7

Estimating Diagnostic Error without a Gold Standard: A Mixed Membership Approach

Elena A. Erosheva

Department of Statistics, University of Washington, Seattle, WA 98195-4320, USA

Cyrille Joutard

Institut de Mathématiques et de Modélisation de Montpellier & Université Montpellier 3, Montpellier Cedex 5, France

CONTENTS

Evaluation of sensitivity and specificity of diagnostic tests in the absence of a gold standard typically relies on latent structure models. For example, two extensions of latent class models in the biostatistics literature, Gaussian random effects (Qu et al., 1996) and finite mixture (Albert and Dodd, 2004), form the basis of several recent approaches to estimating sensitivity and specificity of diagnostic tests when no (or partial) gold standard evaluation is available. These models attempt to account for additional item dependencies that cannot be explained with traditional latent class models, where the classes typically correspond to healthy and diseased individuals.

We propose an alternative latent structure model, namely, the extended mixture Grade of Membership (GoM) model, for evaluation of diagnostic tests without a gold standard. The extended mixture GoM model allows for test results to be dependent on latent degree of disease severity, while also allowing for the presence of some individuals with deterministic response patterns such as all-positive and all-negative test results. We formulate and estimate the model in a hierarchical Bayesian framework. We use a simulation study to compare recovery of true sensitivity and specificity parameters with the extended mixture GoM model, and the latent class, Gaussian random effects, and finite mixture models.

Our findings indicate that when the true generating model contains deterministic mixture components and the sample size is large, all four models tend to underestimate sensitivity and overestimate specificity parameters. These results emphasize the need for sensitivity analyses in real life applications when the data generating model is unknown. Employing a number of latent structure models and examining how the assumptions on latent structure affect conclusions about accuracy of diagnostic tests is a crucial step in analyzing test performance without a gold standard. We illustrate the sensitivity analysis approach using data on screening for *Chlamydia trachomatis*. This example

demonstrates that the extended mixture GoM model not only provides us with new latent structure and the corresponding interpretation to mechanisms that give rise to test results, but also provides new insights for estimating test accuracy without a gold standard.

7.1 Introduction

We consider the problem of estimating sensitivity and specificity of diagnostic or screening tests when results are available from multiple fallible tests but not from gold standard. This could happen when gold standard assessment doesn't exist or when economic or ethical issues in administering the gold standard prevent one from doing so.

Latent class analysis (Lazarsfeld and Henry, 1968; Goodman, 1974) has been at the core of model-based methods for analyzing diagnostic errors in the absence of a gold standard (Hui and Zhou, 1998; Albert and Dodd, 2004; Pepe and Janes, 2007). Recently, two extensions of latent class models, known as Gaussian random effects (Qu et al., 1996) and finite mixture (Albert and Dodd, 2004), have produced several new approaches to estimating sensitivity and specificity of diagnostic tests when no (or partial) gold standard evaluation is available (Hadgu and Qu, 1998; Albert and Dodd, 2004; 2008; Albert, 2007b). See Hui and Zhou (1998) for a comprehensive review of the earlier literature on evaluating diagnostic tests without gold standards. Pepe and Janes (2007) criticize latent class models as a tool for analyzing diagnostic test performance because of the lack of links between biological mechanisms giving rise to test results and dependencies induced by a structure of the model. For example, they assert that most diseases are not dichotomous but occur in varying degrees of severity. Hence, latent class models that employ discrete disease status as a latent variable cannot account for additional correlations induced by disease severity such as occurrences of false negatives for persons with mild disease. One example that Pepe and Janes (2007) provide talks about detection of a particular substance in a biological sample where the amount of substance affects all test results.

The best way to evaluate the performance of tests with unknown characteristics is, undoubtedly, to have at least a partial gold standard assessment (Albert and Dodd, 2008; Albert, 2007a). In the absence of a gold standard, however, having an arsenal of model-based methods can be informative for evaluating sensitivity of scientific conclusions regarding accuracy of diagnostic and screening tests. Because the true data generating mechanism is typically not known, Albert and Dodd (2004) (p. 433) recommends performing sensitivity analysis by using different models: "Although biological plausibility may aid the practitioner in favoring one model over another, a range of estimates from various models of diagnostic error (as well as standard errors) should be reported."

We present an alternative latent structure model for the analysis of test performance when no gold standard is available. Our model is an extension of the Grade of Membership model. The GoM model employs a degree of disease severity as a latent variable, therefore inducing a mixed membership latent structure where individuals can be members of diseased and healthy classes at the same time. This type of latent structure addresses the concerns of Pepe and Janes (2007). We extend the GoM model to obtain the extended mixture GoM model, analogous to the extended finite mixture model by Muthen and Shedden (1999) and the finite mixture model by Albert and Dodd (2004). The extended mixture GoM model allows for a mixture of deterministic and mixed membership responses. For example, some truly positive individuals may have deterministic positive response on every test while others may be subject to diagnostic error according to the GoM model. A version of this model has previously been applied in disability studies (Erosheva et al., 2007), but the extended mixture GoM model is new to the literature on diagnostic testing.

The remainder of the chapter is organized as follows. In Section 7.2 we review latent class (Lazarsfeld and Henry, 1968; Goodman, 1974), latent class random effects (also known as

Gaussian random effects) (Qu et al., 1996), and finite mixture models (Albert and Dodd, 2004) that are commonly used for analysis of diagnostic and screening tests. In Section 7.3 we introduce the GoM model, develop the extended mixture GoM model that allows for deterministic responses, discuss a hierarchical Bayesian framework for estimation of model parameters, and derive sensitivity and specificity estimates. In Section 7.4 we conduct a simulation study examining recovery of specificity and sensitivity parameters under the latent class, the latent class random effects, the finite mixture, and the extended mixture GoM models. We investigate performance of each of these models when the true data-generating model is known, varying the true model among the four alternatives. Our findings further emphasize the need for sensitivity analyses when no gold standard is available. In Section 7.5 we illustrate such a sensitivity analysis using a publicly available dataset on screening for *Chlamydia trachomatis* (CT) (Hadgu and Qu, 1998) . Finally, in Section 7.6 we relate results from our analyses of simulated and real data to prior findings in the literature.

