



Published in final edited form as:

*Stat Methods Med Res.* 2017 April ; 26(2): 914–930. doi:10.1177/0962280214562146.

## A composite likelihood method for bivariate meta-analysis in diagnostic systematic reviews

Yong Chen<sup>\*</sup> [Assistant Professor], Yulun Liu<sup>†</sup> [Ph.D. Candidate], Jing Ning<sup>‡</sup> [Assistant Professor], Lei Nie<sup>§</sup> [Assistant Professor], Hongjian Zhu<sup>¶</sup> [Assistant Professor], and Haitao Chu<sup>||</sup> [Associate Professor]

<sup>\*</sup>Division of Biostatistics, The University of Texas Health Science Center at Houston, 1200 Pressler St, Houston, TX 77030, USA (yong.chen@uth.tmc.edu; Phone: 713-500-9569)

<sup>†</sup>Division of Biostatistics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

<sup>‡</sup>Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA

<sup>§</sup>Department of Epidemiology and Preventive Medicine, University of Maryland Baltimore, Baltimore, MD 21201, USA

<sup>¶</sup>Division of Biostatistics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

<sup>||</sup>Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA

### Abstract

Diagnostic systematic review is a vital step in the evaluation of diagnostic technologies. In many applications, it involves pooling pairs of sensitivity and specificity of a dichotomized diagnostic test from multiple studies. We propose a composite likelihood method for bivariate meta-analysis in diagnostic systematic reviews. This method provides an alternative way to make inference on diagnostic measures such as sensitivity, specificity, likelihood ratios and diagnostic odds ratio. Its main advantages over the standard likelihood method are the avoidance of the non-convergence problem, which is non-trivial when the number of studies are relatively small, the computational simplicity and some robustness to model mis-specifications. Simulation studies show that the composite likelihood method maintains high relative efficiency compared to that of the standard likelihood method. We illustrate our method in a diagnostic review of the performance of contemporary diagnostic imaging technologies for detecting metastases in patients with melanoma.

### Keywords

Bivariate generalized linear mixed effects model; Composite likelihood; Diagnostic accuracy; Diagnostic review; Meta-analysis

## 1 Introduction

Conducting diagnostic reviews is a vital step in the evaluation of diagnostic technologies<sup>1,2</sup>. The majority of diagnostic papers report estimates of sensitivity and specificity and represent data summaries as  $2 \times 2$  tables based on dichotomized test results compared to the gold-standard determination of disease status<sup>3</sup>. The procedure of pooling pairs of sensitivity and specificity is not straightforward because of two important characteristics of this type of data. The first is that the estimated sensitivities and specificities are typically negatively correlated across studies<sup>4</sup>. One of the reasons for such a negative correlation is that the classification of disease status is typically based on a continuum of (at least partly) measurable traits. Suppose the higher trait is associated with positive result. Then higher threshold leads to lower sensitivity but higher specificity. When the threshold varies across studies, the study-specific sensitivity and specificity are negatively correlated<sup>5</sup>. The second important characteristic of the data is the substantial between-study heterogeneity in sensitivities and specificities<sup>6,7,8</sup>. Such heterogeneity may arise due to differences in study population characteristics, variability of assessment, and other factors.

A simple method for conducting diagnostic reviews that has been frequently used in practice is to construct a summary receiver operating characteristic (sROC) curve from the studies using simple linear regression<sup>9,6</sup>. However, serious limitations of this method have been pointed out by several authors<sup>10,8,11</sup>. Specifically, the assumptions of simple linear regression are usually not met and the resulting inference may not be valid. Furthermore, the sROC approach converts each pair of sensitivity and specificity values into a single measure of accuracy, the diagnostic odds ratio, which does not distinguish the ability to detect individuals with disease from the ability to identify healthy individuals. The conation of sensitivity and specificity makes it difficult to determine the optimal use of a test and therefore diminishes the practical utility of this method in clinical practice<sup>4</sup>.

Two statistical methods that have been proposed for conducting diagnostic reviews have overcome the limitations of the sROC approach. One is the hierarchical summary receiver operating characteristic (HSROC) model<sup>12,8</sup>. The other is the bivariate mixed effects models<sup>13,14,4,15,11</sup>. Importantly, Harbord et al. reported that the HSROC model and the bivariate mixed effects model are very closely related, and are even identical in the absence of covariates<sup>16</sup>.

Among the bivariate mixed effects models, the bivariate *general* linear mixed effects model and the bivariate *generalized* linear mixed effects model (BGLMM) are commonly used<sup>13,14,4,15,11,17</sup>. The performances of these two models have been compared by extensive simulation studies<sup>18,19</sup>, and the conclusion is that the BGLMM is preferred due to less bias and better coverage probability performance, especially for studies with small sample sizes, or with sensitivities or specificities close to 1. However, two practical issues in the standard likelihood inference have been reported<sup>18,19</sup>. The first is a non-convergence or non positive definite covariance matrix problem<sup>19</sup>. Such problems are caused mainly by the maximum likelihood estimate of the correlation being close to  $\pm 1$ , and are even more severe when the number of studies is small or moderate. The second practical issue is computational

difficulty caused by a double-integral in the likelihood function. Although modern computational methods such as Laplace or adaptive Gaussian quadrature approximation are easy to implement in software such as NLMIXED in SAS (SAS Institute Inc., Cary, NC) and ADMB (Automatic Differentiation Model Builder)<sup>20</sup>, these approximations may still have non-negligible approximation errors. These computational errors often result in unstable or unreproducible estimates (e.g., results sensitive to initial values)<sup>19</sup>. To our best knowledge, there is no satisfactory solution to these practical problems. More importantly, the standard likelihood inference of BGLMM relies on the bivariate normality assumption on the logit sensitivity and specificity, which may not be appropriate. One scenario is that the logit sensitivity and specificity may follow distributions with heavier tails. Another scenario is that the correlation between sensitivity and specificity may be non-homogeneous across studies. Under these situations, the inference based on the standard likelihood method may lead to biased estimates of diagnostic accuracies and their standard errors.

In this paper, we propose an alternative inference procedure for better computational performance and model robustness. The idea is to construct a composite likelihood (CL) function by using an independent working assumption between sensitivity and specificity<sup>21,22</sup>. Such a CL has been used in longitudinal data analysis and multivariate survival data analysis to account for the correlations between observations<sup>23,24</sup>. There are three immediate advantages of using this CL method. First, the non-convergence or non positive definite covariance matrix problem is resolved since there is no correlation parameter involved in the CL. Secondly, because the two-dimensional integration involved in the standard likelihood is substituted by one-dimensional integrals, the approximation errors are substantially reduced. Thirdly, the inference based on the CL only relies on the marginal normality of logit sensitivity and specificity. Hence the proposed method can be more robust than the standard likelihood inference to mis-specifications of the joint distribution assumption.

