Published in final edited form as:

Stat Med. 2017 January 15; 36(1): 105–121. doi:10.1002/sim.7093.

A simple and robust method for multivariate meta-analysis of diagnostic test accuracy

Yong Chen[a,*,†], **Yulun Liu**[a], **Haitao Chu**[b], **Mei-Ling Ting Lee**[c], and **Christopher H. Schmid**[d]

[a]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

[b]Division of Biostatistics, University of Minnesota School of Public Health, Minneapolis, MN 55455, U.S.A.

[c]Department of Epidemiology and Biostatistics, The University of Maryland School of Public Health, College Park, MD 20742, U.S.A.

[d]Department of Biostatistics, Brown University School of Public Health, Providence, RI 02903, U.S.A.

Abstract

Meta-analysis of diagnostic test accuracy often involves mixture of case–control and cohort studies. The existing bivariate random-effects models, which jointly model bivariate accuracy indices (e.g., sensitivity and specificity), do not differentiate cohort studies from case–control studies and thus do not utilize the prevalence information contained in the cohort studies. The recently proposed trivariate generalized linear mixed-effects models are only applicable to cohort studies, and more importantly, they assume a common correlation structure across studies and trivariate normality on disease prevalence, test sensitivity, and specificity after transformation by some pre-specified link functions. In practice, very few studies provide justifications of these assumptions, and sometimes these assumptions are violated. In this paper, we evaluate the performance of the commonly used random-effects model under violations of these assumptions and propose a simple and robust method to fully utilize the information contained in case–control and cohort studies. The proposed method avoids making the aforementioned assumptions and can provide valid joint inferences for any functions of overall summary measures of diagnostic accuracy. Through simulation studies, we find that the proposed method is more robust to model misspecifications than the existing methods. We apply the proposed method to a meta-analysis of diagnostic test accuracy for the detection of recurrent ovarian carcinoma.

Keywords

composite likelihood; diagnostic accuracy study; diagnostic review; meta-analysis; multivariate beta-binomial model; Sarmanov family

[*]Correspondence to: Yong Chen, Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104, U.S.A. [†] ychen123@mail.med.upenn.edu.

1 Introduction

Multivariate meta-analysis is increasingly popular in biomedical research [1]. One important area of multivariate meta-analysis is the meta-analysis of diagnostic test accuracy, which synthesizes summary statistics on diagnostic accuracy across studies. The summary statistics are often paired indices, such as sensitivity and specificity or positive and negative predictive values (PPV and NPV, respectively) [2]. When pooling the paired indices, at least three important features need to be taken into consideration. The first feature is that the estimated sensitivity and specificity are typically (negatively) correlated across studies [3]. The second feature is that substantial between-study heterogeneity is often found in those paired indices [4–6]. The third feature is the possible dependence between the disease prevalence and diagnostic accuracy (e.g., sensitivity and specificity) [7]. Such dependence has been discussed in the literature extensively and was referred to as ‘spectrum effect’ [8–10].

The first two features of data are well acknowledged in the literature. A variety of methods have been developed to account for them, including the bivariate generalized linear mixed-effects model (BGLMM), which jointly models the study-specific sensitivity and specificity after some transformations (e.g., logit) [3, 11–16], and the hierarchical summary receiver operating characteristic model, which assumes the test results are determined by the latent ‘cutoff’ and ‘accuracy’ values following certain location-scale distributions (e.g., logistic distribution) [5, 14, 17]. It is worth mentioning that these two models are in fact closely related and are identical in the absence of covariates [18]. More recently, Chu *and others* proposed a trivariate generalized linear mixed-effects model (TGLMM) to account for all three aforementioned features. Specifically, the TGLMM jointly models the study-specific disease prevalence, test sensitivity, and specificity as random effects in cohort studies, allowing the dependence between disease prevalence and diagnostic test accuracy. Perhaps most importantly, this model allows inferences on clinically more meaningful indices, such as PPV and NPV [20]. Because the corresponding likelihood quadrature approximation are often used in obtaining the maximum likelihood inference.

The aforementioned methods for meta-analysis of diagnostic tests are based on multivariate normality assumption on the random effects after some pre-specified transformations, or equivalently, assuming the random effects follow certain distributions. Another implicitly made assumption is that the between-study correlations are the same across studies, namely, homogeneity assumption. In other words, the study-specific effects across studies are assumed to follow a common distribution, often a normal distribution. However, in practice, very few studies provide justifications of these assumptions. To the best of our knowledge, there has been no systematic investigation on how sensitive is the inference to misspecified transformations and to heterogeneous correlation structures.

Our motivating example is a systematic review of diagnostic modalities for detection of recurrent ovarian carcinoma. Ovarian cancer is the fifth leading cause of cancer death among American women, more frequently found among women older than 60 years [21]. Although second-look laparotomy is regarded as the ‘gold standard’ for the evaluation of recurrent ovarian carcinoma [22, 23], it is an invasive modality with potential surgical complications and the introduced risk of anesthesia [24]. Even though many studies have been conducted

for evaluating the diagnostic performance of noninvasive methods including cancer antigen 125 (CA125), positron emission tomography (PET) alone, a combination of PET and computed tomography (PET–CT), computed tomography (CT), and MRI, the optimal diagnostic tool has not yet been identified. To this end, Gu *and others* [24] conducted a meta-analysis based on 52 published studies from January 1995 to November 2007 using the fixed-effects regression (summary receiver operating characteristic) models proposed by Moses *and others* [4]. However, the limitations of this fixed-effects regression model such as the ignored uncertainty in regression covariates have been well recognized in the literature. These limitations can lead to biased estimates of diagnostic accuracy measures. For more details, see Rutter and Gatsonis [5] and Arends *and others* [14]. On the other hand, the TGLMM method cannot be applied because the disease prevalence violates the normality assumption after the logit transformation (p-value = 0.008 using the Shapiro–Wilk test). Furthermore, we found that different choices of initial values lead to very different results and sometimes the iterative maximum likelihood algorithm estimates does not converge. For more details, refer to Section 4. This motivates us to develop a method that requires fewer distributional assumptions and is more computationally stable.

The goal of this paper is to first investigate the performance of generalized linear mixed-effects models under various model misspecifications and to propose a simple and robust method that requires fewer model assumptions and is robust to assumptions made by existing methods including specification of transformations and homogeneous between-study correlation structure. Specifically, we first propose the use of the Sarmanov distribution [25] to jointly model study-specific prevalence, sensitivity, and specificity without any transformation while accounting for their between-study correlations. To account for the vital limitation of the Sarmanov distribution that the admissible correlations can be very restrictive [26, 27], we propose a novel composite likelihood method [28] based on a factorization of the multinomial distribution into a product of binomial distributions. The proposed method can be implemented by simply applying standard methods for univariate meta-analysis to make marginal inference for each summary measure (e.g., overall disease prevalence, test sensitivity, and specificity). However, we augment the separate univariate meta-analyses by also estimating the covariance of the univariate pooled estimates. The proposed method does not suffer from any convergence problem and can provide valid inference for joint summary measures (e.g., confidence region for sensitivity and specificity) and for functions of summary measures (e.g., PPV and NPV).

The rest of this paper is organized as follows. In Section 2, we describe the proposed model and the composite likelihood inference. In Section 3, we conduct simulation studies to evaluate the performance of the proposed method and existing methods and examine the impacts of violations of model assumptions. We illustrate our method in Section 4 by applying the proposed method to the motivating example. We provide a brief discussion in Section 5.

2 Statistical methodology

Denote by D the disease status ascertained by a gold standard and by T the result from a diagnostic test under investigation (1: positive; 0: negative). Sensitivity (Se) and Specificity

(Sp) are respectively defined as the probability of a positive test result for a subject with the disease and the probability of a negative result for a subject without the disease; that is, Se = Pr($T = 1|D = 1$) and Sp = Pr($T = 0|D = 0$). PPV and NPV are defined as the probability of having the disease for a subject with a positive test result and the probability of not having the disease for a subject with a negative test result, respectively; that is PPV = Pr(D = 1|T = 1) and NPV = Pr(D = 0|T = 0).

We consider a meta-analysis with m studies. For simplicity of notation, we assume that the first m_1 studies are case–control studies and the remaining $m_2 = m - m_1$ studies are cohort studies. Typically, the data for each study are summarized by a 2 × 2 table. Specifically, denote by n_{i11}, n_{i00}, n_{i01}, and n_{i10} the number of true positives, true negatives, false positives, and false negatives, respectively. We also denote by $n_{i+1} = n_{i01} + n_{i11}$ and $n_{i+0} = n_{i00} + n_{i10}$ the number of subjects with the disease and the number of healthy individuals, respectively, and denote by n_{i++} the total number of subject in the ith study. Let π_i, Se_i, and Sp_i be the study-specific disease prevalence, sensitivity, and specificity, respectively. Note that π_i estimable only in cohort studies; that is, $i = m_1 + 1, \ldots, m$.

In practice, there is often a great heterogeneity in study-specific disease prevalence, sensitivity, and specificity across studies. Such heterogeneity arises owing to differences in study population characteristics, laboratory methods, and so on. On the other hand, it is well understood that sensitivity and specificity are often negatively correlated. In addition, for cohort studies, it has been suggested that sensitivity and specificity can be correlated with disease prevalence as well [19]. This is because for classification of disease status based on continuous traits, the underlying distribution of the continuous traits not only determines disease prevalence but also determines misclassification rates (i.e., sensitivity and specificity). In other words, subjects with true levels close to the cut point are more likely to be misclassified [7].