7.2 Overview of Existing Model-Based Approaches

Sensitivity and specificity are key accuracy parameters of diagnostic and screening tests. The general framework for estimating diagnostic errors without a gold standard starts by assuming a latent structure model and then deriving sensitivity and specificity parameters that correspond to the model formulation. This section introduces a common notation and presents a concise overview of latent class (Lazarsfeld and Henry, 1968; Goodman, 1974), latent class Gaussian random effects (Qu et al., 1996), and finite mixture models (Albert and Dodd, 2004) that are commonly used for analysis of diagnostic and screening tests. For simplicity of the exposition, we omit the subject index.

Let $x = (x_1, x_2, \ldots, x_J)$ be a vector of dichotomous variables, where x_j takes on values $l_j \in \mathcal{L}_j = \{0, 1\}$, $j = 1, 2, \ldots, J$. Let $\mathcal{X} = \prod_{j=1}^{J} \mathcal{L}_j$ be the set of all possible outcomes l for vector x. Denote a positive test result by $x_j = 1$ and a negative result by $x_j = 0$. Denote the disease indicator by δ, with $\delta = 1$ standing for the presence of the disease. Let $\tau = P(\delta = 1)$ denote the disease prevalence parameter for the population of interest.

The latent class approach assumes two classes, the healthy and the sick. The probability to observe response pattern l is the weighted sum of probabilities to observe l from each latent class:

$$\mathrm{P}(x = l) = \mathrm{P}(x = l|\delta = 1)\mathrm{P}(\delta = 1) + \mathrm{P}(x = l|\delta = 0)\mathrm{P}(\delta = 0), \;\; l \in \mathcal{X}.$$

The tests are assumed to be conditionally independent given the true disease status. Test result x_j is a Bernoulli random variable with class conditional probabilities $\lambda_{1j} = P(x_j = 1|\delta = 1)$ and $\lambda_{2j} = P(x_j = 1|\delta = 0)$ for a given true disease status. The conditional probabilities $\lambda_{1j}, \lambda_{2j}, j = 1, \ldots, J$ and the weight $\mathrm{P}(\delta = 1) = 1 - \mathrm{P}(\delta = 0) = \tau$ are the model parameters.

For the jth diagnostic test, its sensitivity is the probability of the positive test result given that the true diagnosis is positive, $\mathrm{P}(x_j = 1|\delta = 1)$, and its specificity is the probability of a negative response given that the true diagnosis is negative, $\mathrm{P}(x_j = 0|\delta = 0) = 1 - \mathrm{P}(x_j = 1|\delta = 0)$. The sensitivity and specificity of test j implied by the latent class model are then simply

$$\mathrm{P}(x_j = 1|\delta = 1) = \lambda_{1j}$$

and

$$\mathrm{P}(x_j = 0|\delta = 0) = 1 - \lambda_{2j}.$$

The Gaussian random effects model of Qu et al. (1996) is an attempt to relax the assumption of independence conditional on the true disease status. This model assumes that test outcomes are independent Bernoulli realizations with probabilities given by the standard normal cdf $\mathbf{\Phi}\left(\beta_{j\delta} + \sigma_\delta b\right)$,

where $\beta_{j\delta}, \delta = 0, 1; j = 1, \ldots, J$ are latent class parameters and b is an individual-specific standard Normal random effect. Under this latent class Gaussian random effects model,

$$\mathrm{P}(x = l|\delta) = \left\{ \int \prod_j \mathbf{\Phi} \left(\beta_{j\delta} + \sigma_\delta b\right)^{l_j} \left(1 - \mathbf{\Phi} \left(\beta_{j\delta} + \sigma_\delta b\right)\right)^{1-l_j} \right\} \phi(b)db,$$

where $\phi(b)$ is the standard normal density. The sensitivity and specificity for test j under the latent class Gaussian random effects model are then

$$\mathrm{P}(x_j = 1|\delta = 1) = \mathbf{\Phi} \left(\frac{\beta_{j1}}{(1 + \sigma_1^2)^{1/2}} \right)$$

and

$$\mathrm{P}(x_j = 0|\delta = 0) = 1 - \mathbf{\Phi} \left(\frac{\beta_{j0}}{(1 + \sigma_0^2)^{1/2}} \right),$$

respectively.

The finite mixture model (Albert and Dodd, 2004) also uses the two-class structure as its basis and adds two point masses for the combinations of all-zero and all-one responses. These point masses correspond to the healthiest and the most severely diseased patients that are always classified correctly. Let t be an indicator that denotes correct classification. Specifically, let $t = 0$ if a healthy subject is always classified correctly (i.e., has the all-zero response pattern with J tests), $t = 1$ if a diseased subject is always classified correctly, and let $t = 2$ otherwise. Thus, subjects are either always classified correctly, when either $t = 0$ or $t = 1$, or a diagnostic error is possible when $t = 2$. Denote the probabilities for correctly classifying diseased and healthy subjects by $\eta_1 = \mathrm{P}(t = 1)$ and $\eta_0 = \mathrm{P}(t = 0)$, respectively. Let also $w_j(\delta_i)$ denote the probability of the jth test making a correct diagnosis when $t = 2$.

The finite mixture model of Albert and Dodd (2004) assumes that the test results x_j are independent Bernouilli random variables, conditional on the true disease status and the classification indicator. Thus,

$$\mathrm{P}(x_j = 1|\delta, t) = \begin{cases} w_j(1), & \text{if } \delta = 1, \text{and } t = 2 \\ 1, & \text{if } \delta = 1, \text{and } t = 1 \\ 1 - w_j(0), & \text{if } \delta = 0, \text{and } t = 2 \\ 0, & \text{if } \delta = 0, \text{and } t = 0. \end{cases}$$

Note that $\mathrm{P}(x_j = 1|\delta = 1, t = 0) = \mathrm{P}(x_j = 1|\delta = 0, t = 1) = 0$. The specificity and sensitivity of the jth test under the finite mixture model are then

$$\mathrm{P}(x_j = 1|\delta = 1) = \eta_1 + (1 - \eta_1)w_j(1)$$

and

$$\mathrm{P}(x_j = 0|\delta = 0) = \eta_0 + (1 - \eta_0)w_j(0),$$

respectively.

