122; (2) prevalence = 125/162, sensitivity (stool) = 38/125, specificity (stool) = 35/37, sensitivity (serology) = specificity (serology) = 100%; or (3) prevalence

Table 1
Results of the serology test and stool examination

		Stool examination ($T1$)		
		+	−	Total
Serology test ($T2$)	+	38	87	125
	−	2	35	37
Total		40	122	162

= 100%, sensitivity (stool) = 40/162, sensitivity (serology) = 125/162, and so on. Note that, under conditional independence, as soon as any two of the five parameters are considered exactly known, unbiased estimates of the other three are immediately available. While the above three scenarios are all equally consistent with the data, they are not all equally plausible given the prior information available about the test properties. In this article, we will make use of this distinction in order to derive final estimates of all parameters that are consistent with both the data and the prior information.

Equal-tailed 95% prior probability intervals for the sensitivities and specificities of the two tests were determined in consultation with faculty from the McGill Centre for Tropical Disease. These values, which were determined from information documented in previous studies and clinical opinion, are presented in Table 2 (see Joseph et al., 1995, and references therein). From the 95% prior probability intervals in Table 2, we can see that there is great uncertainty about the performance parameters of the tests. Using a frequentist approach that assumes conditional independence, two of the five unknown parameters must be assumed known in order to estimate the other three (Walter and Irwig, 1988), as in the hypothetical solutions described above. Constraining any of the parameters at a fixed value gives us a solution for a simpler, identifiable problem but not the one with which we are presented. Moreover, it is not obvious which of these parameters should be held constant or to what values they should be restricted. Furthermore, these methods lead to confidence intervals that are too narrow since the uncertainty in the constrained parameters is ignored.

Ignoring the possible dependence between test results, Joseph et al. (1995) showed that a Bayesian approach can be used to obtain interpretable posterior distributions for each of the unknown parameters relative to a given prior distribution. In summarizing the available information about each parameter in the form of a prior distribution and updating by Bayes theorem, no constrained parameters are required and therefore the uncertainty about each parameter value is fully accounted for. The prior information is used to distinguish between the numerous possible solutions for the nonidentifiable problem. This approach is approximately numerically equivalent to the frequentist approach when a degenerate (point mass) distribution is used that matches the constrained parameter values and diffuse prior distributions are used for the nonconstrained parameters. The Bayesian approach is not limited to this unnatural choice of prior distributions, so it can be viewed as a useful generalization of the standard frequentist approach. The problem of formulating a suitable prior distribution or deriving the posterior density are not worsened by virtue of the nonidentifiability of this diagnostic testing problem since the parameters involved have an easily understood interpretation. Roughly speaking, in order to obtain a useful solution using this approach, informative priors would be needed on at least as many parameters as would be constrained when using the frequentist approach. This is a reasonable requirement in practice since there will usually be considerable experience with at least one of the tests, although the sensitivity and specificity will usually not be known exactly. The posterior inference for a nonidentifiable problem can be greatly affected by altering our choice of prior distributions, as will be illustrated in Section 4. This is true even with an infinite sample size. Therefore, the prior distributions must be elicited with care. Neath and Samaniego (1997) demonstrate that, when using a Bayesian approach to estimate parameters involved in the nonidentifiable problem that arises when results from only one diagnostic test are available, a large subset of the prior parameter space results in posterior mean estimates that are closer to the true values than the prior mean estimates. However, they caution that some priors will result in worse estimates even when the sample size increases without bound.

While Joseph et al. (1995) did not discuss correlations between test results, several models have been proposed that adjust for the conditional dependence between diagnostic tests. Approaches have included latent class modeling (Espeland and Handelman, 1989; Yang and Becker, 1997) and random effects models (Qu et al., 1996; Hadgu and Qu, 1998; Qu and Hadgu, 1998). All of the above methods use a frequen-

Table 2
Elicited 95% prior probability intervals for the sensitivity and specificity of the stool examination and the serology test, along with corresponding prior distribution parameters for the fixed and random effects models

		Elicited 95% prior PI[a]	Prior parameters for fixed effects model[b]		Prior parameters for random effects model[b]			
Stool examination ($T1$)	Sensitivity	0.07–0.47	α_{S_1} 4.44	β_{S_1} 13.31	A_{11} −0.811	$\sigma_{a_{11}}$ 0.380	B_1 0.668	σ_{b_1} 0.5
	Specificity	0.89–0.99	α_{C_1} 71.25	β_{C_1} 3.75	A_{10} 2.171	$\sigma_{a_{10}}$ 0.261	B_0 0.861	σ_{b_0} 0.5
Serology test ($T2$)	Sensitivity	0.63–0.92	α_{S_2} 21.96	β_{S_2} 5.49	A_{21} 1.012	$\sigma_{a_{21}}$ 0.268	B_1 0.668	σ_{b_1} 0.5
	Specificity	0.31–0.96	α_{C_2} 4.1	β_{C_2} 1.76	A_{20} 0.692	$\sigma_{a_{20}}$ 0.560	B_0 0.861	σ_{b_0} 0.5

[a] PI refers to probability interval.

[b] See Section 2 for interpretation of parameters from the fixed effects model and Section 3 for interpretation of parameters from the random effects model.

tist approach to estimate the parameters involved. Despite the fact that they propose different approaches to model dependence, they all require a minimum of four or more tests in order to estimate all parameters of interest uniquely.

In this article, we present two models, which we will call the fixed effects model and the random effects model, to draw simultaneous inferences about the prevalence of disease and all test parameters in the situation where multiple diagnostic tests are used, while adjusting for the conditional dependence between them. In particular, we consider the nonidentifiable situation of less than four tests and propose a Bayesian approach for its solution, thus extending the work of Joseph et al. (1995) to the case of conditionally dependent tests. The fixed effects model is described in Section 2, where we directly model the correlation between the tests. Our random effects model, which indirectly models the correlations between the tests via subjects' specific intensities, is described in Section 3. The main distinction between these models is whether the test properties remain constant (fixed effects) from subject to subject or not (random effects). In Section 4, we analyze the *Strongyloides* data using both of our methods. We close with a discussion in Section 5.

