2015

A pseudo-likelihood approach for estimating diagnostic accuracy of multiple binary medical tests

Wei Liu
Department of Mathematics, Harbin Institute of Technology, Harbin

Bo Zhang
Division of Biostatistics, Office of Surveillance and Biometrics, Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD, bo.zhang@fda.hhs.gov

Zhiwei Zhang
Division of Biostatistics, Office of Surveillance and Biometrics, Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD

Baojiang Chen
University of Nebraska Medical Center, Omaha, NE

Xiao-Hua Zhou
University of Washington

Computational Statistics and Data Analysis 84 (2015) 85–98

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A pseudo-likelihood approach for estimating diagnostic accuracy of multiple binary medical tests

Wei Liu [a], Bo Zhang [b,*], Zhiwei Zhang [b], Baojiang Chen [c], Xiao-Hua Zhou [d]

[a] Department of Mathematics, Harbin Institute of Technology, Harbin 150001, PR China

[b] Division of Biostatistics, Office of Surveillance and Biometrics, Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD 20993, USA

[c] Department of Biostatistics, College of Public Health, University of Nebraska Medical Center, Omaha, NE 68198, USA

[d] Department of Biostatistics, School of Public Health and Community Medicine, University of Washington, HSR&D Center of Excellence, U.S. Department of Veterans Affairs Puget Sound Health Care System, Seattle, WA 98198, USA

ARTICLE INFO

Article history:
Received 23 March 2014
Received in revised form 10 October 2014
Accepted 11 November 2014
Available online 20 November 2014

Keywords:
Sensitivity and specificity
Random effects
Latent class models
Composite likelihood
Imperfect reference standards

ABSTRACT

Latent class models with crossed subject-specific and test(rater)-specific random effects have been proposed to estimate the diagnostic accuracy (sensitivity and specificity) of a group of binary tests or binary ratings. However, the computation of these models are hindered by their complicated Monte Carlo Expectation–Maximization (MCEM) algorithm. In this article, a class of pseudo-likelihood functions is developed for conducting statistical inference with crossed random-effects latent class models in diagnostic medicine. Theoretically, the maximum pseudo-likelihood estimation is still consistent and has asymptotic normality. Numerically, our results show that not only the pseudo-likelihood approach significantly reduces the computational time, but it has comparable efficiency relative to the MCEM algorithm. In addition, dimension-wise likelihood, one of the proposed pseudo-likelihoods, demonstrates its superior performance in estimating sensitivity and specificity.
Published by Elsevier B.V.

1. Introduction

Pharmaceutical and regulatory statisticians who work in the fields of developing state-of-the-art medical devices and radiology diagnostic tests have shown tremendous interests in the strategies for accurately estimating diagnostic accuracy of multiple binary medical tests or raters. It has been widely recognized that sensitivity and specificity are two primary measures that characterize the diagnostic accuracy of binary tests. Statistical methodologies have been proposed to estimate sensitivity and specificity of binary tests (Zhou et al., 2002; Pepe, 2003). When investigators are committed to estimating the average sensitivity and specificity of a group of tests or raters, latent class models (Qu et al., 1996; Hui and Zhou, 1998; Qu and Hadgu, 1998; Albert et al., 2001; Albert and Dodd, 2008), in which the true disease status is considered as a latent variable, have been proved to be an effective and strategic approach. Qu et al. (1996) and Qu and Hadgu (1998) proposed a random-effects latent class model that characterized conditional dependence between tests through random effects. Albert et al. (2001) proposed a latent class model with a finite mixture structure to account for dependence between tests. Albert and Dodd (2008) extended latent class models in Qu and Hadgu (1998) and Albert et al. (2001) to incorporate information from both the verified and nonverified subjects during estimation. More recently, Zhang et al. (2012) introduced a latent

* Corresponding author.
E-mail address: Bo.Zhang@fda.hhs.gov (B. Zhang).

http://dx.doi.org/10.1016/j.csda.2014.11.006
0167-9473/Published by Elsevier B.V.

class model with crossed random effects (with both subject-specific and test-specific random effects) for estimating sensitivity and specificity of a group of binary tests or raters. Despite the success of these random-effects latent class models in diagnostic statistics, computational complexity has been the major flaw of the latent class models with crossed random effects. This article is devoted to proposing a competitive pseudo-likelihood approach that reduces the computational burden of crossed random effects latent class models in estimating diagnostic accuracy.

Statistical inference in random-effects models (or mixed-effects models) (Laird and Ware, 1982; McCulloch and Searle, 2001) starts from parameter estimation step, in which maximum likelihood estimation (MLE) is achieved via numerical optimization methods, such as Newton–Raphson algorithm and its variants. However, these optimization procedures require numerical approximation of the likelihoods of random-effects models that may involve evaluating high-dimensional integrals. In the latent class models with crossed random effects, as we show in Section 2, this computational issue becomes the major obstacle because the high-dimensional integral in the full likelihood of a crossed random effects latent class model is intractable. Computational strategies have been proposed to overcome the numerical difficulties posed by high-dimensional integrations in random-effects models. Breslow and Clayton (1993) proposed a pseudo quasi-likelihood method that seeks the parameter estimation by maximizing the joint distribution of the observed data and the random effects. Approximate MLE computation using the Laplace approximation also proposed by Steele (1996) and Lee and Nelder (2001). Yet, these methods cannot provide generally consistent estimation (Lin and Breslow, 1996). Another option is to develop simulation-based Monte Carlo algorithms for obtaining the MLE, which include Monte Carlo Markov chain (MCMC) algorithms proposed by Zeger and Karim (1991) and McCulloch (1997) and Monte Carlo Expectation–Maximization (MCEM) algorithm proposed by Booth and Hobert (1999) and Booth et al. (2001). These algorithms have drawbacks that include the computational duration and MCEM convergence assessment, so that they cannot provide instant statistical inference results to practitioners. Statistical inference turns to be even harder in the latent class models with crossed random effects. Unlike the conventional random-effects models with independent between-subject observations, the full likelihood of a crossed random effects latent class model is an integral whose dimension increases with the number of subjects and the number of tests. In a crossed diagnostic design with I subjects and J tests, the full likelihood appears to be a $(I \times J)$-dimensional integral embedded in an I-dimensional summation, which particularly deteriorates the statistical inference procedure when using crossed random effects latent class models.

In this article, we propose a pseudo-likelihood approach as a competitive statistical analysis strategy for the crossed random effects latent class models that estimate the sensitivity and specificity of a group of binary medical tests or raters. A class of pseudo-likelihood functions is created in Section 2, including pairwise likelihood (Bellio and Varin, 2005), triple-wise likelihood, hybrid likelihoods, and dimension-wise likelihood. The benefits of using the proposed pseudo-likelihood approach are enormous. The proposed pseudo-likelihood functions contain the integrations with a reduced dimension and possess more succinct integrands than the full likelihood. The implementation of the pseudo-likelihood approach is rather simple with the aid of a numerical optimization package. Parameter estimates obtained from maximizing the proposed pseudo-likelihoods are consistent and asymptotically follow normal distributions (Lindsay, 1988; Molenberghs and Verbeke, 2005; Varin et al., 2011). The parameter variance estimation can be achieved by bootstrapping. Estimation efficiency of the estimators obtained from maximizing dimension-wise likelihoods is comparable to the full likelihood maximized by the MCEM algorithm. Yet, the efficiency of pairwise and hybrid likelihoods may not be satisfactory especially when imperfect reference standards do not exist. In Section 5, we analyze a colon cancer detection data published by Zhou et al. (2002), for the purpose of demonstrating the pseudo-likelihood approach.

2. Methodology

2.1. Crossed random effects models for estimating diagnostic accuracy

Let Y_{ij} denote the binary diagnostic result of a disease ($Y_{ij} = 1$ for having the disease and $Y_{ij} = 0$ for not having the disease) for the ith subject from the jth test (rater), $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$. We denote D_i, a binary latent variable, as the true disease status of the ith subject. Under the circumstances that there is no gold or imperfect reference standard, we consider the following model with two crossed random effects for Y_{ij}:

$$P(Y_{ij} = 1|D_i = d_i, b_i, c_j) = \hbar^{-1}\left(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j\right), \quad \sigma_{d_i}, \tau_{d_i} > 0, \tag{1}$$

where $\hbar^{-1}(\cdot)$ denotes the inverse of a general link function (such as a probit link function or a logit link function), b_i is the subject-specific random effect with probability density distribution (p.d.f.) $g_1(x)$, and c_j is the test-specific random effect with p.d.f. $g_2(x)$. The three unobserved latent variables D_i, b_i, and c_j are assumed to be independent of each other. Let $\pi_{d_i} = P(D_i = d_i)$, and π_1 is consequently the prevalence of disease among the subjects. Under the probit link and the normality assumption of b_i and c_j, (1) can be used to estimate the average sensitivity and specificity of the rater (test) population: $S_e = \Phi(\beta_1/\sqrt{1+\sigma_1^2+\tau_1^2})$ and $S_p = \Phi(-\beta_0/\sqrt{1+\sigma_0^2+\tau_0^2})$, where $\Phi(\cdot)$ denotes the standard normal cumulative density function. Contrast to the latent class models in Qu et al. (1996) and Albert et al. (2001) with a single subject-specific random effect, (1) describes test-specific variation through an additional random effect c_j. Investigators usually assume b_i and

c_j follow a standard normal distribution. Yet, one may also consider the case where $g_1(x)$ and $g_2(x)$ are two-group mixture normal distributions with p.d.f.: $g_m(x) = \lambda_m\phi(x;\, \mu_{1m},\, \nu_{1m}^2) + (1-\lambda_m)\phi(x;\, \mu_{2m},\, \nu_{2m}^2)$, $m = 1, 2$, where λ_m is the proportion of the first group, $0 \le \lambda_m \le 1$, and $\phi(x;\, \mu,\, \nu^2)$ is the p.d.f. of a normal distribution with mean μ and variance ν^2. For model identification, we assume $\mu_{1m} < \mu_{2m}$, $\lambda_m\mu_{1m} + (1-\lambda_m)\mu_{2m} = 0$, and $\lambda_m\nu_{1m}^2 + (1-\lambda_m)\nu_{2m}^2 + \lambda_m(1-\lambda_m)(\mu_{1m}-\mu_{2m})^2 = 1$. Then, the model-based estimates of average sensitivity and specificity can be obtained by marginalizing over the two-group mixture models.

Under the circumstances that there is an imperfect reference standard (Qu and Hadgu, 1998; Valenstein, 1990; Whiting et al., 2004; Albert, 2009), let us denote the binary rating T_i as the test result for the ith subject from the imperfect reference standard. Then, $S^T_{t_i|d_i} = P(T_i = t_i|D_i = d_i)$ characterizes the diagnostic accuracy of the imperfect reference standard. In practice, the results from previous studies are used to obtain estimates of $S^T_{t_i|d_i}$ (including sensitivity $S^T_{1|1} = P(T_i = 1|D_i = 1)$ and specificity $S^T_{0|0} = P(T_i = 0|D_i = 0)$ of the imperfect reference standard), which subsequently are used to estimate $P(Y_{ij}|D_i)$, the diagnostic accuracy of the tests or raters. To incorporate the information of the imperfect reference standard, we consider the following model for Y_{ij}:

$$P(Y_{ij} = 1|T_i = t_i, D_i = d_i, b_i, c_j) = \hbar^{-1}\left(\beta_{d_it_i} + \sigma_{d_it_i}b_i + \tau_{d_it_i}c_j\right), \quad \sigma_{d_it_i},\, \tau_{d_it_i} > 0. \tag{2}$$

Note that the model flexibility of (2) can be reduced. Investigators can assume that, given $D_i = d_i$, $\beta_{d_it_i}$ is not related to t_i (i.e., $\beta_{d_it_i} = \beta_{d_i}$), or $\sigma_{d_it_i}$ and $\tau_{d_it_i}$ are not related to t_i (i.e., $\sigma_{d_it_i} = \sigma_{d_i}$, $\tau_{d_it_i} = \tau_{d_i}$). The two assumptions combined leads to $P(Y_{ij} = 1|T_i = t_i, D_i = d_i) = P(Y_{ij} = 1|D_i = d_i)$ in Zhang et al. (2012) and Albert (2009). Investigators can further assume that the variation is identical in both disease and non-disease groups, which implies $\sigma_{d_it_i} = \sigma$ and $\tau_{d_it_i} = \tau$.