7.3 A Mixed Membership Approach to Estimating Diagnostic Error

The GoM model can be thought of as a different extension of latent class models where random effects are individual-specific grades of membership (Erosheva, 2005). The extended GoM mixture

model combines individuals of mixed membership with those of full membership who have predetermined response patterns. Although the extended mixture GoM model allows for an arbitrary choice of the number and the nature of deterministic response patterns, it is reasonable to assume two deterministic responses in the medical testing context. Analogous to the approach of Albert and Dodd (2004), we use two deterministic components in the extended mixture GoM model to allow for inclusion of some healthy and diseased individuals who have deterministic responses with the all-zero and all-one patterns, respectively. However, to model tests' diagnostic errors for other subjects in the population, our approach is to use the GoM model while the finite mixture model of Albert and Dodd (2004) relies on using the two-class latent class model to model diagnostic errors.

Next, we describe the GoM model before introducing the extended mixture GoM model and deriving a Bayesian estimation algorithm for the extension.

7.3.1 The Grade of Membership Model

As before, let $x = (x_1, x_2, \ldots, x_J)$ be a vector of dichotomous variables, where x_j takes on values $l_j \in \mathcal{L}_j = \{0, 1\}$, $j = 1, 2, \ldots, J$. Let K be the number of mixture components (extreme profiles) in the GoM model. To preserve generality, we will provide notation and estimation algorithms for an arbitrary value of K. However, in the medical testing context, we will assume $K = 2$ to be consistent with the existing literature.

Let $g = (g_1, g_2, \ldots, g_K)$ be a latent partial membership vector of K nonnegative random variables that sum to 1. In what follows, we use notation $p(\ )$ to refer to both probability density and probability mass functions. Each extreme profile is characterized by a vector of conditional response probabilities for manifest variables, given that the kth component of the partial membership vector is 1 and the others are zero, $\lambda_{kj} = p(x_j = 1|g_k = 1),\ k = 1, 2, \ldots, K;\ j = 1, 2, \ldots, J$. Given partial membership vector $g \in [0, 1]^K$, the conditional distribution of manifest variable x_j is given by a convex combination of the extreme profiles' response probabilities, i.e., $p(x_j = 1|g) = \sum_{k=1}^{K} g_k \lambda_{kj}, j = 1, 2, \ldots, J$. Let us denote the distribution of g by $D(g)$. The local independence assumption states that manifest variables are conditionally independent, given the latent variables. Using this assumption and integrating out latent variable g, we obtain the marginal distribution for response pattern l in the form of a continuous mixture

$$p(x = l) = \int \prod_{j=1}^{J} \left(\sum_{k=1}^{K} g_k \lambda_{kj}^{l_j} (1 - \lambda_{kj})^{1-l_j} \right) dD(g),\ l \in \mathcal{X},$$

where $\mathcal{X} = \prod_{j=1}^{J} \mathcal{L}_j$ is the set of all possible outcomes for vector x.

The latent class representation of the GoM model leads naturally to a data augmentation approach (Tanner, 1996). Denote by $\mathbf{x}$ the matrix of observed responses x_{ij} for all subjects. Let $\boldsymbol{\lambda}$ denote the matrix of conditional response probabilities. Augment the observed data for each subject with realizations of the latent classification variables $z_i = (z_{i1}, \ldots, z_{iJ})$. Denote by $\mathbf{z}$ the matrix of latent classifications z_{ij}. Let $z_{ijk} = 1$, if $z_{ij} = k$ and $z_{ijk} = 0$ otherwise.

We assume the distribution of membership scores is Dirichlet with parameters α. The joint probability model for the parameters and augmented data is

$$p(\mathbf{x}, \mathbf{z}, \mathbf{g}, \boldsymbol{\lambda}, \alpha) = p(\boldsymbol{\lambda}, \alpha) \prod_{i=1}^{N} \left[p(z_i|g_i) p(x_i|\boldsymbol{\lambda}, z_i) \cdot Dir(g_i|\alpha) \right],$$

where

$$p(z_i|g_i) = \prod_{j=1}^{J} \prod_{k=1}^{K} g_{ik}^{z_{ijk}},\ \ p(x_i|\boldsymbol{\lambda}, z_i) = \prod_{j=1}^{J} \prod_{k=1}^{K} \left(\lambda_{kj}^{x_{ij}} (1 - \lambda_{kj})^{1-x_{ij}} \right)^{z_{ijk}}$$

and

$$Dir(g_i|\alpha) = \frac{\mathbf{\Gamma}(\sum_k \alpha_k)}{\mathbf{\Gamma}(\alpha_1)...\mathbf{\Gamma}(\alpha_K)} g_{i1}^{\alpha_1 - 1} \dots g_{iK}^{\alpha_K - 1}.$$

We assume the prior on extreme profile response probabilities $\boldsymbol{\lambda}$ is independent of the prior on the hyperparameters α. We further assume that the prior distribution of extreme profile response probabilities treats items and extreme profiles as independent, hence $p(\boldsymbol{\lambda}, \alpha) = p(\alpha) \prod_{k=1}^{K} \prod_{j=1}^{J} p(\lambda_{kj})$. We assume the prior $p(\lambda_{kj})$ is $Beta(1, 1)$. Estimation of the GoM model can be done via a Metropolis-Hastings within Gibbs algorithm as described by Erosheva (2003).

7.3.2 The Extended Mixture GoM Model

To define the extended mixture GoM model, we assume two patterns of deterministic responses that correspond to the all-zero and all-one test results. Similar to Albert and Dodd (2004), we introduce the classification indicator variable t. Let $t = 0$ for the healthiest individuals ($\delta = 0$) who are always classified correctly; let $t = 1$ for the sick individuals ($\delta = 1$) who are always classified correctly; and let $t = 2$ for the other individuals whose distribution of test results is given by the GoM model with parameters α, λ. Denote the respective weights for the multinomial distribution of t by $\theta = (\theta_0, \theta_1, \theta_2)$. The interpretation of θ_0 and θ_1 is similar to that of η_0 and η_1 in the finite mixture model of Albert and Dodd (2004); we are using a different notation symbol to emphasize that the values of those parameters will be different due to differences between the models.

Note that parameter estimation for the extended mixture GoM model would be identical to the estimation for the standard GoM model if we could modify the observed counts for the all-zero and all-one responses by subtracting the numbers of individuals who are always classified correctly. However, these numbers are typically unknown which means that we have to estimate weights of the deterministic components.