To account for both between-study heterogeneity and correlations, we propose the following hybrid beta-binomial (HBB) model, which combines case–control and cohort studies. Given study-specific effects, the results from a case–control study are modeled by the product of binomial distributions, that is, $(n_{i11}, n_{i00})|(\mathrm{Se}_i, \mathrm{Sp}_i) \sim \mathrm{Binomial}(n_{i+1}; \mathrm{Se}_i) \times \mathrm{Binomial}(n_{i+0}, \mathrm{Sp}_i)$, and the results from a cohort study are often modeled by a multinomial distribution, that is, $(n_{i11}, n_{i10}, n_{i01}, n_{i00})|(\pi_i, \mathrm{Se}_i, \mathrm{Sp}_i) \sim \mathrm{Multinomial}(n_{i++}; \pi_i\mathrm{Se}_i, (1 - \pi_i)(1 - \mathrm{Sp}_i), \pi_i(1 - \mathrm{Se}_i), (1 - \pi_i)\mathrm{Sp}_i)$. To model the study-specific effects, we propose to model them on the original scale rather than on a transformed scale because the choice of transformation can be arbitrary. Specifically, we propose to use the following bivariate and trivariate Sarmanov beta distributions [26]: for $i = 1, \ldots, m_1$ (i.e., case–control studies),

$$(\mathrm{Se}_i, \mathrm{Sp}_i) \sim g_2(p_1, p_2; \mu_2, \delta_1, \mu_2, \delta_2, \rho_{12}),$$

$$g_2(p_1, p_2; \mu_1, \delta_1, \mu_2, \delta_2, \rho_{12}) = \mathrm{beta}\,(p_1; \mu_1, \delta_1)\,\mathrm{beta}\,(p_2; \mu_2, \delta_2)\left\{1 + \rho_{12}\frac{(p_1 - \mu_1)(p_2 - \mu_2)}{\delta_1\delta_2}\right\};$$

for $i = m_1 + 1,\ldots,m$ (i.e., cohort studies),

$$(\pi_i, \mathrm{Se}_i, \mathrm{Sp}_i) \sim g_3(p_0, p_1, p_2; \mu_0, \delta_0, \mu_1, \delta_1, \mu_2, \delta_2, \rho_{01}, \rho_{02}, \rho_{12}),$$

$$\begin{aligned} g_3(p_0, p_1, p_2; \mu_0, \delta_0, \mu_1, \delta_1, \mu_2, \delta_2, \rho_{01}, \rho_{02}, \rho_{12}) &= \mathrm{beta}(p_0; \mu_0, \delta_0)\,\mathrm{beta}(p_1; \mu_1, \delta_1)\,\mathrm{beta}(p_2; \mu_2, \delta_2) \\ &\times \left\{1 + \rho_{01}\frac{(p_0 - \mu_0)(p_1 - \mu_1)}{\delta_0\delta_1} + \rho_{02}\frac{(p_0 - \mu_0)(p_2 - \mu_2)}{\delta_0\delta_1} \right. \\ &\left. + \rho_{12}\frac{(p_1 - \mu_1)(p_2 - \mu_2)}{\delta_1\delta_2}\right\}, \end{aligned}$$

where μ_0, μ_1, and μ_2 are the overall disease prevalence, sensitivity, and specificity; δ_0, δ_1, and δ_2 are the standard deviations (heterogeneities) of study-specific effects; and ρ_{01}, ρ_{02}, and ρ_{12} are the pairwise correlations among the study-specific prevalence, sensitivity, and specificity. The Sarmanov beta distributions have the following attractive features. First, the marginal distributions of π_j, Se_j, and Sp_j are beta distributions. Second, the parameters (ρ_{01}, ρ_{02} ρ_{12}) characterize the pairwise correlations among the study-specific effects in the original scale, which have intuitive interpretations. In contrast, the correlation parameters in the TGLMM are correlations in the transformed scale, whose interpretations are transformation dependent. Third, the Sarmanov beta distributions are pseudo-conjugate to binomial distributions, which leads to a closed-form expression of the likelihood function. For more details on properties of Sarmanov distributions, we refer the readers of interest to Lee [26] and Shubina and Lee [27].

Under the assumption of Sarmanov distributions, the model parameters can be estimated by maximiz-ing the full likelihood (hereafter referred to as the HBB-FL method). Denote $\boldsymbol{\theta_0} = (\mu_0, \delta_0)$, $\boldsymbol{\theta_1} = (\mu_1, \delta_1)$, $\boldsymbol{\theta}_2 = (\mu_2, \delta_2)$, and $\boldsymbol{\rho} = (\rho_{01}, \rho_{02}, \rho_{12})$. The log-likelihood function of $(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \boldsymbol{\rho})$ can be written as

$$\log L(\boldsymbol{\theta_0}, \boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \boldsymbol{\rho}) = \sum_{i=1}^{m_1} \log \Pr(n_{i00}, n_{i11}; n_{i+0}, n_{i+1}, \boldsymbol{\theta_1}, \boldsymbol{\theta}_2, \rho_{12}) + \sum_{i=m_1+1}^{m} \log \Pr(n_{i00}, n_{i01};$$

$$n_{i11};, \boldsymbol{\theta_1}, \boldsymbol{\theta}_2, \boldsymbol{\rho})$$

$$= \sum_{i=1}^{m} \log \iint \text{Binomial}\,(n_{i11}; Se_i) \times \text{Binomial}\,(n_{i00}; Sp_i)$$

$$\times g_2(Se_i, Sp_i; \boldsymbol{\theta_1}, \boldsymbol{\theta}_2, \rho_{12})\, dSe_i dSp_i$$

$$+ \sum_{i=m_1+1}^{m} \log \iiint \text{Multinomial}\,(n_{i11}, n_{i10}, n_{i01}, n_{i00}; n_{i++}, \pi_i Se_i, (1-\pi_i)(1-Sp_i),$$

$$\times\ \pi_i(1-Se_i), (1-\pi_i)Sp_i)$$

$$\times g_3(\pi_i, Se_i, Sp_i; \boldsymbol{\theta}_0, \boldsymbol{\theta_1}, \boldsymbol{\theta}_2, \rho)\, d\pi_i dSe_i dSp_i,$$

$$= \sum_{i=1}^{m_1} \log\left[P_{BB}(n_{i11}; n_{i+1}, \boldsymbol{\theta_1}) \mathrm{P}_{\mathrm{BB}}(n_{i00}; n_{i+0}, \boldsymbol{\theta}_2)\left\{1 + \rho_{12}\psi_1(n_{i11}; \boldsymbol{\theta_1})\psi_2(n_{i00}; \boldsymbol{\theta}_2)\right\}\right]$$

$$+ \sum_{i=m_1+1}^{m} \log[\mathrm{P}_{\mathrm{BB}}(n_{i+1}; n_{i++}, \boldsymbol{\theta}_0)\, P_{BB}(n_{i11}; n_{i+1}, \boldsymbol{\theta_1})\, P_{BB}(n_{i00}; n_{i+0}, \boldsymbol{\theta}_2)$$

$$\times \left\{1 + \rho_{01}\psi_0(n_{i+1}; \boldsymbol{\theta}_0)\psi_1(n_{i11}; \boldsymbol{\theta_1}) + \rho_{02}\psi_0(n_{i+1}; \boldsymbol{\theta}_0)\psi_2(n_{i00}; \boldsymbol{\theta}_2)\right.$$

$$\left. + \rho_{12}\psi_1(n_{i11}; \theta_1)\psi_2(n_{i00}; \boldsymbol{\theta}_2)\right\}],$$

(1)

Where

$$\psi_0(n_{i+1}; \boldsymbol{\theta}_0) = \frac{1}{\delta_0}\left[\frac{n_{i+1}\delta_0^2 + \mu_0\left\{\mu_0(1-\mu_0) - \delta_0^2\right\}}{n_{i++}\delta_0^2 + \left\{\mu_0(1-\mu_0) - \delta_0^2\right\}} - \mu_0\right],$$

$$\psi_{\mathbf{1}}(n_{i\mathbf{11}};\boldsymbol{\theta}_{\mathbf{1}}) = \frac{\mathbf{1}}{\delta_{\mathbf{1}}}\left[\frac{n_{i\mathbf{11}}\delta_{\mathbf{1}}^2 + \mu_{\mathbf{1}}\left\{\mu_{\mathbf{1}}(\mathbf{1}-\mu_{\mathbf{1}}) - \delta_{\mathbf{1}}^2\right\}}{n_{i+\mathbf{1}}\delta_{\mathbf{1}}^2 + \left\{\mu_{\mathbf{1}}(\mathbf{1}-\mu_{\mathbf{1}}) - \delta_{\mathbf{1}}^2\right\}} - \mu_{\mathbf{1}}\right],$$

$$\psi_{2}(n_{i00};\boldsymbol{\theta}_{2}) = \frac{\mathbf{1}}{\delta_{\mathbf{1}}}\left[\frac{n_{i00}\delta_{2}^2 + \mu_{2}\left\{\mu_{2}(\mathbf{1}-\mu_{2}) - \delta_{2}^2\right\}}{n_{i+0}\delta_{2}^2 + \left\{\mu_{2}(\mathbf{1}-\mu_{2}) - \delta_{2}^2\right\}} - \mu_{\mathbf{1}}\right],$$

and $P_{BB}(y; n, \mu, \delta)$ is the probability mass function of a beta binomial distribution; that is,

$$P_{BB}(y; n, \mu, \delta) = \binom{n}{y}\frac{B(y+\mu\left\{\mu(\mathbf{1}-\mu)-\delta^2\right\}/\delta^2, n-y+(\mathbf{1}-\mu)\left\{\mu(\mathbf{1}-\mu)-\delta^2\right\}/\delta^2}{B(\mu\left\{\mu(\mathbf{1}-\mu)-\delta^2\right\}/\delta^2, (\mathbf{1}-\mu)\left\{\mu(\mathbf{1}-\mu)-\delta^2\right\}/\delta^2)}.$$

The HBB-FL method accounts for all possible correlations between study-specific prevalence, sen-sitivity, and specificity. Although multi-dimensional integrations are involved in the likelihood of the HBB model, the likelihood has a closed-form expression due to the pseudo-conjugation of Sarmanov beta distributions. Such a computational convenience is advantageous over the commonly used multivariate generalized linear mixed-effects model, which requires numerical integrations and can be unstable, nonconvergent, or sensitive to the choice of initial values [29].

However, an intrinsic limitation of the Sarmanov beta-binomial model (1) is that the admissible values of correlations can be much narrower than [−1, 1] [26, 27, 30, 31]. Such a limitation can be vital because substantial negative correlation between study-specific sensitivity and specificity is often found [2]. For example, the gray region in Figure 1 displays the region of admissible values for (ρ_{02},ρ_{12}) when $\rho_{01} = 0, \mu_0 = \mu_1 = \mu_2 = 0.5$ and $\delta_0 = \delta_1 = \delta_2 = 0.16$, which is too restrictive to model any meaningful correlations.

To circumvent such limitation, we propose the following composite likelihood method [28]. The basic idea of composite likelihood is to construct a likelihood function by multiplying a set of marginal or conditional densities, which avoids specification of the full-likelihood function [28, 32]. The motivation is that the commonly used measures of diagnostic accuracies including overall sensitivity, specificity, PPV, and NPV are functions of $(\mu_0, \mu_{\mathbf{1}}, \mu_2)$ and are not functions of $(\rho_{01}, \rho_{02}, \rho_{\mathbf{12}})$. For example, PPV can be calculated as $PPV = \mu_0\mu_{\mathbf{1}}\left\{\mu_0\mu_{\mathbf{1}} + (\mathbf{1}-\mu_0)(\mathbf{1}-\mu_2)\right\}^{-\mathbf{1}}$. Another motivation is that the likelihood function based on a cohort study for given $(\pi_i, \mathrm{Se}_i, \mathrm{Sp}_i)$ can be factorized as

$$\begin{aligned}&\mathrm{Multinomial}(n_{i\mathbf{11}}, n_{i\mathbf{10}}, n_{i01}, n_{i00}; n_{i++}, \pi_i Se_i, (\mathbf{1}-\pi_i)(\mathbf{1}-Sp_i), \pi_i(\mathbf{1}-Se_i), (\mathbf{1}-\pi_i)Sp_i) \\ &\propto \mathrm{Binomial}(n_{i+\mathbf{1}}|n_{i++}; \pi_i) \times \mathrm{Binomial}(n_{i\mathbf{11}}|n_{i+\mathbf{1}}; Se_i) \times \mathrm{Binomial}(n_{i00}|n_{i+0}; Sp_i),\end{aligned} \quad (2)$$

where $i = m_1 + 1, \ldots, m$. Importantly, each binomial distribution in Equation (2) only involves one random effect. Thus, we propose to construct a composite likelihood by setting all the correlations $(\rho_{01}, \rho_{02}, \rho_{12})$ to 0. We obtain the following composite likelihood:

$$\log L_c(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \log L_0(\boldsymbol{\theta}_0) + \log L_1(\boldsymbol{\theta}_1) + \log L_2(\boldsymbol{\theta}_2), \quad (3)$$

where

$$\log L_0(\boldsymbol{\theta}_0) = \sum_{i=m_1+1}^{M} \log \Pr(n_{i+1}; n_{i++}) = \sum \log \int \text{Binomial}(n_{i+1}; n_{i++}, \pi_i) beta(\pi_i; \boldsymbol{\theta}_0) d\pi_i$$

$$= \sum \log P_{BB}(n_{i+1}; n_{i++}, \boldsymbol{\theta}_0)$$

and similarly

$$log\, L_1(\boldsymbol{\theta}_1) = \sum_{i=1}^{m} \log P_{BB}(n_{i11}; n_{i+1}, \boldsymbol{\theta}_1) \; and \; \log L_2(\boldsymbol{\theta}_2) = \sum_{i=1}^{m} \log P_{BB}(n_{i00}; n_{i+0}, \boldsymbol{\theta}_2).$$

