Published in final edited form as:

Stat Methods Med Res. 2019 ; 28(10-11): 2912–2923. doi:10.1177/0962280218791602.

Copas-Like Selection Model to Correct Publication Bias in Systematic Review of Diagnostic Test Studies

Jin Piao¹, Yulun Liu², Yong Chen², Jing Ning³

¹The University of Southern California, Los Angeles, USA

²The University of Pennsylvania, Philadelphia, USA

³The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

The accuracy of a diagnostic test, which is often quantified by a pair of measures such as sensitivity and specificity, is critical for medical decision making. Separate studies of an investigational diagnostic test can be combined through meta-analysis; however, such an analysis can be threatened by publication bias. To the best of our knowledge, there is no existing method that accounts for publication bias in the meta-analysis of diagnostic tests involving bivariate outcomes. In this paper, we extend the Copas selection model from univariate outcomes to bivariate outcomes for the correction of publication bias when the probability of a study being published can depend on its sensitivity, specificity, and the associated standard errors. We develop an expectation-maximization algorithm for the maximum likelihood estimation under the proposed selection model. We investigate the finite sample performance of the proposed method through simulation studies and illustrate the method by assessing a meta-analysis of 17 published studies of a rapid diagnostic test for influenza.

Keywords

Copas selection model; Diagnostic test accuracy; Linear mixed model; Meta-analysis; Systematic reviews

Introduction

Meta-analysis of diagnostic test studies combines information from multiple publications and investigates whether a test is sufficiently specific or sensitive to fulfill its intended purpose in clinical practice. The accuracy of a diagnostic test is often quantified by a pair of measures, such as sensitivity and specificity, where sensitivity refers to the proportion of subjects correctly identified as having the disease or condition, and specificity refers to the proportion of subjects correctly identified as not having the disease or condition. The sensitivity and specificity within a study are independent as they are estimated from different

Reprints and permission: sagepub.co.uk/journalsPermissions.nav

Corresponding author: Jing Ning, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.
jning@mdanderson.org.

sub-groups, but in a quantitative synthesis via meta-analysis, they are dependent across studies. By simultaneously modeling both outcomes, Reitsma et al.¹ proposed a bivariate meta-regression model to analyze the data based on a linear mixed model (LMM). Instead of pooling information of sensitivity and specificity into a summary measure, the bivariate approach preserves the two-dimensional nature of the original data. Through the use of random effects, the correlation between sensitivity and specificity can be taken into account. Chu et al.² extended this approach using a generalized linear mixed model (GLMM) to jointly model sensitivity and specificity. The GLMM approach yields unbiased and more efficient results, particularly for the situation in which the data are sparse and there are a small number of positive or negative results. In addition to bivariate random-effects models, Rutter et al.³ developed a hierarchical summary receiver operating characteristic model that accounts for both between- and within-study variabilities. Interestingly, Harbord et al.⁴ found that bivariate random-effects meta-analysis and the hierarchical regression approach are very closely related, and even identical under some circumstances. See Ma et al.⁵ for an overview of the existing methods, including LMM and GLMM, and a discussion of their advantages and limitations. Hattori and Zhou⁶ recently proposed a Copas-type selection model for diagnostic studies and proposed a sensitivity analysis approach to evaluate the effects of publication on the summary receiver operating characteristic curve.

Despite the many approaches proposed for meta-analysis of diagnostic tests, the validity of meta-analysis may be threatened by the well-known issue of publication bias.⁷ Publication bias occurs when the decision to publish a study relies on the direction of the results, statistical significance of the results, and other factors.⁸ Such a problem has been extensively studied in the context of meta-analysis of clinical trials, and many methods have been developed to account for publication bias; for instance, the funnel plot,⁹ rank correlation,¹⁰ linear regression,¹¹ “trim and fill” methods,^{12,13} the Copas model,^{14–18} and Bayesian selection models.¹⁹ However, the impact of publication bias has been understudied in the context of systematic reviews of diagnostic test accuracy. One possible reason is that the statistical significance of treatment efficacy is well defined in clinical trials; however, statistical significance is less intuitively defined for measures of diagnostic test accuracy and (even if well defined) is much less relevant in diagnostic studies. Nevertheless, empirical evidence has demonstrated that smaller studies tend to report greater diagnostic test accuracy; thus, ignoring the impact of publication bias when conducting a meta-analysis can lead to the overestimation of the diagnostic test accuracy.²⁰

In systematic reviews of diagnostic test accuracy, it is common practice for investigators to evaluate the evidence of publication bias by evaluating a univariate summary of test accuracy such as the diagnostic odds ratio (DOR); see, for example, Xing et al. and Glas et al.^{21,22} Such a procedure may lead to biased results and loss of efficiency.²³ Deeks et al.²³ conducted comprehensive simulation studies and suggested that Begg,¹⁰ Egger¹¹ and Macaskill tests²⁴ that use the typical DORs are likely to be seriously misleading with inflated type I error rates. Burkner et al. compared the performance of several tests using different univariate measures of diagnostic accuracy.²⁵ They recommended that the trim and fill test combined with the log of DOR has reasonable performance to detect publication bias in diagnostic meta-analyses, while some regression-type methods based on univariate measures may perform poorly with either inflated type I error rates or low power.

The purpose of this paper is to fill this methodological gap by proposing a selection model that can explicitly account for the selective reporting process in the context of studies of diagnostic test accuracy. We extend the Copas selection model¹⁴ to the setting of bivariate outcomes so that the separate impacts of sensitivity and specificity on the publication process are allowed. We then equip the commonly used bivariate random effects model¹ with the proposed bivariate Copas-like model to study the potential impact of publication bias in systematic review of diagnostic accuracy. We further propose an expectation-maximization (EM) algorithm for the maximum likelihood estimation under the bivariate Copas-like selection model. The proposed EM algorithm is computationally efficient and stable.

The remainder of this paper is organized as follows. We introduce the bivariate Copas-like selection model and describe the proposed EM algorithm for parameter estimations in Section 2. In Section 3, we conduct simulation studies to assess the finite sample performance of the proposed method, and compare with existing methods. We apply the proposed method to a meta-analysis of rapid diagnostic tests for influenza in Section 4. We briefly discuss the results and possible extensions of the proposed method in Section 5.

Model and Estimation

Denote D as the disease status ascertained by the gold standard test and T as the disease status ascertained by a diagnostic test under investigation. For a specific study, denote n_{11} , n_{00} , n_{01} and n_{10} as the number of true positives, true negatives, false positives and false negatives, respectively. Let $n_{1+} = n_{11} + n_{10}$ be the number of patients with the disease and $n_{0+} = n_{01} + n_{00}$ be the number of patients without the disease. Estimated sensitivity is $\widehat{Se} = n_{11}/n_{1+}$ and estimated specificity is $\widehat{Sp} = n_{00}/n_{0+}$, as shown in Table 1. The variances of sensitivity and specificity on the logit scale can be estimated by $\text{var}\{\text{logit}(\widehat{Se})\} = 1/n_{11} + 1/n_{10}$ and $\text{var}\{\text{logit}(\widehat{Sp})\} = 1/n_{01} + 1/n_{00}$ if the numbers of patients with and without the disease are large. For studies with entries of zero, e.g., $n_{00} = 0$ or $n_{11} = 0$, a continuity correction can be made by adding 0.5 to each cell in the 2×2 table. Specifically, the estimated sensitivity is $\widehat{Se} = \{n_{11} + 0.5\}/\{n_{1+} + 1\}$, and the estimated specificity is $\widehat{Sp} = \{n_{00} + 0.5\}/\{n_{0+} + 1\}$.