The full likelihood of (1) and (2) is complicated by the two crossed random effects b_i and c_j. Let $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iJ})'$ be the J dichotomous test results on the ith subject. Let $Y = (Y_1', Y_2', \ldots, Y_I')$, $D = (D_1, D_2, \ldots, D_I)$, $T = (T_1, T_2, \ldots, T_I)$, and θ be the vector containing unknown parameters $\beta_{d_it_i}$, $\sigma_{d_it_i}$, $\tau_{d_it_i}$, π_{d_i} and the unknown parameters in $g_1(x)$ and $g_2(x)$. Define a function $\ell^*(\cdot)$ such that $\ell^*(1) = 1$, $\ell^*(0) = -1$. Because the observed data consist of both the test results Y and the imperfect reference standard results T, the full likelihood of model (2) is given by

$$\begin{aligned} L(\theta; y, t) &= \sum_{d_1=0,1}\cdots\sum_{d_I=0,1}\left\{P(Y = y|T = t, D = d)\prod_{i=1}^{I}S^T_{t_i|d_i}\prod_{i=1}^{I}\pi_{d_i}\right\} \\ &= \sum_{d_1=0,1}\cdots\sum_{d_I=0,1}\int\cdots\int\left\{\prod_{i=1}^{I}\prod_{j=1}^{J}\hbar^{-1}\left(\ell^*(y_{ij})(\beta_{d_it_i} + \sigma_{d_it_i}b_i + \tau_{d_it_i}c_j)\right)\right\} \\ &\quad\times\prod_{i=1}^{I}S^T_{t_i|d_i}\pi_{d_i}g_1(b_i)\mathrm{d}b_i\prod_{j=1}^{J}g_2(c_j)\mathrm{d}c_j. \end{aligned} \tag{3}$$

The full likelihood of model (1) can be similarly derived as in (3):

$$L(\theta; y) = \sum_{d_1=0,1}\cdots\sum_{d_I=0,1}\int\cdots\int\left\{\prod_{i=1}^{I}\prod_{j=1}^{J}\hbar^{-1}\left(\ell^*(y_{ij})(\beta_{d_it_i} + \sigma_{d_it_i}b_i + \tau_{d_it_i}c_j)\right)\right\}\prod_{i=1}^{I}\pi_{d_i}g_1(b_i)\mathrm{d}b_i\prod_{j=1}^{J}g_2(c_j)\mathrm{d}c_j. \tag{4}$$

Because of the built-in high-dimensional integration and summation, it is difficult to precisely evaluate (3) and (4) by numerical approximation. Thus, we propose in Section 2.2 a class of pseudo-likelihoods for parameter estimation.

2.2. Maximum pseudo-likelihood estimation

In this section, we propose a class of pseudo-likelihoods for estimating diagnostic accuracy with or without an imperfect reference standard. For the fixed positive integers s_1 and s_2 satisfying $2 \le s_1 \le I$ and $2 \le s_2 \le J$, the proposed pseudo-likelihood for (1) is

$$\begin{aligned} PL_{s_1,s_2}(\theta; y) &= \prod_{i=1}^{I}\prod_{j_1<j_2<\cdots<j_{s_2}}^{J}P(Y_{ij_1} = y_{ij_1}, Y_{ij_2} = y_{ij_2}, \ldots, Y_{ij_{s_2}} = y_{ij_{s_2}}; \theta) \\ &\quad\times\prod_{j=1}^{J}\prod_{i_1<i_2<\cdots<i_{s_1}}^{I}P(Y_{i_1j} = y_{i_1j}, Y_{i_2j} = y_{i_2j}, \ldots, Y_{i_{s_1}j} = y_{i_{s_1}j}; \theta). \end{aligned} \tag{5}$$

In (5), the notation $\prod^J_{j_1<j_2<\cdots<j_{s_2}}$ is a variant of $\prod_{1\le j_1<j_2<\cdots<j_{s_2}\le J}$, representing the product of all such terms satisfying $1 \le j_1 < j_2 < \cdots < j_{s_2} \le J$. In (5),

$$\begin{aligned} &P(Y_{ij_1} = y_{ij_1}, Y_{ij_2} = y_{ij_2}, \ldots, Y_{ij_{s_2}} = y_{ij_{s_2}}; \theta) \\ &= \sum_{d_i=0,1}\pi_{d_i}\int_{b_i}\left\{\prod_{j=j_1,\ldots j_{s_2}}\int_{c_j}\hbar^{-1}\left(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i}b_i + \tau_{d_i}c_j)\right)g_2(c_j)dc_j\right\}g_1(b_i)db_i, \end{aligned} \tag{6}$$

and

$$P(Y_{i_1 j} = y_{i_1 j}, Y_{i_2 j} = y_{i_2 j}, \ldots, Y_{i_{s_1} j} = y_{i_{s_1} j}; \theta) = \int_{c_j} \left\{ \prod_{i=i_1,\ldots,i_{s_1}} \sum_{d_i=0,1} \pi_{d_i} \int_{b_i} \hbar^{-1}\left(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j)\right) g_1(b_i) db_i \right\} g_2(c_j) dc_j. \tag{7}$$

The numerical evaluation process of (6) is composed of conducting s_2 one-dimensional integrations inside for the c_j's and a one-dimensional integration outside for b_i. Thus, the computational complexity of (6) is $O(s_2 \mathcal{N}^2)$, in which $\mathcal{N}$ is the size of this computation problem and it is equal to the number of operations in a one-dimensional integration. Similarly, the numerical evaluation process of (7) is composed of conducting $2s_1$ one-dimensional integrations inside for the b_j's and a one-dimensional integration outside for c_j, which implies its computational complexity is $O(s_1 \mathcal{N}^2)$. When the probit link is used and the two random effects b_i and c_j follow standard normal distributions, the one-dimensional integrals inside (6) and (7) have analytical forms. That is, when $\hbar^{-1} = \Phi$ and $g_1(b_i) \sim N(0, 1)$ and $g_2(c_j) \sim N(0, 1)$, we have $\int_{b_i} \Phi\left(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j)\right) g_1(b_i) db_i = \Phi\left(\ell^*(y_{ij})(\beta_{d_i} + \tau_{d_i} c_j)/\sqrt{1+\sigma_{d_i}^2}\right)$ and $\int_{c_j} \Phi\left(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j)\right) g_2(c_j) dc_j = \Phi\left(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i} b_i)/\sqrt{1+\tau_{d_i}^2}\right)$, which reduces the computational complexity of (6) and (7) to be $O(s_2 \mathcal{N})$ and $O(s_1 \mathcal{N})$, respectively. However, this bonus will disappear if the random effects are not specified to follow standard normal distributions or the probit link is not employed.

Noticeably, $PL_{s_1,s_2}(\theta; y)$ represents a pseudo-likelihood class $\{PL_{s_1,s_2}(\theta; y) \mid s_1 \in \mathbb{Z}^+, s_2 \in \mathbb{Z}^+, 2 \le s_1 \le I, 2 \le s_2 \le J\}$, where $\mathbb{Z}^+$ denotes the set of positive integers. When $s_1 = s_2 = 2$, pseudo-likelihood (5) becomes the pairwise likelihood in Bellio and Varin (2005):

$$PL_{2,2}(\theta; y) = \prod_{i=1}^{I} \prod_{j_1<j_2}^{J} P(Y_{ij_1} = y_{ij_1}, Y_{ij_2} = y_{ij_2}; \theta) \prod_{j=1}^{J} \prod_{i_1<i_2}^{I} P(Y_{i_1 j} = y_{i_1 j}, Y_{i_2 j} = y_{i_2 j}; \theta), \tag{8}$$

in which

$$P(Y_{ij_1} = y_{ij_1}, Y_{ij_2} = y_{ij_2}; \theta) = \sum_{d_i=0,1} \pi_{d_i} \int_{b_i} \left\{ \prod_{j=j_1,j_2} \int_{c_j} \hbar^{-1}\left(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j)\right) g_2(c_j) dc_j \right\} g_1(b_i) db_i, \tag{9}$$

and

$$P(Y_{i_1 j} = y_{i_1 j}, Y_{i_2 j} = y_{i_2 j}; \theta) = \int_{c_j} \left\{ \prod_{i=i_1,i_2} \sum_{d_i=0,1} \pi_{d_i} \int_{b_i} \hbar^{-1}\left(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j)\right) g_1(b_i) db_i \right\} g_2(c_j) dc_j. \tag{10}$$

When $s_1 = s_2 = 3$, pseudo-likelihood (5) extends (Bellio and Varin, 2005) to the triple-wise likelihood:

$$PL_{3,3}(\theta; y) = \prod_{i=1}^{I} \prod_{j_1<j_2<j_3}^{J} P(Y_{ij_1} = y_{ij_1}, Y_{ij_2} = y_{ij_2}, Y_{ij_3} = y_{ij_3}; \theta) \times \prod_{j=1}^{J} \prod_{i_1<i_2<i_3}^{I} P(Y_{i_1 j} = y_{i_1 j}, Y_{i_2 j} = y_{i_2 j}, Y_{i_3 j} = y_{i_3 j}; \theta), \tag{11}$$

for which the probability terms can be similarly derived as in (9) and (10). Some members in the proposed pseudo-likelihood class, such as (8) and (11), suffer from the curse of dimensionality. For example, (8) is the product of $I\binom{J}{2} + J\binom{I}{2}$ probability terms, which implies the computational complexity of (8) is $O(\max\{IJ^2\mathcal{N}^2, JI^2\mathcal{N}^2\})$. Likewise, (11) consists of $I\binom{J}{3} + J\binom{I}{3}$ probability terms, which implies the computational complexity of (8) is $O(\max\{IJ^3\mathcal{N}^2, JI^3\mathcal{N}^2\})$. When either I or J is a large number, the total number of probability terms in (5) becomes unacceptably large and the computational complexity increases exponentially. This makes it not feasible to seek parameter estimates from maximizing such a pseudo-likelihood using the Newton–Raphson algorithm.

There is a solution to the high-dimensional problem. In the empirical applications where there are a moderate number of tests but a large number of subjects, the special case of pseudo-likelihood (5) with $s_1 = I$ and $s_2 = 2$ can be considered:

$$PL_{I,2}(\theta; y) = \prod_{i=1}^{I} P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{iJ} = y_{iJ}; \theta) \prod_{j=1}^{J} \prod_{i_1<i_2}^{I} P(Y_{i_1 j} = y_{i_1 j}, Y_{i_2 j} = y_{i_2 j}; \theta), \tag{12}$$

in which $P(Y_{i_1 j} = y_{i_1 j}, Y_{i_2 j} = y_{i_2 j}; \theta)$ remains as in (10) and

$$P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{ij} = y_{ij}; \theta) = \sum_{d_i=0,1} \pi_{d_i} \int_{b_i} \left\{ \prod_{j=1}^{J} \int_{c_j} \hbar^{-1}\Big(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j)\Big) g_2(c_j) dc_j \right\} g_1(b_i) db_i. \tag{13}$$

Pseudo-likelihood (12) substitutes the density of an entire row in the data for the product of the pairwise or tripe-wise densities of this row, which avoids evaluating the integrations for $I\binom{J}{s_2}$ times. In the cases where there are a moderate number of subjects but a large number of tests, the special case of pseudo-likelihood (5) with $s_1 = 2$ and $s_2 = J$ can be considered:

$$PL_{2,J}(\theta; y) = \prod_{i=1}^{I} \prod_{j_1<j_2}^{J} P(Y_{ij_1} = y_{ij_1}, Y_{ij_2} = y_{ij_2}; \theta) \prod_{j=1}^{J} P(Y_{1j} = y_{1j}, Y_{2j} = y_{2j}, \ldots, Y_{Ij} = y_{Ij}; \theta), \tag{14}$$

in which $P(Y_{ij_1} = y_{ij_1}, Y_{ij_2} = y_{ij_2}; \theta)$ remains as in (9) and

$$P(Y_{1j} = y_{1j}, Y_{2j} = y_{2j}, \ldots, Y_{Ij} = y_{Ij}; \theta) = \int_{c_j} \left\{ \prod_{i=1}^{I} \sum_{d_i=0,1} \pi_{d_i} \int_{b_i} \hbar^{-1}\Big(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j)\Big) g_1(b_i) db_i \right\} g_2(c_j) dc_j. \tag{15}$$

Further, we can take $s_1 = I$ and $s_2 = J$ to ultimately avoid the curse of dimensionality in (8) and (11):

$$PL_{I,J}(\theta; y) = \prod_{i=1}^{I} P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{ij} = y_{ij}; \theta) \prod_{j=1}^{J} P(Y_{1j} = y_{1j}, Y_{2j} = y_{2j}, \ldots, Y_{Ij} = y_{Ij}; \theta), \tag{16}$$

for which the column-wise and row-wise joint densities are specified in (13) and (15), respectively. We name (16) as "dimension-wise likelihood" and (12) and (14) as "hybrid likelihoods". Note that the computational complexity of the joint densities (13) and (15) in the dimension-wise likelihood and the hybrid likelihoods is $O(J\mathcal{N}^2)$ and $O(I\mathcal{N}^2)$, respectively, if maximized by the Newton–Raphson algorithm. The dimension-wise likelihood, with totally $(I + J)$ probability terms, is a problem of $O(\max\{J\mathcal{N}^2, I\mathcal{N}^2\})$, which avoids increasing exponentially with I or J and therefore avoids the curse of dimensionality. Of course, the pseudo-likelihood methods are expected to be less efficient than the MLE, so we investigated the extent of efficiency loss in Section 3.