2. Fixed Effects Model

In this section, we model the conditional dependence between tests using the covariance between tests within the diseased and nondiseased populations. Assume that we have results from two different dichotomous tests T_j, $j = 1, 2$, from a sample of N subjects such that a positive result on the jth test is denoted by $T_j = 1$ and a negative result by $T_j = 0$. Let $N_{t_1 t_2}$ denote the number of subjects who fall into the cross-classification $T_1 = t_1, T_2 = t_2$, $t_1, t_2 = 0, 1$. The parameters of primary interest in the diagnostic testing setup are the prevalence of the disease in the population, which we denote by π, and the sensitivity and specificity of the different tests, which we denote by S_j and C_j, $j = 1, 2$, respectively. Let D denote the (latent) true disease status such that $D = 1$ among diseased subjects and $D = 0$ among nondiseased subjects. The prevalence is defined as the probability of being truly diseased in the population under study, i.e., $\pi = P(D = 1)$. It follows that the probability of being nondiseased is given by $P(D = 0) = 1 - P(D = 1) = 1 - \pi$. The sensitivity of the jth test is the conditional probability that a subject who is truly diseased will be correctly diagnosed by the test as being positive, i.e., $S_j = P(T_j = 1 \mid D = 1)$, $j = 1, 2$. The specificity of the jth test is the conditional probability that the test is negative for a truly nondiseased subject, i.e., $C_j = P(T_j = 0 \mid D = 0)$, $j = 1, 2$. The conditional dependence between tests may be estimated using a measure such as the covariance between tests within each disease class. We denote the covariance between the two tests among the diseased and nondiseased subjects as $covs_{12}$ and $covc_{12}$, respectively. It can be shown (Vacek, 1985) that

$$P(T_1 = 1, T_2 = 1 \mid D = 1) = S_1 S_2 + covs_{12}, \quad \text{and}$$
$$P(T_1 = 1, T_2 = 0 \mid D = 1) = S_1(1 - S_2) - covs_{12}. \quad (1)$$

Therefore, $P(T_1 = 1, T_2 = 1)$ is increased by an amount $covs_{12}$ when two tests are correlated compared with the conditionally independent case. The probabilities of observing the remaining combinations of test results are

$$\begin{aligned}
P(T_1 = 0, T_2 = 1 \mid D = 1) &= (1 - S_1)S_2 - covs_{12},\\
P(T_1 = 0, T_2 = 0 \mid D = 1) &= 1 - S_1)(1 - S_2) + covs_{12},\\
P(T_1 = 1, T_2 = 1 \mid D = 0) &= (1 - C_1)(1 - C_2) + covc_{12},\\
P(T_1 = 1, T_2 = 0 \mid D = 0) &= (1 - C_1)C_2 - covc_{12},\\
P(T_1 = 0, T_2 = 1 \mid D = 0) &= C_1(1 - C_2) - covc_{12},\\
P(T_1 = 0, T_2 = 0 \mid D = 0) &= C_1 C_2 + covc_{12}. \quad (2)
\end{aligned}$$

We take $covs_{12}$ and $covc_{12}$ to be positive since this is the case that arises most frequently in practice. Analogous results to those presented here can be derived for negative correlations.

Let the true (latent) number of diseased subjects for each combination of test results, $(T_1 = t_1, T_2 = t_2)$, be denoted by $Y_{t_1 t_2}$, $t_1, t_2 = 0, 1$. Using equations (1) and (2), we can write the likelihood function of the observed data given the latent data as

$$\begin{aligned}
L &= P(N_{11}, N_{10}, N_{01}, N_{00} \mid \\
&\qquad \pi, S_1, S_2, C_1, C_2, covs_{12}, covc_{12}, Y_{11}, Y_{10}, Y_{01}, Y_{00})\\
&\propto (\pi(S_1 S_2 + covs_{12}))^{Y_{11}} (\pi(S_1(1 - S_2) - covs_{12}))^{Y_{10}}\\
&\quad \times (\pi((1 - S_1)S_2 - covs_{12}))^{Y_{01}}\\
&\quad \times (\pi((1 - S_1)(1 - S_2) + covs_{12}))^{Y_{00}}\\
&\quad \times ((1 - \pi)((1 - C_1)(1 - C_2) + covc_{12}))^{N_{11} - Y_{11}}\\
&\quad \times ((1 - \pi)((1 - C_1)C_2 - covc_{12}))^{N_{10} - Y_{10}}\\
&\quad \times ((1 - \pi)(C_1(1 - C_2) - covc_{12}))^{N_{01} - Y_{01}}\\
&\quad \times ((1 - \pi)(C_1 C_2 + covc_{12}))^{N_{00} - Y_{00}}, \quad (3)
\end{aligned}$$

which is essentially a multinomial likelihood function. We used standard distributional families to represent our prior information. The choice of distributions discussed below is not unique and they may be replaced by other suitable densities, as needed.

(1) The prevalence is assumed to follow a beta prior distribution with parameters α_π and β_π, $\pi \sim \text{beta}(\alpha_\pi, \beta_\pi)$.

(2) The sensitivities and specificities are also assumed to have beta prior densities such that $S_j \sim \text{beta}(\alpha_{S_j}, \beta_{S_j})$, $j = 1, 2$ and $C_j \sim \text{beta}(\alpha_{C_j}, \beta_{C_j})$, $j = 1, 2$.

(3) The feasible range of the covariance is determined by the sensitivities among the diseased subjects and the specificities among the nondiseased subjects. Clearly,

$$(C_1 - 1)(1 - C_2) \le covc_{12} \le \min(C_1, C_2) - C_1 C_2, \quad (4)$$

where $min(a, b)$ is the minimum of a and b. Since we are only interested in the situation when the two tests are positively correlated and the expression for the lowerbound in equation (4) is always negative, the lower bound of $covc_{12}$ was fixed at zero. Therefore, the upper and lower bounds for $covc_{12}$ are given by

$$0 \le covc_{12} \le \min(C_1, C_2) - C_1 C_2,$$

and analogously,

$$0 \le covs_{12} \le \min(S_1, S_2) - S_1 S_2. \quad (5)$$

The covariance parameters are taken to have generalized beta (genbeta) prior distributions, $covs_{12} \sim$

genbeta($\alpha_{covs_{12}}, \beta_{covs_{12}}$), $0 \leq covs_{12} \leq u_s$, where $u_s = \min(S_1, S_2) - S_1 S_2$ and $covc_{12} \sim$ genbeta($\alpha_{covc_{12}}$, $\beta_{covc_{12}}$), $0 \leq covc_{12} \leq u_c$, where $u_c = \min(C_1, C_2) - C_1 C_2$. The generalized beta density is a standard beta density that has been stretched or shrunk to correspond to a range other than the standard $[0, 1]$.