A bivariate Copas selection model

Consider N studies to be reviewed. For study i , let $\widehat{Se}_i$ and $\widehat{Sp}_i$ denote the respective empirical estimates of the sensitivity and specificity of diagnostic test T . Denote s_{11} and s_{12} as the respective standard errors of $\widehat{Se}_i$ and $\widehat{Sp}_i$ on the logit scale. Following the bivariate linear mixed-effects model proposed by Reitsma et al.¹, we assume that the sensitivities and specificities from individual studies (on the logit scale) are approximately normally distributed around the mean values after the ad hoc continuity correction. In the Discussion, we discuss the use of generalized linear mixed models as a possible alternative. Allowing for between-study heterogeneity, we assume the variance component models

$$\begin{aligned}\text{logit}(\widehat{\text{Se}}_i) &= \beta_1 + u_{i1} + s_{i1}\epsilon_{i1}, \\ \text{logit}(\widehat{\text{Sp}}_i) &= \beta_2 + u_{i2} + s_{i2}\epsilon_{i2}.\end{aligned}\quad (1)$$

Here, β_1 and β_2 are respectively the underlying true sensitivity and specificity on the logit scale. The residuals ϵ_{i1} and ϵ_{i2} are independent standard normal random variables, and $i = 1, \dots, N$. The residuals ϵ_{i1} and ϵ_{i2} are assumed to be independent since sensitivity and specificity are estimated from different sub-groups within a study. The random effects u_{i1} and u_{i2} are assumed to follow the respective normal distributions $N(0, \tau_1^2)$ and $N(0, \tau_2^2)$, where τ_1^2 and τ_2^2 measure the between-study heterogeneities. To account for the potential (negative) correlation between the sensitivities and specificities across studies, we let $\text{corr}(u_{i1}, u_{i2}) = \rho_B$.

To account for publication bias in a meta-analysis, selection models that impose parametric distributions to characterize the underlying publication mechanism are usually assumed. Among many selection models for assessing publication bias, the Copas model is popular for its ability to account for both the effect size and the precision¹⁴. However, the Copas selection model cannot be directly applied to the meta-analysis of diagnostic tests because it is based on a univariate outcome. Here, we propose a bivariate Copas-like selection model for the meta-analysis of diagnostic tests. Motivated by Copas et al.¹⁴, we postulate a latent variable Z_i for the publication process, defined as

$$Z_i = \gamma_0 + \gamma_1/s_{i1} + \gamma_2/s_{i2} + \delta_i, \quad (2)$$

where $\delta_i \sim N(0, 1)$, $\text{corr}(\delta_i, \epsilon_{i1}) = \rho_1$, $\text{corr}(\delta_i, \epsilon_{i2}) = \rho_2$. It is assumed that study i is published if the latent variable $Z_i > 0$. The parameter γ_0 controls the overall likelihood of a study being published, and the parameters γ_1 and γ_2 characterize how the chance of publication depends on the precision of the estimated sensitivity and specificity, respectively. In general, γ_1 and γ_2 are positive, so studies with smaller standard errors are more likely to be published. We expect ρ_1 and ρ_2 to be positive such that studies with larger sensitivities and specificities tend to be published. Using the multivariate normality assumption, we can rewrite the models as

$$\begin{pmatrix} Z_i \\ \text{logit}(\widehat{\text{Se}}_i) \\ \text{logit}(\widehat{\text{Sp}}_i) \end{pmatrix} \sim N \left(\begin{pmatrix} \gamma_0 + \gamma_1/s_{i1} + \gamma_2/s_{i2} \\ \beta_1 \\ \beta_2 \end{pmatrix}, \Sigma \right), \quad (3)$$

where

$$\Sigma = \begin{bmatrix} 1 & \rho_1 s_{i1} & \rho_2 s_{i2} \\ \rho_1 s_{i1} & \tau_1^2 + s_{i1}^2 & \rho_B \tau_1 \tau_2 \\ \rho_2 s_{i2} & \rho_B \tau_1 \tau_2 & \tau_2^2 + s_{i2}^2 \end{bmatrix} = \begin{bmatrix} \Sigma_{i11} & \Sigma_{i12} \\ \Sigma_{i21} & \Sigma_{i22} \end{bmatrix}.$$

Note that the model in equation (1) is slightly different from the Copas selection model^{14–17}. Here, we treat the within-study variance as known and use the observed within-study variance to approximate the true within-study variance. Under this modified selection model, the number of parameters is fixed and does not increase with the number of studies in the meta-analysis. Hereafter, we refer to this modified Copas model as a Copas-like model. One advantage of this model is that it takes into consideration both the magnitude of diagnostic test accuracies (i.e., $\widehat{\text{Se}}_i$ and $\widehat{\text{Sp}}_i$) and the standard errors of the estimated diagnostic accuracies (i.e., s_{i1} and s_{i2}), allowing for different effects of these factors (ρ_1 , ρ_2 , γ_1 and γ_2).

Under the selection model, the probability of the i th study being published is $P(Z_i > 0 | s_{i1}, s_{i2}) = P(\delta_i > -\gamma_0 - \gamma_1/s_{i1} - \gamma_2/s_{i2}) = \Phi(\gamma_0 + \gamma_1/s_{i1} + \gamma_2/s_{i2})$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Denote $\eta = (\beta_1, \beta_2, \rho_B, \tau_1, \tau_2, \rho_1, \rho_2, \gamma_0, \gamma_1, \gamma_2)$. For $i = 1, \dots, N$, the log-likelihood of observing $(\widehat{\text{Se}}_i, \widehat{\text{Sp}}_i)$ given $Z_i > 0$, s_{i1} and s_{i2} can be written as

$$\begin{aligned} \log L(\eta) &= \sum_{i=1}^N \left\{ \log P(\logit(\widehat{\text{Se}}_i), \logit(\widehat{\text{Sp}}_i) | Z_i > 0, s_{i1}, s_{i2}) \right\} \\ &= \sum_{i=1}^N \left\{ \log \frac{P(Z_i > 0 | \logit(\widehat{\text{Se}}_i), \logit(\widehat{\text{Sp}}_i), s_{i1}, s_{i2}) f(\logit(\widehat{\text{Se}}_i), \logit(\widehat{\text{Sp}}_i) | s_{i1}, s_{i2})}{P(Z_i > 0 | s_{i1}, s_{i2})} \right\} \\ &= \sum_{i=1}^N \left\{ \log P(Z_i > 0 | \logit(\widehat{\text{Se}}_i), \logit(\widehat{\text{Sp}}_i), s_{i1}, s_{i2}) + \log f(\logit(\widehat{\text{Se}}_i), \logit(\widehat{\text{Sp}}_i) | s_{i1}, s_{i2}) \right. \\ &\quad \left. - \log P(Z_i > 0 | s_{i1}, s_{i2}) \right\} \\ &= \sum_{i=1}^N \left\{ \log \Phi(\mu_{ic} / \sqrt{\sigma_{ic}}) - \frac{1}{2} \log |\Sigma_{i22}| - \frac{1}{2} D_i^T \Sigma_{i22}^{-1} D_i - \log \Phi(\gamma_0 + \gamma_1/s_{i1} + \gamma_2/s_{i2}) \right\} + C, \end{aligned} \quad (4)$$