Because each component of the composite likelihood $\log L_c(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is a true marginal likelihood the corresponding score equation is unbiased, and the maximum composite likelihood estimator $(\widetilde{\boldsymbol{\theta}_0}, \widetilde{\boldsymbol{\theta}_1}, \widetilde{\boldsymbol{\theta}_2})$ can be shown to be consistent and approximately normally distributed with a mean of $(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and covariance matrix of Σ/m, where

$$\sum = \left(\pi^{-1} I_{00}^{-1} \pi^{-1/2} I_{00}^{-1} I_{01} I_{11}^{-1} \pi^{-1/2} I_{11}^{-1} I_{00}^{-1} I_{02} I_{22}^{-1} I_{11}^{-1} I_{12} I_{22}^{-1} I_{22}^{-1}\right),$$

and $m_2/m \to \pi > 0$ *as* $m \to \infty$,

$$I_{00} = E\left\{-\frac{1}{m_2}\frac{\partial^2 \log L_0(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0^2}\right\}, I_{12} = E\left[\frac{1}{m}\left\{\frac{\partial \log L_1(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1}\right\}\left\{\frac{\partial \log L_2(\boldsymbol{\theta}_2}{\partial \boldsymbol{\theta}_2}\right\}^T\right]$$

$$I_{jj} = E\left\{-\frac{1}{m}\frac{\partial^2 \log L_j(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j^2}\right\}, \text{ and } I_{0j} = E\left[\frac{1}{m_2}\left\{\frac{\partial \log L_0(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}\right\}\left\{\frac{\partial \log L_j(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}\right\}^T\right],$$

for $j = 1, 2$. The matrix Σ can be consistently estimated by its empirical counterpart. A detailed proof is provided in Appendix B of the supporting information.

This composite-likelihood method (hereafter referred to as HBB-CL method) circumvents the limitation of restrictive admissible correlations in the HBB-FL method. The maximum composite-likelihood estimator $_j$ $(j = 0, 1, 2)$ can be easily obtained by fitting a univariate beta-binomial model, and the $\tilde{}$ covariance matrix of $(\widetilde{\boldsymbol{\theta}_0}, \widetilde{\boldsymbol{\theta}_1}, \widetilde{\boldsymbol{\theta}_2})$ can be calculated by the preceding formulas. Notice that the off-diagonal matrices in Σ/m properly account for the correlations among the estimated overall prevalence, sensitivity, and specificity and their heterogeneities. This is an important advantage of our proposed HBB-CL method: we reduce the multivariate meta-analysis to separate much simpler univariate meta-analysis

while being able to properly account for the correlations among estimated summary measures. As a by-product, the constant correlation structure assumption is not needed for the HBB-CL method. Hence, it offers additional robustness to such a model assumption. Working code of this method, along with a working example, have been included in our R package 'xmeta', at https://cran.r-project.org/package=xmeta.

3 Simulation study

3.1 Simulation settings

In this section, we conduct a set of simulation studies to assess the performance of the proposed HBB-CL method and to investigate its robustness under misspecifications of the random-effects distributions, correlation structures, and link functions. For comparison purposes, we also consider the full-likelihood methods based on the hybrid TGLMM with logit-normal assumption of random-effects distributions (hereafter referred to as HGLMM-FL method) and the HBB-FL method, respectively. The detail for the HGLMM-FL method is provided in Appendix C of the supporting information. We consider five sce-narios described in Table I. In the first three scenarios, there is no study-level covariate, whereas there are study-level covariates in the remaining two scenarios. In scenario $\mathcal{M}_1$, the random-effects are generated from a multivariate beta distribution with Clayton copula using the *Copula* package in R [33, 34], and the data (n_{i11}, n_{i10}, n_{i01}, n_{i00}) are generated from binomial and multinomial distributions for case– control and cohort studies, respectively. The numbers n_{i+1}, n_{i+0}, and n_{i++} are randomly sampled from the motivating ovarian cancer data. For each scenario, we consider settings with the number of studies, m, being 30 or 50. Each of them has equal numbers of case–control studies and cohort studies. We specify the parameters for overall disease prevalence and diagnostic test sensitivity and specificity as (μ_0, μ_1, μ_2) = (0.55, 0.88, 0.92) and the between-study heterogeneity parameters as (δ_0, δ_1, δ_2) = (0.1, 0.1, 0.1). Here, the overall disease prevalence is obtained from the study population in our motivating example. To investigate the impacts of the correlation structure and non-homogeneous correlations, we set the correlation parameters (ρ_{01}, ρ_{02}, ρ_{12}) at (0.2, −0.2, −0.6), (0.6, −0.6, −0.6), a mixture of (0, 0, −0.6) and (0.2, −0.2, −0.6), and a mixture of (0, 0, −0.6) and (0.6, −0.6, −0.6). For scenario $\mathcal{M}_1$, the assumptions of the proposed HBB-CL and HBB-FL methods are not violated, whereas the assumptions of the HGLMM-FL method are violated. For scenario $\mathcal{M}_2$, random effects are generated from a multivariate logit-normal where both homogenous correlations and non-homogenous correlations are also considered. The assumptions of the HGLMM-FL method are satisfied for homogeneous correlation settings but not satisfied for non-homogeneous correlation settings, whereas the assumptions of the HBB-CL method are not satisfied for both correlation settings. For scenario $\mathcal{M}_3$, we consider the impacts of misspecified link functions where the link function is a complementary log–log for the study-specific prevalence and is logit for the study-specific sensitivity and specificity. In the last two scenarios (i.e., $\mathcal{M}_4$ and $\mathcal{M}_5$), two study-level covariates are incorporated with a binary covariate generated from a Bernoulli distribution with a probability of 0.5 and a continuous covariate generated from a uniform distribution between −1 and 1. The random effects are generated from multivariate beta distributions in scenario $\mathcal{M}_4$ and from multivariate logit-normal distributions in scenario $\mathcal{M}_5$.

For each simulation setting, we generate 1000 datasets. The samples are simulated in R (R Development Core Team, version 3.0.1). The HBB-CL method is implemented in R by using the 'aod' packageBoth [35] Both HGLMM-FL and HBB-FL methods are fitted by the package NLMIXED of SAS (SAS Institute Inc., Cary, NC) where Gaussian quadratures are used to approximate the likelihood function with integrals. Programming codes are provided in Appendix D of the supporting information.

3.2 Simulation results

Each simulated dataset is analyzed by both HBB-CL and HGLMM-FL methods (or the HBB-FL method). Bias (BIAS), empirical standard errors (SE), and coverage probability (CP) of the CIs based on 1000 simulations for the overall disease prevalence, sensitivity, specificity, PPV, and NPV are provided in Table II. Because of limited space, the estimated model-based standard errors are presented in Appendix E of the supporting information where we provide additional simulation results. We note that PPV and NPV are functions of prevalence, sensitivity, and specificity, and hence, their standard errors can be obtained by the delta method. The upper panel of Table II summarizes the results under scenario $\mathcal{M}_1$. As expected, the biases of the HBB-CL method are small, and the range of coverages is 89.7–93.8% under all settings of correlation structures. In addition, we observe that the estimates based on the HBB-FL method are slightly biased, with low to moderate coverage probabilities (the range of coverages is 63.8–90.2%) due to the constraints on correlations. In contrast, for the HGLMM-FL method, all estimates except prevalence estimates are biased and the coverages are poor (range of coverage is 18.2–59.8% for estimates other than prevalence). This suggests that the HGLMM-FL method is very sensitive to the misspecification of random-effects distributions. The non-homogeneous correlation structure seems to have low impact on the coverage. This may be due to the fact that the coverage of the HGLMM-FL method is already quite low.

The middle panel of Table II presents the results under scenario $\mathcal{M}_2$. When the HGLMM-FL method is used, the estimates have small bias, even when the correlations are non-homogeneous, and the range of coverage is 80.2–84.9% under homogeneous correlation settings. However, its coverage quickly deterio-rates to 55.6–65.0% under non-homogeneous settings. When the HBB-CL method is used, the estimates have larger biases than those from the HGLMM-FL method, but the coverage is consistently in the range of 69.7–92.5% for all settings of correlations. The non-homogeneous correlations have little impact on the coverage. We note that the coverages of the HBB-CL method are comparable with the coverage of the HGLMM-FL method when the correlations are high and are better when the correlations are non-homogeneous. The bottom panel of Table II presents the results under scenario $\mathcal{M}_3$ where part of the link functions are misspecified. The HBB-CL method is consistently better than the HGLMM-FL method in coverage under all correlation settings. We note that the estimates of the HBB-FL method under both scenarios $\mathcal{M}_2$ and $\mathcal{M}_3$ have severe biases and low coverages, and the results are summarized in Table S11 of the supporting information. A possible reason is that the estimates based on the HBB-FL method are quite sensitive to model misspecifications. In addition to the results with moderate study sizes, we also observe similar findings on the relative performances of these

methods when the number of studies is large ($m = 50$). The results are summarized in Table S10 of the supporting information.

When study-level covariates are considered, the advantage of the HBB-CL method in coverage is more evident. Figure 2 presents the results of estimated coefficients in meta-regression under scenario $\mathcal{M}_4$. The HGLMM-FL method has sizable biases, and its coverage is below 40%. In contrast, the HBB-CL method has small biases and satisfactory coverage under all correlation settings. Figure 3 presents the results under scenario $\mathcal{M}_5$. Although the HGLMM-FL method has small bias, its coverage is still below 55%. Similar as that in scenario $\mathcal{M}_2$, the HBB-CL method has sizable biases but better coverages (80% or higher) under all correlation settings. The detailed results are summarized in Tables S2–S9 of the supporting information.

In summary, the HBB-CL method performs well with small bias and good coverage when the marginal distributions are beta-binomial. Its performance is robust to the strength of correlations or non-homogeneity of correlations across studies. In addition, the HBB-CL method is found to be more robust than the HGLMM-FL method to model misspecifications under both meta-analysis and meta-regression settings. Finally, the non-convergence rate of the HGLMM-FL method can be up to 56.3%, whereas the non-convergence rate of the HBB-CL method is less than 1% under all settings considered. Considering all these advantages, the HBB-CL method can be a useful and simple alternative to the existing methods.