When the likelihood function in equation (3) is combined with the above prior distributions, we obtain the following expression for the joint posterior distribution of the parameters:

$$
\begin{aligned}
&p(\pi, S_1, C_1, S_2, C_2, covs_{12}, covc_{12}, Y_{11}, Y_{10}, Y_{01}, Y_{00} \mid \\
&\quad N_{11}, N_{10}, N_{01}, N_{00}) \\
&\quad \propto L \times \pi^{\alpha_\pi - 1}(1-\pi)^{\beta_\pi - 1} S_1^{\alpha_{S_1} - 1} \\
&\qquad \times (1-S_1)^{\beta_{S_1} - 1} S_2^{\alpha_{S_2} - 1} (1-S_2)^{\beta_{S_2} - 1} \\
&\qquad \times C_1^{\alpha_{C_1} - 1} (1-C_1)^{\beta_{C_1} - 1} C_2^{\alpha_{C_2} - 1} (1-C_2)^{\beta_{C_2} - 1} \\
&\qquad \times covs_{12}^{\alpha_{covs_{12}} - 1} (u_s - covs_{12})^{\beta_{covs_{12}} - 1} \\
&\qquad \times covc_{12}^{\alpha_{covc_{12}} - 1} (u_c - covc_{12})^{\beta_{covc_{12}} - 1}.
\end{aligned}
$$

Given the complexity of this model, it is not possible to obtain the marginal distributions for the parameters analytically. Therefore, we use a Gibbs sampler algorithm (Gelfand and Smith, 1990) to obtain samples from the marginal posterior distribution of each parameter. The full conditional distributions for each parameter, required by the Gibbs sampler, are listed in the Appendix.

3. Random Effects Model

In this section, we present a Bayesian approach similar to the frequentist approach of Qu et al. (1996), which takes into account the variation in test parameters over the population by modeling the conditional dependence between multiple tests using random effects. The sensitivities and specificities of the tests are modeled as functions of a latent, subject-specific random variable. Applying the same latent value within each patient across all tests induces a dependence between the tests without explicit reference to a covariance parameter.

This situation can be conceptualized as one where the performance of a test in a given subject is a function of a continuous random variable, which we will term the intensity, I_k. This intensity can be considered, e.g., as a summary measure of the severity of illness of the subject that affects the ease of detection of disease in the subject. The sensitivity and specificity of a test for each subject are functions of this underlying intensity, of the form $f(I_k)$, where f is a continuous, monotonically increasing function taking values between zero and one. Here I_k is taken to be a random variable following a normal$(0, 1)$ distribution. While we assume equal variances for diseased and nondiseased subjects in our application, more generally, different variances can be used.

Retaining the notation used in the previous section, we introduce some parameters at the individual subject level. The test result for the kth ($k = 1, \ldots, N$) subject on the jth ($j = 1, 2$) test is denoted by $T_{jk} = 1$ or 0 for a positive or a negative result, respectively. The true disease status of the kth subject is denoted by $D_k = d_k, d_k = 0, 1$. The intensity of the kth subject is denoted by $I_k = i_k$.

The probability that the kth subject has a positive result on the jth test is given by $P(T_{jk} = 1 \mid D_k = d_k, I_k = i_k) = \Phi(a_{jd_k} + b_{jd_k} i_k)$, $d_k = 0, 1, I_k \sim \mathrm{N}(0, 1)$, where Φ represents the cumulative distribution function of the normal$(0, 1)$ distribution and (a_{jd_k}, b_{jd_k}), $j = 1, 2$, $d_k = 0, 1$ are real, unknown parameters. The mean sensitivity of the jth test over all subjects is then given by integrating the expression for the sensitivity of the jth test for the kth subject over all possible values of I_k, so that

$$
\begin{aligned}
S_j &= P(T_j = 1 \mid D = 1) \\
&= \int_{-\infty}^{\infty} P(T_{jk} = 1 \mid D_k = 1, I_k = i_k) d\Phi(i_k) \\
&= \Phi\left(\frac{a_{j1}}{\sqrt{1 + b_{j1}^2}}\right), \qquad j = 1, 2,
\end{aligned}
$$

where D is the true disease status and $d\Phi(\cdot)$ denotes the probability density of the standard normal distribution. Similarly, the specificity of the jth test is given by

$$
C_j = P(T_j = 0 \mid D = 0) = \Phi\left(-\frac{a_{j0}}{\sqrt{1 + b_{j0}^2}}\right), \qquad j = 1, 2.
$$

If $b_{jd_k} = 0, j = 1, 2$, $d_k = 0, 1$, the tests are conditionally independent.

The results of different tests are taken to be independent of each other conditional on the disease status D_k and the latent variable I_k. Therefore, the likelihood for this model can be written as

$$
\begin{aligned}
L &\propto \prod_{k=1}^{N} P(T_{1k} = t_{1k}, T_{2k} = t_{2k} \mid \psi, I_k = i_k, D_k = d_k) \\
&= \prod_{k=1}^{N} \left(\pi \prod_{j=1}^{2} \Phi(a_{j1} + b_{j1} i_k)^{t_{jk}} \right. \\
&\qquad \left. \times (1 - \Phi(a_{j1} + b_{j1} i_k))^{(1 - t_{jk})} \right)^{d_k} \\
&\quad \times \left((1-\pi) \prod_{j=1}^{2} \Phi(a_{j0} + b_{j0} i_k)^{(1 - t_{jk})} \right. \\
&\qquad \left. \times \left(1 - \Phi(a_{j0} + b_{j0} i_k)\right)^{t_{jk}} \right)^{(1 - d_k)},
\end{aligned}
$$

where ψ is the vector of parameters to be estimated, $\psi = (\pi, (a_{1d_k}, b_{1d_k}, a_{2d_k}, b_{2d_k}) d_k = 0, 1)$.

