Published in final edited form as:

Biometrics. 2015 June ; 71(2): 538–547. doi:10.1111/biom.12264.

A unification of models for meta-analysis of diagnostic accuracy studies without a gold standard

Yulun Liu^{*}, Yong Chen[†], and Haitao Chu[‡]

^{*}Division of Biostatistics, The University of Texas School of Public Health, Houston, Texas 77030, USA

[†]Division of Biostatistics, The University of Texas School of Public Health, Houston, Texas 77030, USA

[‡]Division of Biostatistics, University of Minnesota School of Public Health, Minneapolis, Minnesota 55455, USA

Abstract

Several statistical methods for meta-analysis of diagnostic accuracy studies have been discussed in the presence of a gold standard. However, in practice, the selected reference test may be imperfect due to measurement error, non-existence, invasive nature, or expensive cost of a gold standard. It has been suggested that treating an imperfect reference test as a gold standard can lead to substantial bias in the estimation of diagnostic test accuracy. Recently, two models have been proposed to account for imperfect reference test, namely, a multivariate generalized linear mixed model (MGLMM) and a hierarchical summary receiver operating characteristic (HSROC) model. Both models are very flexible in accounting for heterogeneity in accuracies of tests across studies as well as the dependence between tests. In this paper, we show that these two models, although with different formulations, are closely related and are equivalent in the absence of study-level covariates. Furthermore, we provide the exact relations between the parameters of these two models and assumptions under which two models can be reduced to equivalent submodels. On the other hand, we show that some submodels of the MGLMM do not have corresponding equivalent submodels of the HSROC model, and vice versa. With three real examples, we illustrate the cases when fitting the MGLMM and HSROC models leads to equivalent submodels and hence identical inference, and the cases when the inferences from two models are slightly different. Our results generalize the important relations between the bivariate generalized linear mixed model and HSROC model when the reference test is a gold standard.

Keywords

Diagnostic test; Generalized linear mixed model; Hierarchical model; Imperfect reference test; Meta-analysis

Corresponding author: Yong Chen, Division of Biostatistics, The University of Texas School of Public Health, Houston, TX 77030, USA (chenyong1203@gmail.com).
equally contributed.

Supplementary Materials: Web Appendices and the SAS code referenced in Sections 5 and 6 are available with this paper at the *Biometrics* website on Wiley Online Library.

1 Introduction

The rapid growth of evidence-based medicine has led to a dramatic increase in attention to statistical methods for meta-analysis¹. One important area is meta-analysis of diagnostic accuracy studies which combines the measures of test performance (e.g., sensitivity and specificity) across multiple studies. For many applications in diagnostic tests, the test under evaluation (referred to as the index test) is compared to a perfect reference test (i.e., sensitivity=specificity=1), also known as a gold standard. When such a gold standard is available, two categories of statistical methods are popular. The first category consists of methods based on a summary receiver operating characteristic (SROC) curve generated from the study data^{2,3}. Among them, the hierarchical summary receiver operating characteristic (HSROC) model has been recommended³. The second category consists of methods that use bivariate mixed effects models to model sensitivity and specificity simultaneously^{4,5}. Among mixed effects models, the bivariate generalized linear mixed effects model (BGLMM)⁵ has been recommended for better coverage performance and avoidance of continuity correction⁶. Interestingly, Harbord *et al.*⁷ found that the BGLMM and HSROC models are closely related and even equivalent in the absence of covariates. Here, the equivalence means that both models give the same likelihood except different parametrizations. In addition, Harbord *et al.* provided the relations between parameters in both models under the assumption of reference test being a gold standard.

In this paper, we are dealing with a different setting. In practice, the reference test may be imperfect because of measurement error, non-existence, invasive nature, or expensive cost of a gold standard⁸. Despite the imperfection of reference tests in many applications, the imperfect reference tests are simply treated as gold standard tests in many analyses. Such a procedure can lead to biased estimates of diagnostic test accuracies⁹. To account for such bias, a variety of methods, including both frequentist and Bayesian methods, have been proposed. Hui and Walter¹⁰ and Walter *et al.*¹¹ have proposed methods to evaluate the diagnostic accuracy under the assumptions of homogeneous sensitivity and specificity. To fully account for the heterogeneity in test accuracies across studies, Chu *et al.*¹² proposed a flexible random effects model by modeling study-specific disease prevalence, sensitivities and specificities of the index and reference tests in a multivariate generalized linear mixed model (MGLMM) framework. Recently, Dendukuri *et al.*¹³ proposed a hierarchical summary receiver operating characteristic (HSROC) model, extending the HSROC model of Rutter and Gatsonis³ with a gold standard reference test to the situation where no gold standard test is available. This model postulates a study-specific continuous latent variable and a study-specific cutoff and accuracy values for each diagnostic test. Both the MGLMM and HSROC frameworks have the advantages that they account for heterogeneity across studies and allow for dependence between the index test and reference test.

The MGLMM framework by Chu *et al.*¹² and the HSROC framework by Dendukuri *et al.*¹³ are flexible and statistically rigorous, and are expected to be increasingly popular. In this paper, we show that these two models, although with different formulations, are closely related and some of their submodels are equivalent. We provide the exact relations between the parameters of these two models and assumptions under which two models can be reduced to equivalent submodels. On the other hand, we show that some submodels of one

framework do not have corresponding equivalent submodels in the other framework. With three examples, we illustrate a case when fitting the MGLMM and HSROC models lead to equivalent submodels and hence identical inference, and two cases when the inferences from the two models are different. Our results generalize the important relations between the BGLMM and HSROC models established by Harbord *et al.*⁷ when the reference test is a gold standard.

The contributions of this work are three-fold. First, we extend the HSROC model along the line of Dendukuri *et al.*¹³ by allowing the study-specific cutoff and accuracy values for the reference test. Second, we establish the relations between the MGLMM and HSROC frameworks, as well as their corresponding submodels. Third, with real examples, we illustrate the similarities and differences between the MGLMM and HSROC frameworks, and provide new insights on modeling based on the established relations. Throughout this paper, we consider the case when the same type of reference test is used in all studies. The cases when completely different reference tests are used in each study are not considered. In addition, the comparison between the HSROC and MGLMM models is under classical framework. Relations between model parameters and their maximum likelihood estimators under two models are derived. Comparison under Bayesian framework, which involves investigation on prior specifications and posterior distributions, is not considered in this paper.

This paper is organized as follows. Section 2 presents three real examples. We describe the MGLMM framework by Chu *et al.*¹² in Section 3, and extend the HSROC model along the line of Dendukuri *et al.*¹³ in Section 4. We establish the mathematical relationship between these two frameworks in Section 5. We illustrate the similarities and differences between two frameworks by studying three motivating examples in Section 6. A brief discussion is provided in Section 7.

2 Examples

We use three examples of meta-analyses to illustrate the similarities and differences between the MGLMM and HSROC frameworks. In this section, we briefly describe the background of these three examples, which will be revisited in Section 6.