This article is organized as follows. In Section 2, we describe the proposed CL method. In Section 3, we conduct simulation studies to compare the CL method with the standard likelihood method where their biases, coverage probabilities and relative efficiencies are investigated. We illustrate the CL method in Section 4 with a diagnostic review of contemporary diagnostic imaging technologies for detecting metastases in patients with melanoma. We provide a brief discussion in Section 5.

## 2 Statistical Methodology

We consider a diagnostic review with  $m$  studies. For the  $i$ th study, denote  $n_{i11}$ ,  $n_{i00}$ ,  $n_{i01}$ , and  $n_{i10}$  as the number of true positives, true negatives, false positives, and false negatives, respectively,  $i = 1, \dots, m$ . Let  $n_{i1} = n_{i11} + n_{i10}$  and  $n_{i0} = n_{i01} + n_{i00}$  be the number of diseased and healthy subjects, respectively, and  $Se_i$  and  $Sp_i$  be the study-specific sensitivity and specificity, respectively.

To account for the heterogeneity between studies and the correlation between  $Se_i$  and  $Sp_i$ , the following bivariate generalized linear mixed model (BGLMM) approach is commonly

used in diagnostic reviews<sup>15,11</sup>. The BGLMM can be formulated in two stages. Specifically, the first stage of the BGLMM is

$$n_{i11}|Se_i \sim \text{Binomial}(n_{i1}, Se_i), n_{i00}|Sp_i \sim \text{Binomial}(n_{i0}, Sp_i). \quad (1)$$

If  $Se_i$  and  $Sp_i$  are known, the number of true positives  $n_{j11}$  and the number of true negative  $n_{j00}$  are assumed to follow independent binomial distributions. This conditional independence assumption is generally reasonable because the true positives and true negatives are from two different groups of patients. At the second stage, to take the heterogeneity between studies and the correlation between  $Se_i$  and  $Sp_i$  into consideration, a random effects model is assumed,

$$g(Se_i) = X_i^T \beta_1 + \mu_{i1}, g(Sp_i) = Z_i^T \beta_2 + \mu_{i2}. \quad (2)$$

Here  $g(\cdot)$  is a known link function such as a logit function,  $X_i$  and  $Z_i$  are vectors of study-level covariates, possibly overlapping, related to  $Se_i$  and  $Sp_i$ , respectively, and the intercepts  $(\mu_{j1}, \mu_{j2})$  are often assumed to follow a bivariate normal distribution with mean zero and covariate matrix  $\Sigma$ ,

$$\Sigma = \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix},$$

where  $\tau_1^2$  and  $\tau_2^2$  capture the between-study heterogeneity in sensitivities and specificities, respectively, and  $\rho$  describes the correlation between the random effects  $Se_i$  and  $Sp_i$  in the transformed scale.

To keep the notation simple and make our discussion concrete, we assume  $X_i = Z_i = 1$  and choose a logit link function. In this case,  $\beta_1$  and  $\beta_2$  are the respective overall sensitivity and specificity (in a logit scale). For diagnostic accuracy, there are several measures available to help clinicians in decision making. The most popular one is the pair of overall sensitivity and specificity, i.e.,  $\exp(\beta_1)/\{1+\exp(\beta_1)\}$  and  $\exp(\beta_2)/\{1+\exp(\beta_2)\}$ , respectively. Alternatively, likelihood ratios,  $LR+$  and  $LR-$ , have been suggested in the literature<sup>25</sup>, where  $LR+ = \text{sensitivity}/(1 - \text{specificity}) = \text{Pr}(+|D)/\text{Pr}(+|\bar{D}) = \exp(\beta_1)\{1+\exp(\beta_2)\}/\{1+\exp(\beta_1)\}$ , and  $LR- = (1 - \text{sensitivity})/\text{specificity} = \text{Pr}(-|D)/\text{Pr}(-|\bar{D}) = \{1 + \exp(\beta_2)\}/[\exp(\beta_2)\{1 + \exp(\beta_1)\}]$ . Likelihood ratios quantify the extent to which a test result changes the probability of disease. Thus they can be used by clinicians to make decisions in treating patients, conducting further testing, or not evaluating patients further because the prevalence of the disease is low<sup>25</sup>. If a single measure of diagnostic accuracy is preferred, a commonly used one is the diagnostic odds ratio (dOR), where  $\text{dOR} = \{\text{sensitivity}/(1-\text{sensitivity})\} \times \{\text{specificity}/(1-\text{specificity})\} = \exp(\beta_1 + \beta_2)$ . The value of dOR ranges from zero to infinity, with a higher value indicating better discriminatory power. A value of 1 is expected for tests with no difference detected between the diseased group and the healthy group<sup>26</sup>. We note that these descriptive measures are all functions of  $\beta_1$  and  $\beta_2$  only.

For simplicity of notation, denote  $\theta_1=(\beta_1, \tau_1^2)$ ,  $\theta_2=(\beta_2, \tau_2^2)$  and  $\Theta=(\Theta_1, \Theta_2)$ . The log likelihood function of  $(\Theta_1, \Theta_2, \rho)$  is

$$\begin{aligned} \log L(\theta_1, \theta_2, \rho) &= \sum_{i=1}^m \log \Pr(n_{i00}, n_{i11} | n_{i0}, n_{i1}) \\ &= \sum_{i=1}^m \log \int \int \text{Binomial}(n_{i00} | n_{i0}; Sp_i) \\ &\quad \times \text{Binomial}(n_{i11} | n_{i1}; Se_i) \phi(Se_i, Sp_i; \theta_1, \theta_2, \rho) dSe_i dSp_i, \quad (3) \end{aligned}$$

where  $\phi(\cdot, \cdot; \Theta_1, \Theta_2, \rho)$  is the bivariate logit normal distribution indexed by  $(\Theta_1; \Theta_2; \rho)$  and  $\text{Binomial}(\cdot | \cdot; \cdot)$  is the binomial distribution. The integral in equation (3) does not have a closed form and has to be evaluated by numerical methods such as adaptive Gaussian quadrature<sup>27</sup>. In practice, the package NLMIXED in SAS version 9.1 (SAS Institute Inc., Cary, NC) can be used to maximize the approximation to the log likelihood function in equation (3). However, the standard maximum likelihood inference (hereafter referred to as standard likelihood method or SL method) faces the computational difficulties as described in the Introduction section. These problems are due to the two-dimensional integrals in the likelihood function and the need of estimating the correlation parameter  $\rho$ <sup>19</sup>.