where $D_i = (\logit(\widehat{\text{Se}}_i) - \beta_1, \logit(\widehat{\text{Sp}}_i) - \beta_2)^T$, $\mu_{ic} = \gamma_0 + \gamma_1/s_{i1} + \gamma_2/s_{i2} + \Sigma_{i12} \Sigma_{i22}^{-1} D_i$, $\sigma_{ic} = 1 - \Sigma_{i12} \Sigma_{i22}^{-1} \Sigma_{i12}^T$, and C is a constant that does not depend on the parameters. In equation (4), P represents the probability function and f represents the density function. Note that in the above equation, we have

$$P(Z_i > 0 | \logit(\widehat{\text{Se}}_i), \logit(\widehat{\text{Sp}}_i), s_{i1}, s_{i2}) = \Phi(\mu_{ic} / \sqrt{\sigma_{ic}})$$

since $\{Z_i | \text{logit}(\widehat{\text{Se}}_i), \text{logit}(\widehat{\text{Sp}}_i), s_{i1}, s_{i2}\}$ follows $N(\mu_{ic}, \sigma_{ic})$ using the conditional distribution properties of the multivariate normal distribution. When both ρ_1 and ρ_2 equal zero, i.e., $(\widehat{\text{Se}}_i, \widehat{\text{Sp}}_i)$ and Z_i are unrelated, the probability of being published is not impacted by the estimated sensitivity and specificity. Under this scenario, $\Phi(\mu_{ic}/\sqrt{\sigma_{ic}})$ reduces to $\Phi(\gamma_0 + \gamma_1/s_{i1} + \gamma_2/s_{i2})$, and the above likelihood is equivalent to the likelihood obtained from standard random effects models. In other words, when $\rho_1 = \rho_2 = 0$, there is no publication bias.

Estimation: an EM algorithm

For the univariate Copas selection model, as acknowledged by Copas et al.,¹⁷ maximizing the observed likelihood is computationally challenging because the likelihood takes its maximum over a very flat plateau and often has non-convergent results. We expect similar computational challenges for the bivariate Copas-like selection model. To circumvent this computing challenge, we exploit the underlying feature of biased sampling and propose the following EM algorithm.

We denote the observed (published) data as

$$O = \{O_i, i = 1, \dots, N\} = \{\text{logit}(\widehat{\text{Se}}_i), \text{logit}(\widehat{\text{Sp}}_i), s_{i1}, s_{i2}, i = 1, \dots, N\}.$$

For each observed study $i, i = 1, \dots, N$, we assume that there are additional $m_i (m_i > 0)$ studies that are not observed due to publication bias, where the value of m_i is unknown. Based on the symmetry assumption from the random effects models in equation (1), the m_i studies have sensitivity and specificity values that are symmetrical around the underlying true values, and we let

$$O_i^* = \{\text{logit}(\widehat{\text{Se}}_i^*) = 2\beta_1 - \text{logit}(\widehat{\text{Se}}_i), \text{logit}(\widehat{\text{Sp}}_i^*) = 2\beta_2 - \text{logit}(\widehat{\text{Sp}}_i), s_{i1}, s_{i2}, I(Z_{ij} < 0), j = 1, \dots, m_i\}$$

be the corresponding unpublished data. Here Z_{ij} is the latent variable of the j th unpublished study corresponding to the i th published study, which controls for the publication process. By the biased data generating mechanism for the i th study (there are m_i unpublished studies and one published study) and the definition of the geometric distribution, the random integer follows a geometric distribution $(1 - p_i)^m p_i, m = 0, 1, \dots$, where

$$p_i = P(Z_i > 0 | \text{logit}(\widehat{\text{Se}}_i), \text{logit}(\widehat{\text{Sp}}_i), s_{i1}, s_{i2}) = \Phi(\mu_{ic}/\sqrt{\sigma_{ic}})$$

is the probability of the i th study being published given the observed data.

Following the principle of the EM algorithm, we treat $\{O_i, O_i^*, i = 1, \dots, N\}$ as the “complete data” and derive the full likelihood, including the published data and the unpublished data. The corresponding log-likelihood based on the complete data is

$$\begin{aligned} \log L^*(\eta) = & \sum_{i=1}^N \left\{ \log P(Z_i > 0, \text{logit}(\widehat{\text{Se}}_i), \text{logit}(\widehat{\text{Sp}}_i) | s_{i1}, s_{i2}) \right. \\ & \left. + m_i \log P(Z_i < 0, 2\beta_1 - \text{logit}(\widehat{\text{Se}}_i), 2\beta_2 - \text{logit}(\widehat{\text{Sp}}_i) | s_{i1}, s_{i2}) \right\}. \end{aligned}$$

We denote the current parameter value in the EM step by $\eta^* = (\beta_1^*, \beta_2^*, \rho_B^*, \tau_1^*, \tau_2^*, \rho_1^*, \rho_2^*, \gamma_0^*, \gamma_1^*, \gamma_2^*)$. In the E-step, given the observed data and η^* , the conditional expectation of m_i is

$$E(m_i | O; \eta^*) = \frac{1 - P(Z_i > 0 | O; \eta^*)}{P(Z_i > 0 | O; \eta^*)} = \frac{1 - \Phi(\mu_{ic} / \sqrt{\sigma_{ic}} | \eta^*)}{\Phi(\mu_{ic} / \sqrt{\sigma_{ic}} | \eta^*)}.$$