4 Applications

We apply the proposed HBB-CL method and the HGLMM-FL method to the motivating ovarian cancer data. The data contain 52 studies that were reported between January 1995 and November 2007 [24], with 40 case–control and 12 cohort studies. The modalities for diagnosing recurrent ovarian carcinoma include CA125, PET, PET–CT, CT, and MRI. The data for each modality are provided in Table S1 of the supporting information. We note that inference on disease prevalence requires at least two cohort studies because the model for disease prevalence contains two parameters, namely, overall disease prevalence and heterogeneity. Thus, the subgroup MRI is excluded from the analysis because only one cohort study is available.

Before applying the HGLMM-FL method, we evaluate the normality of study-specific disease prevalence, sensitivity, and specificity after a logit transformation. The Q–Q plots in Figure 4 suggest the lack of normality for study-specific disease prevalence. A formal test of normality using Shapiro–Wilk test is conducted with a p-value of 0.008 for study-specific disease prevalence, 0.56 for sensitivity, and 0.13 for specificity after logit transformation. In this case, the HGLMM-CL method with logit link should not be used. To further compare the results from the HGLMM-FL method and the proposed HBB-CL method, we apply both methods to each diagnostic modality. Table III summarizes the results from both methods where a variety of initial values are used. We find that applying the HGLMM-FL method encounters the non-convergence and singular covariance problems. Furthermore, the estimates of overall prevalence, sensitivity, and specificity are quite sensitive to the choice of

initial values. This finding is consistent with the findings in simulation studies. In contrast, the HBB-CL method is very stable and not sensitive to the choice of initial values.

Figure 5 presents the forest plots for estimated overall disease prevalence, sensitivity, specificity, PPV, and NPV from the HBB-CL method for the four diagnostic modalities. We find that among all modalities, patients diagnosed by PET-CT have the highest disease prevalence (73%; 95%CI = 57% to 85%), and PET–CT has the highest estimated sensitivity (91%; 95%CI = 85% to 95%). In addition, PET–CT has the highest estimated PPV (95%; 95%CI = 86% to 99%) and NPV (79%; 95%CI = 58% to 91%) among all four tests. In practice, high NPV is necessary for a diagnostic test to be useful at ruling out diseases, and high PPV is necessary for a diagnostic test to be useful at confirming diseases. Among all modalities, CA125 has the highest estimated specificity (93%; 95%CI = 89% to 95%).

For comparison purpose, we present the results from Gu *and others* [24] as dash lines in Figure 5, where the regression model [4] was used to obtain estimates of sensitivity and specificity. From their analyses, they found that among four diagnostic modalities, PET-CT has the highest sensitivity (91%; 95%CI = 88% to 94%) and CA125 has the highest specificity (93%; 95%CI = 89% to 95%). Such a conclusion is consistent with the results from the proposed HBB-CL method. Although the point estimates of sensitivities and specificities from both methods are consistent, it is interesting to note that the CIs of sensitivities and specificities are considerably narrower than CIs from the HBB-CL method. This may be due to the ignored uncertainty in regression covariates, which has been noticed by Rutter and Gatsonis [5], Walter [17], and Arends *and others* [14]. We note that the regression model-based method [4] cannot provide estimates of disease prevalence, PPV, or NPV because the cohort studies have been treated the same as the case–control studies and information on disease prevalence has not been utilized. In addition, we also present the results based on the BGLMM method as dotted lines in Figure 5, where it incorporates the correlation that exists between sensitivity and specificity. Compared with the HBB-CL method, the BGLMM method produces similar summary estimates of PET–CT sensitivity and CA125 specificity but with wider CIs (93%; 95%CI = 89% to 97% and 94%; 95%CI = 91% to 98%, respectively).

Figure 6 presents the summary points and 95% joint confidence regions for four diagnostic modalities based on our composite-likelihood method. The parametric representation of the boundary of the elliptical Wald-type confidence region for sensitivity and specificity in logit scales is given by [36]

$$S_{\mathbf{1}} = \widehat{S_1} + s_{S_1}\sqrt{(2f_{2,n-2;\alpha})}\cos\phi \text{ and } C_1 = \hat{C}_1 + s_{C_1}\sqrt{(2f_{2,n-2;\alpha})}\cos(\phi + \arccos r),$$

where s_{S_1} and s_{C_1} are the estimated standard errors of $\hat{S}_1$ and $\hat{C}_1$, r is the estimate of their correlation, ϕ runs from 0 to 2π, and $f_{2,n-2;\alpha}$ is the upper 100α% point of the F distribution with degrees of freedom 2 and $n-2$, and n is the number of studies. For joint confidence regions of pairs of measures, such as PPV and NPV, the delta method can be applied to obtain the covariance matrix. As shown in the top left panel of Figure 6, sensitivity and specificity tend to have similar variations. Further more,there is more variation in NPV than

in PPV as displayed in the top right because the number of true positives is much less than the number of negatives. In summary, given the wide range of the confidence regions, more studies are needed to increase the precision of these inferences and to reach more definitive conclusions on the comparisons. Figure 7 presents the estimated PPV and NPV with respect to a given prevalence value, with their pointwise 95% CIs. These plots for the whole range of clinically likely prevalences can provide the usefulness of the tests at different incidences of the condition. For example, we find that CA125 has estimated overall PPV (95%; 95%CI = 92% to 97%) and NPV (60%; 95%CI= 54% to 66%) when the estimated prevalence is given as 63%. This may suggest that CA125 could provide a useful screening test for the patient with recurrent ovarian carcinoma, with only a 5% change of falsely identifying an unaffected patient. By contrast, the estimated prevalence for patients diagnosed by PET–CT is 73%, and the estimated overall PPV and NPV are 95% (95%CI = 91% to 98%) and 80% (95%CI = 69% to 88%), respectively. It is immediately clear that PET–CT may be more useful than CA125 in ruling out clinically well patients.

5. Discussion

In this paper, we proposed a simple and robust method for multivariate meta-analysis of diagnostic test accuracy. This method can be applied to the general setting of data where a mixture of case–control and cohort studies are collected. This method fully utilizes the prevalence information from the cohort studies and can provide inference for prevalence-dependent summary measures such as PPV and NPV. If we are interested in PPV and NPV alone, they can be obtained from fitting the TGLMM model [19] using only the cohort studies. However, in such a case, the information on sensitivity and specificity contained in the case–control studies is not incorporated, which results in loss of efficiency, and the amount of efficiency loss depends on the number of the case–control studies relative to the number of the cohort studies. Perhaps most importantly, the use of composite likelihood, built upon a novel factorization of a multinomial distribution into the product of three binomial distributions, avoids the limitation of restrictive admissible correlations allowed by Sarmanov beta-binomial distributions and enables meta-analyses that perform three separate univariate meta-analyses to make valid joint inferences and also valid inferences for any functions of the summary measures. Through simulation studies, we find that the proposed method is robust to the homogeneous correlation structure assumption and is more robust to model misspecifications than the commonly used TGLMMs. As meta-analyses conventionally favor simple and robust procedures, the proposed method is expected to be widely applicable to practical studies. By using univariate methods for meta-analysis to make marginal inferences for the outcomes, our procedure does not make an attempt to allow any borrowing of strength. Borrowing of strength refers to the potential for multivariate meta-analyses to provide more precise point estimates than multiple univariate meta-analyses of the same outcome data [1]. The borrowing of strength afforded by multivariate meta-analysis has in any case often been found to be small [37, 38]. In summary, the advantage of the proposed HBB-CL method is that sensitivity and specificity can always be estimated, even when computational problems arise or the model is misspecified. In addition, the proposed HBB-CL method can provide the estimates of PPV and NPV via borrowing information from cohort studies. However, there are several intrinsic

limitations of the proposed composite-likelihood-based method: first, because the correlations are not explicitly modeled, the proposed method cannot provide the receiver operation characteristic (ROC) curve; second, by combining case–control studies with cohort studies together, bias may be introduced because the case–control studies may be likely to overestimate diagnostic test accuracy. Existing practice includes exclusion of the case–control studies, which leads to loss of information. Methods that account for such potential bias without discarding the case–control studies are of interest and are currently under investigation.

In addition to the aforementioned limitations of the proposed method, there are several directions for future extensions. It has been acknowledged that case–control studies tend to overestimate the diagnostic accuracy of a test, particularly the studies with the inclusion of non-representative patients or the use of different reference standards [39, 40]. For some diagnostic studies, the gold standard test is given to only a selected subset of patients, the selection of which depends on the results of the test under evaluation. Ignoring this underlying outcome-dependent process can lead to a partial verification bias in the estimation of disease prevalence and diagnostic sensitivity and specificity [8, 41]. Furthermore, the gold standard test in the diagnostic studies may not be available because of measurement errors or other factors. The related work has been discussed in the literature [42–45]. Another important issue is to account for publication bias [46] where publication or non-publication of a diagnostic test depends on the estimated accuracy of this test [47]. This is particularly challenging when the studies are heterogeneous [48]. Methods accounting for these biases are under investigation and will be reported in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Yong Chen was supported in part by grant number R03HS022900 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the ‘Agency for Healthcare Research and Quality. Haitao Chu was supported in part by the US NIAID AI103012, AHRQ R03HS020666, NCI 1P01CA142538, and NCI 2P30CA077598. We want to thank Dan Jackson for helpful discussion, which has improved the presentation of this work.