Example 1 (Papanicolaou test for diagnosis of cervical neoplasia): Fahey, Irwig, and Macaskill¹⁴ reported data from a meta-analysis of the Papanicolaou (Pap) test that diagnoses cervical neoplasia (defined as the growth of abnormal cells on the surface on the cervix). This data is comprised of 59 studies that published between January 1984 and March 1992. The diagnostic accuracy of the Pap test (i.e., index test) is evaluated by comparing with the histology test (i.e., reference test), which is not a perfect test¹⁴.

Example 2 (Rheumatoid factor test for diagnosis of rheumatoid arthritis): Nishimura *et al.*¹⁵ collected data from Rheumatoid factor (RF) test (i.e., index test) for detection of Rheumatoid arthritis (RA), and reported 50 studies published before September 2006 with a total of 15,286 patients. In this meta-analysis, the American College of Rheumatology (ACR) 1987 revised criteria were used as the reference standard of RA. Although the ACR

1987 criteria (i.e., reference test) is widely used as an approximate ‘gold standard’ for RA classification, it may be an imperfect reference test for classification.

Example 3 (Computed tomography for diagnosis of coronary artery disease): Schuetz *et al.*¹⁶ reported data from a systematic review of the computed tomography (CT) test (i.e., index test) for diagnosis of the coronary artery disease. A total of 89 studies were included and collected through MEDLINE that were published before September 2006. In this meta-analysis, the conventional coronary angiography (CAG) test (i.e., reference test) was treated as the ‘gold standard’ for diagnosing the presence of coronary stenoses. However, CAG test may not be perfect due to measurement errors in angiography.

3 The MGLMM framework

Chu *et al.*¹² proposed an MGLMM for diagnostic tests without a gold standard. Following their notations, for study i ($i = 1, 2, \dots, D$), denote π_i as the study-specific disease prevalence, and (Se_{i1}, Sp_{i1}) and (Se_{i2}, Sp_{i2}) as the respective pairs of study-specific sensitivities and specificities for the index test T_1 and the reference test T_2 . Typically, the data for each study are summarized by a 2×2 table, cross-tabulating the test results from T_1 and T_2 . To account for the heterogeneity across studies and correlations among sensitivities and specificities of T_1 and T_2 and disease prevalence, an MGLMM is formulated in two stages. The first stage specifies the cell probabilities in the i th 2×2 table as functions of test sensitivities, specificities and disease prevalence of i th study. At the second stage, a random effects model on test sensitivities, specificities and disease prevalence is assumed.

Denote $H(\cdot)$ as the cumulative distribution function (cdf) of a continuous distribution in the location-scale family, denoted as $h(\cdot; \mu, \sigma)$ with location parameter $\mu = 0$ and scale parameter $\sigma = 1$ (such as standard logistic distribution, and standard normal distribution), and $H^{-1}(\cdot)$ as the inverse of the cdf (e.g., logit and probit functions). The study-specific test sensitivities, specificities and disease prevalence, after an H^{-1} transformation, are assumed to jointly follow a multivariate normal distribution with mean $\mathcal{M} = (P, S_1, C_1, S_2, C_2)^T$ and variance $\mathbf{V}$, where

$$\mathbf{V} = \begin{pmatrix} \sigma_P^2 & \sigma_{PS_1} & \sigma_{PC_1} & \sigma_{PS_2} & \sigma_{PC_2} \\ & \sigma_{S_1}^2 & \sigma_{S_1C_1} & \sigma_{S_1S_2} & \sigma_{S_1C_2} \\ & & \sigma_{C_1}^2 & \sigma_{C_1S_2} & \sigma_{C_1C_2} \\ & & & \sigma_{S_2}^2 & \sigma_{S_2C_2} \\ & & & & \sigma_{C_2}^2 \end{pmatrix}.$$

Here the parameters P , S_j , and C_j are the overall disease prevalence, and overall sensitivity and specificity for diagnostic test j in the transformed scale, $j = 1, 2$. In addition, the variance parameters $\sigma_P^2, \sigma_{S_1}^2, \sigma_{C_1}^2, \sigma_{S_2}^2$ and $\sigma_{C_2}^2$ describe the between-study heterogeneity in disease prevalence, sensitivities and specificities for tests T_1 and T_2 , and the parameters in the off-diagonal of $\mathbf{V}$ account for the dependence among the study-specific prevalence, sensitivities

and specificities. To avoid confusion in notations, we use English letters to represent the parameters in the MGLMM. We note that there are twenty parameters in this model.

4 The HSROC framework

Now we provide an extended HSROC framework which is closely related to the model by Dendukuri *et al.*¹³. Extending and along the lines of the pioneering work by Rutter and Gatsonis³, we assume that the test $T_j, j = 1, 2$, is based on a continuous latent variable, Z_{ij} , which comes from different distributions given different disease status D_i for patients in the i th study. Since the gold standard is absent, the disease status D_i is unknown and has to be treated as a dichotomous latent variable. We assume that the result of the diagnostic test T_j on the patients in the i th study is based on a comparison between the latent variable Z_{ij} and a study-specific “cutoff” value θ_{ij} . The diagnostic test T_j is positive if $Z_{ij} > \theta_{ij}$ and is negative otherwise. The latent variable Z_{ij} follows the location-scale distribution $h(z|\mu, \sigma) = \sigma^{-1}h((z - \mu)/\sigma|0, 1)$ with the location and scale parameters of $-a_{ij}/2$ and $\exp(-\beta_j/2)$ when $D_i = 0$, and $a_{ij}/2$ and $\exp(\beta_j/2)$ when $D_i = 1$.³ refers the study-specific a_{ij} as “accuracy value” and the parameter β_j as a “shape parameter”, because the former quantifies the distance between two possible distributions of the latent variable Z_{ij} , and the latter describes the asymmetry of the ROC curve. To complete the specification of the model, the parameters a_{ij} and θ_{ij} are assumed to be independent as in Rutter and Gatsonis³ and follow normal distributions $N(\Lambda_j, \sigma_{\Lambda_j}^2)$ and $N(\Theta_j, \sigma_{\Theta_j}^2)$ respectively. Here for the test $T_j (j = 1, 2)$, the parameters $(\Lambda_j, \sigma_{\Lambda_j}^2)$ denote the respective overall mean and between-study variation of the accuracy values, and $(\Theta_j, \sigma_{\Theta_j}^2)$ denote the mean and variation of the cutoff values. Furthermore, the disease status D_i is positive if a variable Z_{i0} from location-scale distribution $h(z|0, 1)$ is greater than a cutoff value θ_0 , and is negative otherwise, where θ_0 is distributed as $N(\Theta_0, \sigma_{\Theta_0}^2)$.