It follows that the expected complete-data log-likelihood function, given the current parameter estimates, is

$$\begin{aligned} \log L_E^*(\eta) &= \sum_{i=1}^N \left\{ \log P(Z_i > 0, \text{logit}(\widehat{\text{Se}}_i), \text{logit}(\widehat{\text{Sp}}_i) | s_{i1}, s_{i2}) \right. \\ &\quad \left. + E(m_i | O, \eta^*) \log P(Z_i < 0, \text{logit}(\widehat{\text{Se}}_i^*), \text{logit}(\widehat{\text{Sp}}_i^*) | s_{i1}, s_{i2}) \right\} \\ &= \sum_{i=1}^N \left[\log \{ P(Z_i > 0 | \text{logit}(\widehat{\text{Se}}_i), \text{logit}(\widehat{\text{Sp}}_i), s_{i1}, s_{i2}) f(\text{logit}(\widehat{\text{Se}}_i), \text{logit}(\widehat{\text{Sp}}_i) | s_{i1}, s_{i2}) \} \right. \\ &\quad \left. + E(m_i | O, \eta^*) \log \{ P(Z_i < 0 | \text{logit}(\widehat{\text{Se}}_i^*), \text{logit}(\widehat{\text{Sp}}_i^*), s_{i1}, s_{i2}) f(\text{logit}(\widehat{\text{Se}}_i^*), \text{logit}(\widehat{\text{Sp}}_i^*) | s_{i1}, s_{i2}) \} \right] \\ &= \sum_{i=1}^N \left[\log P(Z_i > 0 | \text{logit}(\widehat{\text{Se}}_i), \text{logit}(\widehat{\text{Sp}}_i), s_{i1}, s_{i2}) + \log f(\text{logit}(\widehat{\text{Se}}_i), \text{logit}(\widehat{\text{Sp}}_i) | s_{i1}, s_{i2}) \right. \\ &\quad \left. + E(m_i | O, \eta^*) \log P(Z_i < 0 | \text{logit}(\widehat{\text{Se}}_i^*), \text{logit}(\widehat{\text{Sp}}_i^*), s_{i1}, s_{i2}) \right. \\ &\quad \left. + E(m_i | O, \eta^*) \log f(\text{logit}(\widehat{\text{Se}}_i^*), \text{logit}(\widehat{\text{Sp}}_i^*) | s_{i1}, s_{i2}) \right] \\ &= \sum_{i=1}^N \left[-\frac{1 + E(m_i | O, \eta^*)}{2} \log |\Sigma_{i22}| - \frac{1 + E(m_i | O, \eta^*)}{2} D_i^T \Sigma_{i22}^{-1} D_i \right. \\ &\quad \left. + \log \Phi(\mu_{ic} / \sqrt{\sigma_{ic}}) + E(m_i | O, \eta^*) \log \Phi(-\mu_{ic} / \sqrt{\sigma_{ic}}) \right] \\ &\quad . \end{aligned} \tag{5}$$

The last equation is derived based on the marginal and conditional properties from the multivariate normal distributions:

$$\begin{pmatrix} \text{logit}(\widehat{\text{Se}}_i) \\ \text{logit}(\widehat{\text{Sp}}_i) \end{pmatrix} \sim N \left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \Sigma_{i22} \right),$$

$$\begin{pmatrix} \text{logit}(\widehat{\text{Se}}_i^*) \\ \text{logit}(\widehat{\text{Sp}}_i^*) \end{pmatrix} = \begin{pmatrix} 2\beta_1 - \text{logit}(\widehat{\text{Se}}_i) \\ 2\beta_2 - \text{logit}(\widehat{\text{Sp}}_i) \end{pmatrix} \sim N \left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \Sigma_{i22} \right),$$

and

$$\{Z_i | \text{logit}(\widehat{\text{Se}}_i), \text{logit}(\widehat{\text{Sp}}_i), s_{i1}, s_{i2}\} \sim N(\mu_{ic}, \sigma_{ic}).$$

In the M-step, we maximize the conditional expectation of the complete data log-likelihood (5) from the E-step, and update the estimates for the unknown parameters. With a given convergence criterion, we can solve η iteratively. The parameters from the selection model are treated as free parameters in the sensitivity analysis for the original Copas selection model^{14–17}. Of note, this is different from the proposed Copas-like selection model. For the proposed method, the parameter vector η , including the parameters from the selection model, is estimable and is updated in each iteration of the EM algorithm. Hereafter, we name the proposed method as EM under Copas-like model (EM-CLM). We use the estimation procedure by Louis et al.²⁶ for the variance estimation. Louis (1982) proposed techniques to derive the observed information matrix of the EM algorithm through the computation of the conditional expectation of the second derivative matrix and the square of the complete-data gradient vector. Specifically,

$$I_{obs} = \sum_{i=1}^n E_{\hat{\eta}} \{ B_i(\hat{\eta} | \mathcal{O}) \} - \sum_{i=1}^n E_{\hat{\eta}} \{ S_i(\hat{\eta} | \mathcal{O}_i) S_i^T(\hat{\eta} | \mathcal{O}_i) \} \quad (6)$$

$$- 2 \sum_{i < j}^n E_{\hat{\eta}} \{ S_i(\hat{\eta} | \mathcal{O}_i) \} E_{\hat{\eta}} \{ S_j(\hat{\eta} | \mathcal{O}_j)^T \}$$

where $B_i(\hat{\eta} | \mathcal{O})$ and $S_i(\hat{\eta} | \mathcal{O})$ are the Hessian matrix and score of the complete-data log-likelihood function. All three terms in (6) can be easily calculated using R functions “hessian” and “jacobian”, or the by-products of the EM algorithm. Similar to the trim and fill method,¹² the proposed EM-CLM can provide an estimate of the total number of non-published studies.

Simulation Studies

We conducted simulation studies to evaluate the finite sample performance of the proposed EM-CLM. For comparison purpose, we also conducted the bivariate LMM (hereafter referred to as Bivariate-LMM) without considering the potential publication bias. Here, we considered three different sizes of meta-analyses, a moderate sample size (N=15) and two large sample sizes (N=50 or 200). Although most meta-analyses have small sample sizes, we include the scenario with 200 studies to empirically validate the large sample properties. We

have performed the simulation studies by varying the values of all parameters under the model. Specifically, we have considered combinations of the following settings: the sensitivity and specificity were respectively set to be (0.5, 0.85), (0.8, 0.8) or (0.9, 0.4); the between-study variances were (0.5, 0.5) or (1, 2); the correlation between sensitivity and specificity ρ_B was -0.3 or -0.6 ; the correlation parameters ρ_1 and ρ_2 were (0.3, 0.3) or (0.2, 0.5); and the non-publishing rate was set to be 15% or 30% by letting $(\gamma_0, \gamma_1, \gamma_2) = (-0.35, 0.3, 0.3)$ or $(\gamma_0, \gamma_1, \gamma_2) = (-0.9, 0.3, 0.3)$, respectively. We generated the study-specific variances for both sensitivity and specificity under logit scales from the square of a normal random variable $N(0.5, 0.25)$. In total, we considered 54 unique simulation settings under different combinations of the parameter values. Due to limited space, we have included some of the results in the Supplementary Materials (Tables S1-S7).

The left panels of Tables 2 and 3 summarize the empirical biases, empirical standard deviations, average estimated standard errors, root mean squared errors (RMSEs) and coverage probabilities of the 95% confidence intervals (CIs) of sensitivity, specificity on the logit scale and between-study heterogeneities based on 1000 simulations. For the EM-CLM, the parameters were well estimated in all scenarios: the biases were small and the standard errors were close to the empirical deviations, which suggested that the proposed EM-CLM can effectively reduce the publication bias. As expected, the standard errors of all estimates decreased as the sample sizes increased. Furthermore, the coverage probabilities obtained from the EM-CLM were close to the nominal level when the sample sizes were large ($N = 200$). The right panels of Tables 2 and 3 summarize the results from the Bivariate-LMM method. The Bivariate-LMM method yielded relatively larger biases for the parameters of main interest β_1 and β_2 (sensitivity and specificity on the logit scale). The coverage probabilities obtained from the Bivariate-LMM method were not close to the nominal level.