Now we propose an alternative inference procedure. We note that the commonly used descriptive measures of diagnostic test (e.g., overall sensitivity, specificity, diagnostic likelihood ratios and diagnostic odds ratio) are all functions of  $\beta_1$  and  $\beta_2$  only, and do not involve the correlation parameter  $\rho$ . In addition, the computational problems are mainly caused by the two-dimensional integrals. We propose to construct a pseudolikelihood by using an independent working assumption. Specifically, setting  $\rho = 0$  in equation (3), we obtain the pseudolikelihood

$$\log L_p(\theta_1, \theta_2) = \log L_1(\theta_1) + \log L_2(\theta_2), \quad (4)$$

where

$$\begin{aligned} \log L_1(\theta_1) &= \sum_{i=1}^m \log \Pr(n_{i11} | n_{i1}; \theta_1) = \sum_{i=1}^m \left\{ \log \int \text{Binomial}(n_{i11} | n_{i1}, Se_i) \phi(Se_i; \theta_1) dSe_i \right\}, \\ \log L_2(\theta_2) &= \sum_{i=1}^m \log \Pr(n_{i00} | n_{i0}; \theta_2) = \sum_{i=1}^m \left\{ \log \int \text{Binomial}(n_{i00} | n_{i0}, Sp_i) \phi(Sp_i; \theta_2) dSp_i \right\}, \end{aligned}$$

and  $\phi(\cdot; \Theta_j)$  is the logit normal distribution indexed by  $\Theta_j$  ( $j = 1, 2$ ). We note that only one-dimensional integrals are involved in the pseudolikelihood. Hence the approximation errors can be reduced. In addition, the non-convergence or non positive definite covariance matrix problem is alleviated since there is no correlation parameter involved in the pseudolikelihood. More importantly, in contrast to the bivariate normality assumption made by the standard likelihood method, the pseudolikelihood relies on the marginal normality of

logit sensitivity and specificity. Hence the pseudolikelihood based inference may be more robust than the standard likelihood inference to misspecifications of the joint distribution assumption.

Such pseudolikelihood  $L_p(\theta_1, \theta_2)$  is a type of CL where (weighted) marginal or conditional densities are multiplied together to form the CL<sup>28,21,29,22</sup>. Since each component of the CL function, i.e.,  $\log L_j(\theta_j)$  ( $j = 1, 2$ ), is a true log marginal likelihood, the corresponding score equation can be shown to be unbiased. Consequently, the maximum CL estimator  $(\tilde{\theta}_1, \tilde{\theta}_2)$  is consistent and asymptotically normal.

By a standard argument using asymptotic theories, we can show that the estimator  $(\tilde{\theta}_1, \tilde{\theta}_2)$  is asymptotically normal with mean zero and covariance matrix

$$\Sigma = \begin{pmatrix} I_{11}^{-1} & I_{11}^{-1} I_{12} I_{22}^{-1} \\ (I_{11}^{-1} I_{12} I_{22}^{-1})^T & I_{22}^{-1} \end{pmatrix},$$

where

$$I_{jj} = E \left\{ -\frac{\partial^2 \log L_j(\theta_j)}{\partial \theta_j^2} \right\} \text{ and } I_{12} = E \left[ \left\{ \frac{\partial \log L_1(\theta_1)}{\partial \theta_1} \right\} \left\{ \frac{\partial \log L_2(\theta_2)}{\partial \theta_2} \right\}^T \right]$$

for  $j = 1, 2$ . Consequently,  $(\tilde{\theta}_1, \tilde{\theta}_2)$  is approximately normal with zero mean and covariance matrix  $\hat{\Sigma}/m$  where  $\hat{\Sigma}$  is

$$\hat{\Sigma} = \begin{pmatrix} \hat{I}_{11}^{-1} & \hat{I}_{11}^{-1} \hat{I}_{12} \hat{I}_{22}^{-1} \\ (\hat{I}_{11}^{-1} \hat{I}_{12} \hat{I}_{22}^{-1})^T & \hat{I}_{22}^{-1} \end{pmatrix}, \quad (5)$$

where

$$\hat{I}_{11} = \frac{1}{m} \sum_{i=1}^m \frac{\partial^2 \log \Pr(n_{i11}|n_{i1};\theta_1)}{\partial \theta_1^2}, \quad \hat{I}_{22} = \frac{1}{m} \sum_{i=1}^m \frac{\partial^2 \log \Pr(n_{i00}|n_{i0};\theta_2)}{\partial \theta_2^2}$$

and

$$\hat{I}_{12} = \frac{1}{m} \sum_{i=1}^m \left\{ \frac{\partial \log \Pr(n_{i11}|n_{i1};\theta_1)}{\partial \theta_1} \right\} \left\{ \frac{\partial \log \Pr(n_{i00}|n_{i0};\theta_2)}{\partial \theta_2} \right\}^T.$$

It is worth noting that the asymptotic results of the maximum CL estimates follow the standard maximum likelihood<sup>30</sup> under model mis-specifications. Specifically, the variance takes the form of a sandwich form<sup>31</sup> as that in generalized estimating equation<sup>23</sup>, which often arises when the information equality does not hold.

The maximum CL estimates can be obtained by conducting two separate univariate meta-analyses with a random effect model based on data  $\{(n_{i11}, n_{i1}) : i = 1, \dots, m\}$  or  $\{(n_{i00}, n_{i0}) : i = 1, \dots, m\}$ . Such univariate random effect model can be easily fitted in most of statistical software. The covariance matrix can be easily calculated using the above formulas, which only involve one-dimensional integrations. We note that although the correlation parameter is set at 0, the off-diagonal matrices in  $\hat{\Sigma}$  still properly account for the covariance between the estimated overall sensitivity and specificity. In contrast, such covariance cannot be properly considered if investigators conduct meta-analysis by univariate meta-analyses only. In other words, although both the CL method and the univariate meta-analysis provide the same and valid inference on sensitivity and specificity alone, they provide different inferences on *functions* of sensitivity and specificity such as diagnostic likelihood ratios and diagnostic odds ratio. The CL method can correctly account for the covariance between the estimated sensitivity and specificity whereas the univariate method cannot due to the ignored covariance between the estimated sensitivity and specificity. In Section 1 of the Supplementary Materials, we numerically demonstrate the advantage of the proposed method over the univariate analysis in estimating functions of sensitivity and specificity, where the univariate method leads to confidence intervals with incorrect coverage probabilities. We consider the CL method as a method between bivariate and univariate meta-analyses, inheriting the ability of inferring functions of overall parameters (such as diagnostic likelihood ratios and diagnostic odds ratio) with correct variance/covariance estimates while not suffering from their limitations.