We assume the study-specific values $(\theta_0, \theta_1, a_{i1}, \theta_2, a_{i2})^T$ jointly follow a multivariate normal distribution with mean $\boldsymbol{\mu} = (\Theta_0, \Theta_1, \Lambda_1, \Theta_2, \Lambda_2)^T$ and variance $\boldsymbol{\Omega}$, where

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_{\Theta_0}^2 & \sigma_{\Theta_0\Theta_1} & \sigma_{\Theta_0\Lambda_1} & \sigma_{\Theta_0\Theta_2} & \sigma_{\Theta_0\Lambda_2} \\ & \sigma_{\Theta_1}^2 & 0 & \sigma_{\Theta_1\Theta_2} & \sigma_{\Theta_1\Lambda_2} \\ & & \sigma_{\Lambda_1}^2 & \sigma_{\Lambda_1\Theta_2} & \sigma_{\Lambda_1\Lambda_2} \\ & & & \sigma_{\Theta_2}^2 & 0 \\ & & & & \sigma_{\Lambda_2}^2 \end{pmatrix}.$$

Here the parameters Θ_0, Θ_j and Λ_j are the overall means of the prevalence cutoff value, the cutoff and accuracy values for tests T_1 and T_2 . The variance parameters $\sigma_{\Theta_0}^2, \sigma_{\Theta_1}^2, \sigma_{\Lambda_1}^2, \sigma_{\Theta_2}^2$ and $\sigma_{\Lambda_2}^2$ describe the between-study variation in the prevalence cutoff value, the cutoff and accuracy values. In addition, $\text{cov}(\theta_1, a_{i1}) = \text{cov}(\theta_2, a_{i2})$ are set as zero for model identifiability, as specified in Dendukuri *et al.*¹³. Such a specification is necessary and an empirical justification is provided in Web Appendix A. The covariance parameters $\sigma_{\Theta_0\Theta_1}, \sigma_{\Theta_0\Lambda_1}, \sigma_{\Theta_0\Theta_2}$ and $\sigma_{\Theta_0\Lambda_2}$ describe the dependence between the prevalence cutoff value θ_0

and test characteristic parameters $(\theta_{i1}, \alpha_{i1}, \theta_{i2}, \alpha_{i2})$. And the covariance parameters $\sigma_{\Theta_1\Theta_2}$, $\sigma_{\Theta_1\Lambda_2}$, $\sigma_{\Lambda_1\Theta_2}$ and $\sigma_{\Lambda_1\Lambda_2}$ describe the dependence between two tests. Hereafter, the above model is referred to as the HSROC model and Greek letters are used to represent its model parameters. We note that there are twenty parameters in the HSROC model. In particular, when Z_{ij} follows a normal distribution, the HSROC model reduces to the model considered by Dendukuri *et al.*¹³, referred to as HSROC-D. For clarification of notations of models, we listed MGLMM, HSROC and HSROC-D in Table 1.

5 Relations between the two frameworks

We establish the exact mathematical relations between parameters in the MGLMM model and those in the HSROC model in subsection 5.1, elucidate the relations under two corresponding submodels in subsection 5.2, and discuss the extension of the relations under meta-regression models in subsection 5.3.

5.1 Reference test with heterogeneous sensitivity and specificity across studies

In this subsection, we consider the situations when the sensitivity and specificity of the reference test are heterogeneous across studies. We note that the formulation of the MGLMM model is based on the study-specific prevalence, sensitivities and specificities, i.e., $(\pi_i, \text{Se}_{i1}, \text{Sp}_{i1}, \text{Se}_{i2}, \text{Sp}_{i2})$, whereas the HSROC model is based on the study-specific cutoff and accuracy values, i.e., $(\theta_{i0}, \theta_{i1}, \alpha_{i1}, \theta_{i2}, \alpha_{i2})$. The parameters of both models are the distribution parameters (mean and variance) of these random effects. Therefore, to establish the relationship between the model parameters, it is sufficient to establish the mathematical relations between the study-specific effects $(\pi_i, \text{Se}_{i1}, \text{Sp}_{i1}, \text{Se}_{i2}, \text{Sp}_{i2})$ and $(\theta_{i0}, \theta_{i1}, \alpha_{i1}, \theta_{i2}, \alpha_{i2})$.

Recall that the latent variable Z_{i1} follows the location-scale distribution $h(z|\mu, \sigma) = \sigma^{-1}h((z - \mu)/\sigma|0, 1)$ with the mean and scale parameters of $\alpha_{i1}/2$ and $\exp(\beta_1/2)$ when $D_i = 1$. By standardizing Z_{i1} , we have

$$\begin{aligned} \Pr(Z_{i1} \geq \theta_{i1}|D_i \\ = 1) &= \Pr\{(Z_{i1} \\ &- \alpha_{i1}/2)\exp(\\ &- \beta_1/2) \geq (\theta_{i1} - \alpha_{i1}/2)\exp(\\ &- \beta_1/2)|D_i \\ &= 1\} = H\{-(\theta_{i1} - \alpha_{i1}/2)\exp(-\beta_1/2)\}, \end{aligned}$$

where $H(\cdot)$ is the cdf of $h(\cdot|0, 1)$. Similarly, $\Pr(Z_{i1} < \theta_{i1}|D_i = 0) = H\{(\theta_{i1} + \alpha_{i1}/2)\exp(\beta_1/2)\}$. Under the HSROC model, the study-specific disease prevalence, sensitivity and specificity of T_1 in the i th study, in the transformed scale, can be calculated as

$$\begin{aligned}
H^{-1}(\pi_i) &= H^{-1}\{\Pr(D_i=1)\} = H^{-1}\{\Pr(Z_{i0} \geq \theta_{i0})\} = H^{-1}\{H(-\theta_{i0})\} = -\theta_{i0}, \\
H^{-1}(\text{Se}_{i1}) &= H^{-1}\{\Pr(T_{i1}=1|D_i=1)\} = H^{-1}\{\Pr(Z_{i1} \geq \theta_{i1}|D_i=1)\} = -(\theta_{i1} - \alpha_{i1}/2)\exp(-\beta_1/2), \\
H^{-1}(\text{Sp}_{i1}) &= H^{-1}\{\Pr(T_{i1}=0|D_i=0)\} = H^{-1}\{\Pr(Z_{i1} < \theta_{i1}|D_i=0)\} = (\theta_{i1} + \alpha_{i1}/2)\exp(\beta_1/2).
\end{aligned}$$

Similarly, the study-specific sensitivity and specificity of T_2 can be calculated as $H^{-1}(\text{Se}_{i2}) = -(\theta_{i2} - \alpha_{i2}/2)\exp(-\beta_2/2)$ and $H^{-1}(\text{Sp}_{i2}) = (\theta_{i2} + \alpha_{i2}/2)\exp(\beta_2/2)$.