The prior distributions we used are given below, but they may be replaced by other suitable densities, as indicated by the prior information. A beta prior distribution was used for the prevalence, π, such that $\pi \sim$ beta(α_π, β_π). A bivariate normal prior distribution, with mean (A_{jd_k}, B_{jd_k}) and a diagonal variance–covariance matrix, was used over the parameter pairs (a_{jd_k}, b_{jd_k}). The prior standard deviations for a_{jd_k} and b_{jd_k} were denoted by $\sigma_{a_{jd_k}}$ and $\sigma_{b_{jd_k}}$, respectively. As in the case of the fixed effects model, we use a Gibbs sampler to obtain samples from the marginal posterior distributions of the parameters in ψ. The required full conditional distributions are listed in the Appendix.

In both the fixed and random effects models, lack of identifiability means that the posterior distribution does not necessarily concentrate on the true parameter values, even as the

sample size tends to infinity. Rather, it concentrates on the set of parameter values consistent with the data, and the prior distributions are used to delineate which sets of parameter values are more plausible than others. Therefore, the influence of the prior distributions does not dissipate even with an infinite sample size.

4. Application to Estimating the Prevalence of *Strongyloides* Infection

We now return to the *Strongyloides* problem introduced in Section 1. We retain the prior distributions discussed by Joseph et al. (1995), which are listed in Table 2. In doing so, it is important to note that the median and 95% probability interval for the sensitivities and specificities given in Table 2 represent marginal prior information since the tests are correlated. In addition, for the random effects model, which allows for subject-to-subject variations in the test properties depending on the intensity, the values given in Table 2 represent marginal prior information for the mean over all subjects in the population. Subject-specific sensitivities and specificities vary about this mean, as discussed below. Since very little was known *a priori* about the prevalence of *Strongyloides* infection in a Cambodian population, a uniform (beta($\alpha_\pi = 1$, $\beta_\pi = 1$)) prior was used.

Elicitation of prior distribution parameters for the fixed effects model. The parameters for the beta(α, β) prior distributions over the sensitivities and specificities (Table 2) were determined by solving the two equations that set the center of the elicited 95% probability intervals to the mean of the corresponding distribution, $\alpha/(\alpha+\beta)$, and a quarter of the 95% prior probability interval to its standard deviation,

$$\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}.$$

In order to obtain a meaningful solution for a nonidentifiable problem, we need to have informative prior distributions on at least as many parameters as would need to be constrained in a frequentist approach. For the fixed effects model, this means we must have informative distributions on at least $7-(2^2-1)=4$ parameters. In this particular example, we were able to determine informative prior distributions for the sensitivities and specificities. Since we did not have useful information about the covariances between tests, we used uniform (genbeta(1, 1) over the feasible range) prior distributions over the two covariance parameters.

Elicitation of prior distribution parameters for the random effects model. For simplicity, we consider a specific case of the random effects model, where $b_{jd_k} = b_{d_k}$, $j = 1, 2$. This means that a change in the value of I_k will cause the sensitivities and specificities of all tests for the kth subject to change by the same amount on the probit scale. As in the case of the fixed effects model, the prior mean values for the sensitivities and specificities were set equal to the center of their elicited 95% probability intervals in Table 2. Using the expression derived in equation (5) and the prior mean values of the sensitivities, we can estimate the range of values in which the mean covariance among the diseased subjects lies as $0 \le covs_{12} \le \min(S_1, S_2) - S_1 S_2 = \min(0.25, 0.8) - (0.25)(0.8) = 0.05$. For purposes of estimating the prior densities of a_{11}, a_{21}, and b_1, we arbitrarily fixed $covs_{12} = 0.025$ since this value lies in the middle of the range in the above equation. We discuss later that this choice has little effect on the final prior parameter values. By relating the expressions for the mean values of the sensitivities and for the covariance to their elicited values, we have three equations in three unknown parameters, (A_{11}, A_{21}, B_1), i.e.,

$$S_1 = \Phi\left(\frac{A_{11}}{\sqrt{1+B_1^2}}\right) = 0.25,$$

$$S_2 = \Phi\left(\frac{A_{21}}{\sqrt{1+B_1^2}}\right) = 0.8,$$

$$\int_{-\infty}^{\infty} \Phi(A_{11}+B_1 i_k)\,\Phi(A_{21}+B_1 i_k) d\Phi(i_k) - (0.25)(0.8) = 0.025.$$

We used a bisectional search algorithm (Thisted, 1988, p. 170) to solve for (A_{11}, A_{21}, B_1).

Similarly, the possible range of values for the mean covariance among the nondiseased subjects, $covc_{12}$, was determined using the two specificities such that $0 \le covc_{12} \le \min(C_1, C_2) - C_1 C_2 = \min(0.95, 0.70) - (0.95)(0.70) = 0.035$. Once again we arbitrarily set $covc_{12} = 0.0175$, which is the midpoint of the range in the above equation. The following set of equations was used to estimate (A_{10}, A_{20}, B_0):

$$C_1 = \Phi\left(\frac{A_{10}}{\sqrt{1+B_0^2}}\right) = 0.95,$$

$$C_2 = \Phi\left(\frac{A_{20}}{\sqrt{1+B_0^2}}\right) = 0.70,$$

$$\int_{-\infty}^{\infty} \Phi(A_{10}+B_0 i_k)\Phi(A_{20}+B_0 i_k) d\Phi(i_k) - (0.95)(0.7) = 0.0175.$$

The values we have selected for the covariance parameters are by no means unique. In the absence of any information about the covariance between the tests, it seems sensible to use the midpoint of the range as the prior mean so that the prior distribution can easily cover the feasible range. Another approach may be to first run the fixed effects model and then use the mean values of the posterior distributions for the covariance parameters obtained there.

To determine the approximate prior standard deviations for a_{11}, a_{21}, and b_1, we used a contour plot of S_1 on the (a_{11}, b_1) plane. From the contour plot in Figure 1, we determined that, as S_1 ranges from 0.07 to 0.47 (its 95% prior probability interval), a_{11} ranges approximately from -1.653 to -0.132. The standard deviation of a_{11} was taken to be a quarter of this range, namely,

$$\sigma_{a_{11}} = \frac{-0.132-(-1.653)}{4} = 0.380.$$

The range of b_1 is less obvious since the same value of b_1 could correspond to the entire range of values of S_1. We can deduce from this that the value of S_1 is mainly determined by a_{11} while the value of b_1 has a greater bearing on the value of the covariance between the tests.