One potential limitation of the EM algorithm is that the algorithm may converge to a local maxima of the likelihood function depending on the starting values. To assess this issue, we maximized the likelihood by using 10 different sets of initial values, in which the final estimators are defined as the maximizer over the 10 resulting estimators. We compared the simulation results obtained by using a single set of initial values to those obtained by using 10 different sets of initial values. The results were almost identical, suggesting that the EM algorithm is quite robust to the choice of the initial values under the Copas selection model.

Data Application

Influenza, also known as the flu, is an infectious disease caused by the influenza virus, of which there are three known types, influenza viruses A, B, and C.²⁷ According to the estimates from the World Health Organization, these influenza viruses cause about 3 to 5 million cases of severe illness and approximately 250,000 to 500,000 deaths each year throughout the world.^{28,27} Therefore, the early diagnosis of influenza is essential for infection control. Many rapid diagnostic tests for influenza are available. They can quickly detect influenza viruses and so are commonly used to screen patients with suspected influenza infection.²⁸ However, even though such diagnostic tests can produce quick results, they may have inconsistent accuracy, especially regarding the sensitivity of the test, which may range from 10% to 80%.^{29,28}

We considered the EM-CLM and Bivariate-LMM applied to a systematic review of published studies that evaluated one of the rapid diagnostic tests for influenza, the BinaxNOW influenza A and B, compared with the gold standard of using reverse transcriptase polymerase chain reaction (RT-PCR) to identify the virus. This systematic review reported by Chartrand et al.²⁸ combined 17 studies of 7619 patients with influenza between 2008 and 2011. We conducted the meta-analysis based on the Bivariate-LMM method, which took into account the negative association between sensitivity and specificity, but without adjusting for potential publication bias. The resulting estimate of sensitivity was 48.0% (95% CI = 39.2% to 56.7%) and that of specificity was 98.0% (95% CI = 96.1% to 99.9%). Considering the potential publication bias, we applied the proposed EM-CLM to this meta-analysis. The estimated sensitivity of the BinaxNOW influenza A and B test was 43.4% (95% CI: 36.1% to 50.7%) and the estimated specificity was 98.0% (95% CI: 96.0% to 99.9%). The between-study heterogeneities were respectively 0.75 (95% CI: 0.26 to 1.31) and 0.42 (95% CI: 0.13 to 0.88) for sensitivity and specificity on the logit scale. The estimated negative correlation between sensitivity and specificity was -0.48 (95% CI: -0.99 to -0.01). For the selection model, the estimated parameters associated with correlation ρ_1 was 0.37 (95% CI: 0.10 to 0.48) and ρ_2 was 0.40 (95% CI: 0.25 to 0.71). The estimated parameters $\hat{\gamma}$ were $(-0.40, 0.48, 0.66)^T$ with 95% CIs $(-0.71, 0.08)$, $(0.38, 0.57)$ and $(0.34, 0.98)$, respectively. The mean estimate for the total number of unpublished studies was 2.1. Compared to the results of the bivariate-LMM without adjusting for publication bias, the estimated results we obtained after adjusting for publication bias suggested that the diagnostic test had lower sensitivity, but similar specificity.

Figure 1 presents the summary points and their 95% confidence regions for the pair of true positive rate (TPR) and false positive rate (FPR) under logit scale, where the regions are calculated as Wald-based confidence regions.³⁰ As shown in Figure 1, these confidence regions are elliptical because the estimated overall sensitivity and specificity are correlated. In addition, we observe that the sensitivity was low with wide confidence interval. Due to the lower sensitivity of this test, physicians may need to verify negative tests with other alternative tests, especially during periods of peak influenza activity (i.e., high prevalence).

In practice, the positive predictive value (PPV) and negative predictive value (NPV) can assist physicians in making decisions in the context of different levels of influenza activity. Figure 2 shows the results on the PPV and NPV from methods that do and do not adjust for publication bias given 10% and 25% disease prevalence, and their associated 95% CIs. In the left panel of Figure 2, given the prevalence of influenza of 10%, the estimate of PPV from the EM-CLM was slightly different from the result obtained without adjusting for publication bias. When the prevalence increased to 25%, both methods had similar results. For the NPV, both methods had similar results regardless of the prevalence level, as shown in the right panel of Figure 2. Figure 3 presents the estimated PPV and NPV with respect to a given disease prevalence, with their pointwise 95% CIs. Plots for the range of clinically likely prevalence can indicate the usefulness of the test at different levels of incidence of the condition. For example, in the left panel of Figure 3, when the prevalence of influenza is 10%, the estimated overall PPV and NPV are 73% (95% CI = 69.1% to 76.5%) and 94.4% (95% CI = 87% to 97.7%) based on the method that does not adjust for publication bias;

whereas the EM-CLM provides the estimated overall PPV of 70.3% (95%CI = 66.6% to 73.8%) and NPV of 94.0% (95% CI = 86.0% to 97.5%). These results suggest that the method that ignores publication bias tended to overestimate the NPV in the meta-analysis of the BinaxNOW influenza A and B test. Interestingly, the impact of publication bias on the NPV was less severe than that on the PPV in our application, which can be explained by the high specificity and small impact of publication bias on the specificity of the BinaxNOW influenza A and B test. In this particular example, it seems that there is no strong evidence of severe publication bias. In practice, investigators are encouraged to perform the proposed method as a sensitivity analysis, and compare how the results will vary with or without adjusting for the potential publication bias.

Discussion

Bivariate measures such as sensitivity and specificity are commonly used in the meta-analysis of diagnostic tests. The existing literature of methods to account for publication bias has mainly focused on univariate meta-analysis.^{10,11,31} Intuitively, there are at least two advantages to considering publication bias in the multivariate meta-analysis framework. The first advantage is the potential gain in power: multivariate outcomes are often correlated, which allows for borrowing of information across outcomes. The second advantage is the avoidance of inconsistent inference: applying univariate methods separately to each outcome may lead to different conclusions on the number of unpublished studies. In this article, we propose a bivariate Copas-like selection model to account for the effect of the potential publication bias in the meta-analysis of diagnostic test accuracy. Our model has two components: the first component is the joint modeling of sensitivity and specificity (on the logit scale), which accounts for the heterogeneity and negative correlation through a LMM;¹ the second component is the selection model accounting for the probability process. We use a latent variable, the distribution of which is related to the estimated sensitivity and specificity (and related standard errors) of the test, to describe the probability of publication for each study. We further develop an EM algorithm to facilitate the model fitting, which is computationally stable as suggested through simulation studies.