For estimating diagnostic accuracy with the imperfect reference standard T_i, we similarly propose a class of pseudo-likelihoods that can incorporate T_i as follows:

$$\begin{aligned} PL_{s_1,s_2}(\theta; y, t) = \prod_{i=1}^{I} \prod_{j_1<j_2<\cdots<j_{s_2}}^{J} & P(T_i = t_i, Y_{ij_1} = y_{ij_1}, Y_{ij_2} = y_{ij_2}, \ldots, Y_{ij_{s_2}} = y_{ij_{s_2}}; \theta) \\ & \times \prod_{j=1}^{J} \prod_{i_1<i_2<\cdots<i_{s_1}}^{I} P(T_{i_1} = t_{i_1}, T_{i_2} = t_{i_2}, \ldots, T_{i_{s_1}} = t_{i_{s_1}}, Y_{i_1 j} = y_{i_1 j}, \\ & Y_{i_2 j} = y_{i_2 j}, \ldots, Y_{i_{s_1} j} = y_{i_{s_1} j}; \theta). \end{aligned} \tag{17}$$

In (17),

$$P(T_i = t_i, Y_{ij_1} = y_{ij_1}, Y_{ij_2} = y_{ij_2}, \ldots, Y_{ij_{s_2}} = y_{ij_{s_2}}; \theta) = \sum_{d_i=0,1} \pi_{d_i} S^T_{t_i|d_i} \int_{b_i} \left\{ \prod_{j=j_1,\ldots,j_{s_2}} \int_{c_j} \hbar^{-1}\Big(\ell^*(y_{ij})(\beta_{d_i t_i} + \sigma_{d_i t_i} b_i + \tau_{d_i t_i} c_j)\Big) g_2(c_j) dc_j \right\} g_1(b_i) db_i,$$

and

$$P(T_{i_1} = t_{i_1}, T_{i_2} = t_{i_2}, \ldots, T_{i_{s_1}} = t_{i_{s_1}}, Y_{i_1 j} = y_{i_1 j}, Y_{i_2 j} = y_{i_2 j}, \ldots, Y_{i_{s_1} j} = y_{i_{s_1} j}; \theta) = \int_{c_j} \left\{ \prod_{i=i_1,\ldots,i_{s_1}} \sum_{d_i=0,1} \pi_{d_i} S^T_{t_i|d_i} \int_{b_i} \hbar^{-1}\Big(\ell^*(y_{ij})(\beta_{d_i t_i} + \sigma_{d_i t_i} b_i + \tau_{d_i t_i} c_j)\Big) g_1(b_i) db_i \right\} g_2(c_j) dc_j.$$

The methodologies proposed above for $PL_{s_1,s_2}(\theta; y)$ can be directly applied to (17) to resolve the high-dimensional problem when the imperfect reference standard is used.

2.3. Monte Carlo Expectation–Maximization algorithm

In this section, an MCEM algorithm is developed to obtain the MLE of the unknown parameters θ in (2). Although the MCEM algorithm is developed based upon model (2), it can be easily extended to (1).

We take the latent true disease status $D = (D_1, \ldots, D_I)'$, subject-specific random effects $b = (b_1, \ldots, b_I)'$, and test-specific random effects $c = (c_1, \ldots, c_J)'$ as missing variables in the EM algorithm. Denote $X^* = (Y_1', \ldots, Y_I', T_1, \ldots, T_I)'$ the observed data, $Z^* = (D', b', c')'$ the missing data, and $Y^* = (X^{*\prime}, Z^{*\prime})'$ the complete data.

E-STEP. At the $(r+1)$th iteration of EM algorithm, the E-step involves calculation of the Q-function

$$Q(\theta|\theta^{(r)}) = \int \left\{\log f(y^*|\theta)\right\} f(z^*|x^*, \theta^{(r)}) \mathrm{d}z^*, \tag{18}$$

where $\theta^{(r)}$ denotes the parameter vector from the rth iteration in the EM algorithm, $f(z^*|x^*, \theta^{(r)})$ is the conditional distribution of missing data given the observed data and $\theta^{(r)}$, and

$$f(y^*|\theta) = \prod_{i=1}^{I}\prod_{j=1}^{J} \hbar^{-1}\left(\ell^*(y_{ij})(\beta_{d_i t_i} + \sigma_{d_i t_i} b_i + \tau_{d_i t_i} c_j)\right) \prod_{i=1}^{I} S^T_{t_i|d_i} \pi_{d_i} g_1(b_i)\mathrm{d}b_i \prod_{j=1}^{J} g_2(c_j)\mathrm{d}c_j$$

is the complete-data likelihood. In the following, a Monte Carlo numerical integration algorithm is developed for (18) by using Gibbs sampler.

Note that, the full conditional distributions of D_i, b_i, and c_j are

$$P(D_i = d_i|x^*, b, c, \theta^{(r)}) \propto S^T_{t_i|d_i}\pi_{d_i} \prod_{j=1}^{J} \hbar^{-1}\left(\ell^*(y_{ij})(\beta_{d_i t_i} + \sigma_{d_i t_i} b_i + \tau_{d_i t_i} c_j)\right),$$

$$f(b_i|x^*, D, c, \theta^{(r)}) \propto g_1(b_i) \prod_{j=1}^{J} \hbar^{-1}\left(\ell^*(y_{ij})(\beta_{d_i t_i} + \sigma_{d_i t_i} b_i + \tau_{d_i t_i} c_j)\right),$$

and

$$f(c_j|x^*, D, b, \theta^{(r)}) \propto g_2(c_j) \prod_{i=1}^{I} \hbar^{-1}\left(\ell^*(y_{ij})(\beta_{d_i t_i} + \sigma_{d_i t_i} b_i + \tau_{d_i t_i} c_j)\right),$$

respectively. To draw a Markov chain $\{z^{*(1)}, z^{*(2)}, \ldots, z^{*(N)}\}$ from the conditional distribution $f(z^*|x^*, \theta^{(r)})$ using Gibbs sampler, we start from the starting values $z^{*(0)}$ with $f(z^{*(0)}) > 0$. At the nth step of the Gibbs sampling procedure, $z^{*(n)}$ is drawn by sequentially sampling from the conditional distributions. Specifically, if $z^{*(n)} = (d^{(n)\prime}, b^{(n)\prime}, c^{(n)\prime})'$, then we run the following updating steps iteratively:

I. Update $D_i^{(n)}$ for $i = 1, \ldots, I$ simultaneously by

$$P(D_i^{(n)} = 1|x^*, b^{(n-1)}, c^{(n-1)}, \theta^{(r)}) = \frac{P(D_i^{(n)} = 1|x^*, b^{(n-1)}, c^{(n-1)}, \theta^{(r)})}{P(D_i^{(n)} = 1|x^*, b^{(n-1)}, c^{(n-1)}, \theta^{(r)}) + P(D_i^{(n)} = 0|x^*, b^{(n-1)}, c^{(n-1)}, \theta^{(r)})}.$$

II. Update $b_i^{(n)}$ for $i = 1, \ldots, I$ simultaneously using the Metropolis–Hasting algorithm. Sample $\tilde{b}_i$ from the proposal distribution $\tilde{f}(\tilde{b}_i|b_i^{(n-1)})$, which is conditional on $b_i^{(n-1)}$. The sampled $\tilde{b}_i$ is accepted as $b_i^{(n)}$ with the probability

$$\min\left\{1, \frac{f(\tilde{b}_i|x^*, D^{(n)}, c^{(n-1)}, \theta^{(r)})\tilde{f}(b_i^{(n-1)}|\tilde{b}_i)}{f(b_i^{(n-1)}|x^*, D^{(n)}, c^{(n-1)}, \theta^{(r)})\tilde{f}(\tilde{b}_i|b_i^{(n-1)})}\right\};$$

otherwise $b_i^{(n)} = b_i^{(n-1)}$.

III. Update $c_j^{(n)}$ for $j = 1, \ldots, J$ simultaneously using the Metropolis–Hasting algorithm. Sample $\tilde{c}_j$ from the proposal distribution $\tilde{f}(\tilde{c}_j|c_j^{(n-1)})$ that is conditional on $c_j^{(n-1)}$. The sampled $\tilde{c}_j$ is accepted as $c_j^{(n)}$ with the probability

$$\min\left\{1, \frac{f(\tilde{c}_j|x^*, D^{(n)}, b^{(n)}, \theta^{(r)})\tilde{f}(c_j^{(n-1)}|\tilde{c}_j)}{f(c_j^{(n-1)}|x^*, D^{(n)}, b^{(n)}, \theta^{(r)})\tilde{f}(\tilde{c}_j|c_j^{(n-1)})}\right\};$$

otherwise $c_j^{(n)} = c_j^{(n-1)}$.

The normal distributions conditional on the previous Gibbs sampler draws are suggested to be used as proposal distributions of b_i in Step (II) and c_j in Step (III); that is, $\tilde{b}_i^{(n)}$ and $\tilde{c}_j^{(n)}$ are sampled from $\phi(\tilde{b}_i; b_i^{(n-1)}, \nu_b^2)$ and $\phi(\tilde{c}_j; c_j^{(n-1)}, \nu_c^2)$, respectively, where ν_b^2 and ν_c^2 are the pre-specified variances for the proposal distributions. Through the three updating steps (I)–(III), the

Markov chain $\{z^{*(1)}, z^{*(2)}, \ldots, z^{*(N)}\}$ is produced. If the first N_0 draws $z^{*(1)}, \ldots, z^{*(N_0)}$ are discarded as the burn-in period, then the approximation of $Q(\theta|\theta^{(r)})$ is then

$$\hat{Q}(\theta|\theta^{(r)}) = \frac{1}{N-N_0}\sum_{n=N_0+1}^{N}\left[\sum_{i=1}^{I}\sum_{j=1}^{J}\log\hbar^{-1}\left(\ell^*(y_{ij})(\beta_{d_i^{(n)}t_i}+\sigma_{d_i^{(n)}t_i}b_i^{(n)}+\tau_{d_i^{(n)}t_i}c_j^{(n)})\right)\right.$$
$$\left.+\sum_{i=1}^{I}\log(S^T_{t_i|d_i^{(n)}})+\sum_{i=1}^{I}\log(\pi_{d_i^{(n)}})+\sum_{i=1}^{I}\log\left\{g_1(b_i^{(n)})\right\}+\sum_{j=1}^{J}\log\left\{g_2(c_j^{(n)})\right\}\right]$$

M-STEP: Choose $\theta^{(r+1)}$ such that $\theta^{(r+1)} = \arg_\theta \max \hat{Q}(\theta|\theta^{(r)})$.

To obtain the MLE of θ, we consider to run R iterations of the MCEM algorithm until it converges. Convergence of the MCEM algorithm is demonstrated in Section 3.2 by numerical examples.