Denote $b_1 = \exp(\beta_1/2)$ and $b_2 = \exp(\beta_2/2)$. The relations between $(\pi_i, \text{Se}_{i1}, \text{Sp}_{i1}, \text{Se}_{i2}, \text{Sp}_{i2})$ and $(\theta_{i0}, \theta_{i1}, \alpha_{i1}, \theta_{i2}, \alpha_{i2})$ can be written in matrix form as

$$H^{-1} \begin{pmatrix} \pi_i \\ \text{Se}_{i1} \\ \text{Sp}_{i1} \\ \text{Se}_{i2} \\ \text{Sp}_{i2} \end{pmatrix} = \mathbf{S}^{-1} \begin{pmatrix} \theta_{i0} \\ \theta_{i1} \\ \alpha_{i1} \\ \theta_{i2} \\ \alpha_{i2} \end{pmatrix}, \text{ where } \mathbf{S}^{-1} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -b_1^{-1} & \frac{1}{2}b_1^{-1} & 0 & 0 \\ 0 & b_1 & \frac{1}{2}b_1 & 0 & 0 \\ 0 & 0 & 0 & -b_2^{-1} & \frac{1}{2}b_2^{-1} \\ 0 & 0 & 0 & b_2 & \frac{1}{2}b_2 \end{pmatrix}. \quad (1)$$

By taking the expectation and variance of both sides of equation (1), we obtain the following relationship between the parameters in the MGLMM framework and the HSROC framework

$$\mathcal{M} = \mathbf{S}^{-1} \cdot \mathcal{H}, \quad (2)$$

$$\mathbf{V} = \mathbf{S}^{-1} \cdot \mathbf{\Omega} \cdot (\mathbf{S}^{-1})^T, \quad (3)$$

where $\mathcal{M} = (P, S_1, C_1, S_2, C_2)^T$ is the mean vector of disease prevalence, sensitivities and specificities of the two tests in the transformed scale (i.e., H^{-1} -transformation), and $\mathcal{H} = (\Theta_0, \Theta_1, \Lambda_1, \Theta_2, \Lambda_2)^T$ is the mean vector of prevalence cutoff value, cutoff values and accuracy values of the two tests. By equations (2) and (3), we can write the MGLMM model parameters as functions of parameters in the HSROC model, and vice versa by multiplying the matrix $\mathbf{S}$ (or $\mathbf{S}^T$) to both hands of equation (2) and to the left and right of matrices in equation (3). The detailed results of such mathematical relations between two sets of 20 model parameters are provided in Web Appendix B. Here we only highlight some interesting findings.

The first interesting result is $\beta_j = 0.5 \log(\sigma_{C_j}^2 / \sigma_{S_j}^2)$ for $j = 1, 2$. The shape parameter β_j , which characterizes the asymmetry of the ROC curve for test T_j , is determined solely by the ratio of variances of sensitivity and specificity of the test in the transformed scale. Recall the equation (4.11) of Harbord *et al.*⁷ under the assumption of a gold standard reference test reveals the same finding. Our result generalizes the previous finding where the reference test is a gold standard. Secondly, by equation (3), we have

$$\sigma_{S_j}^2 = \sigma_{C_j}^2 = 0 \iff \sigma_{\Theta_j}^2 = \sigma_{\Lambda_j}^2 = 0, \text{ for } j=1, 2. \quad (4)$$

This confirms our intuition that the homogeneity in sensitivity and specificity of the test T_j under the MGLMM framework is equivalent to the homogeneity in cutoff and accuracy values of that test. Thirdly, equation (3) implies

$$\sigma_{PS_j} = \sigma_{PC_j} = 0 \iff \sigma_{\Theta_0\Theta_j} = \sigma_{\Theta_0\Lambda_j} = 0, \text{ for } j=1, 2. \quad (5)$$

This agrees with our intuition. If disease prevalence is independent of sensitivity and specificity of the test j under the MGLMM framework, then cutoff value for the prevalence is also independent of cutoff and accuracy values of that test under the HSROC framework, and vice versa. Lastly, we have

$$\sigma_{S_1S_2} = \sigma_{C_1S_2} = \sigma_{S_1C_2} = \sigma_{C_1C_2} = 0 \iff \sigma_{\Theta_1\Theta_2} = \sigma_{\Lambda_1\Lambda_2} = \sigma_{\Theta_1\Lambda_2} = \sigma_{\Lambda_1\Theta_2} = 0. \quad (6)$$

The relation (6) justifies that the independence assumption between two tests can be equivalently made by imposing the constraints on $(\sigma_{S_1S_2}, \sigma_{C_1S_2}, \sigma_{S_1C_2}, \sigma_{C_1C_2})$ in the MGLMM framework and on $(\sigma_{\Theta_1\Theta_2}, \sigma_{\Lambda_1\Lambda_2}, \sigma_{\Theta_1\Lambda_2}, \sigma_{\Lambda_1\Theta_2})$ in the HSROC framework.

5.2 Reference test with homogeneous sensitivity and specificity across studies

In this subsection, we consider the submodel of the MGLMM when sensitivity and specificity of the reference test are homogeneous across studies, i.e., $\sigma_{S_2}^2 = \sigma_{C_2}^2 = 0$, and disease prevalence is independent of sensitivity and specificity of the test T_1 , i.e., $\sigma_{PS_1} = \sigma_{PC_1} = 0$. By equations (4) and (5), the corresponding submodel of the HSROC model can be obtained by letting $\sigma_{\Theta_2}^2 = \sigma_{\Lambda_2}^2 = 0$ and $\sigma_{\Theta_0\Theta_1} = \sigma_{\Theta_0\Lambda_1} = 0$. We denote both reduced submodels by MGLMM_R and HSROC_R. Both submodels have 7 parameters and are summarized in the lower panel of Table 1. The functional relations between parameters of these two submodels can be represented as

$$\begin{array}{cc} \text{MGLMM}_R \Rightarrow \text{HSROC}_R & \text{HSROC}_R \Rightarrow \text{MGLMM}_R \\ \left\{ \begin{array}{l} \Theta_0 = -P \\ \Theta_1 = -\frac{1}{2} \left(\sigma_{C_1}^2 / \sigma_{S_1}^2 \right)^{1/4} S_1 + \frac{1}{2} \left(\sigma_{S_1}^2 / \sigma_{C_1}^2 \right)^{1/4} C_1 \\ \Lambda_1 = \left(\sigma_{C_1}^2 / \sigma_{S_1}^2 \right)^{1/4} S_1 + \left(\sigma_{S_1}^2 / \sigma_{C_1}^2 \right)^{1/4} C_1 \\ \beta_1 = \frac{1}{2} \log \left(\sigma_{C_1}^2 / \sigma_{S_1}^2 \right) \\ \sigma_{\Theta_0}^2 = \sigma_P^2 \\ \sigma_{\Theta_1}^2 = \frac{1}{2} \left\{ \left(\sigma_{C_1}^2 \sigma_{S_1}^2 \right)^{1/2} - \sigma_{S_1C_1} \right\} \\ \sigma_{\Lambda_1}^2 = 2 \left\{ \left(\sigma_{C_1}^2 \sigma_{S_1}^2 \right)^{1/2} + \sigma_{S_1C_1} \right\} \end{array} \right. & \left\{ \begin{array}{l} P = -\Theta_0 \\ S_1 = - \left(\Theta_1 - \frac{1}{2} \Lambda_1 \right) \exp(-\beta_1/2) \\ C_1 = \left(\Theta_1 + \frac{1}{2} \Lambda_1 \right) \exp(\beta_1/2) \\ \sigma_P^2 = \sigma_{\Theta_0}^2 \\ \sigma_{S_1}^2 = \left(\sigma_{\Theta_1}^2 + \frac{1}{4} \sigma_{\Lambda_1}^2 \right) \exp(-\beta_1) \\ \sigma_{C_1}^2 = \left(\sigma_{\Theta_1}^2 + \frac{1}{4} \sigma_{\Lambda_1}^2 \right) \exp(\beta_1) \\ \sigma_{S_1C_1} = -\sigma_{\Theta_1}^2 + \frac{1}{4} \sigma_{\Lambda_1}^2 \end{array} \right. \end{array} \quad (7)$$