To derive the Markov chain Monte Carlo (MCMC) sampling algorithm for the extended mixture GoM model, we further augment data with individual classification indicators. Let N be the total number of individuals in the sample, and let $n_0^{(m)}$ and $n_1^{(m)}$ be the expected values of the all-zero cell count and the all-one cell count, respectively, for the mixed membership individuals (with $t = 2$) at the m-th iteration. Denote the number of individuals with at least one positive and at least one zero response in their response pattern by n_{mix}. The total number of individuals with $t = 2$ at the mth iteration is then $n_{GoM}^{(m)} = n_0^{(m)} + n_1^{(m)} + n_{mix}$. Let the prior distribution for weights θ be uniform on the simplex and update θ at the end of the posterior step with:

$$\theta_0^{(m+1)} = \theta_0^{(m)} + \frac{n_0^{(m)} - n_0^{(m+1)}}{N}, \quad \theta_1^{(m+1)} = \theta_1^{(m)} + \frac{n_1^{(m)} - n_1^{(m+1)}}{N},$$

and

$$\theta_2^{(m+1)} = \frac{n_0^{(m+1)} + n_1^{(m+1)} + n_{mix}}{N} = 1 - \theta_0^{(m+1)} - \theta_1^{(m+1)}.$$

Given the number of individuals subject to classification error, $n_{GoM}^{(m)}$, the estimation of model parameters for the stochastic GoM compartment is identical to that used in the case of the standard GoM model. We use a reparameterization of $\alpha = (\alpha_1, \dots, \alpha_K)$ with $\xi = (\xi_1, \dots, \xi_K)$ and α_0, which reflect proportions of the item responses that belong to each mixture category and the spread of the membership distribution. The closer α_0 is to zero, the more probability is concentrated near the mixture categories; similarly, the larger α_0 is, the more probability is concentrated near the population average membership score. We assume that α_0 and ξ are independent since they govern two unrelated qualities of the distribution of the GoM scores. In the absence of a strong prior opinion about hyperparameters α_0 and ξ, we take the prior distribution $p(\xi)$ to be uniform on the simplex

and $p(\alpha_0)$ to be a proper diffuse gamma distribution. We also assume that the prior distribution on the GoM scores is independent of the prior distribution on the structural parameters. The joint distribution of the parameters and augmented data for the mixed membership component is

$$p(\boldsymbol{\lambda})p(\alpha_0)p(\xi)\left(\prod_{i=1}^{n_{GoM}} D(g_i|\alpha)\right)\prod_{i=1}^{n_{GoM}}\prod_{j=1}^{J}\prod_{k=1}^{K}\left(g_{ik}\lambda_{kj}^{x_{ij}}(1-\lambda_{kj})^{1-x_{ij}}\right)^{z_{ijk}},$$

where z_{ijk} is the latent class indicator as before. We sample from the posterior distribution of $\xi = (\xi_1, \ldots, \xi_K)$ and α_0 by using the Gibbs sampler with two Metropolis-Hastings steps (see Erosheva, 2003). The modified sampling algorithm for the extended mixture GoM model can be easily generalized to a number of deterministic response patterns greater than two.

7.3.3 Sensitivity and Specificity with the Extended Mixture GoM Model

Here we derive sensitivity and specificity estimates under the extended mixture GoM model. As before, let us denote the true diagnosis of a subject by $\delta = 1$ or $\delta = 0$ for the presence or absence of the disease, respectively. If patient i has the disease, then, under the extended mixture GoM model, this patient either belongs to the deterministic compartment with the clear positive diagnosis, $t = 1$, or they belongs to the stochastic compartment with the classification indicator $t = 2$. In terms of probability, this translates into $\mathrm{P}(\delta = 1) = \theta_1 + \theta_2\xi_1$. As a consequence, the sensitivity of item j can be expressed as follows :

$$\begin{aligned}\mathrm{P}(x_j = 1|\delta = 1) &= \mathrm{P}(x_j = 1, t = 1|\delta = 1) + Pr(x_j = 1, t = 2|\delta = 1)\\ &= [\mathrm{P}(t = 1) + \mathrm{P}(x_j = 1, t = 2, \delta = 1)]\,/\mathrm{P}(\delta = 1).\end{aligned}$$

Noticing that

$$\mathrm{P}(x_j = 1, t = 2, \delta = 1) = \mathrm{P}(x_j = 1|t = 2, \delta = 1)\mathrm{P}(\delta = 1|t = 2)\mathrm{P}(t = 2),$$

we obtain a parametric form for sensitivity of item j under the extended mixture GoM model:

$$\mathrm{P}(x_j = 1|\delta = 1) = \frac{\theta_1}{\theta_1 + \theta_2\xi_1} + \lambda_{1j}\frac{\theta_2\xi_1}{\theta_1 + \theta_2\xi_1}.$$

Similarly, the absence of the disease for patient i means that either i belongs to the deterministic compartment, $t = 0$ of a clear negative diagnosis, or he/she belongs to the stochastic compartment with classification indicator $t = 2$. Therefore, $\mathrm{P}(\delta = 0) = \theta_0 + \theta_2\xi_2$ and we obtain

$$\begin{aligned}\mathrm{P}(x_j = 1|\delta = 0) &= \mathrm{P}(x_j = 1, t = 0|\delta = 0) + \mathrm{P}(x_j = 1, t = 2|\delta = 0)\\ &= \mathrm{P}(x_j = 1, t = 0|\delta = 0) + \mathrm{P}(x_j = 1, t = 2, \delta = 0)/\mathrm{P}(\delta = 0).\end{aligned}$$

Because $\mathrm{P}(x_j = 1, t = 0|\delta = 0) = 0$, we have

$$\mathrm{P}(x_j = 1|\delta = 0) = \lambda_{2j}\frac{\theta_2\xi_2}{\theta_0 + \theta_2\xi_2},$$

and the specificity estimate of item j under the extended mixture GoM model can be obtained as $1 - \mathrm{P}(x_j = 1|\delta = 0)$.