References

1. Jackson D, Riley R, White I. Multivariate meta-analysis: potential and promise. Statistics in Medicine 2011; 30(20): 2481–2498. [PubMed: 21268052]
2. Zhou X, Obuchowski N, McClish D. Statistical Methods in Diagnostic Medicine John Wiley & Sons: New York, 2002.
3. Reitsma J, Glas A, Rutjes A, Scholten R, Bossuyt P, Zwinderman A. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of Clinical Epidemiology 2005; 58(10):982–990. [PubMed: 16168343]
4. Moses L, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Statistics in Medicine 1993; 12(14):1293–1316. [PubMed: 8210827]
5. Rutter C, Gatsonis C. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Statistics in Medicine 2001; 20(19):2865–2884. [PubMed: 11568945]

6. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. Journal of Clinical Epidemiology 1995; 48(1):119–130. [PubMed: 7853038]
7. Leeflang M, Deeks J, Gatsonis C, Bossuyt P, Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Annals of Internal Medicine 2008; 149(12):889–897. [PubMed: 19075208]
8. Ransohoff D, Feinstein A. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. New England Journal of Medicine 1978; 299(17):926–930. [PubMed: 692598]
9. Levy D, Labib S, Anderson K, Christiansen J, Kannel W, Castelli W. Determinants of sensitivity and specificity of electrocardiographic criteria for left ventricular hypertrophy. Circulation 1990; 81(3): 815–820. [PubMed: 2137733]
10. Mulherin S, Miller W. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Annals of Internal Medicine 2002; 137(7):598–602. [PubMed: 12353947]
11. Van Houwelingen H, Zwinderman K, Stijnen T. A bivariate approach to meta-analysis. Statistics in Medicine 1993; 12(24):2273–2284. [PubMed: 7907813]
12. Van Houwelingen H, Arends L, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in Medicine 2002; 21(4):589–624. [PubMed: 11836738]
13. Chu H, Cole S, Reitsma J, Zwinderman A. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. Journal of Clinical Epidemiology 2006; 59(12):1331–1333. [PubMed: 17098577]
14. Arends L, Hamza T, van Houwelingen J, Heijenbrok-Kal M, Hunink M, Stijnen T. Bivariate random effects meta-analysis of ROC curves. Medical Decision Making: An International Journal of the Society for Medical Decision Making 2007; 28(5):621–638.
15. Hamza T, Reitsma J, Stijnen T. Meta-analysis of diagnostic studies: a comparison of random intercept, normal–normal, and binomial–normal bivariate summary ROC approaches. Medical Decision Making 2008; 28(5):639–649. [PubMed: 18753684]
16. Chu H, Guo H, Zhou Y. Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. Medical Decision Making 2010; 30(4):499–508. [PubMed: 19959794]
17. Walter S Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. Statistics in Medicine 2002; 21(9):1237–1256. [PubMed: 12111876]
18. Harbord R, Deeks J, Egger M, Whiting P, Sterne J. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 2007; 8(2):239–251. [PubMed: 16698768]
19. Chu H, Nie L, Cole S, Poole C. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: alternative parameterizations and model selection. Statistics in Medicine 2009; 28(18):2384–2399. [PubMed: 19499551]
20. Ma X, Nie L, Cole SR, Chu H. Statistical methods for multivariate meta-analysis of diagnostic tests: an overview and tutorial. Statistical Methods in Medical Research 2016; 25(4):1596–1619. [PubMed: 23804970]
21. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA: A Cancer Journal for Clinicians 2014; 64(1):9–29. [PubMed: 24399786]
22. Megibow A, Bosniak M, Ho A, Beller U, Hulnick D, Beckman E. Accuracy of CT in detection of persistent or recurrent ovarian carcinoma: correlation with second-look laparotomy. Radiology 1988; 166(2):341–345. [PubMed: 3336709]
23. Rubin S, Randall T, Armstrong K, Chi D, Hoskins W. Ten-year follow-up of ovarian cancer patients after second-look laparotomy with negative findings. Obstetrics & Gynecology 1999; 93(1):21–24. [PubMed: 9916949]
24. Gu P, Pan L, Wu S, Sun L, Huang G. CA 125, PET alone, PET–CT, CT and MRI in diagnosing recurrent ovarian carcinoma: a systematic review and meta-analysis. European Journal of Radiology 2009; 71(1):164–174. [PubMed: 18378417]
25. Sarmanov OV. Generalized normal correlation and two-dimensional Fréchet classes. Soviet Mathematics Doklady 1966; 7:596–599.
26. Lee MLT. Properties and applications of the Sarmanov family of bivariate distributions. Communications in Statistics-Theory and Methods 1996; 25(6):1207–1222.

27. Shubina M, Lee MLT. On maximum attainable correlation and other measures of dependence for the Sarmanov family of bivariate distributions. Communications in Statistics—Theory and Methods 2004; 33(5):1031–1052.
28. Lindsay B Composite likelihood methods. Contemporary Mathematics 1988; 80(1):221–239.
29. Hamza TH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. Journal of Clinical Epidemiology 2008; 61(1):41–51. [PubMed: 18083461]
30. Chen Y, Chu H, Luo S, Nie L, Chen S. Bayesian analysis on meta-analysis of case–control studies accounting for within-study correlation. Statistical Methods in Medical Research 2015; 24(6):836–855. [PubMed: 22143403]
31. Kuss O, Hoyer A, Solms A. Meta-analysis for diagnostic accuracy studies: a new statistical model using beta-binomial distributions and bivariate copulas. Statistics in Medicine 2014; 33(1):17–30. [PubMed: 23873593]
32. Varin C, Reid N, Firth D. An overview of composite likelihood methods. Statistica Sinica 2011; 21(1):5–42.
33. Yan J Enjoy the joy of copulas: with a package copula. Journal of Statistical Software 2007; 21(4): 1–21.
34. Kojadinovic I, Yan J. Modeling multivariate distributions with continuous margins using the copula R package. Journal of Statistical Software 2010; 34(9):1–20.
35. Lesnoff M, Lancelot R. AOD: analysis of overdispersed data. R package version 13–1, 2012.
36. Douglas J Confidence regions for parameter pairs. The American Statistician 1993; 47(1):43–45.
37. Sohn S Multivariate meta analysis with potentially correlated marketing study results. Naval Research Logistics (NRL) 2000; 47(6):500–510.
38. Trikalinos TA, Hoaglin DC, Schmid C. An empirical comparison of univariate and multivariate meta-analyses for categorical outcomes. Statistics in Medicine 2014; 33(9):1441–1459. [PubMed: 24285290]
39. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282(11): 1061–1066. [PubMed: 10493205]
40. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Annals of Internal Medicine 2004; 140(3): 189–202. [PubMed: 14757617]
41. Rutjes A, Reitsma J, Di Nisio M, Smidt N, Van Rijn J, Bossuyt P. Evidence of bias and variation in diagnostic accuracy studies. Canadian Medical Association Journal 2006; 174(4):469–476. [PubMed: 16477057]
42. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. American Journal of Epidemiology 1995; 141(3):263–272. [PubMed: 7840100]
43. Chu H, Chen S, Louis TA. Random effects models in a meta-analysis of the accuracy of two diagnostic tests without a gold standard. Journal of the American Statistical Association 2009; 104(486):512–523. [PubMed: 19562044]
44. Dendukuri N, Schiller I, Joseph L, Pai M. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. Biometrics 2012; 68(4):1285–1293. [PubMed: 22568612]
45. Liu Y, Chen Y, Chu H. A unification of models for meta-analysis of diagnostic accuracy studies without a gold standard. Biometrics 2014; 71(2):538–547. [PubMed: 25358907]
46. Rothstein H, Sutton A, Borenstein M. Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments John Wiley & Sons: Chichester, England, 2006.
47. Bürkner PC, Doebler P. Testing for publication bias in diagnostic meta-analysis: a simulation study. Statistics in Medicine 2014; 33(18):3061–3077. [PubMed: 24753050]
48. Terrin N, Schmid C, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. Statistics in Medicine 2003; 22(13):2113–2126. [PubMed: 12820277]

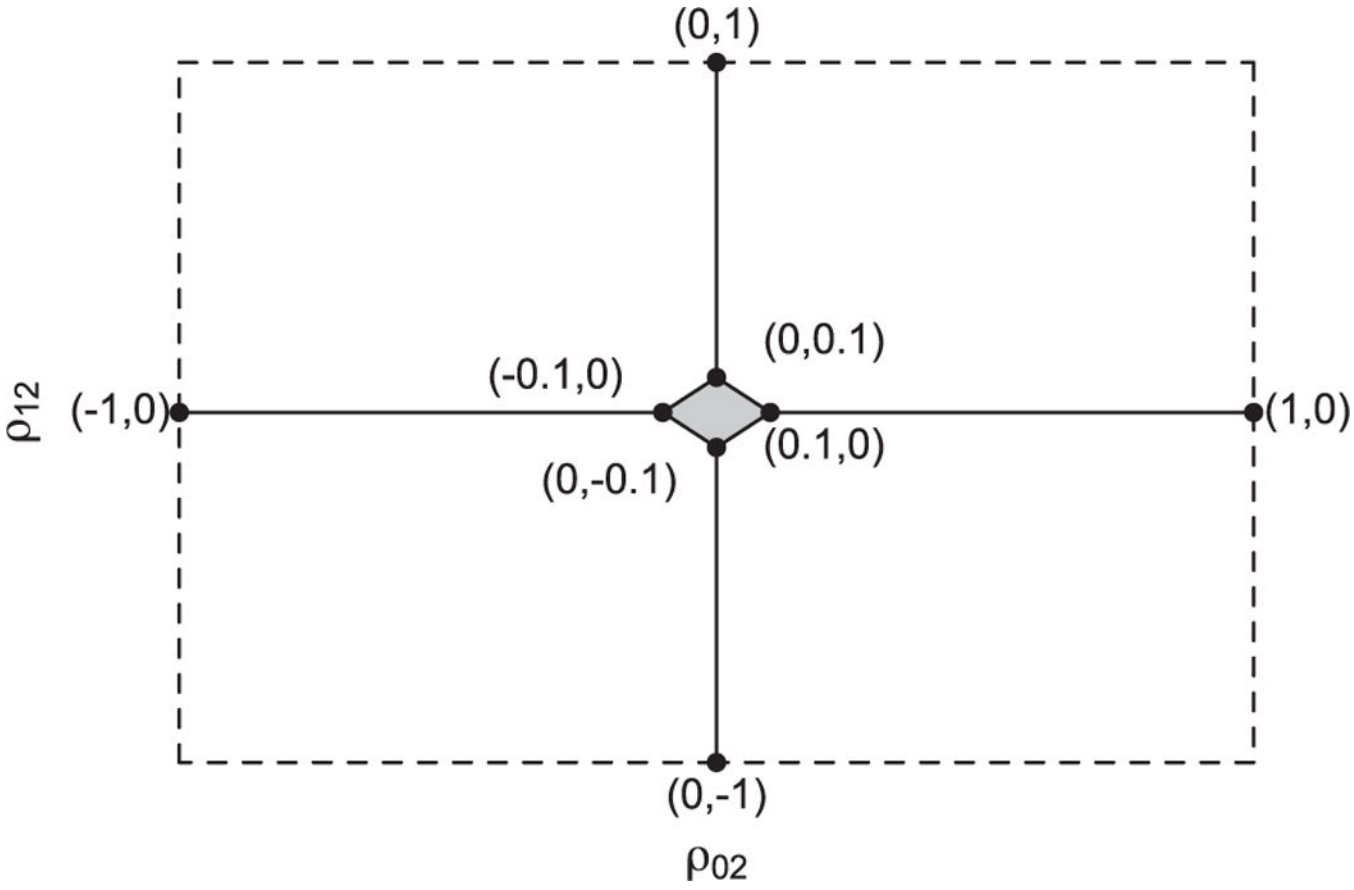

Figure 1.

The region of admissible values of (ρ_{02}, ρ_{12}) given the correlation $\rho_{01} = 0$. The parameters $\mu_0 = \mu_1 = \mu_2 = 0.5$ and $\delta_0 = \delta_1 = \delta_2 = 0.16$. Derivation of the admissible region defined by a system of linear inequalities is provided in Appendix A of the supporting information.