### 3 Simulation Study

We conduct simulation studies to evaluate the finite sample performance of the CL method and compare it to that of the SL method. The data are generated from a two-stage procedure as specified by equations (1) and (2). We consider five different scenarios including settings with or without study level covariates, normal and t distributions for random effects, and logit and complementary log-log (c-log-log) link functions. Table 1 provides a description of these five different scenarios. For the scenario with study-level covariates, we consider two covariates: a binary covariate (e.g. 1 for regional cancer and 0 for distant cancer), and a continuous covariate sampled from a uniform distribution (e.g., QUADAS score with range of [1, 14]). We consider three different sizes of meta-analysis with a relatively small number ( $m = 10$ ), a moderate number ( $m = 25$ ), and a relatively large number ( $m = 50$ ) of studies. The number of subjects in each study is randomly sampled from an application of meta-analysis for diagnostic tests of metastases, which will be described in details in Section 4. We consider two settings of sensitivity and specificity:  $(Se, Sp) = (0.70, 0.80)$  and  $(Se, Sp) = (0.90, 0.95)$ , reflecting scenarios of low accuracy and high accuracy tests. Five thousand datasets are generated for each simulation setting. For each generated dataset, we apply both the CL and SL methods to obtain the estimates of sensitivity (Se), specificity (Sp), LR+ and LR-. Specifically, the CL method is implemented in R (R Development Core Team, Version 2.14.1) by using the glmmML package<sup>32</sup>. The model-based standard errors of the CL estimates are obtained by equation (5). To obtain the SL estimates, we use the adaptive Gaussian quadrature method in the SAS NLMIXED procedure (SAS Institute Inc., Cary, NC). The model-based standard errors of the SL estimates are obtained by the inverse of the



negative hessian matrix of the log likelihood. Programming codes are provided in the Appendix.

In Table 2 we show the empirical bias (BIAS), empirical standard error (SE), average of model-based standard error estimates (MBSE) and the coverage probability (CP) of the 95% confidence intervals for the estimates when data is generated from  $\mathcal{M}_1$  with (Se, Sp)=(0.90, 0.95). When the number of studies is moderate ( $m = 25$ ) or relatively large ( $m = 50$ ), both the CL and SL methods provide approximately unbiased estimates (relative bias < 3%), MBSEs close to true SEs, and CPs close to nominal levels, with the CP from the CL method being slightly better than that from the SL method. When the number of studies is relatively small ( $m = 10$ ), both methods have approximately 10% relative bias and under-estimated standard errors in the estimation of LR+ and LR-. The CP of the CL method is still acceptable with the range of [87.8%, 90.7%] and is not influenced by the degree of correlation. In contrast, the CP of the SL method deteriorates as the magnitude of correlation increases, and has a range of [73.0%, 90.3%]. When the SL method is applied, around 5 ~ 11% of the simulated replicates have the non-convergence problem (i.e., number of iterations reaches the default number of 200 iterations while the relative gradient convergence criterion  $< 1e - 10$  is not satisfied), or non-positive definite covariance matrix problem. The fitting results under non-convergence were excluded when summarizing simulation results from replications. Simulation results with lower sensitivity, i.e., (Se, Sp)=(0.70, 0.80), are summarized in Table S1 of the Supplemental Material. Similar findings are obtained with the CL method achieving better CPs compared to the SL method.

One interesting finding seen from Table 2 is that the empirical SE of the estimates from the SL method does not decrease as the correlation increases (in magnitude) and the empirical SE from the SL method is very close to that from the CL method. This suggests that the efficiency gain in the joint analysis of sensitivity and specificity when conducting diagnostic reviews may not be large. Such an observation has been previously reported by Simel and Bossuyt<sup>25</sup>. To investigate the relative efficiency of the CL method compared to the SL method, we plot the relative efficiency (defined by the sample variance of the SL estimates, divided by that of the CL estimates) against the correlation  $\rho$ , as shown in Figure 1. All four panels in Figure 1 show that the relative efficiency of the CL method is at least 90%. Note that when the number of studies is 25, the CL method tends to be more efficient than the SL method; whereas when the number of studies is 50, the SL method tends to be more efficient.

In Table 3 we evaluate the robustness of the CL and SL methods under various model misspecifications when the number of studies is 50. Under the setting  $\mathcal{M}_2$  where the study-specific sensitivity and specificity (in logit) are generated from bivariate  $t$  distribution with 4 degree of freedom, both methods produce satisfactory inferences with unbiased estimates and CP close to the nominal level. Interestingly, under the setting  $\mathcal{M}_3$  where c-log-log (instead of logit) link is used, the CP of the CL method remains satisfactory, whereas the CP of the SL method deteriorates rapidly as the magnitude of correlation increases. This suggests that the SL method (which requires to estimate the correlation) is not sensitive to the heavy tail distribution under logit link, but is very sensitive under the asymmetric c-log-log link function. In contrast, the CL method is quite robust to the heavy tail distribution



under both link functions. Under the setting  $\mathcal{M}_4$ , we consider the heterogeneity in correlation where the correlation takes one value in half of the studies and takes a different value in the remaining half. Under this setting, the likelihood of the SL method is mis-specified, whereas the likelihood of the CL method is not because the CL method does not assume homogeneous correlation across studies. As expected, the CL method leads to unbiased estimates with CP close to the nominal level, whereas the SL method underestimates the standard errors and has poor CPs (range of CP: [80.7%, 83.8%]).

Table 4 summarizes the simulation results when study-level covariates are available and the number of studies is 30 or 50 (i.e., setting  $\mathcal{M}_5$ ). In this case, the regression coefficients are parameters of interest. Similar to the findings from Table 2, both methods provide unbiased estimates and CPs close to the nominal level. The CL method has up to 23.4% of efficiency loss. In this setting, the SL method is recommended if there is no convergence problem, where the CL method can be considered as an alternative.

In summary, simulation studies suggest that the CL method outperforms the SL method when the number of studies is relatively small, and when the correlation is heterogeneous across studies. The CL method is also more robust than the SL method under various model mis-specification settings considered. When study-level covariates are available, the CL method can be less efficient than the SL method, which should be used as an alternative when the SL method encounters convergence problem.

## 4 Applications

Melanoma is a malignant tumor of melanocytes, the cells that produce the skin pigment, melanin. Less common than other types of skin cancer, melanoma is much more dangerous when not found early, and causes the majority (75%) of deaths related to skin cancer<sup>33</sup>. Although sentinel lymph node biopsy is the acknowledged gold standard for the pathological staging of melanoma in patients whose lymph nodes are clinically negative, imaging technology has also been used in some clinical settings for preoperative lymph node assessment and postoperative surveillance. Imaging technology can be used for the early detection of melanoma metastasis, and provides a cost-effective surveillance approach<sup>34</sup>. Currently, the diagnostic imaging technologies most commonly used for melanoma include ultrasonography (US), computed tomography (CT), positron emission tomography (PET) and a combination of the latter two technologies (PET-CT). It is critical to evaluate the performance of these contemporary diagnostic imaging technologies when used for patients with melanoma.