7.4 Simulation Study

In this section, we present a simulation study with the primary aim to examine recovery of sensitivity and specificity parameters under the four different latent structure models: the latent class (Lazarsfeld and Henry, 1968; Goodman, 1974), the latent class random effects (Qu et al., 1996), the finite

mixture (Albert and Dodd, 2004), and the extended mixture GoM models introduced earlier. We investigate performance of each of these models when the true model is known, varying the true model among the four alternatives under two sample sizes, $N = 1000$ and $N = 4000$. We also report the comparative fit of the models in each case, however, model fit was not a primary goal of our study. Earlier work demonstrated difficulties in distinguishing between models with different dependence structures (Albert and Dodd, 2004), and pointed out that even equally well-fitting models may result in different accuracy estimates (Begg and Metz, 1990).

For the simulations we considered $J = 6$ and set the true specificities and sensitivities to be the same for all 6 items. Specifically, we used the value of 0.9 for the sensitivity parameter and 0.95 for the specificity. This setting allowed us to examine the recovery of accuracy parameters for a given model by simply computing the respective average sensitivity and specificity estimates across all items. and the hyperparameter was set at $\alpha_0 = 0.25$

We selected data generating designs under each model to reflect important features of biomedical screening and diagnostic data. Most noticeably, contingency tables formed on the basis of this type of data often contain many zeros and small observed cell counts but also have several large cell counts. The large observed cell counts typically include the all-zero and the all-one response patterns. The following parameter choices produced simulated data with many zeros and large all-zero and all-one counts and items with 0.9 sensitivity and 0.95 specificity:

1. For data generated under the latent class model, we chose: $\tau = 0.1$ and $\lambda_{1j} = 0.9$, $\lambda_{2j} = 0.05$, for all j.

2. For data generated under the latent class random effects model, we chose: $\sigma_0 = \sigma_1 = 1.5$, $\tau = 0.1$, and $\beta_{j0} = -2.965$, $\beta_{j1} = 2.31$, for all j.

3. For data generated under the finite mixture model, we chose: $\tau = 0.1$, $\eta_0 = 0.2$, $\eta_1 = 0.5$, and $w_j(0) = 0.9375$, $w_j(1) = 0.8$, for all j.

4. For data generated under the extended mixture Grade of Membership model, we chose the following parameter values: $\theta = (0.85, 0.05, 0.10)$, $\alpha = (0.02, 0.06)$, and $\lambda_{1j} = 0.7$, $\lambda_{2j} = 0.6$, for all j.

Among the four models, the latent class is the least complex with 13 independent parameters; the latent class random effects and the finite mixture models both have 15 independent parameters, and the extended mixture GoM model is the most complex with 16 independent parameters. We used BUGS (Bayesian inference using Gibbs sampling) to estimate the latent class, latent class random effects, and finite mixture models. We used a C code for estimation of the extended mixture GoM model.

Tables 7.1–7.4 report posterior means and standard errors of the sensitivity and specificity parameters, averaged over the six items for each model value of the log-likelihood, as well as goodness-of-fit criteria. We report the log-likelihood, the G^2 likelihood ratio criteria (Bishop et al., 1975), and the truncated sum of squared Pearson residuals (SSPR) χ^2 (Erosheva et al., 2007) computed for observed counts larger than 1 (i.e., the sum did not include residuals for the cells with zero observed counts). The log-likelihood and the goodness-of-fit criteria were evaluated as the posterior means of the parameters for each model.

TABLE 7.1
Results for the LCM generating model.

			N=1000	
Criterion	LCM	LCRE	FM	ExtM-GoM
SSPR χ^2_{tr}	23.44	22.68	18.47	22.14
G^2	30.24	31.11	34.59	54.54
Log-likelihood	-1528.19	-1528.62	-1530.36	-1545.39
Sensitivity	0.894 (0.032)	0.890 (0.033)	0.893 (0.032)	0.910 (0.029)
Specificity	0.951 (0.007)	0.951 (0.007)	0.950 (0.007)	0.953 (0.006)
			N=4000	
Criterion	LCM	LCRE	FM	ExtM-GoM
SSPR χ^2_{tr}	49.42	44.61	40.90	178.23
G^2	55.73	61.70	56.19	280.96
Log-likelihood	-6418.15	-6421.13	-6418.38	-6525.79
Sensitivity	0.902 (0.015)	0.895 (0.016)	0.901 (0.016)	0.913 (0.009)
Specificity	0.948 (0.004)	0.949 (0.004)	0.948 (0.004)	0.953 (0.005)

TABLE 7.2
Results for the LCRE generating model.

			N=1000	
Criterion	LCM	LCRE	FM	ExtM-GoM
SSPR χ^2_{tr}	308.75	54.98	36.45	34.55
G^2	232.28	62.60	62.02	59.57
Log-likelihood	-1301.07	-1216.24	-1215.95	-1221.95
Sensitivity	0.838 (0.031)	0.886 (0.040)	0.878 (0.039)	0.905 (0.010)
Specificity	0.969 (0.005)	0.958 (0.009)	0.971 (0.007)	0.976 (0.004)
			N=4000	
Criterion	LCM	LCRE	FM	ExtM-GoM
SSPR χ^2_{tr}	710.79	60.92	65.74	61.13
G^2	514.76	73.25	76.32	62.11
Log-likelihood	-5283.37	-5062.62	-5064.15	-5069.51
Sensitivity	0.868 (0.016)	0.938 (0.013)	0.876 (0.019)	0.936 (0.008)
Specificity	0.973 (0.003)	0.955 (0.005)	0.970 (0.003)	0.976 (0.005)

TABLE 7.3
Results for the FM generating model.

			N=1000	
Criterion	LCM	LCRE	FM	ExtM-GoM
SSPR χ^2_{tr}	159.19	55.07	43.91	43.40
G^2	100.70	71.37	75.01	80.61
Log-likelihood	-1517.72	-1503.05	-1504.87	-1513.80
Sensitivity	0.928 (0.027)	0.898 (0.053)	0.907(0.034)	0.915 (0.021)
Specificity	0.949 (0.007)	0.953 (0.008)	0.950 (0.008)	0.954 (0.006)
			N=4000	
Criterion	LCM	LCRE	FM	ExtM-GoM
SSPR χ^2_{tr}	109.61	44.53	36.38	93.08
G^2	91.77	53.30	43.89	108.38
Log-likelihood	-5287.40	-5268.16	-5263.46	-5292.61
Sensitivity	0.838 (0.019)	0.841 (0.023)	0.842 (0.020)	0.864 (0.011)
Specificity	0.969 (0.003)	0.968 (0.003)	0.968 (0.003)	0.969 (0.002)

TABLE 7.4
Results for the extended mixture GoM generating model.