Of note, there are differences between the proposed method and the Copas-type method proposed by Hattori and Zhou⁶, although both are extensions of the Copas selection model for diagnostic tests. First, the model assumptions are different. Hattori and Zhou⁶ assumed a bivariate binomial model on the pair of true positive and false positive rates, and modeled the correlation between them due to sharing the study-specific cutoff values by using shared random effects. We assumed a random effects model on the pair of sensitivity and specificity values explicitly. The selection model from the proposed method is more intuitive since we can easily correlate the publication propensity with the diagnostic accuracies. Second, the maximum likelihood estimators of the parameters under the proposed models can be obtained using the EM algorithm, while a sensitivity approach was used to evaluate the effects of publication on the summary receiver operating characteristic curve in the paper by Hattori and Zhou.

There are some assumptions and limitations of the proposed method. First, we assume parametric assumptions of the bivariate random effects model in Eq(1) and the selection

process in Eq(2). Particularly, the latent variable Z for the publication process is assumed to follow a normal distribution. With this normal distribution and the symmetric assumption implied in Eq(1), we can determine the distribution of the latent variable m_i , the number of unpublished studies corresponding with the i th published studies. We have conducted sensitivity studies to evaluate the robustness of the proposed method with respect to violations of model assumptions under two scenarios: (i) violation of the normality assumptions; and (ii) violation of the publication process assumption and using an alternative process to introduce publication bias. For both scenarios, the proposed method is relatively robust and the estimated results for the parameters of interest (β_1, β_2) remain in a reasonable range. More details can be found in the Supplementary Materials. Second, due to the bivariate nature of the analysis, the number of parameters under the diagnostic meta-analysis is more than that under the univariate meta-analysis. Also, we need several additional parameters to characterize the publication process. Hence, a moderate sample size is required, considering the number of parameters that need to be estimated.

A Bayesian method provides an alternative way to fit the proposed Copas-like selection model. For meta-analysis of univariate outcomes, Mavridis et al.¹⁹ implemented a Bayesian method for model fitting under the Copas model. As stated, the major advantage of using the Bayesian framework is that the prior knowledge on the probability of publication can be utilized in the specification of the prior distributions, and may help model fitting in the scenario with small sample sizes. For the univariate setting, the performances of the Bayesian method and the EM algorithm have been compared, and are comparable in many scenarios.¹⁸ To the best of our knowledge, there is no available literature about adjusting for publication bias in diagnostic meta-analysis using a Bayesian framework. Further performance comparisons between the Bayesian method and EM algorithm under a bivariate meta-regression model is beyond the scope of this article, and is a worthy objective for future research.

In this article, we restrict our attention to the setting in which there are no study-level covariates. To accommodate the possibility of study-level covariates, we can incorporate these covariates into the first component of the proposed model in equation (1). To model the pairs of sensitivity and specificity, we used the LMM proposed by Reitsma et al.¹ instead of the GLMM for model fitting simplicity. Reitsma et al.³² commented that the differences in fitting results between GLMM and LMM are likely to be small from the clinical perspective, and fitting the “exact” binomial model requires a more complex fitting algorithm. The bivariate LMM is subject to two limitations: the required normality assumption and the needed continuity correction. Following the bivariate GLMM proposed by Chu et al.,³³ we may generalize the logit transformation to a general link function and consider the misspecification issues of the link function. Also, the selection model can be extended as $Z_i = \gamma_0 + \gamma_1/f(s_i) + \delta_i$, where $f(s_i)$ represents a function combining the variation from logit sensitivity and logit specificity. For example, it could be the trace or determinant³⁴ of the within-study variance-covariance matrix of logit sensitivity and specificity. Determining how to choose the best function to summarize this information is worthy of future work and beyond the scope of this paper.

Another extension of our proposed method is to accommodate the case in which a gold standard does not exist. Although several papers have discussed this issue for both the traditional univariate meta-analysis and multivariate meta-analysis,^{35–38} it is not clear how to incorporate an adjustment for this scenario in the selection model for the publication bias adjustment. This is beyond the scope of this paper and will be investigated in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, and Zwinderman AH, “Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews,” *Journal of Clinical Epidemiology*, vol. 58, no. 10, pp. 982–990, 2005. [PubMed: 16168343]
2. Chu H and Cole S, “Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach,” *Journal of Clinical Epidemiology*, vol. 59, no. 12, p. 1331, 2006.
3. Rutter C and Gatsonis C, “A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations,” *Statistics in Medicine*, vol. 20, no. 19, pp. 2865–2884, 2001. [PubMed: 11568945]
4. Harbord RM, Deeks JJ, Egger M, Whiting P, and Sterne JA, “A unification of models for meta-analysis of diagnostic accuracy studies,” *Biostatistics*, vol. 8, no. 2, pp. 239–251, 2007. [PubMed: 16698768]
5. Ma X, Nie L, Cole SR, and Chu H, “Statistical methods for multivariate meta-analysis of diagnostic tests: an overview and tutorial,” *Statistical Methods in Medical Research*, vol. 25, no. 4, pp. 1596–1619, 2016. [PubMed: 23804970]
6. Hattori S and Zhou X-H, “Sensitivity analysis for publication bias in meta-analysis of diagnostic studies for a continuous biomarker,” *Statistics in Medicine*, vol. 37, no. 3, pp. 327–342, 2018. [PubMed: 28990209]
7. Rothstein HR, Sutton AJ, and Borenstein M, *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons, 2006.
8. Dickersin K, “Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm,” *Publication Bias in Meta-Analysis*, pp. 11–33, 2005.
9. Light R and Pillemer DB, *Summing up*. Harvard University Press, 1984.
10. Begg CB and Mazumdar M, “Operating characteristics of a rank correlation test for publication bias,” *Biometrics*, vol. 50, no. 4, pp. 1088–1101, 1994. [PubMed: 7786990]
11. Egger M, Smith GD, Schneider M, and Minder C, “Bias in meta-analysis detected by a simple, graphical test,” *BMJ*, vol. 315, no. 7109, pp. 629–634, 1997. [PubMed: 9310563]
12. Duval S and Tweedie R, “A nonparametric trim and fill method of accounting for publication bias in metaanalysis,” *Journal of the American Statistical Association*, vol. 95, no. 449, pp. 89–98, 2000.
13. Duval S and Tweedie R, “Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis,” *Biometrics*, vol. 56, no. 2, pp. 455–463, 2000. [PubMed: 10877304]
14. Copas JB and Li H, “Inference for non-random samples,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 59, no. 1, pp. 55–95, 1997.
15. Copas J, “What works?: selectivity models and meta-analysis,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 162, no. 1, pp. 95–109, 1999.
16. Copas J and Shi JQ, “Meta-analysis, funnel plots and sensitivity analysis,” *Biostatistics*, vol. 1, no. 3, pp. 247–262, 2000. [PubMed: 12933507]