The functional relations in equations (7) will be empirically validated through a meta-analysis for diagnosis of cervical neoplasia in Section 6.1. We note that the relations in equations (7) are identical to the main results in Harbord *et al.*⁷ after a reparametrization where the reference test is a gold standard, i.e., $S_2 = C_2 = 1$. It is interesting that exact relations hold even when the reference test is not a gold standard, as long as the sensitivity

and specificity of the reference test are homogeneous across studies, and the sensitivity and specificity of the index test are independent of disease prevalence. For details of connections between the results in equation (7) and the main results in Harbord *et al.*⁷, please refer to Web Appendix C.

5.3 Extensions to models with covariates

In some meta-analyses, study-level covariates such as the study quality, race of the study population and type of recruitment (e.g., family-based versus otherwise) are available. Including study-level covariates can reduce unexplained heterogeneity and correlations. In general, different, but possibly overlapping, covariates may be allowed to affect the study-specific disease prevalence, and sensitivities and specificities of diagnostic test differently in the MGLMM model and to affect the study-specific cutoff and accuracy values differently in the HSROC model. By a similar argument as in Section 5.1 and in Section 4.2 of Harbord *et al.*⁷, we can show that the MGLMM model with common covariates affecting prevalence, sensitivities and specificities is equivalent to a HSROC model with the same covariates affecting both accuracy and cutoff values. The relations between variance parameters in both models are the same as described in equation (3) and the relations between coefficient parameters in both models are provided in Web Appendix D. However, the MGLMM model with different covariates affecting prevalence, sensitivities and specificities is not equivalent to any HSROC model with covariates.

6 Similarities and differences: examples revisited

In this section, we revisit the three examples described in Section 2. We illustrate a case when fitting the MGLMM and HSROC models leads to equivalent submodels and hence identical inference, and two cases when the inferences from two models are slightly different. We also use example 1 to verify the derived relations between the MGLMM_R and HSROC_R submodels as described in equation (7). For each of the three examples, we fit submodels in both MGLMM and HSROC frameworks, and conduct model selection based on two commonly used information criteria, Akaike's information criterion (AIC) and the Bayesian information criterion (BIC)¹⁷. For all three examples, the transformation $H^{-1}(\cdot)$ is taken as logit-function, acknowledging that other link functions may also be considered. The model implementation is through fitting the non-linear mixed effects model using PROC NLMIXED via the adaptive Gaussian quadrature approximation to the likelihood integrated over the random effects in SAS version 9.3 (SAS Institute Inc., Cary, NC). The selection process is summarized in Table 2, including the $-2 \log$ likelihood statistic, AIC and BIC for the MGLMM framework (top panel) and the HSROC framework (bottom panel) respectively.

6.1 Example 1

To analyze the data on the Pap test and histology test, we begin with a fixed effect model (submodel I), and then add any random effect to improve the goodness of fit under all criteria. The results in the third column of Table 2 suggest that the largest improvement is achieved by allowing for the study-specific prevalence (referred to as submodel IIa). Interestingly, such submodel IIa is identical to the model considered by Walter *et al.*¹¹,

which only allows random disease prevalence but not random sensitivities and specificities of both tests. In other words, the model considered by Walter *et al.*¹¹ is the best one random component model. Finally, both AIC and BIC suggest the use of submodel IVe (i.e., MGLMM_R discussed in Section 5.2) which includes the random effects on prevalence, both sensitivity and specificity of the Pap test, and fixed effects on sensitivity and specificity of the histology test (reference test). Following a similar model selection procedure in HSROC framework, the submodel 4b (i.e., HSROC_R discussed in Section 5.2) is selected with allowing for the study-specific prevalence cutoff values, and study-specific cutoff and accuracy values of the Pap test. Therefore, in this example, the best fitted submodels are in fact equivalent according to our results in Section 5.2. To verify that both submodels provide the same inference and validate our derived relations in equation (7), we calculate the estimates and standard errors of parameters in the submodel IVe from results of the submodel 4b using the relations in equation (7), and vice versa. Table 3 presents the parameter estimates obtained from both submodels, and the results of applying equation (7) to transform estimates from one submodel to the other. As shown in Table 3, the results of different parametrizations of two models are identical. In this case, either of the two submodels can be used without any discrepancy.

The summary estimates of overall prevalence, sensitivity and specificity based on MGLMM_R model are computed by taking inverse logit transforms of P , S_1 and C_1 estimates. As a result, the estimates of overall disease prevalence is 0.64 (95% CI: 0.54 to 0.73), and the overall sensitivity and specificity of the Pap test are estimated as 0.65 (95% CI: 0.57 to 0.74) and 0.83 (95% CI: 0.77 to 0.90). The sensitivity and specificity of the histology test are estimated as 0.90 (95% CI: 0.88 to 0.93) and 0.99 (95% CI: 0.96 to 1.00) respectively. Furthermore, the covariance between sensitivity and specificity of Pap test, $\sigma_{S_1 C_1}$, is estimated to be negative, as would be expected due to the trade-off between sensitivity and specificity when the cutoff value varies across studies. In contrast,¹¹ fitted a model (i.e., submodel IIa) with fixed sensitivities and specificities for both Pap and histology tests but with study-specific disease prevalence. The estimated sensitivity and specificity for the Pap test are 0.75 (95% CI: 0.74 to 0.76) and 0.79 (95% CI: 0.78 to 0.81), and that for the histology test are 0.86 (95% CI: 0.84 to 0.88) and 0.90 (95% CI: 0.88 to 0.92), respectively. Both AIC and BIC suggest that the MGLMM_R model (or equivalently, the HSROC_R model) provides a much better fit of the data compared to the submodel IIa in Walter *et al.*¹¹ (likelihood ratio statistic between these two nested submodels = 2,130, $p < 0.001$), suggesting non-negligible heterogeneity in study-specific sensitivity and specificity of Pap test across studies. Specifically, AIC and BIC for the MGLMM_R model (or equivalently, the HSROC_R model) are 39,770 and 39,789 respectively, whereas AIC and BIC for the submodel IIa in Walter *et al.*¹¹ are 41,894 and 41,906, which are substantially larger.