Xing et al.<sup>35</sup> conducted a diagnostic review based on 98 published studies of 10, 528 patients carried out between January 1, 1990 and June 30, 2009. The number of studies for each diagnostic imaging technology and each type of cancer (i.e., regional and distant) are cross-tabulated in Table S3 in the Supplemental Material. To apply the composite likelihood method, we fit a sequence of nested meta-regression models where the smallest model includes a variable for stage of cancer (i.e., 1 for regional and 0 for distant), and three dummy variables for types of imaging modalities with PET-CT as the reference group. Larger models were considered sequentially by adding interaction terms between types of

cancer and imaging modalities. To select models fitted by CL method, modifications of Akaike's information criterion (AIC) and Bayesian information criterion (BIC) can be used<sup>36,37</sup>. Specifically, the composite likelihood version of AIC is defined as

CL-AIC =  $-2\log L_c + 2d_s^*$ , where  $d_s^* = \text{trace} \left\{ \hat{J}(\hat{H})^{-1} \right\}$ ,  $\hat{J}$  is the estimated covariance matrix of  $L_c(\theta_1, \theta_2) / (\theta_1, \theta_2)$  evaluated at  $(\hat{\theta}_1, \hat{\theta}_2)$  and  $\hat{H}$  is  $-2 \log L_c(\theta_1, \theta_2) / (\theta_1, \theta_2)^2$  evaluated at  $(\hat{\theta}_1, \hat{\theta}_2)$ . It is easy to show that  $d_s^*$  equals to the number of parameters in the model. The composite likelihood version of BIC is defined as

CL-BIC =  $-2\log L_c + d_s^* \log(m) + 2\gamma d_s^* \log(P)$ , where  $P$  is the number of model parameters, and  $\gamma$  is a tuning parameter and is taken as 0 when  $P$  is relatively small compared to the number of studies as suggested in Gao and Song<sup>37</sup>. The results of fitting the sequence of nested models are summarized in Table 5.

Both CL-AIC and CL-BIC suggest the use of the baseline model with 12 model parameters. However, model assumptions such as normality and equal variance need to be investigated. QQ-plots of the logit sensitivity and specificity are produced in Figure S2 of the supplementary material, and Shapiro-Wilk test<sup>38</sup> for normality is conducted. Both the plots and the test suggest the normality assumption is appropriate. To evaluate the equal variance assumption across different subgroups (i.e., different diagnostic technology and types of cancer), we apply Bartlett's test<sup>39</sup>. This test suggests that homogeneity assumption is appropriate for logit specificity, but not for logit sensitivity ( $p < 0.001$ ). To study the sensitivity of the results from the baseline model (as recommended by both CL-AIC and CL-BIC) on the equal variance assumption, we conduct a separate analysis for each subgroups, and find that the results from the subgroup analyses are generally similar to those from the baseline meta-regression model.

Figure 2 presents the results from the subgroup analyses with the CL method on the overall diagnostic sensitivity, specificity and diagnostic odds ratio, and the associated 95% confidence intervals for the four diagnostic imaging modalities. The results from the SL method are displayed as the dashed lines in Figure 1 for comparison. We note that for the subgroup with regional cancer, there are only 5 studies with PET-CT, and 3 studies with CT. The inference based on the SL method is sensitive to the choice of initial values, and singular covariance matrix is encountered. Thus, the confidence intervals for these two subgroups are not provided. In general, the results from the CL method are consistent with those from the SL method. For the surveillance of regional lymph node metastasis, US has the highest sensitivity (64%; 95% CI=40% to 82%), specificity (98%; 95% CI=95% to 99%) and diagnostic odds ratio (77.3; 95% CI=22.8 to 262.0) among all four imaging modalities. For the surveillance of distant lymph node metastasis, PET-CT has the highest sensitivity (85%; 95% CI=68% to 94%), specificity (94%; 95% CI = 86% to 97%) and diagnostic odds ratio (83.8; 95% CI=23.2 to 303.1). The results from the CL method suggest that US is a more accurate imaging modality for diagnosing regional lymph node involvement and PET-CT is the preferred imaging modality to diagnose distant metastasis. On the other hand, the confidence intervals for sensitivity, specificity and diagnostic odds ratio overlap substantially. Hence more studies are required to draw more definitive conclusions in regard to these imaging modalities and their roles for detection of metastasis.

Since the estimated bivariate summary measures are often correlated, separate confidence intervals that do not account for such a correlation may be misleading<sup>16</sup>. Figure 3 presents the summary points and 95% confidence region for sensitivity versus 1 minus specificity without stratification on stages of metastasis (i.e. regional or distant) using the CL method. Specifically, the parametric representation of the boundary of the elliptical Wald-type confidence region for sensitivity and specificity (in logit scale) is obtained as<sup>40</sup>,

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

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,  $\phi$  runs from 0 to  $2\pi$ , and  $f_{2,n-2;\alpha}$  is the upper 100 $\alpha$ % point of the  $F$  distribution with degrees of freedom 2 and  $n-2$ , and  $n$  is the number of studies. Similar as shown in Figure 1, the wide ranges of those confidence regions suggest that more studies are needed to increase the precision of those estimates, and to reach definitive conclusions comparing those imaging modalities.

## 5 Discussion

In this paper, we proposed a composite likelihood method for the bivariate analysis of sensitivity and specificity in diagnostic reviews. The main advantages of this method over the standard likelihood method are the avoidance of the non-convergence problem, computational simplicity and some robustness to model mis-specifications such as distributions with heavy tails and non-homogeneous correlations across studies. Furthermore, our simulation studies suggested that the composite likelihood method maintains a high relative efficiency compared to the standard likelihood method. Other bivariate random effects models have been considered in the literature. For example, Sarmanov beta-binomial models have been studied by Chu et al.<sup>41</sup> and Chen et al.<sup>42</sup>. The composite likelihood with working independence assumption can be used to deal with the limitations of constrained correlation parameter space<sup>43,44</sup>. It is worth mentioning that the composite likelihood approach is analogous to two other approaches under specific applications: the generalized estimating equation (GEE) by Liang and Zeger<sup>23</sup> for the inference of marginal models for longitudinal data with an independence working correlation structure, and the pseudo-partial likelihood function by Lin<sup>24</sup> for the inference of marginal models for multivariate survival data when failure times within the same unit are treated as independent random variables. We believe our method can be a useful alternative to the standard likelihood method for bivariate analysis in diagnostic reviews.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