Keeping in mind that we have no prior information on $covs_{12}$, we used a wide prior distribution for b_1. The stan-

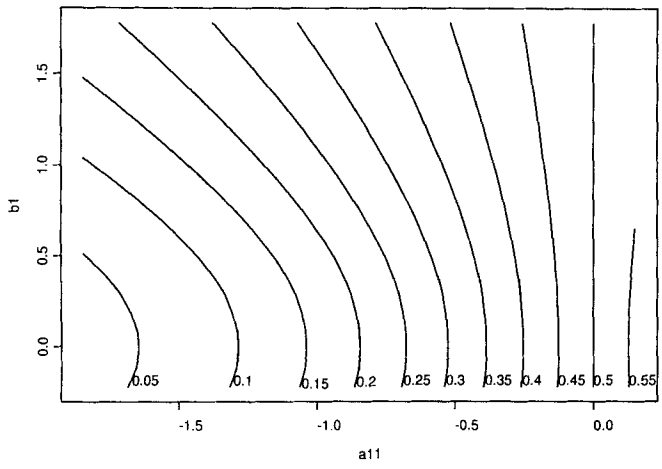

Figure 1. Contour plot of S_1 on the (a_{11}, b_1) plane.

dard deviation for the remaining parameters was determined in a similar fashion (see Table 2). We validated our somewhat *ad hoc* method of prior elicitation by generating a random sample of 10,000 observations from the prior distributions of the (a_{jd_k}, b_{d_k}) pairs using the parameter values in Table 2 and calculating the mean sensitivities and specificities. The means and 95% probability intervals of these samples were found to be very close to the desired values.

In order to study the robustness of our assumptions about the covariance, we also ran both models with degenerate prior distributions over the median and two endpoints of the feasible range for the two covariance parameters. This amounts to using a point mass prior distribution for b_{d_k} in the random effects model.

Results. The results obtained by Joseph et al. (1995) when applying the conditional independence model are presented along with the results obtained from the fixed and random effects models in Table 3. The medians of the posterior distribution for the prevalence (π) estimated by the fixed effects model (0.85) and the random effects model (0.82) were greater than that obtained when assuming conditional independence (0.76) between the two tests. The 95% posterior probability intervals for the prevalence obtained using the fixed and random effects models are also shifted to the right and are somewhat wider compared with the conditional independence model. Therefore, adjusting for conditional dependence here has led to increased prevalence estimates that are slightly less precise compared with the independence model.

Considering degenerate priors over the two covariance parameters at $covs_{12} = 0$ and $covc_{12} = 0$, we obtain results similar to those obtained by Joseph et al. (1995), as expected. For the prevalence, we obtain posterior medians of 0.77 and 0.76, with 95% posterior probability intervals of 0.50–0.91 and 0.48–0.90, for the fixed and the random effects models, respectively. When the two covariances are held fixed at their prior mean values, we obtain posterior medians virtually identical to those in Table 3 but with somewhat tighter probability intervals. Finally, when the two covariances are fixed at their maximum possible values, we find that the posterior median prevalence is considerably increased, to 0.91 and 0.95 for the fixed and random effects models, respectively. From these results, we see that covariances whose values are known with greater precision lead to narrower widths of the 95% intervals and that higher covariances are associated with higher prevalences, at least in this example.

Since the model is nonidentifiable, it is also important to reanalyze the data with different prior values on the test properties. The most uncertain parameter *a priori* was the specificity of serology. The wide prior interval, which ranged from 0.31 to 0.96 (see Table 2), indicated that some researchers were more optimistic about the performance of serology in truly negative subjects than others. We therefore ran two additional analyses, where we assumed ranges of 0.35–0.70 and 0.70–1, corresponding to the prior opinions held by pessimistic and optimistic experts, respectively. The 95% posterior probability intervals for the prevalence obtained using the fixed effects model were (0.53, 0.99) and (0.76, 0.99) for the pessimistic and optimistic priors, respectively. The corresponding

Table 3

Posterior medians and 95% probability intervals of the prevalence and test parameters obtained from three different models

		Conditional independence model		Fixed effects model		Random effects model	
	Variable	Median	95% PI	Median	95% PI	Median	95% PI
	π	0.76	0.52–0.91	0.85	0.54–0.99	0.82	0.52–0.98
Stool	Sensitivity	0.31	0.22–0.44	0.27	0.19–0.39	0.27	0.07–0.61
examination	Specificity	0.96	0.91–0.99	0.93	0.86–0.97	0.98	0.81–0.99
Serology	Sensitivity	0.89	0.80–0.95	0.83	0.73–0.92	0.80	0.66–0.92
test	Specificity	0.67	0.36–0.95	0.67	0.30–0.93	0.66	0.18–0.92
	$covs_{12}$	—	—	0.03	0.01–0.05	—	—
	$covc_{12}$	—	—	0.02	0.00–0.06	—	—
	a_{11}	—	—	—	—	−0.84	−2.13–0.46
	a_{21}	—	—	—	—	1.38	0.84–2.09
	b_1	—	—	—	—	1.30	0.56–2.27
	a_{10}	—	—	—	—	3.04	1.98–41.33
	a_{20}	—	—	—	—	0.26	−1.05–2.17
	b_0	—	—	—	—	0.95	−0.23–2.98

intervals from the random effects model were (0.50, 0.99) and (0.72, 0.99) for the pessimistic and optimistic priors, respectively. These posterior distributions were typically skewed toward one. Thus, if we were sure that the specificity of serology was at least 0.7, we could reasonably assert that the prevalence of *Strongyloides* is at least 0.72. In any case, we are quite certain that the prevalence is at least 50% and prevalences as high as 99% cannot be ruled out, a conclusion of substantive importance to public health officials.

We ran a large number of iterations (20,000) in order to obtain accurate inferences from the Gibbs sampler. Standard diagnostic procedures (Gelman and Rubin, 1992; Raftery and Lewis, 1992) revealed no convergence problems.