6.2 Examples 2 and 3

Table 2 also presents the results from model selection procedure applied to examples 2 and 3. For example 2, both AIC and BIC suggest that submodel 4b provides the best fit in the HSROC framework, suggesting that both cutoff and accuracy values of Rheumatoid factor (RF) test should be considered as random effects across studies. In contrast, the

corresponding submodel IVe with random effects for prevalence, sensitivity and specificity of RF test, and their correlation in MGLMM framework does not provide the best fit. The submodel IVb without the correlation provides a slightly better fit. In fact, a likelihood ratio test comparing the submodels IVb and IVe suggests that incorporating the correlation between sensitivity and specificity of the RF test does not improve the model fit (p -value = 0.08). In this case, submodel IVb is the best model under MGLMM framework, and there is no corresponding submodel under HSROC framework. The parameter estimates and standard errors of the best fitted submodels IVb and 4b are displayed in Table 4. To enable a direct comparison, we calculate the estimates of disease prevalence and sensitivities and specificities of RF test through results from submodel 4b using the equation (1). It so happens that results from model IVb and 4b are similar despite them not being equivalent. This is because the submodel 4b is equivalent to submodel IVe, which only differs from IVb by a correlation parameter. It is worthy mentioning that both submodels yield estimates of sensitivity and specificity of the ACR 1987 criteria (reference test) being 1, suggesting that such test is in fact a gold standard.

In example 3, the best fitted submodels are the submodel IVd with random effects for prevalence, sensitivities for both CT and CAG tests, and their correlation under the MGLMM framework, and the submodel 4f with random effects for prevalence, accuracy of CT test and cutoff of CAG test, and their correlation under the HSROC framework. The parameter estimates and standard errors of submodels IVd and 4f are summarized in Table 4, which show slight differences between these two submodels.

7 Discussion

Multivariate meta-analysis is gaining its popularity recently, especially in meta-analysis of diagnostic accuracy studies¹. In diagnostic accuracy studies, the reference test may be imperfect because it is subject to measurement error or a gold standard is not available in practice. In this paper, we established the equivalence between two recently proposed models that account for the imperfect reference test. Exact relations between the parameters of the two models are established and are empirically validated by an example of meta-analysis for the Papanicolaou test for diagnosis of cervical neoplasia. As we have seen in this example, although seemingly very different parametrizations of these two models, both models lead to equivalent submodels, and hence identical inferences. On the other hand, with two other examples of meta-analysis, we illustrated that there are some differences between two frameworks. In practice, the complexity of the models that should be considered depends on whether the reference test is a gold standard, the number of studies in the meta-analysis, and the degree of heterogeneity of the studies. The choice between MGLMM and HSROC models can be based on the nature of the available data. As suggested by subsection 10.5.4 of the Cochrane handbook¹⁸, when the available studies used a common cut-off value on a continuous or ordinal scale for defining test positivity (e.g. in a commercial test), then the MGLMM model can provide an appropriate framework for test comparisons; if the included studies used different cut-off values for defining positive results, then the HSROC model is a recommended approach. In this paper, we did not consider the situation where the two tests may be conditional dependent given the latent

disease status and study-specific random effects. Further study is needed for the relations between two models under such conditional dependence.

Both the MGLMM and HSROC models, as implemented in SAS PROC NLMIXED, involve maximizing an approximation to the likelihood integrated over the multi-dimensional random effects. Implicitly, the NLMIXED procedure approximates the integrated likelihood function by dual quasi-Newton optimization techniques. When study-specific random effects are allowed for both index and reference tests, the likelihood function involves five dimensional integrals and the maximum likelihood inference may suffer from non-convergence and the approximation to the likelihood may have non-negligible errors, which can result in unstable or unreproducible estimates⁶. Bayesian methods and Monte Carlo Markov Chain techniques with proper priors may circumvent these numerical issues. As pointed out by the associate editor and an anonymous referee, it would be of interest to investigate the relations between two prior specifications, and between two matching posteriors under the MGLMM and HSROC frameworks for future research. Finally, we want to emphasize that the model frameworks considered in this paper are two of many possible formulations. It is worthy to study other model frameworks for diagnostic accuracy studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the editor Jeanine Houwing-Duistermaat, the Associate Editor and two anonymous referees for their constructive comments which have greatly improved the presentation of this paper. Yong Chen was supported by grant number R03HS022900 from the AHRQ. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. Haitao Chu was supported in part by the U.S. AHRQ R03HS020666, the U.S. NIAID AI103012 and the NCI P01CA142538. We want to thank Stacia DeSantis and Jose-Miguel Yamal for their helpful comments.

References

1. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. *Statistics in Medicine*. 2011; 30(20):2481–2498. [PubMed: 21268052]
2. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary roc curve: Data-analytic approaches and some additional considerations. *Statistics in Medicine*. 1993; 12(14):1293–1316. [PubMed: 8210827]
3. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*. 2001; 20(19):2865–2884. [PubMed: 11568945]
4. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*. 2002; 21(4):589–624. [PubMed: 11836738]
5. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology*. 2006; 59(12):1331. [PubMed: 17098577]
6. Hamza TH, Reitsma JB, Stijnen T. Meta-analysis of diagnostic studies: A comparison of random intercept, normal-normal, and binomial-normal bivariate summary ROC approaches. *Medical Decision Making*. 2008; 28(5):639–649. [PubMed: 18753684]
7. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*. 2007; 8(2):239–251. [PubMed: 16698768]

8. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology*. 1995; 141(3): 263–272. [PubMed: 7840100]
9. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *Canadian Medical Association Journal*. 2006; 174(4):469–476. [PubMed: 16477057]
10. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. *Biometrics*. 1980; 36(1):167–171. [PubMed: 7370371]
11. Walter S, Irwig L, Glasziou P, et al. Meta-analysis of diagnostic tests with imperfect reference standards. *Journal of Clinical Epidemiology*. 1999; 52(10):943. [PubMed: 10513757]
12. Chu H, Chen S, Louis TA. Random effects models in a meta-analysis of the accuracy of two diagnostic tests without a gold standard. *Journal of the American Statistical Association*. 2009; 104(486):512–523. [PubMed: 19562044]
13. Dendukuri N, Schiller I, Joseph L, Pai M. Bayesian Meta-Analysis of the Accuracy of a Test for Tuberculous Pleuritis in the Absence of a Gold Standard Reference. *Biometrics*. 2012; 68(4): 1285–1293. [PubMed: 22568612]
14. Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. *American Journal of Epidemiology*. 1995; 141(7):680–689. [PubMed: 7702044]
15. Nishimura K, Sugiyama D, Kogata Y, Tsuji G, Nakazawa T, Kawano S, et al. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. *Annals of Internal Medicine*. 2007; 146(11):797–808. [PubMed: 17548411]
16. Schuetz GM, Zacharopoulou NM, Schlattmann P, Dewey M. Meta-analysis: noninvasive coronary angiography using computed tomography versus magnetic resonance imaging. *Annals of Internal Medicine*. 2010; 152(3):167–177. [PubMed: 20124233]
17. Burnham, KP.; Anderson, DR. Model selection and multi-model inference: a practical information-theoretic approach. Springer; 2002.
18. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy: Chapter 10 Analysing and Presenting Results*. Version 10 The Cochrane Collaboration. 2010

Table 1

Summary of the MGLMM and HSROC models, and their submodels.