			N=1000	
Criterion	LCM	LCRE	FM	ExtM-GoM
SSPR χ^2_{tr}	174.34	48.77	30.94	35.70
G^2	138.55	80.85	66.67	62.61
Log-likelihood	-887.45	-858.59	-851.50	-854.46
Sensitivity	0.791 (0.033)	0.843 (0.050)	0.857 (0.067)	0.835 (0.023)
Specificity	0.998 (0.002)	0.998 (0.004)	0.974 (0.014)	0.979(0.002)
			N=4000	
Criterion	LCM	LCRE	FM	ExtM-GoM
SSPR χ^2_{tr}	394.06	193.19	42.32	41.57
G^2	328.31	270.28	53.14	51.48
Log-likelihood	-3463.84	-3443.51	-3326.26	-3330.02
Sensitivity	0.772 (0.018)	0.874 (0.019)	0.789 (0.024)	0.858 (0.006)
Specificity	0.999 (0.001)	0.969 (0.003)	0.991 (0.005)	0.962 (0.002)

A first remark concerning the simulation results is that the fit criteria do not always favor a generating model. For example, perhaps not surprisingly, the truncated sum of squares of Pearson's residuals tends to be better for the finite mixture and the extended mixture GoM model. These models provide perfect fit for the two largest observed counts, even when they are not the true data-generating models. In general, however, we observe that different latent structure models can produce similar fit; this finding confirms that it can be difficult to distinguish between models with different dependence structures (Albert and Dodd, 2004).

Examining bias for the sensitivity and specificity estimates, we observe that when the latent class model is used to generate the data, all of the models successfully recover the accuracy parameters for both cases, $N = 1000$ and $N = 4000$.

The success in recovery of the accuracy parameters is not that great when more complex models are used to generate the data. Thus, when the latent class random effects model generates the data, even the true model recovers only the specificity parameter, but not the sensitivity parameter. For the latent class random effects as the generating model, the extended mixture GoM does best in recovering the sensitivity for the smaller sample size, but the finite mixture model does best in recovering the sensitivity for the larger sample size. When the finite mixture model is the generating model, all models perform well in recovering the sensitivity and specificity for the $N = 1000$ case, however, all models perform poorly for the $N = 4000$ case. In the latter scenario, we see that all the models considered, including the true model, underestimate the sensitivity and overestimate the specificity parameters. When the extended mixture GoM model is the generating model, we observe that the latent class random effects model does better with recovering the true value of the sensitivity parameter, even though the fit of this model is not as good compared to others. We also observe that all models tend to overestimate the specificity parameter for data generated with the extended mixture GoM model.

Finally, we observe that in the most difficult cases, the sensitivity parameter was underestimated to various degree by all of the models and the specificity—overestimated by all of the models. In such cases, models that provide larger sensitivity estimates and smaller specificity estimates could be considered especially informative in a sensitivity analysis with respect to latent structure assumptions. We also note that the sensitivity and especially specificity estimates under the extended mixture GoM model had smaller standard errors as compared to those of the other models, independently of the true model.

7.5 Analysis of *Chlamydia trachomatis* Data

In this section, we provide sensitivity analysis for a dataset on testing for *Chlamydia trachomatis*, originally considered by Hadgu and Qu (1998). *Chlamydia trachomatis* (CT) is the most common sexually transmitted bacterial infection in the U.S. The data contain binary outcomes on $J = 6$ tests for 4,583 women where positive responses correspond to detection of the disease. The six tests are Syva-DFA, Syva-EIA, Abbott-EIA, GenProbe, Sanofi-EIA, and a culture test (for more information, see Hadgu and Qu, 1998). Table 7.5 provides response patterns with positive observed counts. In our analyses, we follow Hadgu and Qu (1998), who only retained complete response patterns and assumed that patterns with zero observed counts were sampling and not structural zeros.

We assumed two basis latent classes and analyzed the CT data with the latent class, the latent class random effects, the finite mixture, and the extended mixture GoM models. For comparative purposes, we also provide results obtained with the GoM model, although we did not expect it to perform well based on our earlier experience with disability data (Erosheva et al., 2007).

To obtain draws from the joint posterior distribution, we used the same prior and proposal distributions for the standard GoM as well as for the extended mixture GoM models. We chose Gamma

as the prior distribution on α_0, with the shape and the inverse scale parameter equal to 1, and chose the uniform prior for ξ-parameters. We chose the shape parameter for the proposal distribution on α_0 to be equal to $C_1 = 100$, and the sum of the parameters of the proposal distribution for ξ to be equal to $C_2 = 1$. We set starting values for λ to the estimated conditional response probabilities from the latent class model with two classes, and selected the starting value for hyperparameters α to be $(0.001, 0.099)$.

We monitored the convergence of MCMC chains using Geweke convergence diagnostics (Geweke, 1992), and Heidelberger and Welch stationarity and interval halfwidth tests (Heidelberger and Welch, 1983). Furthermore, we visually examined plots of successive iterations. With our choices of starting values and parameters for prior and proposal distributions, all these methods indicated favorable convergence of MCMC chains for both the standard and the extended mixture GoM models.

Under the extended mixture GoM model, the estimated proportion θ_0 of women belonging to the deterministic compartment of a clear negative diagnosis was $\hat{\theta_0} = 0.917$ (sd = 0.009). The observed probability of all-zero response was 0.944. The estimated proportion θ_1 of women belonging to the deterministic compartment of a clear positive diagnosis was $\hat{\theta_1} = 0.017$ (sd=0.009), while the observed probability of all-one response was 0.019. Thus, consistently with the idea of screening tests in medicine, about 97% of individuals with negative results on all six tests were healthy with probability 1 while about 89% of individual with positive results on all six tests were diseased with probability 1. Table 7.6 shows the posterior means and standard deviation estimates of parameters λ_{kj}, ξ, α_0 for the GoM portion of the extended mixture GoM model. We observe that the extreme profile $k = 1$ represents women with likely positive CT diagnosis, while extreme profile $k = 2$ represents women with likely negative CT diagnosis.