17. Copas J and Shi J, "A sensitivity analysis for publication bias in systematic reviews," *Statistical Methods in Medical Research*, vol. 10, no. 4, pp. 251–265, 2001. [PubMed: 11491412]
18. Ning J, Chen Y, and Piao J, "Maximum likelihood estimation and em algorithm of copas-like selection model for publication bias correction," *Biostatistics*, vol. 18, no. 3, pp. 495–504, 2017. [PubMed: 28334132]
19. Mavridis D, Sutton A, Cipriani A, and Salanti G, "A fully bayesian application of the copas selection model for publication bias extended to network meta-analysis," *Statistics in Medicine*, vol. 32, no. 1, pp. 51–66, 2013. [PubMed: 22806991]
20. Song F, Eastwood AJ, Gilbody S, Duley L, and Sutton AJ, "Publication and related biases.," *Health technology assessment (Winchester, England)*, vol. 4, no. 10, p. 1, 2000.
21. Xing Y, Bronstein Y, Ross M, Askew R, Lee J, Gershenwald J, Royal R, and Cormier J, "Contemporary diagnostic imaging modalities for the staging and surveillance of melanoma patients: a meta-analysis," *Journal of the National Cancer Institute*, vol. 103, no. 2, pp. 129–142, 2011. [PubMed: 21081714]
22. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, and Bossuyt PM, "The diagnostic odds ratio: a single indicator of test performance," *Journal of Clinical Epidemiology*, vol. 56, no. 11, pp. 1129–1135, 2003. [PubMed: 14615004]
23. Deeks JJ, Macaskill P, and Irwig L, "The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed," *Journal of Clinical Epidemiology*, vol. 58, no. 9, pp. 882–893, 2005. [PubMed: 16085191]
24. Macaskill P, Walter SD, and Irwig L, "A comparison of methods to detect publication bias in meta-analysis," *Statistics in Medicine*, vol. 20, no. 4, pp. 641–654, 2001. [PubMed: 11223905]
25. Burkner P-C and Doebler P, "Testing for publication bias in diagnostic meta-analysis: a simulation study," *Statistics in Medicine*, vol. 33, no. 18, pp. 3061–3077, 2014. [PubMed: 24753050]
26. Louis TA, "Finding the observed information matrix when using the em algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 44, no. 2, pp. 226–233, 1982.
27. World Health Organization and others, "Influenza (seasonal) fact sheet no. 211, march 2014," 2014.
28. Chartrand C, Leeflang MM, Minion J, Brewer T, and Pai M, "Accuracy of rapid influenza diagnostic tests: a meta-analysis," *Annals of Internal Medicine*, vol. 156, no. 7, pp. 500–511, 2012. [PubMed: 22371850]
29. Centers for Disease Control and Prevention and others, "Rapid diagnostic testing for influenza: information for clinical laboratory directors," <http://www.cdc.gov/flu/professionals/diagnosis/rapidlab.htm>, 2007.
30. Douglas J, "Confidence regions for parameter pairs," *The American Statistician*, vol. 47, no. 1, pp. 43–45, 1993.
31. Peters JL, Sutton AJ, Jones DR, Abrams KR, and Rushton L, "Comparison of two methods to detect publication bias in meta-analysis," *Jama*, vol. 295, no. 6, pp. 676–680, 2006. [PubMed: 16467236]
32. Reitsma J and Zwinderman A, "Response to chu and cole: Bivariate meta-analysis of sensitivity and specificity with sparse data," *Journal of Clinical Epidemiology*, vol. 59, no. 12, pp. 1332–1333, 2006.
33. Chu H, Guo H, and Zhou Y, "Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models," *Medical Decision Making*, vol. 30, no. 4, 2009.
34. Mavridis D, Welton NJ, Sutton A, and Salanti G, "A selection model for accounting for publication bias in a full network meta-analysis," *Statistics in Medicine*, vol. 33, no. 30, pp. 5399–5412, 2014. [PubMed: 25316006]
35. Sadatsafavi M, Shahidi N, Marra F, FitzGerald MJ, Elwood KR, Guo N, and Marra CA, "A statistical method was used for the meta-analysis of tests for latent tb in the absence of a gold standard, combining random-effect and latent-class methods to estimate test accuracy," *Journal of Clinical Epidemiology*, vol. 63, no. 3, pp. 257–269, 2010. [PubMed: 19692208]
36. Chu H, Chen S, and 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*, vol. 104, no. 486, pp. 512–523, 2009. [PubMed: 19562044]

37. Dendukuri N, Schiller I, Joseph L, and Pai M, "Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference," *Biometrics*, vol. 68, no. 4, pp. 1285–1293, 2012. [PubMed: 22568612]
38. Liu Y, Chen Y, and Chu H, "A unification of models for meta-analysis of diagnostic accuracy studies without a gold standard," *Biometrics*, vol. 71, no. 2, pp. 538–547, 2015. [PubMed: 25358907]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Summary points and 95% confidence regions of TPR (sensitivity) versus FPR (1 - specificity) for the BinaxNOW influenza A and B test: Filled circle: summary point; solid line: boundary of 95% confidence region for the summary point.

Figure 2.

Estimated PPVs (left panel) and NPVs (right panel) for the given disease prevalence of 10% and 25%, and their corresponding 95% confidence intervals by methods that do and do not adjust for publication bias. Solid line: without adjusting for the publication bias; dashed line: the EM-CLM, which adjusts for publication bias.

Figure 3. Overall PPV and NPV plots for a given disease prevalence based on the meta-analysis with and without adjusting for publication bias.

Table 1.Data from an individual study in a 2×2 table

		<u>Gold standard (D)</u>	
		Positive	Negative
Diagnostic test (T)	Positive	n_{11}	n_{01}
	Negative	n_{10}	n_{00}
	Total	n_{1+}	n_{0+}

Table 2.

Summary statistics when $(\beta_1, \beta_2, \tau_1, \tau_2, \rho_B, \rho_1, \rho_2) = (0, 1.735, 0.5, 0.5, -0.3, 0.3, 0.3)$. NP: non-publishing rate; N: sample size; PAR: parameter; TRUE: true parameter value; MEAN: point estimate; SD: empirical standard deviation; SE: model-based standard error; RMSE: root mean squared error; CP: 95% coverage probability.