5. Discussion

Time, cost, and other constraints often create situations where results from less than four diagnostic tests are available. Faced with this nonidentifiable situation, the analyst is presented with a choice of methods. Attempting to estimate all unknown parameters using maximum likelihood methods will result in the infinite number of possible solutions discussed in the introduction, which is not useful in practice. More simply (but also unrealistically), one can select two of the five unknown parameters to be exactly known and, assuming conditional independence, solve as reviewed by Walter and Irwig (1988). This results in estimates of the other three parameters, but the 95% confidence intervals are artificially narrow since the uncertainty inherent in the constrained parameters is ignored. Furthermore, the estimates can be biased if conditional independence does not hold. The Bayesian approach of Joseph et al. (1995), while retaining the assumption of conditional independence, replaces the need for arbitrarily constrained parameters with more realistic prior distributions over all parameters. While this means that the prior information about the tests needs to be quantified, the method results in 95% intervals that include the uncertainty inherent in all five parameters.

In this article, we have further extended these methods to account for conditional dependence. If the degree of correlation between the tests is known *a priori* with high precision, then our methods adjust for the dependence between the tests while estimating the prevalence and test parameters. If the degree of correlation is not known, then our methods provide adjusted, typically wider inferences that incorporate all of the uncertainty inherent in the problem. Our methods could also be extended to include more than two tests by the addition of more covariance terms in the fixed model or a larger number of (a_{jd_k}, b_{jd_k}) parameters in the random effects model, as in Qu et al. (1996). Note that, by appropriately selecting the prior distribution, the methods described in this article can be made to numerically correspond to the Bayesian conditionally independent model or to both the frequentist constrained and unconstrained models. If, however, the tests may be correlated and it is unreasonable to assume that two or more parameters are exactly known, the methods presented here allow one to more realistically model the situation by using more appropriate prior distributions. For a different viewpoint, see the work by Gastwirth and colleagues (Gastwirth, 1987; Gastwirth, Johnson, and Reneau, 1991; Johnson and Gastwirth, 1991).

Of course, given the nonidentifiability of the model, this extra modeling flexibility comes at the price of having to specify a prior distribution, which may not be straightforward and whose influence does not diminish with increasing sample size. Thus, in deciding which model to use, the analyst must trade off the possibility of bias and optimistically narrow confidence intervals versus the need to derive a realistic prior distribution over all parameters.

Both the models discussed here assume the true disease status to be latent. When there is no gold standard, one can choose either to use an explicit, but imperfect, pseudogold standard and accept its imperfections or to allow for a model to implicitly define what disease-positive means through the use of statistical techniques. Alonzo and Pepe (1999) propose the use of a composite reference standard (CRS) that combines information from several imperfect references to formulate a pseudogold standard. This has several advantages in that the CRS is explicitly defined rather than latent, can handle dependencies between tests, and does not depend on any new tests that are under development. In the situation considered in this article, it can be very difficult to derive a reasonable CRS, especially since we have only two tests and the properties of both tests are relatively poor so that no combination of their results comes close to being a gold standard. Therefore, a model-based approach may be a reasonable way to proceed.

The application of these methods to the problem of estimating the prevalence of *Strongyloides* infection shows that adjusting for the possibility of conditional dependence between diagnostic tests may have a substantial effect on the posterior estimates of the prevalence and test properties. Unless something is known *a priori* about the strength of the dependence, however, we also widen the posterior probability intervals. In the absence of any information about the possible dependence between the tests, this method might serve as a sensitivity analysis. By varying the values of the covariance between the tests between plausible limits, we can estimate the extent of the possible bias in our conclusions due to assuming conditional independence between the tests. This example illustrates that posterior inferences strongly depend on the available prior information and on how that information is quantified into prior distributions. Different investigators can reasonably come to different conclusions, depending on their prior views. Some statisticians have considered the ability of Bayesian analysis to mirror the diversity of clinical opinion following a study as an advantage (Spiegelhalter, Freedman, and Parmar, 1994).

The methodology developed here may be viewed as a mapping from a given set of prior distributions to the corresponding set of posterior distributions. Therefore, the posterior density can always be interpreted as a coherent updating of the prior distribution upon seeing the data, but any extrapolation to the truth involves a leap of faith. Thus, the accurate elicitation of prior distributions is very important. We note that this is not a limitation of our methodology per se but rather a limitation of the situation when results from only two tests are available. Though there is a growing literature on the general problem of elicitation, elicitation for diagnostic test properties remains to be addressed. Not all tests perform uniformly well across different populations, and this is difficult to quantify.

Acknowledgements

We thank the referees and the associate editor for their valuable suggestions, which led to a much improved article. This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

Résumé

Beaucoup d'analyses de résultats à partir de tests diagnostiques multiples supposent que les tests soient statistiquement indépendants conditionnellement au vrai état de maladie chez le sujet. Cette supposition peut être violée en pratique, spécialement dans les situations où aucun des tests est un test de référence parfaitement exact. L'inférence classique pour des modèles tenant compte de la dépendance conditionnelle entre tests exige que les résultats d'au moins quatre tests différents soient utilisés afin d'obtenir une solution identifiable, mais il n'est pas toujours faisable d'avoir des résultats d'autant de tests. Nous utilisons une approche bayésienne pour inférer sur la prévalence de la maladie et les propriétés des tests en ajustant sur la possibilité d'une dépendance conditionnelle entre tests, particulièrement quant on n'a que deux tests. Nous proposons des modèles à effets fixes et des modèles à effets aléatoires. Dans la mesure où avec moins de quatre tests le problème n'est pas identifiable, les distributions a posteriori sont fortement dépendantes de l'information a priori concernant les propriétés des tests et la prévalence de la maladie, même dans le cas de gros échantillons. Si le degré de corrélation entre les tests est connu a priori avec une grande précision, alors nos méthodes ajustent pour la dépendance entre tests. Dans le cas contraire, nos méthodes fournissent des inférences ajustées qui incorporent la totalité de l'incertitude inhérente au problème, résultant typiquement en des estimations d'intervalles élargis. Nous illustrons nos méthodes en utilisant des données d'une étude sur la prévalence d'infection à *Strongyloïdes* parmi des réfugiés Cambodgiens au Canada.

References

Alonzo, T. A. and Pepe, M. S. (1999). Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Statistics in Medicine* **22,** 2987–3003.

Brenner, H. (1996). How independent are multiple 'independent' diagnostic classifications? *Statistics in Medicine* **15,** 1377–1386.

Espeland, M. A. and Handelman, S. L. (1989). Using latent class models to characterize and assess relative error in discrete measurements. *Biometrics* **45,** 587–599.