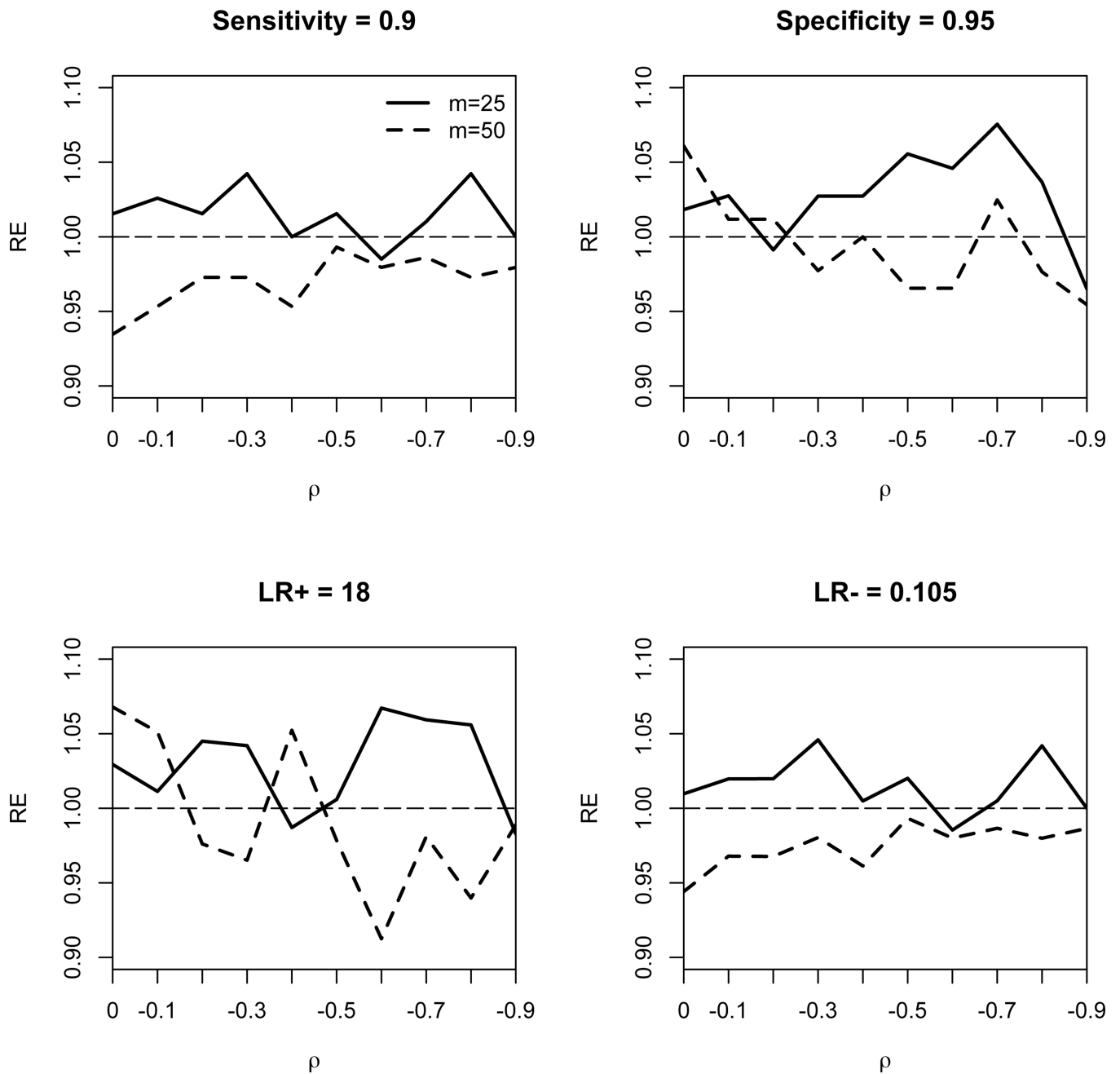
The authors thank the editor and two referees for their valuable comments during the improvement of this paper. Yong Chen was supported in part by grant number R03HS022900 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. Jing Ning was partially supported by start-up funds from the

University of Texas MD Anderson Cancer Center. Haitao Chu was supported in part by the US NIAID A1103012, NCI P01CA142538, NCI P30CA077598, and U54-MD008620. This article reflects the views of the authors and should not be construed to represent views or policies of the government. The authors want to thank Dr. Janice Cormier, Dr. Yan Xing, Dr. Chunyan Cai and Ms. Yining Du for help with the melanoma data.