EM-CLM									Bivariate-LMM					
NP(%)	N	PAR	TRUE	MEAN	SD	SE	RMSE	CP	MEAN	SD	SE	RMSE	CP	
15	15	β_1	0	0.008	0.185	0.174	0.185	0.902	0.030	0.182	0.165	0.184	0.902	
		β_2	1.735	1.745	0.181	0.175	0.181	0.925	1.766	0.177	0.164	0.180	0.900	
		τ_1	0.5	0.446	0.180	0.160	0.188	0.825	0.440	0.176	0.151	0.186	0.805	
		τ_2	0.5	0.452	0.176	0.158	0.182	0.900	0.444	0.177	0.153	0.186	0.850	
		ρ_B	-0.3	-0.454	0.389	0.750	0.418	0.964	-0.465	0.389	0.694	0.423	0.970	
	50	β_1	0	0.010	0.097	0.099	0.098	0.944	0.033	0.095	0.093	0.101	0.923	
		β_2	1.735	1.748	0.097	0.098	0.098	0.933	1.769	0.095	0.093	0.101	0.921	
		τ_1	0.5	0.487	0.085	0.089	0.086	0.937	0.484	0.085	0.083	0.086	0.939	
		τ_2	0.5	0.487	0.084	0.087	0.085	0.934	0.484	0.083	0.083	0.085	0.934	
		ρ_B	-0.3	-0.334	0.235	0.283	0.237	0.930	-0.345	0.236	0.246	0.240	0.916	
	200	β_1	0	0.008	0.048	0.050	0.049	0.955	0.031	0.046	0.047	0.055	0.903	
		β_2	1.735	1.744	0.049	0.049	0.050	0.940	1.766	0.048	0.047	0.057	0.889	
		τ_1	0.5	0.495	0.044	0.044	0.044	0.929	0.492	0.043	0.041	0.044	0.921	
		τ_2	0.5	0.495	0.042	0.043	0.042	0.945	0.493	0.041	0.041	0.042	0.943	
		ρ_B	-0.3	-0.303	0.124	0.127	0.124	0.943	-0.316	0.120	0.121	0.121	0.938	
	30	15	β_1	0	0.016	0.189	0.197	0.190	0.902	0.051	0.181	0.164	0.188	0.877
			β_2	1.735	1.745	0.185	0.196	0.185	0.884	1.780	0.173	0.163	0.179	0.883
			τ_1	0.5	0.449	0.170	0.157	0.177	0.884	0.442	0.167	0.153	0.177	0.890
			τ_2	0.5	0.460	0.169	0.157	0.174	0.870	0.450	0.166	0.148	0.173	0.887
			ρ_B	-0.3	-0.440	0.376	0.711	0.401	0.930	-0.450	0.379	0.638	0.408	0.951
50		β_1	0	0.015	0.098	0.105	0.099	0.945	0.052	0.093	0.091	0.107	0.911	
		β_2	1.735	1.746	0.097	0.107	0.098	0.949	1.781	0.091	0.092	0.102	0.913	
		τ_1	0.5	0.487	0.084	0.086	0.085	0.919	0.483	0.080	0.080	0.082	0.941	
		τ_2	0.5	0.490	0.086	0.088	0.087	0.923	0.485	0.082	0.080	0.083	0.928	
		ρ_B	-0.3	-0.321	0.228	0.269	0.229	0.924	-0.330	0.226	0.238	0.228	0.925	
200		β_1	0	0.012	0.051	0.054	0.052	0.936	0.049	0.047	0.046	0.068	0.810	
		β_2	1.735	1.749	0.053	0.055	0.055	0.928	1.784	0.049	0.046	0.069	0.794	
		τ_1	0.5	0.496	0.044	0.045	0.044	0.936	0.493	0.042	0.040	0.043	0.930	
		τ_2	0.5	0.496	0.043	0.045	0.043	0.936	0.494	0.042	0.040	0.042	0.928	
		ρ_B	-0.3	-0.304	0.122	0.136	0.122	0.949	-0.315	0.114	0.117	0.115	0.948	

Table 3.

Summary statistics when $(\beta_1, \beta_2, \tau_1, \tau_2, \rho_B, \rho_1, \rho_2) = (0, 1.735, 1, 2, -0.3, 0.3, 0.3)$. NP: non-publishing rate; N: sample size; PAR: parameter; TRUE: true parameter value; MEAN: point estimate; SD: empirical standard deviation; SE: model-based standard error; RMSE: root mean squared error; CP: 95% coverage probability.

NP(%)	N	PAR	TRUE	EM-CLM					Bivariate-LMM				
				MEAN	SD	SE	RMSE	CP	MEAN	SD	SE	RMSE	CP
15	15	β_1	0	0.011	0.301	0.289	0.301	0.914	0.039	0.297	0.274	0.300	0.903
		β_2	1.735	1.753	0.548	0.535	0.548	0.926	1.781	0.540	0.514	0.542	0.918
		τ_1	1	0.934	0.242	0.225	0.251	0.888	0.930	0.240	0.216	0.250	0.888
		τ_2	2	1.918	0.389	0.384	0.398	0.896	1.916	0.387	0.377	0.396	0.898
		ρ_B	-0.3	-0.335	0.254	0.282	0.256	0.880	-0.338	0.255	0.268	0.258	0.884
	50	β_1	0	0.013	0.162	0.161	0.163	0.942	0.043	0.159	0.155	0.165	0.927
		β_2	1.735	1.752	0.300	0.299	0.300	0.937	1.783	0.296	0.289	0.300	0.926
		τ_1	1	0.981	0.122	0.125	0.123	0.939	0.979	0.121	0.122	0.123	0.937
		τ_2	2	1.979	0.209	0.217	0.210	0.944	1.976	0.207	0.211	0.208	0.942
		ρ_B	-0.3	-0.308	0.155	0.154	0.155	0.926	-0.311	0.154	0.148	0.154	0.922
	200	β_1	0	0.010	0.080	0.081	0.081	0.944	0.040	0.078	0.078	0.088	0.919
		β_2	1.735	1.748	0.150	0.150	0.151	0.945	1.780	0.148	0.146	0.155	0.932
		τ_1	1	0.993	0.064	0.063	0.064	0.934	0.990	0.063	0.061	0.064	0.931
		τ_2	2	1.993	0.106	0.109	0.106	0.948	1.991	0.105	0.106	0.105	0.940
		ρ_B	-0.3	-0.298	0.077	0.078	0.077	0.955	-0.302	0.076	0.075	0.076	0.947
30	15	β_1	0	0.020	0.306	0.312	0.307	0.911	0.064	0.296	0.271	0.303	0.899
		β_2	1.735	1.746	0.558	0.628	0.558	0.918	1.793	0.540	0.513	0.543	0.916
		τ_1	1	0.930	0.236	0.225	0.246	0.868	0.927	0.234	0.213	0.245	0.884
		τ_2	2	1.916	0.396	0.419	0.405	0.882	1.916	0.386	0.376	0.395	0.906
		ρ_B	-0.3	-0.334	0.245	0.299	0.247	0.883	-0.336	0.247	0.276	0.250	0.891
	50	β_1	0	0.019	0.163	0.182	0.164	0.947	0.066	0.155	0.154	0.168	0.932
		β_2	1.735	1.744	0.302	0.350	0.302	0.950	1.795	0.289	0.289	0.295	0.942
		τ_1	1	0.978	0.124	0.135	0.126	0.922	0.976	0.119	0.120	0.121	0.930
		τ_2	2	1.978	0.217	0.247	0.218	0.943	1.975	0.210	0.211	0.211	0.937
		ρ_B	-0.3	-0.301	0.153	0.175	0.153	0.927	-0.303	0.149	0.147	0.149	0.939
	200	β_1	0	0.015	0.084	0.091	0.085	0.946	0.063	0.079	0.078	0.101	0.871
		β_2	1.735	1.751	0.162	0.170	0.163	0.952	1.802	0.154	0.145	0.168	0.904
		τ_1	1	0.993	0.065	0.069	0.065	0.935	0.991	0.063	0.060	0.064	0.929
		τ_2	2	1.993	0.110	0.124	0.110	0.965	1.992	0.106	0.106	0.106	0.947
		ρ_B	-0.3	-0.298	0.077	0.087	0.077	0.954	-0.301	0.073	0.074	0.073	0.950