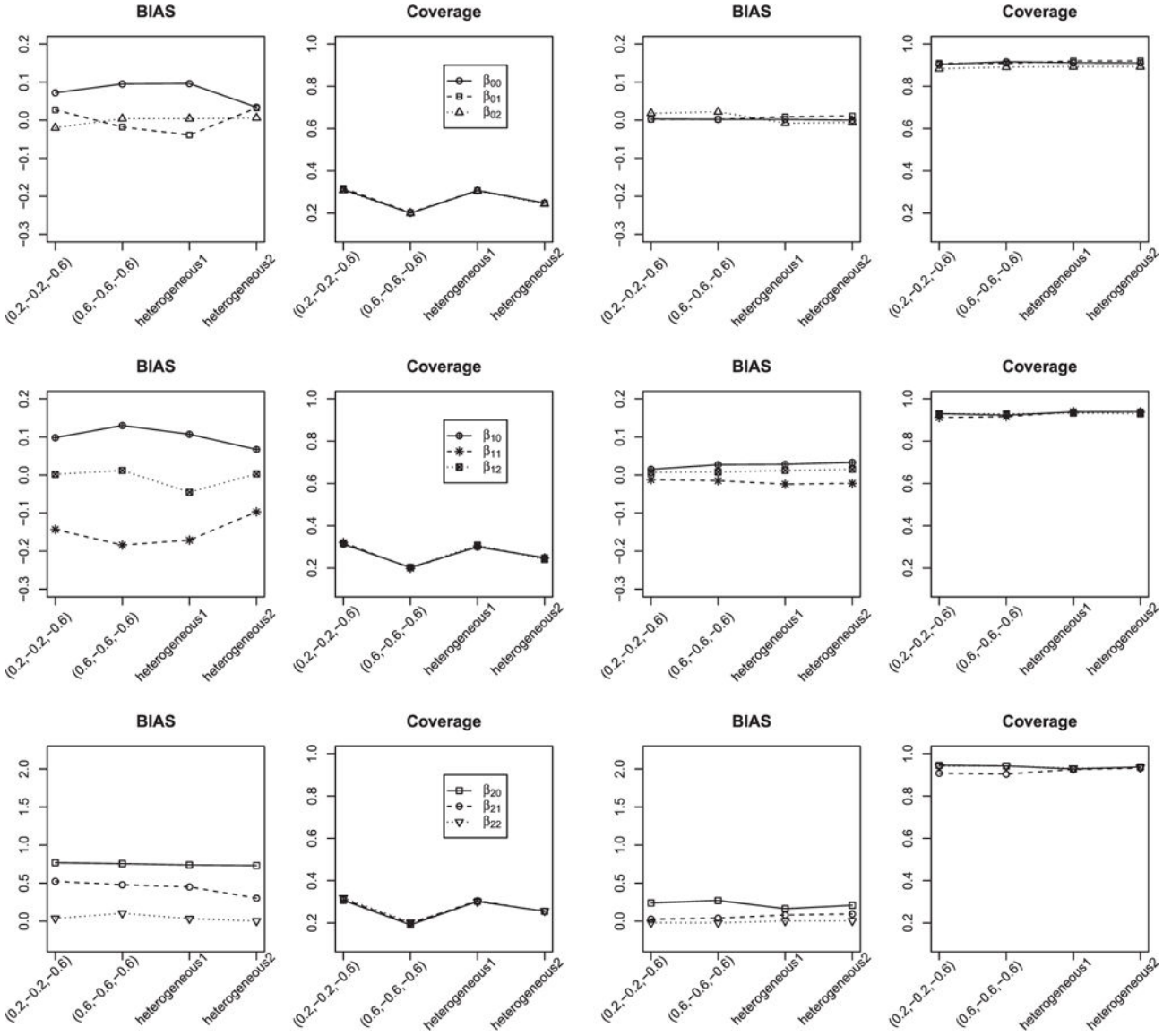

Figure 2.

Bias and coverage for estimated meta-regression parameters from the full-likelihood hybrid trivariate generalized linear mixed-effects model (the two leftmost columns) and composite-likelihood hybrid beta-binomial methods (the two rightmost columns). Each row represents the results for each set of meta-regression parameters (row 1: regression parameters for disease prevalence; row 2: regression parameters for sensitivity; row 3: regression parameters for specificity). The true values of regression parameters for disease prevalence are (β_{00}, β_{01}, β_{02}) = (0.596, 0.000, 0.000), those of regression parameters for sensitivity are(β_{10}, β_{11}, β_{12}) = (0.691, −1.000, 0.000), and those of regression parameters for specificity are (β_{20}, β_{21}, β_{22}) = (2.472, 1.000, 0.000). The data are generated from bivariate beta-binomial distribution for case–control studies and trivariate beta-binomial distribution for cohort studies (i.e., scenario $\mathcal{M}_4$). Results are summarized from 1000 simulations. The *x*-axis represents the different settings of pairwise correlations among study-specific prevalence, sensitivity, and specificity.

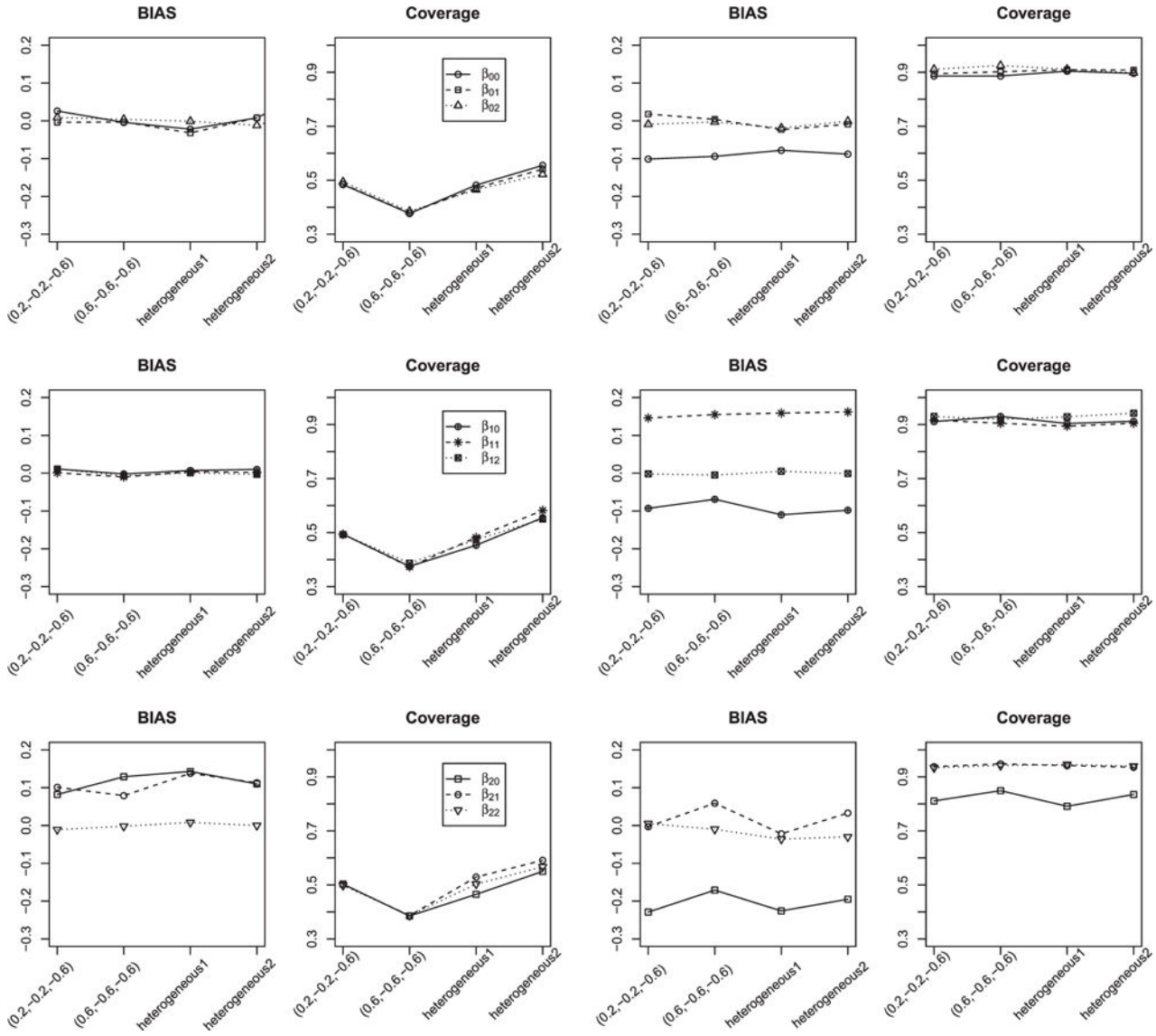

Figure 3.

Bias and coverage for estimated meta-regression parameters from the full-likelihood hybrid trivariate generalized linear mixed-effects model (GLMM) (the two leftmost columns) and the composite-likelihood hybrid beta-binomial methods (the two rightmost columns). Each row represents the results for each set of meta-regression parameters (row 1: regression parameters for disease prevalence; row 2: regression parameters for sensitivity; row 3: regression parameters for specificity). The true values of regression parameters for disease prevalence are (β_{00}, β_{01}, β_{02}) = (0.596, 0.000, 0.000), those of regression parameters for sensitivity are (β_{10}, β_{11}, β_{12}) = (0.691, −1.000, 0.000), and those for regression parameters for specificity are (β_{20}, β_{21}, β_{22}) = (2.472, 1.000, 0.000). The data are generated from bivariate GLMM for case–control studies and trivariate GLMM for cohort studies (i.e., scenario $\mathcal{M}_5$). Results are summarized from 1000 simulations. The *x*-axis represents the different settings of pairwise correlations among study-specific prevalence, sensitivity, and specificity.

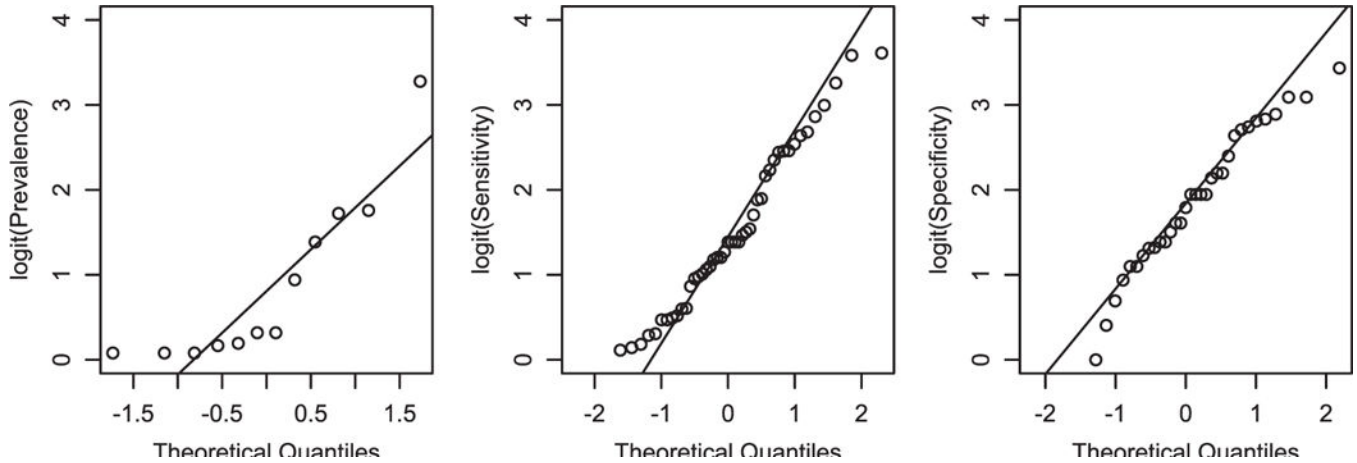

Figure 4.