Fryback, D. G. (1978). Bayes' theorem and conditional nonindependence of data in medical diagnosis. *Computers and Biomedical Research* **11,** 423–434.

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

Gastwirth, J. L., Johnson, W. O., and Reneau, D. M. (1991). Bayesian analysis of screening data: Application to AIDS in blood donors. *The Canadian Journal of Statistics* **19,** 135–150.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85,** 398–409.

Gelman, A. and Rubin, D. B. (1992). Inferences from iterative simulation and multiple sequences (with discussion). *Statistical Science* **7,** 457–511.

Hadgu, A. and Qu, Y. (1998). A biomedical application of latent class models with random effects. *Applied Statistics* **47,** 603–616.

Johnson, W. O. and Gastwirth, J. L. (1991). Bayesian inference for medical screening tests: Approximations useful for the analysis of acquired immune deficiency syndrome. *Journal of the Royal Statistical Society, Series B* **53,** 427–439.

Joseph, L., Gyorkos, T., and Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology* **141,** 263–272.

Neath, A. and Samaniego, F. J. (1997). On the efficacy of Bayesian inference for nonidentifiable models. *American Statistician* **51,** 225–234.

Qu, Y. and Hadgu, A. (1998). A model for evaluating the sensitivity and specificity for correlated diagnostic tests in efficacy studies with an imperfect reference test. *Journal of the American Statistical Association* **93,** 920–928.

Qu, Y., Tan, M., and Kutner, M. H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* **52,** 797–810.

Raftery, A. E. and Lewis, S. M. (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics,* Volume 4, J. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds), 765–776. Oxford: Oxford University Press.

Rubin, D. (1988). Using the SIR algorithm to simulate posterior distributions. *Bayesian Statistics* **3,** 395–402.

Spiegelhalter, D. J., Freedman, L. S., and Parmar, M. K. B. (1994). Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society, Series A* **157,** 357–416.

Thisted, R. A. (1988). *Elements of Statistical Computing.* New York: Chapman and Hall.

Torrance-Rynard, V. L. and Walter, S. D. (1997). Effects of dependent errors in the assessment of diagnostic test performance. *Statistics in Medicine* **16,** 2157–2175.

Vacek, P. M. (1985). The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics* **41,** 959–968.

Walter, S. D. and Irwig, L. M. (1988). Estimation of error rates, disease prevalence, and relative risk misclassified data: A review. *Journal of Clinical Epidemiology* **41,** 923–937.

Yang, I. and Becker, M. P. (1997). Latent variable modeling of diagnostic accuracy. *Biometrics* **53,** 948–958.

Received June 1999. *Revised July* 2000.
Accepted July 2000.

APPENDIX

The full conditional distributions for the fixed and random effects models are listed below. When a parameter had a full conditional distribution that was not of a known form, we used a sampling-importance resampling (SIR) algorithm (Rubin, 1988) to sample its values at each iteration of the Gibbs sampler.

Full Conditional Distributions of Parameters in the Fixed Effects Model

In the fixed effects model, the full conditional distributions of the parameters are as follows:

$$\pi \mid N, Y_{11}, Y_{10}, Y_{01}, Y_{00}, \alpha_\pi, \beta_\pi \sim \text{beta}(\alpha_\pi + Y_{11} + Y_{10} + Y_{01} + Y_{00}, \beta_\pi + N - (Y_{11} + Y_{10} + Y_{01} + Y_{00})),$$

$$\begin{aligned}
&p(S_j \mid S_{3-j}, covs_{12}, Y_{11}, Y_{10}, Y_{01}, Y_{00}, \alpha_{S_j}, \beta_{S_j}, u_s, \beta_{covs_{12}}) \\
&\quad \propto (S_1S_2 + covs_{12})^{Y_{11}}(S_1(1-S_2) - covs_{12})^{Y_{10}}((1-S_1)S_2 - covs_{12})^{Y_{01}} \\
&\quad\quad \times ((1-S_1)(1-S_2) + covs_{12})^{Y_{00}} S_j^{\alpha_{S_j}-1}(1-S_j)^{\beta_{S_j}-1}(u_s - covs_{12})^{\beta_{covs_{12}}-1}, \qquad j = 1, 2, \\
&p(C_j \mid C_{3-j}, covc_{12}, N_{11}, N_{10}, N_{01}, N_{00}, Y_{11}, Y_{10}, Y_{01}, Y_{00}, \alpha_{C_j}, \beta_{C_j}, u_c, \beta_{covc_{12}}) \\
&\quad \propto ((1-C_1)(1-C_2) + covc_{12})^{N_{11}-Y_{11}}((1-C_1)C_2 - covc_{12})^{N_{10}-Y_{10}}(C_1(1-C_2) - covc_{12})^{N_{01}-Y_{01}} \\
&\quad\quad \times (C_1C_2 + covc_{12})^{N_{00}-Y_{00}} C_j^{\alpha_{C_j}-1}(1-C_j)^{\beta_{C_j}-1}(u_c - covc_{12})^{\beta_{covc_{12}}-1}, \qquad j = 1, 2,
\end{aligned}$$

$$\begin{aligned}
&p(covs_{12} \mid S_1, S_2, Y_{11}, Y_{10}, Y_{01}, Y_{00}, u_s, \alpha_{covs_{12}}, \beta_{covs_{12}}) \\
&\quad \propto (S_1S_2 + covs_{12})^{Y_{11}}(S_1(1-S_2) - covs_{12})^{Y_{10}}((1-S_1)S_2 - covs_{12})^{Y_{01}}((1-S_1)(1-S_2) + covs_{12})^{Y_{00}} \\
&\quad\quad \times covs_{12}^{\alpha_{covs_{12}}-1}(u_s - covs_{12})^{\beta_{covs_{12}}-1},
\end{aligned}$$

where $u_s = \min(S_1, S_2) - S_1S_2$;

$$\begin{aligned}
&p(covc_{12} \mid C_1, C_2, N_{11}, N_{10}, N_{01}, N_{00}, Y_{11}, Y_{10}, Y_{01}, Y_{00}, u_c, \alpha_{covc_{12}}, \beta_{covc_{12}}) \\
&\quad \propto ((1-C_1)(1-C_2) + covc_{12})^{N_{11}-Y_{11}}((1-C_1)C_2 - covc_{12})^{N_{10}-Y_{10}}(C_1(1-C_2) - covc_{12})^{N_{01}-Y_{01}} \\
&\quad\quad \times (C_1C_2 + covc_{12})^{N_{00}-Y_{00}} covc_{12}^{\alpha_{covc_{12}}-1}(u_c - covc_{12})^{\beta_{covc_{12}}-1},
\end{aligned}$$