3. Numerical studies

3.1. Simulation 1: Competition among pseudo-likelihood peers

This subsection aims at comparing the finite sample performance of the four pseudo-likelihoods $PL_{2,2}(\theta; y)$, $PL_{2,J}(\theta; y)$, $PL_{I,2}(\theta; y)$, and $PL_{I,J}(\theta; y)$ for model (1), in estimating sensitivity and specificity and in estimating the relevant parameters. We also compare the pseudo-likelihoods $PL_{2,2}(\theta; y, t)$, $PL_{2,J}(\theta; y, t)$, $PL_{I,2}(\theta; y, t)$, and $PL_{I,J}(\theta; y, t)$ for model (2) with imperfect reference standards. A simulation study was conducted with various combinations of I and J, representing several circumstances with different numbers of subjects and tests (raters). Simulation data were generated from $P(Y_{ij} = 1|D_i = d_i, b_i, c_j) = \Phi(\beta_{d_i} + \sigma_{d_i}b_i + \tau_{d_i}c_j)$, with $b_i \overset{\text{i.i.d.}}{\sim} N(0, 1)$, $c_j \overset{\text{i.i.d.}}{\sim} N(0, 1)$, and $\pi_1 = 0.5$. The rest of parameters were specified as either (i) $\beta_1 = 1.795$, $\beta_0 = -1.795$, $\sigma_1 = \sigma_0 = 1$, $\tau_1 = \tau_0 = 1$, or (ii) $\beta_1 = 1.269$, $\beta_0 = -1.269$, $\sigma_1 = \sigma_0 = 0.5$, $\tau_1 = \tau_0 = 0.5$, both of which generated the true average sensitivity 0.85 and true average specificity 0.85. Three (I, J) combinations were examined: $(I, J) = (100, 25)$, $(200, 50)$, and $(500, 25)$. The total number of simulation replications was set to be 500. After data generation, each of the simulation data sets was fit to model (1) via the pseudo-likelihoods. In the second part of the simulation study, simulation data was generated from $P(Y_{ij} = 1|T_i = t_i, D_i = d_i, b_i, c_j) = \Phi(\beta_{d_i} + \sigma_{d_i}b_i + \tau_{d_i}c_j)$ with the same parameter specifications that are described above. The realizations of the imperfect reference standard T_i were generated from $S^T_{1|1} = S^T_{0|0} = 0.95$, given the true disease status D_i. After data generation, each of the simulation data sets was fit to model (2) via the pseudo-likelihoods. Here, we make an assumption that, given the true disease status is available, the observed test results (ratings) are independent of the imperfect reference standard. This is an assumption that has been made in Zhang et al. (2012) and Albert (2009), and is usually true in practice.

Table 1 (for $\sigma = \tau = 1$) and Table 2 (for $\sigma = \tau = 0.5$) report the estimation biases of sensitivity and specificity, as well as the relevant parameters, obtained from maximizing the pseudo-likelihoods in this simulation study. The empirical standard derivations are reported in the parentheses. Regardless of the true values of σ and τ and the combinations of (I, J), Tables 1 and 2 are delivering the same messages. When the imperfect reference standard T_i is not available, the dimension-wise likelihood $PL_{I,J}(\theta; y)$ usually has the smallest estimation bias and variance among the four pseudo-likelihoods in estimating sensitivity and specificity, the pairwise likelihood $PL_{2,2}(\theta; y)$ possesses the largest estimation bias and variance, and the hybrid likelihoods $PL_{2,J}(\theta; y)$ and $PL_{I,2}(\theta; y)$ reside in between. When the imperfect reference standard T_i is available, the four pseudo-likelihoods ($PL_{2,2}(\theta; y, t)$, $PL_{2,J}(\theta; y, t)$, $PL_{I,2}(\theta; y, t)$, and $PL_{I,J}(\theta; y, t)$) perform equally well in terms of estimation bias and variance of sensitivity and specificity. With the aid of the imperfect reference standard, the estimation biases of the pairwise likelihood and the hybrid likelihoods are decreased considerably while their variances reflect no significant change. This indicates that the imperfect reference verification plays a key role in improving estimation precision when using the proposed pseudo-likelihood approach. In Tables 1 and 2, dimension-wise likelihood appears to have persistent estimation capacity regardless of whether there is an imperfect reference standard. However, this is generally not true. Previous research (Zhang et al., 2012; Albert, 2009) showed that reliable imperfect reference standards can improve the robustness of diagnostic accuracy estimation in latent class random-effects models. Thus, it is highly recommended to incorporate the imperfect reference verification as long as it is available.

3.2. Simulation 2: Pseudo-likelihood versus MCEM

This subsection aims at evaluating the efficiency of MPLE via the Newton–Raphson algorithm and the MLE via the MCEM algorithm in models (1) and (2). Prior to the comparison of estimation efficiency, we took one simulated data set as an example, to assess the convergence of the MCEM algorithm described in Section 2.3 in seeking MLEs. One data set was simulated from (2), as in Simulation 1, with $I = 100$, $J = 25$, $b_i \overset{\text{i.i.d.}}{\sim} N(0, 1)$, $c_j \overset{\text{i.i.d.}}{\sim} N(0, 1)$, $\pi_1 = 0.5$, $\beta_1 = 1.795$, $\beta_0 = -1.795$, $\sigma_1 = \sigma_0 = 1$, and $\tau_1 = \tau_0 = 1$. We set $N = 2000$ and $N_0 = 1000$ and started the algorithm from three different combinations of parameter initial values. Fig. 1 shows the change of parameter values along with $R = 200$ MCEM iterations. It is observed in Fig. 1 that the three executions with different initial values converge at around 100th iteration and become stable

Table 1
The estimation biases of sensitivity and specificity, as well as the related parameters, obtained from maximizing the pseudo-likelihoods in the Simulation 1 under $\sigma = \tau = 1$. Standard derivations are reported in the parentheses.

	S_e	S_p	β_1	β_0	σ	τ	π_1
	$I = 100, J = 25$						
$PL_{2,2}(\theta; y)$	$\mathbf{-0.032}_{(\mathbf{0.073})}$	$\mathbf{-0.021}_{(\mathbf{0.069})}$	$-0.023_{(0.626)}$	$-0.078_{(0.697)}$	$0.257_{(0.199)}$	$-0.066_{(0.175)}$	$0.010_{(0.046)}$
$PL_{I,2}(\theta; y)$	$\mathbf{-0.024}_{(\mathbf{0.051})}$	$\mathbf{-0.011}_{(\mathbf{0.044})}$	$0.048_{(0.386)}$	$-0.146_{(0.370)}$	$0.227_{(0.130)}$	$0.100_{(0.102)}$	$0.011_{(0.034)}$
$PL_{2,J}(\theta; y)$	$\mathbf{0.012}_{(\mathbf{0.052})}$	$\mathbf{0.013}_{(\mathbf{0.048})}$	$0.291_{(0.511)}$	$-0.278_{(0.427)}$	$0.119_{(0.197)}$	$0.086_{(0.177)}$	$0.001_{(0.032)}$
$PL_{I,J}(\theta; y)$	$\mathbf{0.001}_{(\mathbf{0.044})}$	$\mathbf{0.002}_{(\mathbf{0.045})}$	$0.050_{(0.342)}$	$-0.058_{(0.330)}$	$-0.013_{(0.156)}$	$0.017_{(0.184)}$	$0.002_{(0.062)}$
$PL_{2,2}(\theta; y, t)$	$\mathbf{-0.000}_{(\mathbf{0.042})}$	$\mathbf{-0.001}_{(\mathbf{0.045})}$	$0.149_{(0.330)}$	$-0.145_{(0.326)}$	$0.109_{(0.216)}$	$0.070_{(0.170)}$	$0.001_{(0.054)}$
$PL_{I,2}(\theta; y, t)$	$\mathbf{-0.001}_{(\mathbf{0.041})}$	$\mathbf{-0.002}_{(\mathbf{0.043})}$	$0.135_{(0.296)}$	$-0.135_{(0.314)}$	$-0.012_{(0.336)}$	$0.158_{(0.160)}$	$0.000_{(0.053)}$
$PL_{2,J}(\theta; y, t)$	$\mathbf{0.000}_{(\mathbf{0.043})}$	$\mathbf{0.000}_{(\mathbf{0.045})}$	$0.185_{(0.367)}$	$-0.180_{(0.351)}$	$0.169_{(0.178)}$	$0.054_{(0.177)}$	$0.000_{(0.054)}$
$PL_{I,J}(\theta; y, t)$	$\mathbf{-0.003}_{(\mathbf{0.039})}$	$\mathbf{-0.004}_{(\mathbf{0.041})}$	$0.047_{(0.330)}$	$-0.048_{(0.340)}$	$0.025_{(0.147)}$	$0.036_{(0.204)}$	$0.000_{(0.053)}$
	$I = 200, J = 50$						
$PL_{2,2}(\theta; y)$	$\mathbf{-0.024}_{(\mathbf{0.045})}$	$\mathbf{-0.019}_{(\mathbf{0.046})}$	$0.007_{(0.363)}$	$-0.048_{(0.369)}$	$0.251_{(0.150)}$	$-0.001_{(0.116)}$	$0.006_{(0.029)}$
$PL_{I,2}(\theta; y)$	$\mathbf{-0.017}_{(\mathbf{0.031})}$	$\mathbf{-0.010}_{(\mathbf{0.029})}$	$0.059_{(0.181)}$	$-0.115_{(0.192)}$	$0.201_{(0.108)}$	$0.095_{(0.064)}$	$0.005_{(0.027)}$
$PL_{2,J}(\theta; y)$	$\mathbf{0.012}_{(\mathbf{0.041})}$	$\mathbf{0.010}_{(\mathbf{0.042})}$	$0.276_{(0.548)}$	$-0.242_{(0.392)}$	$0.088_{(0.124)}$	$0.104_{(0.131)}$	$0.000_{(0.024)}$
$PL_{I,J}(\theta; y)$	$\mathbf{0.003}_{(\mathbf{0.030})}$	$\mathbf{0.003}_{(\mathbf{0.029})}$	$0.096_{(0.235)}$	$-0.092_{(0.224)}$	$0.018_{(0.085)}$	$0.068_{(0.109)}$	$0.001_{(0.042)}$
$PL_{2,2}(\theta; y, t)$	$\mathbf{-0.001}_{(\mathbf{0.032})}$	$\mathbf{0.002}_{(\mathbf{0.031})}$	$0.113_{(0.245)}$	$-0.140_{(0.240)}$	$0.104_{(0.159)}$	$0.064_{(0.112)}$	$-0.002_{(0.038)}$
$PL_{I,2}(\theta; y, t)$	$\mathbf{0.000}_{(\mathbf{0.029})}$	$\mathbf{-0.001}_{(\mathbf{0.028})}$	$0.136_{(0.213)}$	$-0.123_{(0.205)}$	$0.032_{(0.226)}$	$0.142_{(0.111)}$	$0.004_{(0.036)}$
$PL_{2,J}(\theta; y, t)$	$\mathbf{0.001}_{(\mathbf{0.031})}$	$\mathbf{-0.000}_{(\mathbf{0.032})}$	$0.161_{(0.236)}$	$-0.154_{(0.231)}$	$0.157_{(0.130)}$	$0.066_{(0.109)}$	$-0.002_{(0.040)}$
$PL_{I,J}(\theta; y, t)$	$\mathbf{-0.003}_{(\mathbf{0.029})}$	$\mathbf{-0.003}_{(\mathbf{0.028})}$	$0.076_{(0.224)}$	$-0.072_{(0.218)}$	$0.063_{(0.107)}$	$0.071_{(0.128)}$	$0.003_{(0.037)}$
	$I = 500, J = 25$						
$PL^*_{2,2}(\theta; y)$	$\mathbf{-0.023}_{(\mathbf{0.049})}$	$\mathbf{-0.018}_{(\mathbf{0.048})}$	$0.001_{(0.374)}$	$-0.043_{(0.377)}$	$0.277_{(0.094)}$	$-0.068_{(0.144)}$	$0.007_{(0.021)}$
$PL_{I,2}(\theta; y)$	$\mathbf{-0.017}_{(\mathbf{0.032})}$	$\mathbf{-0.010}_{(\mathbf{0.029})}$	$0.063_{(0.177)}$	$-0.119_{(0.200)}$	$0.207_{(0.118)}$	$0.102_{(0.073)}$	$0.005_{(0.027)}$
$PL^*_{2,J}(\theta; y)$	$\mathbf{0.010}_{(\mathbf{0.039})}$	$\mathbf{0.017}_{(\mathbf{0.031})}$	$0.253_{(0.347)}$	$-0.299_{(0.266)}$	$0.132_{(0.112)}$	$0.087_{(0.132)}$	$0.001_{(0.016)}$
$PL_{I,J}(\theta; y)$	$\mathbf{0.001}_{(\mathbf{0.034})}$	$\mathbf{0.002}_{(\mathbf{0.036})}$	$0.109_{(0.284)}$	$-0.110_{(0.270)}$	$0.018_{(0.083)}$	$0.102_{(0.118)}$	$0.000_{(0.032)}$
$PL^*_{2,2}(\theta; y, t)$	$\mathbf{0.002}_{(\mathbf{0.034})}$	$\mathbf{-0.000}_{(\mathbf{0.035})}$	$0.157_{(0.273)}$	$-0.140_{(0.265)}$	$0.116_{(0.104)}$	$0.074_{(0.125)}$	$-0.000_{(0.025)}$
$PL_{I,2}(\theta; y, t)$	$\mathbf{0.001}_{(\mathbf{0.032})}$	$\mathbf{-0.001}_{(\mathbf{0.032})}$	$0.130_{(0.244)}$	$-0.111_{(0.234)}$	$0.038_{(0.139)}$	$0.127_{(0.103)}$	$-0.001_{(0.023)}$
$PL^*_{2,J}(\theta; y, t)$	$\mathbf{0.005}_{(\mathbf{0.033})}$	$\mathbf{0.000}_{(\mathbf{0.033})}$	$0.213_{(0.276)}$	$-0.175_{(0.259)}$	$0.180_{(0.092)}$	$0.060_{(0.137)}$	$-0.001_{(0.023)}$
$PL_{I,J}(\theta; y, t)$	$\mathbf{-0.003}_{(\mathbf{0.032})}$	$\mathbf{-0.005}_{(\mathbf{0.032})}$	$0.073_{(0.249)}$	$-0.056_{(0.243)}$	$0.047_{(0.082)}$	$0.083_{(0.134)}$	$0.000_{(0.024)}$