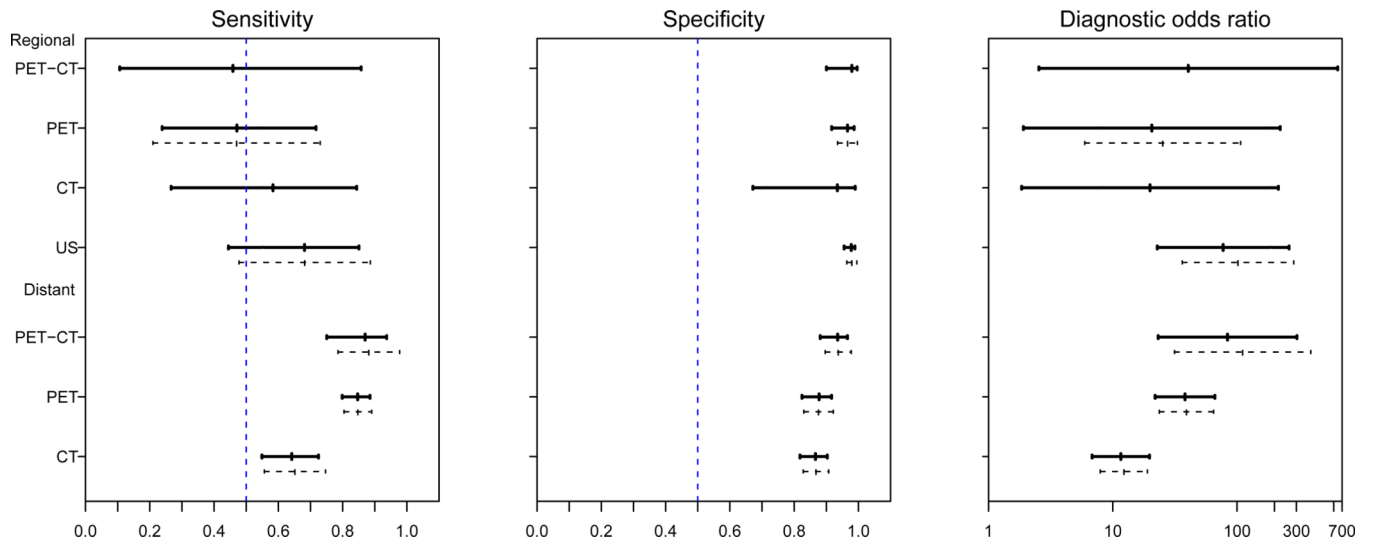
## References

1. Sackett DL, Haynes RB. The architecture of diagnostic research. *Bmj*. 2002; 324(7336):539–541. [PubMed: 11872558]
2. Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. *Journal of clinical epidemiology*. 2003; 56(11):1118–1128. [PubMed: 14615003]
3. Honest H, Khan K. Reporting of measures of accuracy in systematic reviews of diagnostic literature. *BMC health services research*. 2002; 2(1):4. [PubMed: 11884248]
4. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of clinical epidemiology*. 2005; 58(10):982–990. [PubMed: 16168343]
5. Ma X, Nie L, Cole SR, Chu H. Statistical methods for meta-analysis of diagnostic tests: an Overview and Tutorial. *Statistical Methods in Medical Research*. (in press).
6. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary roc curve: Data-analytic approaches and some additional considerations. *Statistics in medicine*. 1993; 12(14):1293–1316. [PubMed: 8210827]
7. 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]
8. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in medicine*. 2001; 20(19):2865–2884. [PubMed: 11568945]
9. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports. *Medical Decision Making*. 1993; 13(4):313. [PubMed: 8246704]
10. 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]
11. Arends L, Hamza T, Van Houwelingen J, Heijnenbrok-Kal M, Hunink M, Stijnen T. Bivariate random effects meta-analysis of ROC curves. *Medical Decision Making*. 2008; 28(5):621. [PubMed: 18591542]
12. Rutter C, Gatsonis C. Regression methods for meta-analysis of diagnostic test data. *Academic radiology*. 1995; 2:S48. [PubMed: 9419705]
13. Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. *Statistics in Medicine*. 1993; 12(24):2273–2284. [PubMed: 7907813]
14. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in medicine*. 2002; 21(4):589–624. [PubMed: 11836738]
15. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of clinical epidemiology*. 2006; 59(12):1331. [PubMed: 17098577]
16. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JAC. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*. 2007; 8(2):239–251. [PubMed: 16698768]
17. 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]
18. 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]
19. Hamza TH, Reitsma JB, Stijnen T. Meta-analysis of diagnostic studies: A comparison of random intercept, normal-normal, and binomial-normal bivariate summary ROC approaches. *Medical Decision Making*. 2008; 28(5):639–649. [PubMed: 18753684]

20. Fournier DA, Skaug HJ, Ancheta J, Ianelli J, Magnusson A, Maunder MN, et al. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*. 2012; 27(2):233–249.
21. Lindsay BG. Composite likelihood methods. *Contemporary Mathematics*. 1988; 80(1):221–39.
22. Varin C, Reid N, Firth D. An overview of composite likelihood methods. *Statistica Sinica*. 2011; 21(1):5–42.
23. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73(1):13–22.
24. Lin D. Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in medicine*. 1994; 13(21):2233–2247. [PubMed: 7846422]
25. Simel DL, Bossuyt PMM. Differences between univariate and bivariate models for summarizing diagnostic accuracy may not be large. *Journal of clinical epidemiology*. 2009; 62(12):1292–1300. [PubMed: 19447007]
26. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. *Journal of clinical epidemiology*. 2003; 56(11):1129–1135. [PubMed: 14615004]
27. Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*. 1995:12–35.
28. Besag J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B (Methodological)*. 1974:192–236.
29. Cox D, Reid N. A note on pseudolikelihood constructed from marginal densities. *Biometrika*. 2004; 91(3):729–737.
30. Kent JT. Robust properties of likelihood ratio tests. *Biometrika*. 1982; 69(1):19–27.
31. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*. 1980:817–838.
32. Broström G, Holmberg H. Generalized linear models with clustered data: Fixed and random effects models. *Computational Statistics & Data Analysis*. 2011; 55(12):3123–3134.
33. Jerant AF, Johnson JT, Sheridan C, Caffrey TJ, et al. Early detection and treatment of skin cancer. *American Family Physician*. 2000; 62(2):357–386. [PubMed: 10929700]
34. Wurm EM, Soyer HP. Scanning for melanoma. *the sun*. 2010; 5:6.
35. Xing Y, Bronstein Y, Ross MI, Askew RL, Lee JE, Gershenwald JE, et al. Contemporary diagnostic imaging modalities for the staging and surveillance of melanoma patients: a meta-analysis. *Journal of the National Cancer Institute*. 2011; 103(2):129–142. [PubMed: 21081714]
36. Varin C, Vidoni P. A note on composite likelihood inference and model selection. *Biometrika*. 2005; 92(3):519–528.
37. Gao X, Song PXX. Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *Journal of the American Statistical Association*. 2010; 105(492):1531–1540.
38. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965; 52(3/4):591–611.
39. Bartlett MS. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London Series A-Mathematical and Physical Sciences*. 1937; 160(901):268–282.
40. Douglas J. Confidence regions for parameter pairs. *The American Statistician*. 1993; 47(1):43–45.
41. Chu H, Nie L, Chen Y, Huang Y, Sun W. Bivariate random effects models for meta-analysis of comparative studies with binary outcomes: methods for the absolute risk difference and relative risk. *Statistical Methods in Medical Research*. 2012; 21(6):621–633. [PubMed: 21177306]
42. 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*. (in press).
43. Chen Y, Luo S, Chu H, Su X, Nie L. An Empirical Bayes Method for Multivariate Meta-analysis with Application in Clinical Trials. *Communications in Statistics - Theory and Methods*. (in press).
44. Chen Y, Luo S, Chu H, Wei P. Bayesian inference on risk differences: an application to multivariate meta-analysis of adverse events in clinical trials. *Statistics in Biopharmaceutical Research*. 2013; 23(5):1042–1053.