Model	Reference tests	Latent variables and their distributions	Random effects and their distributions
HSROC-D (Derdikari et al., 2012)	common cutoff and accuracy value	$Z_{i1} \sim \begin{cases} N\left(\frac{\alpha_{i1}}{2}, \exp\left(\frac{\beta_1}{2}\right)\right) & \text{if } D_i=1, \\ N\left(-\frac{\alpha_{i1}}{2}, \exp\left(-\frac{\beta_1}{2}\right)\right) & \text{if } D_i=0, \end{cases}$ where β_1 : scale parameter; $i = 1, \dots, I$	$\begin{pmatrix} \theta_{i1} \\ \alpha_{i1} \end{pmatrix} \sim N\left(\begin{pmatrix} \Theta_1 \\ \Lambda_1 \end{pmatrix}, \begin{pmatrix} \sigma_{\Theta_1}^2 & 0 \\ 0 & \sigma_{\Lambda_1}^2 \end{pmatrix}\right)$
Extended Models HSROC $\equiv$ MGLMM			
HSROC	study-specific cutoff and accuracy values	$Z_{i1} \sim \begin{cases} h\left(\frac{\alpha_{i1}}{2}, \exp\left(\frac{\beta_1}{2}\right)\right) & \text{if } D_i=1, \\ h\left(-\frac{\alpha_{i1}}{2}, \exp\left(-\frac{\beta_1}{2}\right)\right) & \text{if } D_i=0, \end{cases}$ $Z_{i2} \sim \begin{cases} h\left(\frac{\alpha_{i2}}{2}, \exp\left(\frac{\beta_2}{2}\right)\right) & \text{if } D_i=1, \\ h\left(-\frac{\alpha_{i2}}{2}, \exp\left(-\frac{\beta_2}{2}\right)\right) & \text{if } D_i=0, \end{cases}$ where $h(\cdot; \cdot)$: location-scale family; β_1, β_2 : scale parameters; $i = 1, \dots, I$	$\begin{pmatrix} \theta_{i0} \\ \theta_{i1} \\ \alpha_{i1} \\ \theta_{i2} \\ \alpha_{i1} \end{pmatrix} \sim N\left(\begin{pmatrix} \Theta_0 \\ \Theta_1 \\ \Lambda_1 \\ \Theta_2 \\ \Lambda_2 \end{pmatrix}, \begin{pmatrix} \sigma_{\Theta_0}^2 & \sigma_{\Theta_0\Theta_1} & \sigma_{\Theta_0\Lambda_1} & \sigma_{\Theta_0\Theta_2} & \sigma_{\Theta_0\Lambda_2} \\ \sigma_{\Theta_1}^2 & \sigma_{\Theta_1\Theta_2} & 0 & \sigma_{\Theta_1\Theta_2} & \sigma_{\Theta_1\Lambda_2} \\ \sigma_{\Lambda_1}^2 & \sigma_{\Lambda_1\Theta_2} & \sigma_{\Lambda_1}^2 & \sigma_{\Lambda_1\Theta_2} & \sigma_{\Lambda_1\Lambda_2} \\ \sigma_{\Theta_2}^2 & \sigma_{\Theta_2}^2 & \sigma_{\Theta_2}^2 & \sigma_{\Theta_2}^2 & 0 \\ \sigma_{\Lambda_2}^2 & \sigma_{\Lambda_2}^2 & \sigma_{\Lambda_2}^2 & \sigma_{\Lambda_2}^2 & \sigma_{\Lambda_2}^2 \end{pmatrix}\right)$
MGLMM	study-specific sensitivity and specificity	NA	$H^{-1} \begin{pmatrix} \pi_i \\ Se_{i1} \\ Sp_{i1} \\ Se_{i2} \\ Sp_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} p \\ S_1 \\ C_1 \\ S_2 \\ C_2 \end{pmatrix}, \begin{pmatrix} \sigma_P^2 & \sigma_{PS_1} & \sigma_{PC_1} & \sigma_{PS_2} & \sigma_{PC_2} \\ \sigma_{S_1}^2 & \sigma_{S_1}^2 & \sigma_{S_1C_1} & \sigma_{S_1S_2} & \sigma_{S_2C_2} \\ \sigma_{C_1}^2 & \sigma_{C_1}^2 & \sigma_{C_1}^2 & \sigma_{C_1S_2} & \sigma_{C_1C_2} \\ \sigma_{S_2}^2 & \sigma_{S_2}^2 & \sigma_{S_2}^2 & \sigma_{S_2}^2 & \sigma_{S_2C_2} \\ \sigma_{C_2}^2 & \sigma_{C_2}^2 & \sigma_{C_2}^2 & \sigma_{C_2}^2 & \sigma_{C_2}^2 \end{pmatrix}\right)$
Reduced Models HSROCr $\equiv$ MGLMMr			
HSROCr	common cutoff and accuracy values	$Z_{i1} \sim \begin{cases} h\left(\frac{\alpha_{i1}}{2}, \exp\left(\frac{\beta_1}{2}\right)\right) & \text{if } D_i=1, \\ h\left(-\frac{\alpha_{i1}}{2}, \exp\left(-\frac{\beta_1}{2}\right)\right) & \text{if } D_i=0, \end{cases}$ where $h(\cdot; \cdot)$: location-scale family; β_1 : scale parameters; $i = 1, \dots, I$	$\begin{pmatrix} \theta_{i0} \\ \theta_{i1} \\ \alpha_{i1} \end{pmatrix} \sim N\left(\begin{pmatrix} \Theta_0 \\ \Theta_1 \\ \Lambda_1 \end{pmatrix}, \begin{pmatrix} \sigma_{\Theta_0}^2 & 0 & 0 \\ \sigma_{\Theta_1}^2 & 0 & 0 \\ \sigma_{\Lambda_1}^2 & \sigma_{\Lambda_1}^2 & \sigma_{\Lambda_1}^2 \end{pmatrix}\right)$
MGLMMr	common sensitivity and specificity	NA	$H^{-1} \begin{pmatrix} \pi_i \\ Se_{i1} \\ Sp_{i1} \end{pmatrix} \sim N\left(\begin{pmatrix} p \\ S_1 \\ C_1 \end{pmatrix}, \begin{pmatrix} \sigma_P^2 & 0 & 0 \\ \sigma_{S_1}^2 & \sigma_{S_1C_1}^2 & \sigma_{S_1C_1}^2 \\ \sigma_{C_1}^2 & \sigma_{C_1}^2 & \sigma_{C_1}^2 \end{pmatrix}\right)$

Table 2

Selection of random effects using a forward selection procedure. The upper panel shows the results from fitting submodels in the MGLMM framework; and the lower panel shows the results from fitting submodels in the HSROC framework.