thereafter with tolerable Monte Carlo variations, which provide evidences that the MCEM algorithm is not sensitive to the initial values. In the implementation of the MCEM algorithm, we fixed $\nu_b^2 = \nu_c^2 = 0.25$, which in this context creates favorable proposal distributions for the Gibbs sampling. Numerical experiments showed that the specification of ν_b^2 and ν_c^2 had negligible impact on the MLE results, but detrimental specification prolonged the computation duration. Details on selecting optimal proposal distributions for the Metropolis–Hastings algorithm can be found in Rosenthal (2011) and references within.

To assess the relative efficiency of different methods, simulation data were generated from $P(Y_{ij} = 1|D_i = d_i, b_i, c_j) = \Phi(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j)$ and $P(Y_{ij} = 1|T_i = t_i, D_i = d_i, b_i, c_j) = \Phi(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j)$ with the same parameter specifications in Simulation 1. Due to the extended computational duration of the MCEM algorithm, only one combination $(I, J) = (100, 25)$ was examined, and the total number of simulation replications remained to be 500. The realizations of the imperfect reference standard T_i were generated from $S^T_{1|1} = S^T_{0|0} = 0.95$, given the true disease status D_i. Each of the simulation data sets was fit to the true model by the pseudo-likelihood approach as in Simulation 1 and by the MCEM algorithm. Table 3 reports relative efficiency results for pairwise likelihood, hybrid likelihood and full likelihood (via the MCEM algorithm). The relative efficiency results were calculated as inverse mean squared error ratios to the dimension-wise likelihood that serves as a reference. Table 3 shows that the dimension-wise likelihood is more efficient than other pseudo-likelihoods in estimating sensitivity and specificity. Yet, the dimension-wise likelihood appears to be less efficient in estimating test-specific variation τ and disease prevalence π. Distinction between the pseudo-likelihoods in estimation efficiency is narrowed when the imperfect reference standard is present. The MCEM algorithm mostly possesses higher efficiency than the pseudo-likelihoods. However, compared to the MLE obtained by the MCEM algorithm, the maximum dimension-wise likelihood estimates can provide fairly comparable estimation efficiency.

4. Additional issues

4.1. Subject-specific and test-specific covariates

In diagnostic medicine, auxiliary covariates, either subject-specific or test(rater)-specific, may be the factors that affect the estimation of diagnostic accuracy. The vector of subject-specific covariates, denoted by W_i, represents the characteristics of subjects such as their biomarker levels. The vector of test-specific covariates, denoted by U_j, represents the characteristics of diagnostic tests or raters. A test-specific covariate can be a binary treatment indicator that distinguishes experimental

Table 2
The estimation biases of sensitivity and specificity, as well as the related parameters, obtained from maximizing the pseudo-likelihoods in the Simulation 1 under $\sigma = \tau = 0.5$. Standard derivations are reported in the parentheses.

	S_e	S_p	β_1	β_0	σ	τ	π_1
	$I = 100, J = 25$						
$PL_{2,2}(\theta; y)$	$\mathbf{-0.016}_{(\mathbf{0.063})}$	$\mathbf{-0.008}_{(\mathbf{0.063})}$	$0.013_{(0.412)}$	$-0.071_{(0.497)}$	$0.118_{(0.100)}$	$-0.054_{(0.095)}$	$0.007_{(0.040)}$
$PL_{I,2}(\theta; y)$	$\mathbf{-0.016}_{(\mathbf{0.037})}$	$\mathbf{-0.004}_{(\mathbf{0.034})}$	$-0.015_{(0.194)}$	$-0.050_{(0.193)}$	$0.113_{(0.064)}$	$0.009_{(0.080)}$	$0.011_{(0.031)}$
$PL_{2,J}(\theta; y)$	$\mathbf{0.006}_{(\mathbf{0.039})}$	$\mathbf{0.005}_{(\mathbf{0.037})}$	$0.070_{(0.224)}$	$-0.065_{(0.197)}$	$0.056_{(0.096)}$	$-0.022_{(0.098)}$	$0.001_{(0.035)}$
$PL_{I,J}(\theta; y)$	$\mathbf{0.000}_{(\mathbf{0.026})}$	$\mathbf{0.000}_{(\mathbf{0.029})}$	$-0.019_{(0.135)}$	$0.019_{(0.144)}$	$-0.023_{(0.067)}$	$-0.056_{(0.095)}$	$0.002_{(0.051)}$
$PL_{2,2}(\theta; y, t)$	$\mathbf{0.000}_{(\mathbf{0.034})}$	$\mathbf{-0.001}_{(\mathbf{0.037})}$	$0.052_{(0.173)}$	$-0.046_{(0.182)}$	$0.042_{(0.168)}$	$0.017_{(0.095)}$	$0.001_{(0.054)}$
$PL_{I,2}(\theta; y, t)$	$\mathbf{-0.002}_{(\mathbf{0.032})}$	$\mathbf{-0.003}_{(\mathbf{0.034})}$	$0.057_{(0.160)}$	$-0.053_{(0.170)}$	$-0.045_{(0.294)}$	$0.097_{(0.099)}$	$0.001_{(0.052)}$
$PL_{2,J}(\theta; y, t)$	$\mathbf{0.001}_{(\mathbf{0.035})}$	$\mathbf{-0.000}_{(\mathbf{0.038})}$	$0.066_{(0.189)}$	$-0.060_{(0.198)}$	$0.084_{(0.091)}$	$0.004_{(0.094)}$	$0.000_{(0.054)}$
$PL_{I,J}(\theta; y, t)$	$\mathbf{-0.001}_{(\mathbf{0.028})}$	$\mathbf{-0.001}_{(\mathbf{0.031})}$	$0.010_{(0.162)}$	$-0.009_{(0.174)}$	$-0.006_{(0.071)}$	$0.002_{(0.112)}$	$0.000_{(0.054)}$
	$I = 200, J = 50$						
$PL_{2,2}(\theta; y)$	$\mathbf{-0.010}_{(\mathbf{0.035})}$	$\mathbf{-0.008}_{(\mathbf{0.036})}$	$0.003_{(0.201)}$	$-0.021_{(0.239)}$	$0.108_{(0.074)}$	$-0.026_{(0.062)}$	$0.004_{(0.023)}$
$PL_{I,2}(\theta; y)$	$\mathbf{-0.011}_{(\mathbf{0.023})}$	$\mathbf{-0.003}_{(\mathbf{0.021})}$	$0.006_{(0.102)}$	$-0.049_{(0.106)}$	$0.101_{(0.051)}$	$0.037_{(0.042)}$	$0.004_{(0.022)}$
$PL_{2,J}(\theta; y)$	$\mathbf{0.011}_{(\mathbf{0.027})}$	$\mathbf{0.011}_{(\mathbf{0.027})}$	$0.101_{(0.162)}$	$-0.100_{(0.141)}$	$0.043_{(0.052)}$	$0.019_{(0.057)}$	$-0.000_{(0.019)}$
$PL_{I,J}(\theta; y)$	$\mathbf{0.000}_{(\mathbf{0.018})}$	$\mathbf{0.001}_{(\mathbf{0.017})}$	$-0.011_{(0.092)}$	$0.008_{(0.091)}$	$-0.013_{(0.035)}$	$-0.027_{(0.061)}$	$0.002_{(0.034)}$
$PL_{2,2}(\theta; y, t)$	$\mathbf{-0.001}_{(\mathbf{0.025})}$	$\mathbf{-0.000}_{(\mathbf{0.025})}$	$0.038_{(0.128)}$	$-0.044_{(0.128)}$	$0.053_{(0.138)}$	$0.016_{(0.060)}$	$0.002_{(0.038)}$
$PL_{I,2}(\theta; y, t)$	$\mathbf{-0.001}_{(\mathbf{0.022})}$	$\mathbf{-0.002}_{(\mathbf{0.022})}$	$0.055_{(0.108)}$	$-0.051_{(0.109)}$	$-0.010_{(0.228)}$	$0.089_{(0.074)}$	$-0.001_{(0.038)}$
$PL_{2,J}(\theta; y, t)$	$\mathbf{-0.000}_{(\mathbf{0.023})}$	$\mathbf{0.000}_{(\mathbf{0.026})}$	$0.053_{(0.123)}$	$-0.058_{(0.138)}$	$0.093_{(0.059)}$	$0.011_{(0.060)}$	$0.001_{(0.037)}$
$PL_{I,J}(\theta; y, t)$	$\mathbf{-0.000}_{(\mathbf{0.020})}$	$\mathbf{-0.001}_{(\mathbf{0.019})}$	$0.008_{(0.115)}$	$-0.005_{(0.115)}$	$-0.001_{(0.038)}$	$0.007_{(0.079)}$	$0.001_{(0.038)}$
	$I = 500, J = 25$						
$PL^*_{2,2}(\theta; y)$	$\mathbf{-0.013}_{(\mathbf{0.039})}$	$\mathbf{0.003}_{(\mathbf{0.043})}$	$-0.011_{(0.208)}$	$-0.082_{(0.256)}$	$0.123_{(0.048)}$	$-0.045_{(0.084)}$	$0.008_{(0.019)}$
$PL_{I,2}(\theta; y)$	$\mathbf{-0.010}_{(\mathbf{0.019})}$	$\mathbf{-0.003}_{(\mathbf{0.017})}$	$0.011_{(0.085)}$	$-0.051_{(0.086)}$	$0.099_{(0.051)}$	$0.041_{(0.042)}$	$0.004_{(0.020)}$
$PL^*_{2,J}(\theta; y)$	$\mathbf{0.010}_{(\mathbf{0.025})}$	$\mathbf{0.007}_{(\mathbf{0.023})}$	$0.087_{(0.146)}$	$-0.071_{(0.120)}$	$0.061_{(0.049)}$	$-0.011_{(0.074)}$	$-0.002_{(0.015)}$
$PL_{I,J}(\theta; y)$	$\mathbf{0.001}_{(\mathbf{0.021})}$	$\mathbf{-0.001}_{(\mathbf{0.022})}$	$0.014_{(0.109)}$	$-0.004_{(0.117)}$	$-0.009_{(0.029)}$	$0.012_{(0.078)}$	$-0.000_{(0.023)}$
$PL^*_{2,2}(\theta; y, t)$	$\mathbf{0.001}_{(\mathbf{0.024})}$	$\mathbf{-0.002}_{(\mathbf{0.024})}$	$0.053_{(0.125)}$	$-0.036_{(0.126)}$	$0.063_{(0.067)}$	$0.023_{(0.064)}$	$0.003_{(0.025)}$
$PL_{I,2}(\theta; y, t)$	$\mathbf{-0.000}_{(\mathbf{0.024})}$	$\mathbf{-0.002}_{(\mathbf{0.023})}$	$0.039_{(0.118)}$	$-0.031_{(0.119)}$	$-0.017_{(0.151)}$	$0.072_{(0.063)}$	$0.003_{(0.023)}$
$PL^*_{2,J}(\theta; y, t)$	$\mathbf{0.003}_{(\mathbf{0.024})}$	$\mathbf{-0.001}_{(\mathbf{0.023})}$	$0.071_{(0.130)}$	$-0.049_{(0.123)}$	$0.091_{(0.042)}$	$0.010_{(0.070)}$	$0.000_{(0.025)}$
$PL_{I,J}(\theta; y, t)$	$\mathbf{-0.003}_{(\mathbf{0.021})}$	$\mathbf{0.001}_{(\mathbf{0.020})}$	$-0.007_{(0.113)}$	$-0.013_{(0.113)}$	$-0.005_{(0.033)}$	$0.007_{(0.088)}$	$-0.001_{(0.025)}$

Table 3
Relative efficiency of dimension-wise likelihood to pairwise and hybrid likelihoods, as well as to the MCEM algorithm, obtained in the Simulation 2.