Table 7.5 shows the observed and expected cell counts as well as the number of parameters, the degrees of freedom, and the values of the likelihood ratio statistic G^2 (see Bishop et al., 1975) for all five models considered. As expected, we observe that the fits of the standard GoM model and the latent class model are rather poor. The extended mixture GoM model provides a comparable performance in fit compared to the latent class random effects model of Hadgu and Qu (1998) and the finite mixture model of Albert and Dodd (2004), however, it has one less degree of freedom.

Table 7.7 provides the sensitivity and specificity estimates for the six tests under the five models considered. We see that the sensitivity and specificity estimates are rather high. Given our findings from the simulation study that pointed towards widespread overestimation of specificity and underestimation of sensitivity, we should pay particular attention to the smallest specificity and largest sensitivity estimates. For specificity, the extended mixture GoM model and the finite mixture model provide the smallest values that range from about 0.993 for the culture test to about 0.997 for the Syva-DFA test. For sensitivity, the extended mixture GoM model provides the largest values ranging from about 0.757 for the Abbott-EIA test to about 0.985 for the culture test. Thus, it appears that the extended mixture GoM model contributes new information beyond that available from other latent structure models, for estimating accuracy parameters of diagnostic tests.

TABLE 7.5
Observed and expected cell counts under LCM, LCRE, GoM, ExtM-GoM, and FM models. χ^2 is the truncated sum of squared Pearson residuals for observed cell counts > 1; n is the number of independent parameters; df is degrees of freedom.

	Response pattern	Observed counts	Expected LCM	Expected LCRE	Expected GoM	Expected ExtM-GoM	Expected FM
1	111111	87	53.4	87.30	437.31	87	87.18
2	111110	2	0.85	0.12	11.53	0.50	0.40
3	111101	9	17.26	8.19	146.20	7.64	7.07
4	111011	2	8.47	2.93	73.43	2.60	2.64
5	110111	9	21.2	11.79	167.21	10.27	9.52
6	110101	6	6.85	5.03	56.78	6.98	6.68
7	110011	1	3.36	1.98	29.64	2.42	2.5
8	110001	6	1.09	2.09	10.27	1.91	1.85
9	101111	7	11.57	4.91	91.33	4.44	4.05
10	101001	1	0.59	0.93	6.93	0.97	0.87
11	100111	4	4.59	3.15	36.12	4.08	3.84
12	100101	1	1.48	3.26	13.57	3.08	2.80
13	100011	2	0.73	1.32	8.89	1.20	1.11
14	100001	5	0.27	3.15	4.17	2.09	1.87
15	100000	7	7.50	6.92	6.27	6.91	7.23
16	011111	6	10.58	4.14	93.3	4.02	3.72
17	011101	1	3.42	1.93	33.37	2.80	2.62
18	011011	1	1.68	0.76	15.54	1.00	1.00
19	011001	1	0.54	0.81	5.88	0.97	0.87
20	011000	5	0.07	4.96	1.13	1.26	1.22
21	010111	1	4.20	2.70	36.34	3.72	3.53
22	010101	5	1.36	2.81	13.30	2.89	2.64
23	010001	2	0.28	2.71	3.93	2.59	2.41
24	010000	9	13.56	8.92	7.29	10.34	10.91
25	001111	2	2.29	1.21	20.76	1.65	1.51
26	001101	1	0.74	1.25	9.46	1.47	1.28
27	001001	1	0.20	1.18	2.78	2.54	2.49
28	001000	14	18.46	13.88	10.24	12.72	13.64
29	000111	2	0.91	1.77	9.53	1.82	1.61
30	000101	4	0.37	4.25	5.12	3.39	3.12
31	000100	16	16.54	15.82	11.46	12.45	13.15
32	000011	3	0.22	1.70	3.33	2.43	2.31
33	000010	15	15.73	14.85	9.84	11.40	11.98
34	000001	17	20.02	17.04	11.45	19.50	20.50
35	000000	4328	4321.11	4328.30	3123.48	4328	4322.71
χ^2			604.95	48.61	1529.89	43.29	48.70
G^2			179.68	42.25	1141.70	61.44	76.24
n			13	15	14	16	15
df			50	48	49	47	48

TABLE 7.6
Posterior mean (standard deviation) estimates for the diagnostic error component of the extended mixture GoM model with two extreme profiles (stochastic compartment).

	$k = 1$	$k = 2$
$\lambda_{k,1}$	0.750 (0.058)	0.050 (0.022)
$\lambda_{k,2}$	0.732 (0.058)	0.076 (0.027)
$\lambda_{k,3}$	0.534 (0.065)	0.093 (0.027)
$\lambda_{k,4}$	0.822 (0.057)	0.089 (0.029)
$\lambda_{k,5}$	0.608 (0.068)	0.084 (0.026)
$\lambda_{k,6}$	0.970 (0.021)	0.134 (0.042)
α_0	0.075 (0.057)	
ξ_k	0.278 (0.044)	0.722 (0.044)

TABLE 7.7
Sensitivity and specificity (standard deviation) of the diagnostic tests for LCM, LCRE, GoM, ExtM-GoM, and FM models.

Test	LCM	LCRE	GoM	ExtM-GoM	FM
Specificity					
Syva-DFA	0.998 (0.001)	0.999 (0.001)	0.999 (0.001)	0.998 (0.001)	0.997 (0.001)
Syva-EIA	0.997 (0.001)	0.997 (0.001)	0.999 (0.001)	0.996 (0.001)	0.996 (0.001)
Abbott-EIA	0.996 (0.001)	0.997 (0.001)	0.998 (0.001)	0.995 (0.001)	0.995 (0.001)
GenProbe	0.996 (0.001)	0.997 (0.001)	0.998 (0.001)	0.996 (0.001)	0.995 (0.001)
Sanofi-EIA	0.996 (0.001)	0.997 (0.001)	0.997 (0.001)	0.996 (0.001)	0.996 (0.001)
Culture	0.995 (0.001)	0.998 (0.001)	0.998 (0.001)	0.994 (0.001)	0.993 (0.001)
Sensitivity					
Syva-DFA	0.835 (0.030)	0.747 (0.048)	0.825 (0.034)	0.870 (0.030)	0.860 (0.030)
Syva-EIA	0.822 (0.031)	0.731 (0.048)	0.824 (0.034)	0.861 (0.030)	0.851 (0.032)
Abbott-EIA	0.716 (0.036)	0.636 (0.047)	0.721 (0.037)	0.757 (0.035)	0.748 (0.037)
GenProbe	0.863 (0.028)	0.777 (0.047)	0.864 (0.031)	0.907 (0.030)	0.892 (0.029)
Sanofi-EIA	0.756 (0.034)	0.678 (0.048)	0.751 (0.037)	0.796 (0.036)	0.786 (0.036)
Culture	0.984 (0.011)	0.935 (0.034)	0.978 (0.013)	0.985 (0.010)	0.980 (0.011)