Normal Q–Q plots with respect to logit prevalence (left panel), logit sensitivity (middle panel), and logit specificity (right panel).

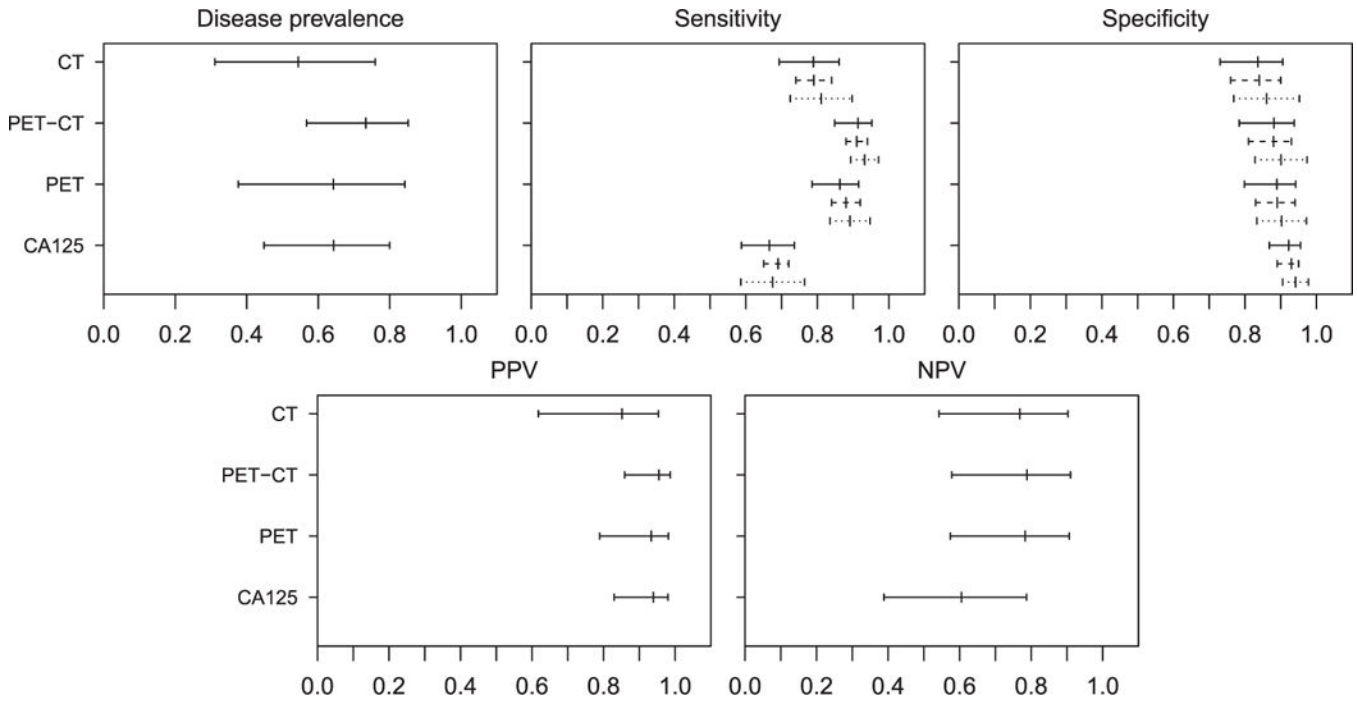

Figure 5.

Upper panels: the estimated disease prevalence, sensitivities, and specificity and 95% CIs of four diag-nostic imaging modalities using the composite-likelihood method. Lower panels: the estimated positive predictive values (PPVs) and negative predictive values (NPVs) and their 95% CIs. Solid segments are the results from the composite-likelihood hybrid beta-binomial method. Dashed segments are the results from Gu *and others* [24]. Dotted segments are the results from the bivariate generalized linear mixed-effects model method. CA125, cancer antigen 125; PET, positron emission tomography; CT, computed tomography.

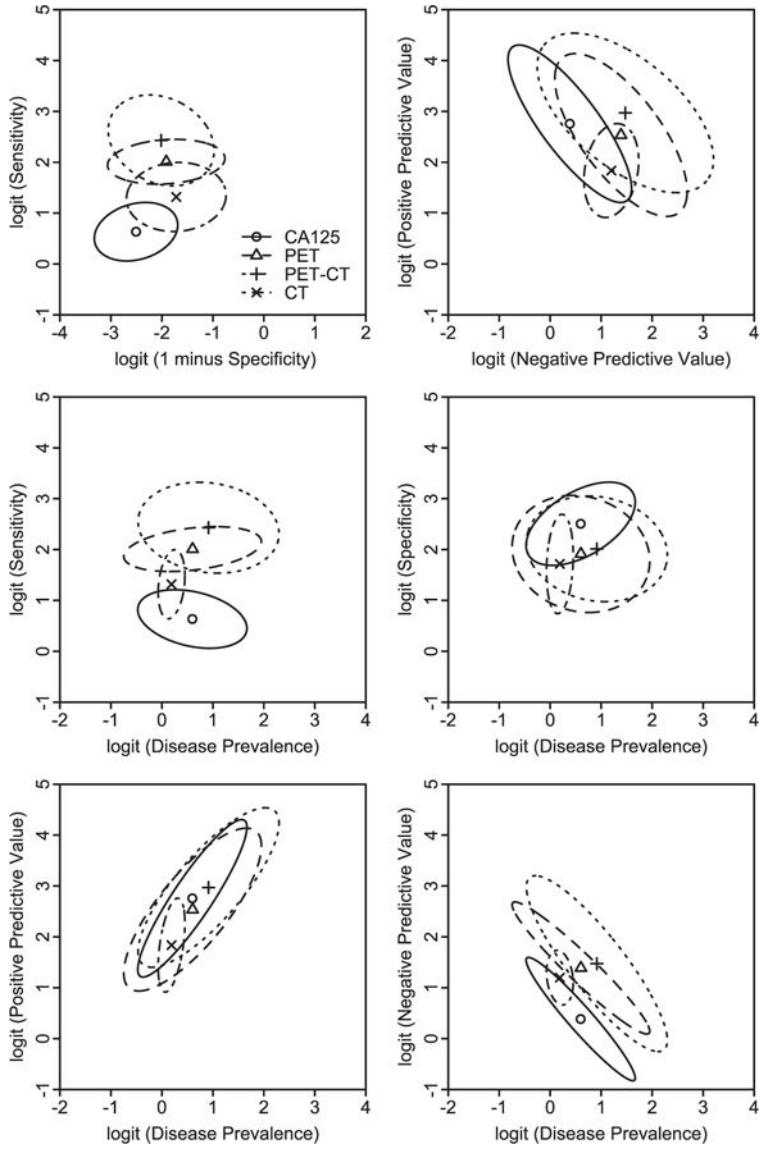

Figure 6.

Summary points and 95% confidence regions of logit(sensitivity) versus logit(1 minus specificity) (top left panel), positive predictive value versus negative predictive value (top right panel), logit(sensitivity) and logit(specificity) versus logit(prevalence) (middle panels), logit(predictive values) versus logit(prevalence) (bot-tom panels) for four diagnostic imaging modalities. Filled circle: summary point. Solid line: boundary of 95% confidence region for the summary point. CA125, cancer antigen 125; PET, positron emission tomography; CT, computed tomography.

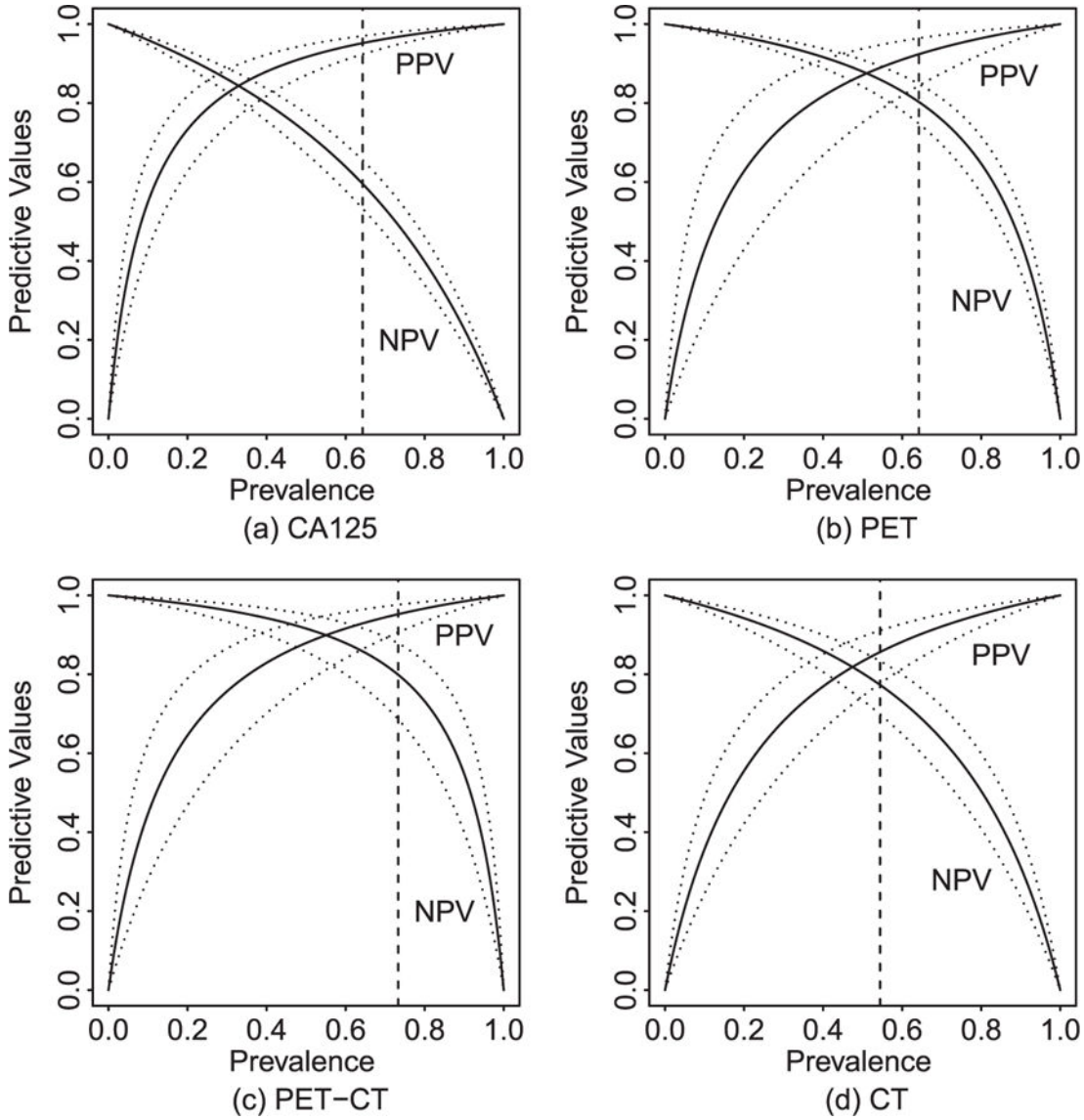

Figure 7.