	S_e	S_p	β_1	β_0	σ	τ	π_1
	$I = 100, J = 25, \sigma = \tau = 1$						
$PL_{2,2}(\theta; y)$	**0.344**	**0.397**	0.256	0.336	0.196	0.970	1.699
$PL_{I,2}(\theta; y)$	**0.660**	**1.018**	0.984	0.479	0.327	1.586	2.738
$PL_{2,J}(\theta; y)$	**0.785**	**0.920**	0.325	0.530	0.466	0.752	3.761
$PL_{I,J}(\theta; y)$	**1.000**	**1.000**	1.000	1.000	1.000	1.000	1.000
MCEM	**1.430**	**1.584**	1.982	1.687	1.444	1.414	1.463
$PL_{2,2}(\theta; y)$	**0.829**	**0.852**	0.723	0.852	0.348	1.274	0.971
$PL_{I,2}(\theta; y)$	**0.911**	**0.908**	0.886	0.931	0.164	0.721	1.029
$PL_{2,J}(\theta; y)$	**0.800**	**0.825**	0.516	0.682	0.321	1.198	0.960
$PL_{I,J}(\theta; y)$	**1.000**	**1.000**	1.000	1.000	1.000	1.000	1.000
MCEM	**1.202**	**1.241**	1.692	1.512	1.518	1.760	1.097
	$I = 100, J = 25, \sigma = \tau = 0.5$						
$PL_{2,2}(\theta; y)$	**0.193**	**0.160**	0.083	0.086	0.182	0.988	1.670
$PL_{I,2}(\theta; y)$	**0.527**	**0.615**	0.511	0.457	0.282	1.846	2.349
$PL_{2,J}(\theta; y)$	**0.527**	**0.538**	0.311	0.421	0.330	1.105	2.232
$PL_{I,J}(\theta; y)$	**1.000**	**1.000**	1.000	1.000	1.000	1.000	1.000
MCEM	**0.961**	**0.960**	0.943	0.980	0.853	1.761	0.934
$PL_{2,2}(\theta; y)$	**0.663**	**0.692**	0.716	0.778	0.162	1.262	0.978
$PL_{I,2}(\theta; y)$	**0.801**	**0.783**	0.884	0.884	0.058	0.521	1.053
$PL_{2,J}(\theta; y)$	**0.620**	**0.660**	0.557	0.628	0.345	1.536	0.963
$PL_{I,J}(\theta; y)$	**1.000**	**1.000**	1.000	1.000	1.000	1.000	1.000
MCEM	**1.252**	**1.279**	1.560	1.444	1.018	1.705	1.101

tests from the predicate tests. A rater-specific covariate can be an ordinal variable characterizing raters' precision levels. Subject-specific and test-specific covariates can be incorporated into (1) or (2) in the same fashion:

$$P(Y_{ij} = 1|D_i = d_i, b_i, c_j) = \hbar^{-1}\left(\beta_{d_i} + \xi'_{d_i} W_i + \zeta'_{d_i} U_j + \sigma_{d_i} b_i + \tau_{d_i} c_j\right), \tag{19}$$

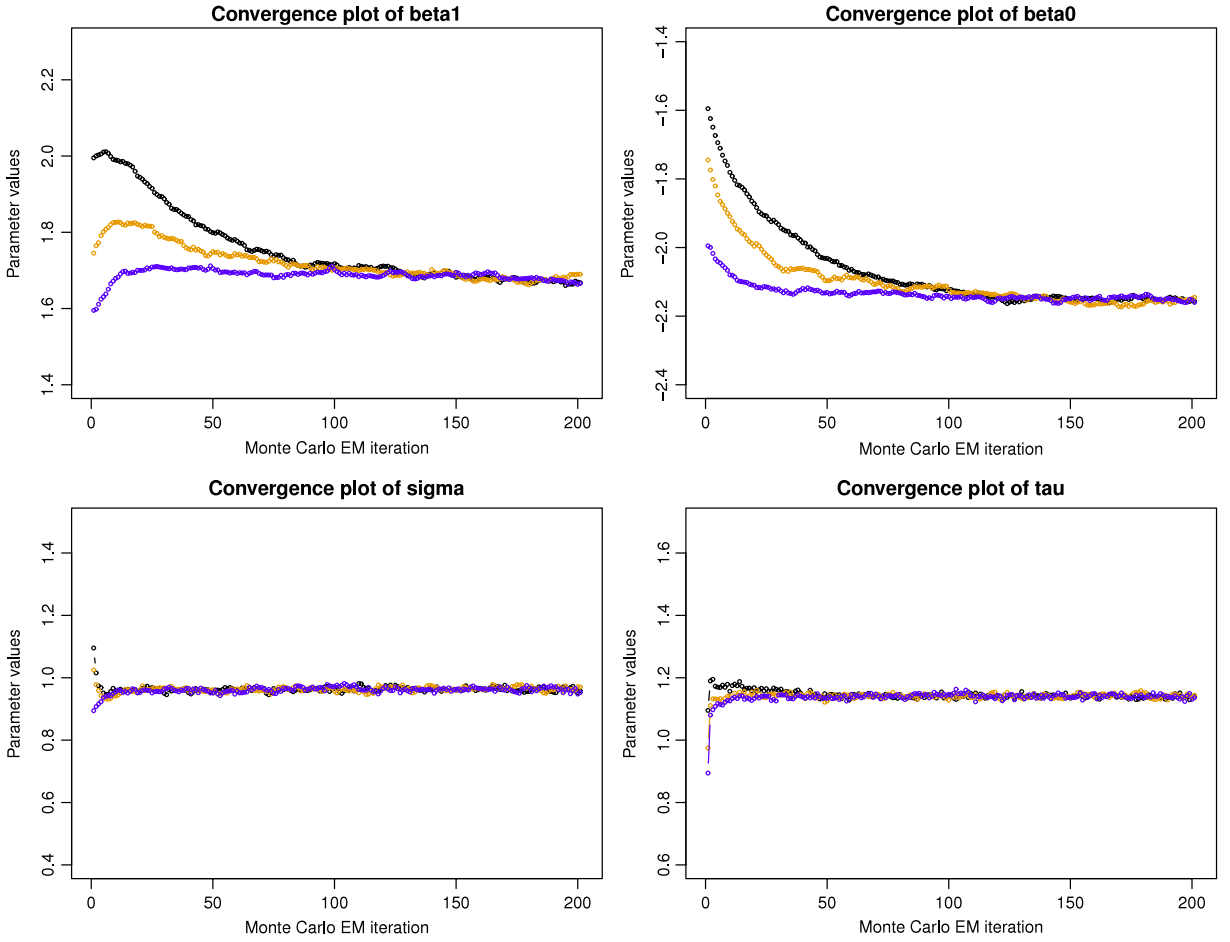

Fig. 1. Convergence plots of the MCEM algorithm.

and

$$P(Y_{ij} = 1|T_i = t_i, D_i = d_i, b_i, c_j) = \hbar^{-1}\left(\beta_{d_i t_i} + \xi'_{d_i t_i} W_i + \zeta'_{d_i t_i} U_j + \sigma_{d_i t_i} b_i + \tau_{d_i t_i} c_j\right). \tag{20}$$

The pseudo-likelihood approach can be implemented for (19) and (20) as described in Section 2.2 to estimate the covariate effects.

4.2. Variance estimation

Inference on θ, whose estimate $\hat{\theta}^c$ is derived from maximizing pseudo-likelihood (5) or (17), can be made based on existing asymptotic theories in Lindsay (1988), Molenberghs and Verbeke (2005) and Varin et al. (2011). The maximum pseudo-likelihood estimator $\hat{\theta}^c$ can be found by solving the pseudo-likelihood score equations $\nabla_\theta p\ell(\theta) = 0$, where $p\ell(\theta)$ is a unified notation that represents $\log PL_{s_1,s_2}(\theta; y)$ and $\log PL_{s_1,s_2}(\theta; y, t)$ with various (s_1, s_2) combinations. Suppose $\hat{\theta}^c_{I,J}$ is the maximum pseudo-likelihood estimator obtained from a $I \times J$ crossed design data matrix. Under sufficient regularity conditions, $\hat{\theta}^c_{I,J}$ asymptotically follows $N(\theta, B(\theta)^{-1}A(\theta)B(\theta)^{-1})$ as $I, J \to \infty$, where $B(\theta) = E\{-\nabla^2_\theta p\ell(\theta)\}$ is the sensitivity matrix, $A(\theta) = \text{var}\{\nabla_\theta p\ell(\theta)\}$ is the variability matrix, and $G(\theta) = B(\theta)A(\theta)^{-1}B(\theta)$ is the Godambe information matrix (Godambe, 1960). Note that, the score statistic $\nabla_\theta p\ell(\theta)$ cannot be written as a sum of independent random components. Yet, the asymptotic normality can still be proved as in Appendix 2 in Lin (1997). To obtain the parameter variance estimation, we need to separately estimate $A(\theta)$ and $B(\theta)$. As discussed in Bellio and Varin (2005), $B(\theta)$ can be estimated by the Hessian of $-p\ell(\theta)$ evaluated by $\hat{\theta}^c$. However, it is impossible to find a close-form estimator for $A(\theta)$ due to the correlation data structure in (1) and (2). Bellio and Varin (2005) suggested a Monte Carlo approach to estimate $A(\theta)$. Given M data sets $Y^*_{(1)}, \ldots, Y^*_{(M)}$ generated under the assumed model (1) or (2) with the true parameter θ replaced by $\hat{\theta}^c$, the Monte Carlo estimate of $A(\theta)$ is given by $\hat{A}(\hat{\theta}^c) = \frac{1}{M}\sum_{m=1}^{M} \nabla_\theta p\ell(\hat{\theta}^c; Y^*_{(m)}) \nabla_\theta p\ell(\hat{\theta}^c; Y^*_{(m)})^T$.

However, a rather simple approach is to estimate the sampling variance of $\hat{\theta}^c$ using a parametric bootstrap approach. To implement the parametric bootstrap, we generate M parametric bootstrap samples $Y^*_{(1)}, \ldots, Y^*_{(M)}$ from (1) or (2) with the true parameter θ replaced by $\hat{\theta}^c$. Specifically, the true disease status is generated according the estimate of π_{d_i}. And, b_i and c_j are generated from standard normal distribution. Then, the response values can be subsequently generated from (1) or (2). After obtaining M parametric bootstrap samples $Y^*_{(1)}, \ldots, Y^*_{(M)}$ and their estimates $\hat{\theta}^c_{(1)}, \hat{\theta}^c_{(2)}, \ldots, \hat{\theta}^c_{(M)}$, parametric bootstrap variance can be calculated as the sample variance of $\{\hat{\theta}^c_{(1)}, \ldots, \hat{\theta}^c_{(M)}\}$. Compared to the variance estimation method in Bellio and Varin (2005), which also requires Monte Carlo simulation under the estimated model, this parametric bootstrap is simpler and its implementation is more straightforward because it does not require calculating the Hessian of $-p\ell(\theta)$ anymore.

4.3. Partial imperfect reference standards

The evaluation of imperfect reference standards may be expensive, time consuming, or unethical to perform on all patients. This is why in some circumstances it is difficult to obtain imperfect reference standards for all subjects. Alternatively, investigators may be able to obtain an imperfect reference standard on a sub-group of subjects, or equivalently, to obtain a "partial imperfect reference standard". We extend the pseudo-likelihood approach in Section 2.2 to estimate diagnostic accuracy with a partial imperfect reference standard, and report the details in the Appendix.