7.6 Conclusion

We presented the extended mixture GoM model as an alternative latent structure model for analyzing accuracy of diagnostic and screening tests. For the medical testing case, the extended mixture GoM model accommodates two types of individuals, those whose diagnosis is certain, independently of the test, and those who are subject to diagnostic error. The latter individuals can be thought of as stochastic "movers" because they may change in their disease status depending on the diagnostic test considered; this idea is analogous to longitudinal mover-stayer models (Blumen et al., 1955). In the extended mixture GoM model, the "stayers" have predetermined response patterns that correspond to particular cells in a contingency table. The extended GoM mixture model can also be seen as a combination of latent class and GoM mixture modeling, analogous to the extended finite mixture model (Muthen and Shedden, 1999). Similar to the finite mixture model of Albert and Dodd (2004), the extended GoM mixture model allows for some individuals to always be diagnosed correctly, however, it relies on the assumption of disease severity for modeling diagnostic error for the other individuals while the finite mixture model of Erosheva (2005) relies on the assumption of two latent classes for the same purpose.

To estimate the extended mixture GoM model, we used a hierarchical Bayesian approach to estimation, drawing on earlier work (Erosheva, 2003). Although we did not use informative priors in our examples, Pfeiffer and Castle (2005) and Albert and Dodd (2004) specifically mention the promise of Bayesian approach in estimating diagnostic error without a gold standard when good prior information is available.

Our findings with the simulation study and with the *Chlamydia trachomatis* (Hadgu and Qu, 1998) data further emphasize the need of carrying out sensitivity analyses when no gold standard is available (Albert and Dodd, 2004) . When the underlying latent structure is unknown, it is important to examine sensitivity of scientific conclusions regarding the estimated accuracy of diagnostic and screening tests to the latent structure assumptions.

This could be done by comparing test accuracy estimates across a number of different latent structure models. Our results demonstrate that the extended mixture GoM model can provide us with new information on estimating test sensitivity and specificity beyond that provided by the latent class (Lazarsfeld and Henry, 1968; Goodman, 1974), latent class Gaussian random effects (Qu et al., 1996), and finite mixture models (Albert and Dodd, 2004). In addition, the extended mixture GoM model offers a plausible interpretation for diagnostic and screening test results by combining the idea of diagnostic error that depends on disease severity for some individuals with the idea of certain diagnosis for the other, typically the most healthy and the most sick individuals.

Finally, the flexible framework of mixed membership models can be used to modify the extended mixture GoM model and to address, for example, diagnostic cases when disease status is ordinal and test results do not come in binary form (Wang and Zhou, 2012). Drawing on the recent development in the class of mixed membership models that includes the GoM model as a special case, one can also modify the model to accommodate, for example, outcomes of mixed types, multiple basis categories in the latent structure, and correlations among membership scores.

References

Albert, P. S. (2007a). Imputation approaches for estimating diagnostic accuracy for multiple tests from partially verified designs. *Biometrics* 63: 947–957.

Albert, P. S. (2007b). Random effects modeling approaches for estimating ROC curves from repeated ordinal tests without a gold standard. *Biometrics* 63: 593–602.

Albert, P. S. and Dodd, L. E. (2004). A cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. *Biometrics* 60: 427–435.

Albert, P. S. and Dodd, L. E. (2008). On estimating diagnostic accuracy from studies with multiple raters and partial gold standard evaluation. *Journal of the American Statistical Association* 103: 61–73.

Begg, C. B. and Metz, C. E. (1990). Consensus diagnoses and "gold standards." *Medical Decision Making* 10: 29–30.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: The MIT press.

Blumen, J., Kogan, M., and Holland, P. W. (1955). *The Industrial Mobility of Labor as a Probability Process*. Cornell Studies of Industrial and Labor Relations *6*. Ithaca, NY: Cornell Univesity Press.

Erosheva, E. A. (2003). Bayesian estimation of the Grade of Membership model. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., (eds), *Bayesian Statistics 7*. New York, NY: Oxford University Press, 501–510.

Erosheva, E. A. (2005). Comparing latent structures of the Grade of Membership, Rasch, and latent class models. *Psychometrika* 70: 619–628.

Erosheva, E. A., Fienberg, S. E., and Joutard, C. (2007). Describing disability through individual-level mixture models for multivariate binary data. *Annals of Applied Statistics* 1: 502–537.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M. (eds), *Bayesian Statistics 4*. Oxford, UK: Oxford University Press, 162–193.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 61: 215–231.

Hadgu, Y. and Qu, A. (1998). A biomedical application of latent class model with random effects. *Journal of the Royal Statistical Society (Series C): Applied Statistics* 47: 603–613.

Heidelberger, P. and Welch, P. (1983). Simulation run length control in the presence of an initial transient. *Operations Research* 31: 1109–1144.

Hui, S. L. and Zhou, X. H. (1998). Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research* 7: 354–370.

Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent Structure Analysis*. Boston, MA: Houghton Mifflin.

Muthén, B. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* 55: 463–469.

Pepe, M. S. and Janes, H. (2007). Insights into latent class analysis of diagnostic test performance. *Biostatistics* 8: 474–484.

Pfeiffer, R. M. and Castle, P. E. (2005). With or without a gold standard. *Epidemiology* 16: 595–597.

Qu, Y., Tan, M., and Kutner, M. H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 55: 258–263.

Tanner, M. A. (1996). *Tools for Statistical Inference. Methods for the Exploration of Posterior Distributions and Likelihood Functions*. New York, NY: Springer, 3rd edition.

Wang, Z. and Zhou, X. (2012). Random effects models for assessing diagnostic accuracy of traditional chinese doctors in absence of a gold standard. *Statistics in Medicine* 31: 661–671.