The overall PPV and NPV plots based on the bivariate random-effects model. The solid and dashed lines denote the estimate and 95% CI, respectively. PPV, positive predictive value; NPV, negative predictive value; CA125, cancer antigen 125; PET, positron emission tomography; CT, computed tomography.

Table I.

Configurations of five different simulation scenarios: $\mathscr{M}_1$–$\mathscr{M}_5$.

Model	Covariates	Link function	Random-effects distribution	Correlation structure
$\mathscr{M}_1$	No	Identity	Multivariate beta	Fixed + mixture
$\mathscr{M}_2$	No	Logit	Multivariate normal	Fixed + mixture
$\mathscr{M}_3^*$	No	Mixture of complementary log–log and logit	Multivariate normal	Fixed + mixture
$\mathscr{M}_4$	Yes	Identity	Multivariate beta	Fixed + mixture
$\mathscr{M}_5$	Yes	Logit	Multivariate normal	Fixed + mixture

In this scenario, the study-specific disease prevalence (after a complementary log–log transformation) and sensitivity and specificity (after a logit transformation) are jointly generated from a multivariate normal distribution.

Table II.

Summary of 1000 simulations with data generated from $\mathcal{M}_1$ (upper panel), $\mathcal{M}_2$ (middle panel), and $\mathcal{M}_3$ (bottom panel) with 30 studies.

$(\rho_{01}, \rho_{02}, \rho_{12})$		BIAS	SE	CP	BIAS	SE	CP	BIAS	SE	CP	BIAS	SE	CP	BIAS	SE	CP
		Prevalence = 55.0			Sensitivity = 88.0			Specificity = 92.0			PPV = 93.1			NPV = 86.2		
Data generated from $\mathcal{M}_1$																
(0.2,−0.2,−0.6)	FL	0.3	5.1	88.5	3.9	2.0	39.5	3.8	1.5	25.6	3.3	1.4	28.9	4.2	2.8	51.6
	CL	−0.2	4.7	91.5	0.1	2.3	91.5	0.1	2.1	91.3	0.0	2.0	92.6	0.2	3.0	90.9
	HBB-FL	−0.1	4.8	70.8	0.3	2.2	73.0	0.3	1.9	74.7	0.2	1.9	72.6	0.3	3.0	71.6
(0.6,−0.6,−0.6)	FL	0.4	5.1	88.4	4.3	2.0	34.5	4.3	1.5	18.2	3.7	1.3	20.2	4.7	2.6	46.5
	CL	−0.1	4.7	90.5	0.4	2.3	90.6	0.4	2.1	89.8	0.3	1.8	91.7	0.5	2.8	89.7
	HBB-FL	−0.1	4.7	63.8	0.3	2.2	65.7	0.3	2.0	67.6	0.2	1.9	66.9	0.4	2.9	66.3
Heterogeneous 1	FL	0.5	5.4	89.3	4.4	2.2	38.5	4.5	1.8	23.5	4.0	1.6	23.2	4.8	2.9	47.5
	CL	−0.1	4.6	92.0	0.1	2.3	91.8	0.1	2.1	91.6	0.0	2.0	93.8	0.1	3.0	93.6
	HBB-FL	0.0	5.0	75.6	0.3	2.4	77.6	0.2	2.2	78.8	0.2	2.1	75.2	0.3	3.2	76.7
Heterogeneous 2	FL*	0.5	5.2	85.0	3.3	2.2	54.2	3.1	1.9	48.7	2.8	1.7	47.0	3.5	2.9	59.8
	CL	0.0	4.7	91.5	0.2	2.3	91.4	0.2	2.1	91.0	0.1	2.0	93.4	0.2	2.9	92.0
	HBB-FL	0.0	5.0	89.0	0.2	2.4	89.1	0.2	2.2	89.3	0.3	2.0	89.1	0.3	3.1	90.2
Data generated from $\mathcal{M}_2$																
(0.2,−0.2,−0.6)	FL†	0.3	6.8	84.0	−0.1	2.5	85.3	−0.1	2.2	84.9	−0.2	2.5	83.2	−0.5	4.0	84.2
	CL	−0.6	5.8	91.5	−3.5	2.7	77.1	−3.0	2.5	80.7	−3.0	2.8	71.4	−3.6	3.9	79.6
(0.6,−0.6,−0.6)	FL‡	0.2	6.7	80.2	−0.1	2.5	81.3	−0.1	2.1	81.7	−0.1	2.2	80.1	−0.3	3.5	82.0
	CL	−0.9	5.6	92.5	−3.2	2.7	79.1	−2.5	2.4	85.5	−2.6	2.5	71.9	−3.0	3.6	79.0
Heterogeneous 1	FL§	−0.2	9.9	59.3	−0.3	3.9	60.9	−0.4	3.3	65.0	−0.7	3.8	61.1	−0.9	6.5	55.6
	CL	−1.1	5.5	91.8	−3.2	2.8	75.0	−2.9	2.5	81.6	−3.2	2.8	69.7	−3.3	3.9	83.9
Heterogeneous 2	FL¶	−0.5	9.9	58.0	−0.3	3.6	64.9	−0.3	3.2	61.9	−0.6	3.6	60.3	−0.6	5.9	56.3
	CL	−1.0	5.6	91.8	−3.4	2.8	77.9	−2.7	2.5	81.6	−2.9	2.7	72.0	−3.2	3.8	81.3
Data generated from $\mathcal{M}_3$																
(0.2, −0.2, −0.6)	FL**	4.1	9.6	81.6	0.2	2.6	88.6	−0.1	2.3	88.1	0.6	2.8	78.9	−3.1	6.3	89.1
	CL	1.5	6.8	91.3	−3.4	2.7	78.0	−2.9	2.5	81.9	−2.3	2.9	91.3	−4.9	4.8	85.7

$(\rho_{01}, \rho_{02}, \rho_{12})$		BIAS	SE	CP	BIAS	SE	CP	BIAS	SE	CP	BIAS	SE	CP	BIAS	SE	CP
		Prevalence = 55.0			Sensitivity = 88.0			Specificity = 92.0			PPV = 93.1			NPV = 86.2		
(0.6, −0.6, −0.6)	FL[††]	4.1	9.1	76.9	−0.2	2.6	82.0	−0.1	2.2	82.3	0.7	2.5	73.9	−2.9	5.3	83.2
	CL	1.3	6.7	92.2	−3.0	2.6	83.5	−2.2	2.4	89.4	−1.7	2.7	90.7	−4.2	4.4	85.7
Heterogeneous 1	FL[‡‡]	3.8	13.1	54.4	−0.1	3.8	63.2	−0.4	3.3	64.6	0.1	4.2	57.7	−3.7	9.2	59.2
	CL	0.8	6.5	91.4	−3.5	2.8	75.4	−2.9	2.5	82.9	−2.5	2.9	91.7	−4.5	4.6	88.9
Heterogeneous 2	FL[§§]	2.8	12.8	60.3	−0.2	3.9	64.1	−0.4	3.4	65.0	−0.1	4.2	59.8	−3.0	8.7	62.0
	CL	0.9	6.6	91.8	−3.3	2.8	79.0	−2.6	2.5	85.0	−2.2	2.8	92.1	−4.3	4.5	88.3

All entries are multiplied by 100.

BIAS, bias; SE, empirical standard error; CP, coverage probability of estimates of overall disease prevalence and diagnostic test sensitivity, speci-ficity, positive predictive value, and negative predictive value. FL, HGLMM-FL method (i.e., the full-likelihood method based on the hybrid generalized linear mixed-effects model); CL, HBB-CL method (i.e., the composite-likelihood method based on the hybrid beta-binomial model; HBB-FL, the full-likelihood method based on the hybrid beta-binomial model; Heterogeneous, (0, 0, −0.6) and (0.2, −0.2, −0.6); Heterogeneous2, (0, 0, −0.6) and (0.6, −0.6, −0.6).

[*] The non-convergence rate was 24.9%.

[†] The non-convergence rate was 28.6%.

[‡] The non-convergence rate was 33.8%.

[§] The non-convergence rate was 56.3%.

[¶] The non-convergence rate was 53.6%.

[**] The non-convergence rate was 22.8%.

[††] The non-convergence rate was 31.1%.

[‡‡] The non-convergence rate was 52.2%.

[§§] The non-convergence rate was 46.8%.

Table III.

Impact of initial values on the estimates using the full likelihood based on hybrid gener-alized linear mixed-effects model (FL) and the composite likelihood (CL) methods when analyzing the data from systematic reviews of recurrent ovarian carcinoma in [24].

Imaging modality	Initial value ($\beta_0, \beta_1, \beta_2, \tau_0, \tau_1, \tau_2, \rho_{01}, \rho_{02}, \rho_{12}$)	FL method ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$)	CL method ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$)
CA125 ($m1 = 12$, $m2 = 3$)	(0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0)	Non-convergent	(0.60, 0.64, 2.51)
	(0.7, 0.7, 2.0, 1.0, 1.0, 1.0, 0.1, 0.1, 0.1)	(0.65, 0.75, 2.93)*	(0.60, 0.64, 2.51)
	(1.0, 1.0, 2.0, 1.1, 1.1, 1.1, 0.2, 0.2, 0.2)	(0.81, 0.70, 3.00)*	(0.60, 0.64, 2.51)
PET ($m1 = 9$, $m2 = 2$)	(0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0)	Non-convergent	(0.48, 2.00, 1.91)
	(0.5, 0.5, 1.0, 1.0, 1.0, 1.0, 0.1, 0.1, 0.1)	(0.63, 2.15, 2.78)*	(0.48, 2.00, 1.91)
	(0.6, 1.9, 2.2, 1.0, 1.0, 1.0, 0.2, 0.2, 0.2)	(0.47, 2.01, 2.33)	(0.48, 2.00, 1.91)
PET–CT ($m1 = 7$, $m2 = 4$)	(0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0)	Non-convergent	(0.92, 2.43, 2.01)
	(0.8, 2.5, 2.3, 1.1, 1.1, 1.1, 0.2, 0.2, 0.2)	(0.58, 2.69, 2.28)*	(0.92, 2.43, 2.01)
	(1.0, 3.0, 3.0, 1.1, 1.1, 1.1, 0.2, 0.2, 0.2)	(0.71, 2.66, 2.22)*	(0.92, 2.43, 2.01)
CT ($m1 = 8$, $m2 = 2$)	(0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0)	(0.19, 1.44, 1.77)*	(0.18, 1.34, 1.65)
	(0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.1, 0.1, 0.1)	(0.18, 1.43, 1.69)	(0.18, 1.34, 1.65)
	(0.1, 1.0, 1.0, 1.1, 1.1, 1.1, 0.1, 0.1, 0.1)	(0.19, 1.43, 1.77)	(0.18, 1.34, 1.65)

CA125, cancer antigen 125; PET, positron emission tomography; CT, computed tomography; m_1, number of case–control studies; m_2, number of cohort studies.

*Singular covariance matrix