5. Application: Colon cancer detection data

Colorectal cancer is one of the leading causes of cancer-related deaths in the United States (ACS, 2013; Jemal et al., 2003). The American Cancer Society has estimated that approximately 142,820 new cases of colon cancer and 50,830 deaths from the disease occurred in 2013 (ACS, 2013). Because most colorectal cancers arise from benign or malignant polyps, detection and removal of polyps have been proved to reduce the incidence and the mortality of colorectal cancer (Mandel et al., 1993). Virtual colonoscopy, or CT colonography, has been shown very appealing as a computer-aided screening tool for polyp detection (Perumpillichira et al., 2005). Zhou et al. (2002) reported a case study on colon cancer detection, in which each of the 14 readers (physicians) made diagnosis on whether a colon segment from a patient had polyps (0 for no polyp; 1 for having polyps). Totally 130 colon segments were evaluated. Among 14 readers, 7 of them were using conventional colonoscopy, whereas other 7 physicians were using computer-aided CT colonography. In addition to the scores given by 14 readers, the reference standard diagnosis was included for the each colon segment, where "F" represents that no polyp existed in the segment, and "T" denotes one or more polyps were present in the segment.

In this section, we report the numerical results from analyzing the colon cancer detection data, for the purpose of demonstrating the proposed methodologies. Model (1) (without the aid of the reference standard diagnosis) and model (2) (with the reference standard diagnosis) with a probit link were applied to the colon cancer detection data with linear predictor $\beta_{d_i} + \zeta_{d_i} U_j + \sigma b_i + \tau c_j$, in which b_i and c_i were assumed to follow standard normal distributions, σ and τ were assumed to be independent of latent disease classes, and U_j was the indicator for the use of colonography technology ($U_j = 1$ if the jth reader used computer-aided CT colonography; $U_j = 0$ if the jth reader used conventional colonoscopy). The linear predictor representation can help to test whether there was distinction in diagnostic accuracy between using the conventional and new colonography technologies. Parameter estimation was achieved by the Newton–Raphson algorithm that maximized the proposed pseudo-likelihood functions, including pairwise likelihood, dimension-wise likelihood, and two hybrid likelihoods from Section 2. Parametric bootstrap with 100 bootstrap samples was applied for variance estimation.

Table 4 reports the analysis results for colon cancer detection data, including estimated sensitivities and specificities for conventional colonoscopy and computer-aid CT colonoscopy and related parameter estimates (standard errors in the parentheses). The data analysis results effectively confirm the conclusions in Section 3 regarding the performance of the pseudo-likelihoods. The diagnostic measures estimated by the pseudo-likelihood methods without using the reference standard evaluation show some discrepancy from one another. Compared to the scenario with no imperfect reference standard, the sensitivities and specificities estimated by the pseudo-likelihood approach with the aid of the imperfect reference verification are more consistent with each other. Benefits of using the imperfect reference standard are very limited in the bootstrap variance estimation. The sensitivity from using conventional colonoscopy is generally higher than computer-aided CT colonography, while the specificity of computer-aided CT colonography is higher than conventional colonoscopy. However, this conclusion is limited to the current colon segment sample, and therefore cannot be generalized.

6. Discussion

This article proposes a class of pseudo-likelihoods for estimating diagnostic accuracy of a group of tests or raters when using crossed random effects latent class models. It has been shown that the maximum pseudo-likelihood estimates possess satisfactory relative efficiency, compared to the maximum likelihood estimates obtained by the MCEM algorithm. The maximum pseudo-likelihood estimation and the MCEM algorithm for maximum likelihood estimation proposed in this article were implemented in R, and the R code is available upon request. The proposed methodologies to obtain maximum pseudo-likelihood estimates are not limited to latent class models (1) and (2) for estimating sensitivity and specificity. The methods can also be applied to other crossed random effects models that fit to the data collected from a crossed design.

Table 4
Analysis results of colon cancer detection data: estimated sensitivities (S_e's) and specificities (S_p's) for conventional colonoscopy (cvc) and computer-aid CT colonoscopy (ccc) and related parameter estimates. Standard errors are reported in the parentheses.

	S_e(cvc)	S_p(cvc)	S_e(ccc)	S_p(ccc)
$PL_{2,2}(\theta; y)$	$\mathbf{0.910}_{(0.016)}$	$\mathbf{0.714}_{(0.040)}$	$\mathbf{0.814}_{(0.071)}$	$\mathbf{0.758}_{(0.040)}$
$PL_{I,2}(\theta; y)$	$\mathbf{0.905}_{(0.020)}$	$\mathbf{0.758}_{(0.038)}$	$\mathbf{0.828}_{(0.077)}$	$\mathbf{0.803}_{(0.034)}$
$PL_{2,J}(\theta; y)$	$\mathbf{0.863}_{(0.013)}$	$\mathbf{0.741}_{(0.035)}$	$\mathbf{0.809}_{(0.026)}$	$\mathbf{0.791}_{(0.036)}$
$PL_{I,J}(\theta; y)$	$\mathbf{0.913}_{(0.057)}$	$\mathbf{0.828}_{(0.041)}$	$\mathbf{0.838}_{(0.070)}$	$\mathbf{0.857}_{(0.041)}$
$PL_{2,2}(\theta; y, t)$	$\mathbf{0.938}_{(0.044)}$	$\mathbf{0.784}_{(0.031)}$	$\mathbf{0.821}_{(0.055)}$	$\mathbf{0.817}_{(0.031)}$
$PL_{I,2}(\theta; y, t)$	$\mathbf{0.914}_{(0.045)}$	$\mathbf{0.783}_{(0.033)}$	$\mathbf{0.806}_{(0.060)}$	$\mathbf{0.817}_{(0.027)}$
$PL_{2,J}(\theta; y, t)$	$\mathbf{0.941}_{(0.052)}$	$\mathbf{0.784}_{(0.032)}$	$\mathbf{0.823}_{(0.065)}$	$\mathbf{0.817}_{(0.032)}$
$PL_{I,J}(\theta; y, t)$	$\mathbf{0.917}_{(0.042)}$	$\mathbf{0.785}_{(0.036)}$	$\mathbf{0.821}_{(0.064)}$	$\mathbf{0.818}_{(0.027)}$

	β_1	β_0	ζ_1	ζ_0	σ	τ	π_1
$PL_{2,2}(\theta; y)$	$2.147_{(0.223)}$	$-0.905_{(0.189)}$	$-0.718_{(0.575)}$	$-0.216_{(0.112)}$	$1.237_{(0.053)}$	$0.162_{(0.056)}$	$0.125_{(0.059)}$
$PL_{I,2}(\theta; y)$	$1.866_{(0.195)}$	$-0.996_{(0.152)}$	$-0.518_{(0.612)}$	$-0.216_{(0.141)}$	$1.000_{(0.236)}$	$0.162_{(0.058)}$	$0.184_{(0.049)}$
$PL_{2,J}(\theta; y)$	$1.801_{(0.065)}$	$-1.063_{(0.175)}$	$-0.360_{(0.093)}$	$-0.268_{(0.130)}$	$1.295_{(0.059)}$	$0.183_{(0.056)}$	$0.173_{(0.056)}$
$PL_{I,J}(\theta; y)$	$2.005_{(0.711)}$	$-1.394_{(0.206)}$	$-0.548_{(0.357)}$	$-0.182_{(0.166)}$	$1.057_{(0.195)}$	$0.241_{(0.090)}$	$0.236_{(0.053)}$
$PL_{2,2}(\theta; y, t)$	$2.073_{(0.784)}$	$-1.058_{(0.150)}$	$-0.831_{(0.678)}$	$-0.161_{(0.126)}$	$0.883_{(0.119)}$	$0.201_{(0.086)}$	$0.204_{(0.043)}$
$PL_{I,2}(\theta; y, t)$	$1.857_{(0.409)}$	$-1.063_{(0.154)}$	$-0.685_{(0.302)}$	$-0.166_{(0.158)}$	$0.901_{(0.108)}$	$0.198_{(0.081)}$	$0.208_{(0.037)}$
$PL_{2,J}(\theta; y, t)$	$2.580_{(0.958)}$	$-1.297_{(0.208)}$	$-1.048_{(0.801)}$	$-0.197_{(0.145)}$	$1.289_{(0.159)}$	$0.259_{(0.094)}$	$0.204_{(0.041)}$
$PL_{I,J}(\theta; y, t)$	$1.967_{(0.390)}$	$-1.122_{(0.188)}$	$-0.659_{(0.260)}$	$-0.173_{(0.177)}$	$0.992_{(0.093)}$	$0.206_{(0.094)}$	$0.205_{(0.046)}$

For instance, we may apply the strategies of constructing the pseudo-likelihoods to multireader, multicase receiver operating characteristic analysis (Obuchowski et al., 2004) as an aid in estimating the area under the curve from a model with crossed random effects. However, flexible model specification in the pseudo-likelihoods may require more sophisticated maximization algorithm such as the composite likelihood EM algorithm in Gao and Song (2011).

Noticeably, Molenberghs et al. (2011) proposed a partitioned pseudo-likelihood (PPL) approach for partitioning high-dimensional correlated data to achieve the goal of reducing the computational complexity. This method can also be adopted to prevent (5) and (17) from suffering the high-dimensional curse. Suppose we split I subjects into κ_I subgroups and split J tests into κ_J subgroups. The PPL approach breaks the original $I \times J$ data frame into $\kappa_I\kappa_J$ blocks. Then, the pseudo-likelihood approach proposed in this article can be directly applied to each block. Denote $\hat{\theta}_k^c$ the vector of parameter estimates obtained from maximizing the pseudo-likelihood using the data from the kth block, where $k = 1, 2, \ldots, \kappa_I\kappa_J$. Molenberghs et al. (2011) derived the overall estimates to be $\hat{\theta}^c = \sum_{k=1}^{\kappa_I\kappa_J} \hat{\theta}_k^c$. Combined with the PPL approach, the pseudo-likelihoods discussed in this article can manage the diagnostic data with large dimensions. Molenberghs et al. (2011) showed that the PPL approach is fully efficient for independent partitions, which is not the case for the diagnostic data collected from a crossed design. They also showed, with dependent partitions, the PPL estimators are sometimes, but not always, fully efficient. Data analyst is advised to study efficiency loss, perhaps using a simulation study designed after the real application at hand. More details on the efficiency loss from using the PPL approach can be found in Section 5 in Molenberghs et al. (2011), where the authors carefully discussed the strategies to adjust the number of subsamples in the partition, to achieve the goal of only having mild efficiency loss when using the PPL approach.

Although this article has discussed several aspects on the pseudo-likelihood approach, there are still related issues that have not been tackled. More recently, it has been shown that latent class models for estimating diagnostic accuracy may be problematic in many practical situations (Albert and Dodd, 2004; Pepe and Janes, 2006; Xie et al., 2013). Albert and Dodd (2004) showed that with a small number of binary tests, estimates of diagnostic accuracy with random-effects latent class models are biased under a misspecified dependence structure. Zhang et al. (2012) showed that imperfect reference standards can substantially increase the robustness of these models. Further investigation is necessary on whether the proposed pseudo-likelihood approach enjoys the similar robustness when an imperfect reference standard is used. Furthermore, in some diagnostic studies, investigators fail to collect all the data that they expect. Missing data can be response or covariate values, potentially mixed with complex missing mechanism. Whether the proposed pseudo-likelihood approach can appropriately adopted to handle the missing data problems is another area that requires further investigation.

Acknowledgments

The views expressed in this article are those of the authors and do not necessarily represent the views of US Food and Drug Administration. We sincerely thank the anonymous reviewers, Associate Editor, and Editors for their valuable comments, which had substantially improved the manuscript.