where $u_c = \min(C_1, C_2) - C_1C_2$;

$$Y_{11} \mid \pi, S_1, S_2, C_1, C_2, covs_{12}, covc_{12}, N_{11} \sim \text{binomial}(N_{11}, p11),$$

where

$$p_{11} = \frac{\pi(S_1S_2 + covs_{12})}{\pi(S_1S_2 + covs_{12}) + (1-\pi)((1-C_1)(1-C_2) + covc_{12})};$$

$$Y_{10} \mid \pi, S_1, S_2, C_1, C_2, covs_{12}, covc_{12}, N_{10} \sim \text{binomial}(N_{10}, p_{10}),$$

where

$$p_{10} = \frac{\pi(S_1(1-S_2) - covs_{12})}{\pi(S_1(1-S_2) - covs_{12}) + (1-\pi)((1-C_1)C_2 - covc_{12})};$$

$$Y_{01} \mid \pi, S_1, S_2, C_1, C_2, covs_{12}, covc_{12}, N_{01} \sim \text{binomial}(N_{01}, p_{01}),$$

where

$$p_{01} = \frac{\pi((1-S_1)S_2 - covs_{12})}{\pi((1-S_1)S_2 - covs_{12}) + (1-\pi)(C_1(1-C_2) - covc_{12})};$$

$$Y_{00} \mid \pi, S_1, S_2, C_1, C_2, covs_{12}, covc_{12}, N_{00} \sim \text{binomial}(N_{00}, p_{00}),$$

where

$$p_{00} = \frac{\pi((1-S_1)(1-S_2) + covs_{12})}{\pi((1-S_1)(1-S_2) + covs_{12}) + (1-\pi)(C_1C_2 + covc_{12})}.$$

Full Conditional Distributions of Parameters in the Random Effects Model

For the random effects model, the full conditional distributions of the parameters are as follows:

$$p(\pi \mid d_1, \ldots, d_N, \alpha_\pi, \beta_\pi) \propto \pi^{(\Sigma_{k=1}^N d_k)+\alpha_\pi-1}(1-\pi)^{(N-\Sigma_{k=1}^N d_k)+\beta_\pi-1},$$

$$\Rightarrow \pi \mid d_1, \ldots, d_N \sim \text{beta}\left(\left(\sum_{k=1}^{N} d_k\right) + \alpha_\pi, \left(N - \sum_{k=1}^{N} d_k\right) + \beta_\pi\right),$$

$$p(d_k \mid t_{1k}, t_{2k}, \psi, i_k)$$

$$\propto \left(\pi \prod_{j=1}^{2} \Phi(a_{j1} + b_{j1} i_k)^{t_{jk}} (1 - \Phi(a_{j1} + b_{j1} i_k))^{(1-t_{jk})}\right)^{d_k}$$

$$\times \left((1-\pi) \prod_{j=1}^{2} \Phi(a_{j0} + b_{j0} i_k)^{(1-t_{jk})} (1 - \Phi(a_{j0} + b_{j0} i_k))^{t_{jk}}\right)^{(1-d_k)}$$

where $\psi = (\pi, (a_{1d_k}, b_{1d_k}, a_{2d_k}, b_{2d_k}) d_k = 0, 1)$, $k = 1, \ldots, N$;

$$\Rightarrow d_k \mid \psi, i_k \sim \text{Bernoulli}(p_k), \qquad k = 1, \ldots, N,$$

where

$$p_k = \frac{\pi \prod_{j=1}^{2} \Phi(a_{j1} + b_{j1} i_k)^{t_{jk}} (1 - \Phi(a_{j1} + b_{j1} i_k))^{(1-t_{jk})}}{Q}, \qquad k = 1, \ldots, N;$$

and

$$Q = \pi \prod_{j=1}^{2} \Phi(a_{j1} + b_{j1} i_k)^{t_{jk}} (1 - \Phi(a_{j1} + b_{j1} i_k))^{(1-t_{jk})} + (1-\pi) \prod_{j=1}^{2} \Phi(a_{j0} + b_{j0} i_k)^{(1-t_{jk})} (1 - \Phi(a_{j0} + b_{j0} i_k))^{t_{jk}},$$

$$p(a_{j1}, b_{j1} \mid (t_{jk}, d_k, i_k), k = 1, \ldots, N) \propto \prod_{k=1}^{N} \Phi(a_{j1} + b_{j1} i_k)^{d_k t_{jk}} (1 - \Phi(a_{j1} + b_{j1} i_k))^{d_k (1-t_{jk})} d\Phi(a_{j1}, b_{j1}), \qquad j = 1, 2,$$

where $d\Phi(a_{j1}, b_{j1})$ is the bivariate normal density of (a_{j1}, b_{j1});

$$p(a_{j0}, b_{j0} \mid (t_{jk}, d_k, i_k), k = 1, \ldots, N) \propto \prod_{k=1}^{N} \Phi(a_{j0} + b_{j0} i_k)^{(1-d_k)(1-t_{jk})} (1 - \Phi(a_{j0} + b_{j0} i_k))^{(1-d_k) t_{jk}} d\Phi(a_{j0}, b_{j0}), \qquad j = 1, 2,$$

where $d\Phi(a_{j0}, b_{j0})$ is the bivariate normal density of (a_{j0}, b_{j0});

$$p(i_k \mid t_{1k}, t_{2k}, a_{10}, b_{10}, a_{20}, b_{20}, a_{11}, b_{11}, a_{21}, b_{21}, d_k)$$

$$\propto \prod_{j=1}^{2} \Phi(a_{j1} + b_{j1} i_k)^{d_k t_{jk}} (1 - \Phi(a_{j1} + b_{j1} i_k))^{d_k (1-t_{jk})}$$

$$\times \prod_{j=1}^{2} \Phi(a_{j0} + b_{j0} i_k)^{(1-d_k)(1-t_{jk})} (1 - \Phi(a_{j0} + b_{j0} i_k))^{(1-d_k) t_{jk}}, \qquad k = 1, \ldots, N.$$