Models	Random effects	Example 1: Pap test			Example 2: Nishimura-RF			Example 3: Schuetz		
		-2logL	AIC	BIC	-2logL	AIC	BIC	-2logL	AIC	BIC
I	NA	45277	45287	45297	36062	36072	36082	17359	17369	17382
IIa	π_i	41882	41894	41906	33623	33635	33646	16509	16521	16537
IIb	Se_{i1}	43329	43341	43353	35480	35492	35504	17235	17247	17263
IIc	Se_{i2}	43398	43410	43423	34700	34712	34723	17064	17076	17092
IId	Sp_{i1}	43520	43532	43544	35527	35539	35551	17035	17047	17063
IIf	Sp_{i2}	42838	42850	42863	34582	34594	34605	17166	17178	17194
IIIa	π_i, Se_{i1}	40510	40524	40539	33133	33147	33161	16391	16405	16424
IIIb	π_i, Se_{i2}	40888	40902	40917	33109	33123	33137	16243	16257	16276
IIIc	π_i, Sp_{i1}	40894	40908	40922	33103	33117	33130	16247	16261	16280
IIId	π_i, Sp_{i2}	40520	40534	40548	33139	33153	33167	16408	16422	16441
IVa	π_i, Se_{i1}, Se_{i2}	39777	39793	39810	32629	32645	32661	16122	16138	16159
IVb	π_i, Se_{i1}, Sp_{i1}	39762	39778	39795	32624	32640	32655	16124	16140	16162
IVc	π_i, Se_{i1}, Sp_{i2}	40506	40522	40539	33138	33154	33169	16390	16406	16428
IVd	$\pi_i, Se_{i1}, Se_{i2}, \rho Se_{i1} Se_{i2}$	39777	39795	39814	32629	32647	32664	16113	16131	16155
IVe	$\pi_i, Se_{i1}, Sp_{i1}, \rho Se_{i1} Sp_{i1}$	39752	39770	39789	32621	32639	32656	16115	16133	16157
IVf	$\pi_i, Se_{i1}, Sp_{i2}, \rho Se_{i1} Sp_{i2}$	40503	40521	40540	33133	33151	33169	16390	16408	16432
I ($\equiv$ I)	NA	45277	45291	45306	36062	36076	36090	17359	17373	17392
2a ($\equiv$ IIa)	θ_0	42085	42101	42118	33623	33639	33654	16510	16526	16547
2b	θ_1	43106	43122	43139	35355	35371	35386	17082	17098	17120
2c	θ_2	42704	42720	42736	34338	34354	34370	17063	17079	17100
2d	α_1	43329	43345	43361	35480	35496	35511	17009	17025	17046
2e	α_2	43398	43414	43431	34700	34716	34731	17063	17079	17101
3a	θ_0, θ_1	39921	39939	39958	33008	33026	33043	16243	16261	16285
3b	θ_0, θ_2	39934	39952	39971	32876	32894	32911	16259	16277	16301

Models	Random effects	Example 1: Pap test			Example 2: Nishimura-RF			Example 3: Schuetz		
		-2logL	AIC	BIC	-2logL	AIC	BIC	-2logL	AIC	BIC
3c	θ_0, a_{11}	40667	40685	40704	33455	33473	33491	16207	16225	16249
3d	θ_0, a_{12}	40880	40898	40916	33113	33131	33148	16246	16264	16288
3e	θ_{11}, a_{11}	42904	42922	42940	30940	30958	30972	16905	16923	16947
4a	$\theta_0, \theta_{11}, \theta_{12}$	39842	39862	39883	32728	32748	32767	16145	16165	16192
4b ($\equiv$ IVe)	$\theta_0, \theta_{11}, a_{11}$	39752	39772	39793	32621	32641	32660	16121	16141	16168
4c	$\theta_0, \theta_{11}, a_{12}$	39776	39796	39816	32629	32649	32668	16138	16158	16185
4d	$\theta_0, \theta_{11}, \theta_{12}, \rho\theta_{11}\theta_{12}$	39773	39795	39818	32681	32703	32724	16145	16167	16197
4e	$\theta_0, \theta_{11}, a_{12}, \rho\theta_{11}a_{12}$	39772	39794	39816	32637	32659	32680	16123	16145	16174
4f	$\theta_0, a_{11}, \theta_{12}, \rho a_{11}\theta_{12}$	39755	39777	3979916	32627	32649	32670	16115	16137	16166

Table 3Results of fitting the MGLMM_R and HSROC_R models to the Papanicolaou (Pap) test data.

Parameter	Estimate (SE) from MGLMM _R model	Results of applying equation (7) to HSROC _R estimates below
P	0.56 (0.21)	0.56 (0.21)
S_1	0.64 (0.18)	0.64 (0.18)
C_1	1.62 (0.23)	1.62 (0.23)
σ_P^2	2.15 (0.48)	2.15 (0.48)
$\sigma_{S_1}^2$	1.67 (0.35)	1.67 (0.35)
$\sigma_{C_1}^2$	1.61 (0.42)	1.61 (0.42)
$\alpha_{S_1 C_1}$	-0.88 (0.31)	-0.88 (0.31)
	Estimate (SE) from HSROC _R model	Results of applying equation (7) to MGLMM _R estimates above
Θ_o	-0.56 (0.21)	-0.56 (0.21)
Θ_1	0.50 (0.18)	0.50 (0.18)
Λ_1	2.27 (0.25)	2.27 (0.25)
β_1	-0.02 (0.15)	-0.02 (0.15)
$\sigma_{\Theta_o}^2$	2.15 (0.48)	2.15 (0.48)
$\sigma_{\Theta_1}^2$	1.26 (0.28)	1.26 (0.28)
$\sigma_{\Lambda_1}^2$	1.51 (0.45)	1.51 (0.45)

Table 4

Summary of the parameter estimates (standard errors) of the final fitted models to Examples 2 and 3 in the MGLMM and HSROC frameworks.

	Example 2: Nishimura-RF		Example 3: Schuetz	
	MGLMM	HSROC	MGLMM	HSROC
	IVb	4b	IVd	4f
S_1	0.68 (0.03)	0.68 (0.03)	0.96 (0.01)	0.96 (0.01)
C_1	0.87(0.02)	0.88(0.02)	0.97 (0.01)	0.97 (0.02)
S_2	1.00 (-)	1.00 (-)	0.91 (0.01)	0.92 (0.02)
C_2	1.00 (-)	1.00 (-)	1.00 (-)	1.00 (-)
P	0.44 (0.03)	0.44 (0.03)	0.64 (0.02)	0.64 (0.02)
σ_P	0.72 (0.08)	NA	0.86 (0.07)	NA
σ_{S_1}	0.83 (0.10)	NA	1.08 (0.13)	NA
σ_{C_1}	1.04 (0.12)	NA	NA	NA
σ_{S_2}	NA	NA	1.02 (0.13)	NA
$\rho_{S_1 S_2}$	NA	NA	0.46 (0.13)	NA
σ_{Θ_0}	NA	0.73 (0.08)	NA	0.86 (0.07)
σ_{Θ_1}	NA	0.74 (0.08)	NA	NA
σ_{Λ_1}	NA	1.14 (0.13)	NA	1.52 (0.51)
σ_{Θ_2}	NA	NA	NA	1.97 (3.01)
$\rho_{\Lambda_1 \Theta_2}$	NA	NA	NA	-0.31 (0.18)