Appendix. Partial imperfect reference standards

Denote V_i an indicator of whether the ith patient is verified by the imperfect reference standard T_i, in which $V_i = 1$ represents the ith patient is verified by the imperfect reference standard (i.e., T_i exists) and $V_i = 0$ otherwise (i.e., T_i is

not observed). Let $V = (V_1, V_2, \ldots, V_I)'$ and let $v = (v_1, v_2, \ldots, v_I)'$ be the set of realizations of V. We now introduce a notation for the event of observing a set of partial imperfect reference standard verifications: $\{T^v = t^v\} = \{T_1^{v_1} = t_1^{v_1}, \ldots, T_I^{v_I} = t_I^{v_I}\}$. When $v_i = 0$, $\{T_i^{v_i} = t_i^{v_i}\} = \{1 = 1\}$ becomes a nuisance condition. When $v_i = 1$, then $\{T_i^{v_i} = t_i^{v_i}\} = \{T_i = t_i\}$. As a consequence, $\{T^v = t^v\}$ precisely denotes the event that the current set of partial imperfect reference standard verifications are observed. Following (1) and (2), the likelihood with a partial imperfect reference standard is

$$\begin{aligned} L(\theta; y, t, v) = \sum_{d_1=0,1} \cdots \sum_{d_I=0,1} P(V = v|Y = y, T^v = t^v, D = d) \int \cdots \int \Bigg[\prod_{i=1}^{I} \prod_{j=1}^{J} \Bigg\{ \hbar^{-1}\Big(\ell^*(y_{ij})(\beta_{d_i t_i} \\ + \sigma_{d_i t_i} b_i + \tau_{d_i t_i} c_j) \Big) \Bigg\}^{v_i} \Bigg\{ \hbar^{-1}\Big(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j) \Big) \Bigg\}^{1-v_i} \Bigg] \prod_{i=1}^{I} \pi_{d_i} \Big\{ S^T_{t_i|d_i} \Big\}^{v_i} \\ \times g_1(b_i) \mathrm{d}b_i \prod_{j=1}^{J} g_2(c_j) \mathrm{d}c_j. \end{aligned}$$

If the evaluation mechanism of the partial imperfect reference standard is completely random (i.e., the patients are randomly assigned to be evaluated), then $p = P(V_i = 1)$ is the proportion of subjects that are evaluated by the imperfect reference standard, and $P(V = v|Y = y, T^v = t^v, D = d) = P(V = v) = \prod_{i=1}^{I} p^{v_i}(1-p)^{1-v_i}$. Another type of verification process occurs, when the probability of verification on the ith subject depends on Y_i; that is, $P(V = v|Y = y, T^v = t^v, D = d) = \prod_{i=1}^{I} P(V_i = 1|Y_i = y_i)$. A special case of this verification process is verification biased sampling (Pepe, 2003), in which the probability of imperfect reference standard evaluation depends on the number of positive tests: $P(V = v|Y = y, T^v = t^v, D = d) = \prod_{i=1}^{I} P(V_i = 1|\bar{Y}_i = \sum_{j=1}^{J} y_{ij})$. Another important special case is extreme verification biased sampling (Albert and Dodd, 2008; Walter, 1999; van der Merwe and Maritz, 2002), in which the imperfect reference standard test is obtained only for the subjects that received all positive test results or ratings. Extreme verification biased sampling usually occurs when the imperfect reference verification is an invasive procedure that is unethical to perform on all subjects if any test result is negative.

If we follow the same format of (17), then the pseudo-likelihood with a partial imperfect reference standard should be proposed as

$$\begin{aligned} PL_{s_1,s_2}(\theta; y, t, v) = \prod_{i=1}^{I} \prod_{j_1<j_2<\cdots<j_{s_2}}^{J} P(V_i = v_i, T_i^{v_i} = t_i^{v_i}, Y_{ij_1} = y_{ij_1}, \ldots, Y_{ij_{s_2}} = y_{ij_{s_2}}; \theta) \\ \times \prod_{j=1}^{J} \prod_{i_1<i_2<\cdots<i_{s_1}}^{I} P(V_{i_1} = v_{i_1}, \ldots, V_{i_{s_1}} = v_{i_{s_1}}, T_{i_1}^{v_{i_1}} = t_{i_1}^{v_{i_1}}, \ldots, T_{i_{s_1}}^{v_{s_1}} = t_{i_{s_1}}^{v_{s_1}}, \\ Y_{i_1 j} = y_{i_1 j}, \ldots, Y_{i_{s_1} j} = y_{i_{s_1} j}; \theta). \end{aligned} \tag{21}$$

However, (21) is applicable only when $P(V_i = v_i|Y_{ij} = y_{ij})$ is identifiable for any j, which is true in completely randomized reference standard evaluation, but is apparently not true in verification biased sampling. We therefore suggest to use the following pseudo-likelihood in verification biased sampling:

$$\begin{aligned} PL_{I,J}(\theta; y, t, v) = \prod_{i=1}^{I} P(V_i = v_i, T_i^{v_i} = t_i^{v_i}, Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}; \theta) \\ \times \prod_{j=1}^{J} P(T_1^{v_1} = t_1^{v_1}, \ldots, T_I^{v_I} = t_{v_I}^{v_I}, Y_{1j} = y_{1j}, \ldots, Y_{Ij} = y_{Ij}; \theta). \end{aligned} \tag{22}$$

In (22),

$$\begin{aligned} P(T_1^{v_1} = t_1^{v_1}, \ldots, T_I^{v_I} = t_{v_I}^{v_I}, Y_{1j} = y_{1j}, \ldots, Y_{Ij} = y_{Ij}; \theta) \\ = \int_{c_j} \Bigg[\prod_{i=1}^{I} \sum_{d_i=0,1} \pi_{d_i} \Big\{ S^T_{t_i|d_i} \Big\}^{v_i} \int_{b_i} \Bigg\{ \hbar^{-1}\Big(\ell^*(y_{ij})(\beta_{d_i t_i} + \sigma_{d_i t_i} b_i + \tau_{d_i t_i} c_j) \Big) \Bigg\}^{v_i} \\ \times \Bigg\{ \hbar^{-1}\Big(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j) \Big) \Bigg\}^{1-v_i} g_1(b_i) db_i \Bigg] g_2(c_j) dc_j \end{aligned}$$

is the marginal density of all observed imperfect reference evaluations and all ratings from the jth test or rater, and

$$\begin{aligned} P(V_i = v_i, T_i^{v_i} = t_i^{v_i}, Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}; \theta) \\ = \sum_{d_i=0,1} P(V_i = v_i|Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}) \pi_{d_i} \Big\{ S^T_{t_i|d_i} \Big\}^{v_i} \int_{b_i} \Bigg[\prod_{j=1}^{J} \int_{c_j} \Bigg\{ \hbar^{-1}\Big(\ell^*(y_{ij})(\beta_{d_i t_i} \\ + \sigma_{d_i t_i} b_i + \tau_{d_i t_i} c_j) \Big) \Bigg\}^{v_i} \Bigg\{ \hbar^{-1}\Big(\ell^*(y_{ij})(\beta_{d_i} + \sigma_{d_i} b_i + \tau_{d_i} c_j) \Big) \Bigg\}^{1-v_i} g_2(c_j) dc_j \Bigg] g_1(b_i) db_i. \end{aligned}$$

The advantage of employing (22), instead of (21), is that the probability of imperfect standard verification V_i that depends on Y_i can be appropriately integrated into the pseudo-likelihood.

References

Albert, P.S., 2009. Estimating diagnostic accuracy of multiple binary tests with an imperfect reference standard. Stat. Med. 28, 780–797.
Albert, P.S., 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., Dodd, L.E., 2008. On estimating diagnostic accuracy from studies with multiple raters and partial gold standard evaluation. J. Amer. Statist. Assoc. 103, 61–73.
Albert, P.S., McShane, L.M., Shih, J.H., 2001. Latent class modeling approaches for assessing diagnostic error without a gold standard: with applications to p53 immunohistochemical assays in bladder tumors. Biometrics 57, 610–619.
American Cancer Society 2013. Cancer facts and figures. Available from: http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2013/index.
Bellio, R., Varin, C., 2005. A pairwise likelihood approach to generalized linear models with crossed random effects. Stat. Model. 5, 217–227.
Booth, J.G., Hobert, J.P., 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. J. R. Stat. Soc. Ser. B 61, 265–285.
Booth, J.G., Hobert, J.P., Jank, W., 2001. A survey of Monte Carlo algorithms for maximizing the likelihood of a two-stage hierarchical model. Stat. Model. 1, 333–349.
Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Amer. Statist. Assoc. 88, 9–25.
Gao, X., Song, P.X., 2011. Composite likelihood EM algorithm with applications to multivariate hidden Markov model. Statist. Sinica 21, 165–185.
Godambe, V., 1960. An optimum property of regular maximum likelihood estimation. Ann. Math. Statist. 31, 1208–1211.
Hui, S.L., Zhou, X.H., 1998. Evaluation of diagnostic tests without a gold standard. Stat. Methods Med. Res. 7, 354–370.
Jemal, A., Murray, T., Samuels, A., Ghafoor, A., Ward, E., Thun, M.J., 2003. Cancer statistics, 2003. Cancer J. Clin. 53, 5–26.
Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974.
Lee, Y., Nelder, J.A., 2001. Hierarchical generalized linear models: A synthesis of generalised linear models, random-effects models and structured dispersions. Biometrika 88, 987–1006.
Lin, X., 1997. Variance component testing in generalised linear models with random effects. Biometrika 84, 309–326.
Lin, X., Breslow, N.E., 1996. Bias correction in generalized linear mixed models with multiple components of dispersion. J. Amer. Statist. Assoc. 91, 1007–1016.
Lindsay, B., 1988. Composite likelihood methods. Contemp. Math. 80, 220–239.
Mandel, J.S., Bond, J.H., Church, T.R., et al., 1993. Reducing mortality from colorectal cancer by screening for fecal occult blood. N. Engl. J. Med. 328, 1365–1371.
McCulloch, C.E., 1997. Maximum likelihood algorithms for generalized linear mixed models. J. Amer. Statist. Assoc. 92, 162–170.
McCulloch, C.E., Searle, S.R., 2001. Generalized, Linear, and Mixed Models. Wiley, New York.
Molenberghs, G., Verbeke, G., 2005. Models for Discrete Longitudinal Data. Springer-Verlag, New York.
Molenberghs, G., Verbeke, G., Iddi, S., 2011. Pseudo-likelihood methodology for partitioned large and complex samples. Statist. Probab. Lett. 81, 892–901.
Obuchowski, N.A., Beiden, S.V., Berbaum, K.S., Hillis, S.L., Ishwaran, H., Song, H.H., Wagner, R.F., 2004. Multireader, multicase receiver operating characteristic analysis: An empirical comparison of five methods. Acad. Radiol. 11, 980–995.
Pepe, M.S., 2003. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press, New York.
Pepe, M.S., Janes, H., 2006. Insights into latent class analysis of diagnostic test performance. Biostatistics 8, 474–484.
Perumpillichira, J.J., Yoshida, H., Sahani, D.V., 2005. Computer-aided detection for virtual colonoscopy. Cancer Imaging 5, 1–16.
Qu, Y., Hadgu, A., 1998. A model for evaluating sensitivity and specificity for correlated diagnostic tests in efficacy studies with an imperfect reference standard. J. Amer. Statist. Assoc. 93, 920–928.
Qu, Y., Tan, M., Kutner, M.H., 1996. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. Biometrics 52, 797–810.
Rosenthal, J.S., 2011. Optimal proposal distributions and adaptive MCMC. In: Handbook of Markov Chain Monte Carlo. pp. 93–112.
Steele, B.M., 1996. A modified EM algorithm for estimation in generalized mixed models. Biometrics 52, 1295–1310.
Valenstein, P.N., 1990. Evaluating diagnostic tests with imperfect standards. Am. J. Clin. Path. 93, 252–258.
van der Merwe, L., Maritz, J.S., 2002. Estimating the conditional false-positive rate for semi-latent data. Epidemiology 13, 424–430.
Varin, C., Reid, N., Firth, D., 2011. An overview of composite likelihood methods. Statist. Sinica 21, 5–42.
Walter, S.D., 1999. Estimation of test sensitivity and specificity when disease confirmation is limited to positive results. Epidemiology 10, 67–72.
Whiting, P., Rutjes, A.W.S., Reitsma, J.B., Bossuyt, P.M., Kleijnen, J., 2004. Sources of variation and bias in studies of diagnostic accuracy. Ann. Intern. Med. 140, 189–202.
Xie, Y., Chen, Z., Albert, P.S., 2013. A crossed random effects modeling approach for estimating diagnostic accuracy from ordinal ratings without a gold standard. Stat. Med. 32, 3472–3485.
Zeger, S.L., Karim, M.R., 1991. Generalized linear models with random effects: a Gibbs sampling approach. J. Amer. Statist. Assoc. 86, 79–86.
Zhang, B., Chen, Z., Albert, P.S., 2012. Estimating diagnostic accuracy of raters without a gold standard by exploiting a group of experts. Biometrics 68, 1294–1302.
Zhou, X.H., Obuchowski, N.A., McClish, D.K., 2002. Statistical Methods in Diagnostic Accuracy. Wiley, New York.