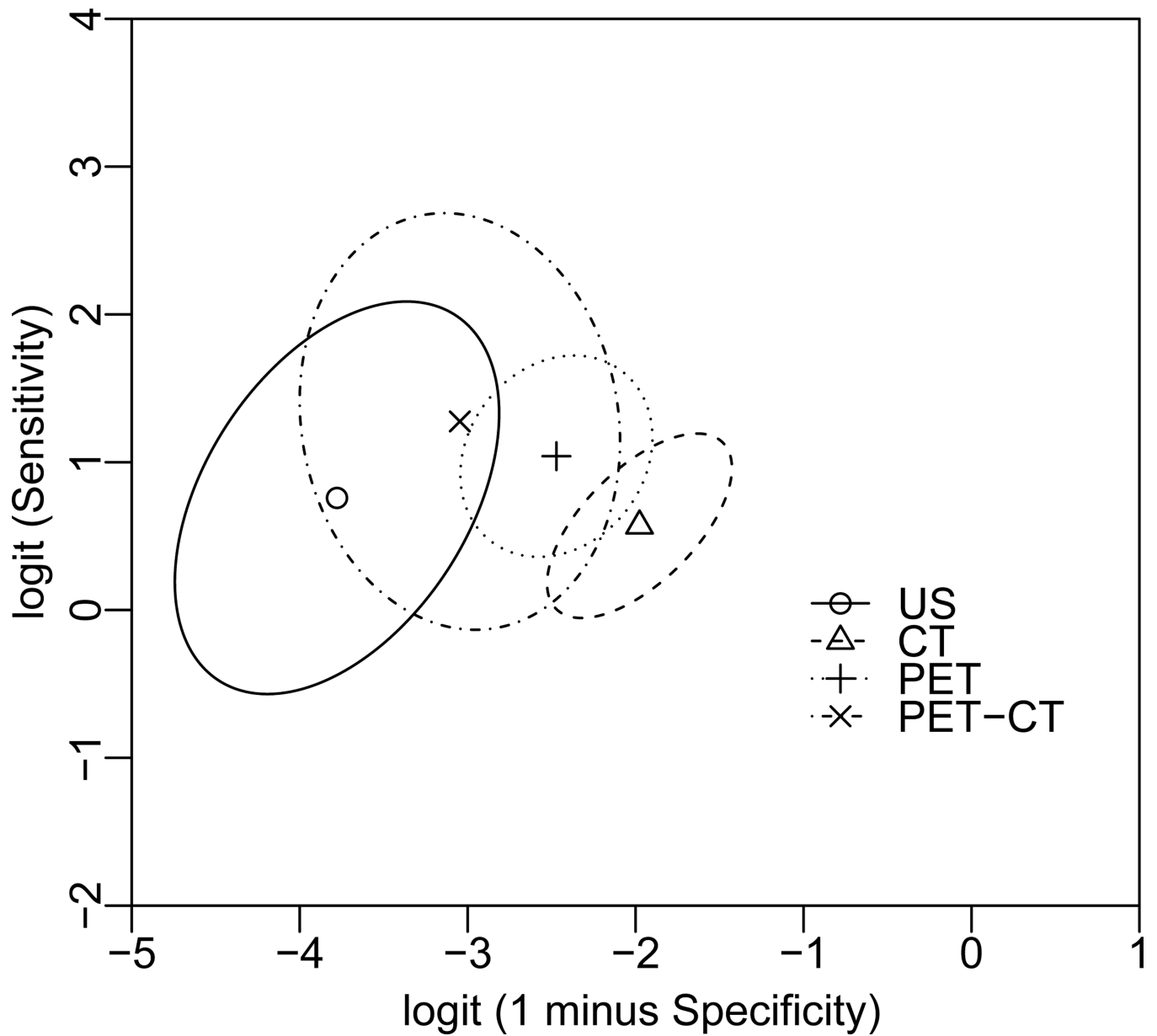


**Figure 1.** Relative efficiency (RE) of maximum composite likelihood estimator of sensitivity (upper left panel), specificity (upper right panel), LR+ (lower left panel) and LR- (lower right panel) compared to the maximum likelihood estimator under various values of correlation  $\rho$ .



**Figure 2.** Estimated accuracies and 95% confidence intervals of four diagnostic imaging technologies when identifying the pathological stage of distant and regional lymph nodes in patients with melanoma. Solid segments: results from the CL method; dashed segments: results from the SL method.





**Figure 3.** Summary points and 95% confidence regions of sensitivity versus 1-minus-specificity for four diagnostic imaging modalities. Filled circle: summary point; solid line: boundary of 95% confidence region for the summary point.

**Table 1**Configurations of five different simulation scenarios:  $\mathcal{M}_1 \sim \mathcal{M}_5$ .

Model	Covariates	Link function	Random effects distribution	Correlation structure
$\mathcal{M}_1$	No	logit	bivariate normal	fixed
$\mathcal{M}_2$	No	logit	bivariate t	fixed
$\mathcal{M}_3$	No	c-log-log	bivariate t	fixed
$\mathcal{M}_4$	No	logit	bivariate normal	mixture
$\mathcal{M}_5$	Yes	logit	bivariate normal	fixed

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Summary of 5,000 simulations with data generated from  $\theta_1$ : bias (BIAS), standard errors (SE), model-based standard errors (MBSE) and coverage probabilities (CP) of estimates. True values of model parameters are: sensitivity = 95.0, LR+ = 18.00, and LR- = 0.11. All entries for sensitivity and specificity are multiplied by 100.

m	p	SL method <sup>†</sup>					CL method*				
		BIAS	SE	MBSE	CP(%)	CP(%)	BIAS	SE	MBSE	CP(%)	
10	0.0	Se	-0.5	3.5	3.2	90.3	-0.5	3.6	3.2	89.9	
		Sp	-0.3	2.0	1.8	90.1	-0.3	1.9	1.8	90.9	
		LR+	1.82	16.49	8.61	87.2	1.63	9.84	7.92	88.3	
		LR-	0.01	0.04	0.03	90.3	0.01	0.04	0.04	90.5	
-0.3		Se	-0.5	3.6	3.3	81.5	-0.5	3.5	3.2	89.9	
		Sp	-0.3	2.0	1.8	81.5	-0.3	2.0	1.8	89.9	
		LR+	1.88	12.23	8.38	79.2	1.46	9.06	7.41	87.8	
		LR-	0.01	0.04	0.03	81.7	0.01	0.04	0.03	90.7	
-0.6		Se	-0.5	3.6	3.2	75.2	-0.6	3.5	3.2	89.5	
		Sp	-0.3	2.0	1.8	74.6	-0.3	2.0	1.8	90.0	
		LR+	1.68	8.71	7.72	73.0	1.51	8.96	7.15	88.4	
		LR-	0.01	0.04	0.03	75.4	0.01	0.04	0.03	90.5	
25	0.0	Se	-0.2	2.2	2.1	93.1	-0.2	2.2	2.1	93.4	
		Sp	-0.1	1.2	1.2	93.2	-0.1	1.2	1.2	92.6	
		LR+	0.56	4.87	4.51	91.7	0.64	4.91	4.51	91.9	
		LR-	0.00	0.02	0.02	92.9	0.00	0.02	0.02	93.6	
-0.3		Se	-0.2	2.2	2.1	92.8	-0.1	2.2	2.1	93.0	
		Sp	-0.1	1.2	1.2	92.6	-0.1	1.2	1.2	93.6	
		LR+	0.54	4.70	4.40	91.2	0.51	4.66	4.35	92.2	
		LR-	0.00	0.02	0.02	92.6	0.00	0.02	0.02	93.0	
-0.6		Se	-0.2	2.2	2.1	90.6	-0.2	2.2	2.1	92.6	
		Sp	-0.1	1.2	1.2	91.0	-0.1	1.2	1.2	93.8	

m	p	SL method <sup>†</sup>						CL method*											
		BIAS	SE	MBSE	CP(%)	BIAS	SE	MBSE	CP(%)	BIAS	SE	MBSE	CP(%)						
50	0.0	LR+	0.57	4.61	4.30	89.2	0.47	4.36	4.24	92.5	LR+	0.57	4.61	4.30	89.2	0.47	4.36	4.24	92.5
		LR-	0.00	0.02	0.02	90.8	0.00	0.02	0.02	92.7	LR-	0.00	0.02	0.02	90.8	0.00	0.02	0.02	92.7
50	0.0	Se	-0.2	1.6	1.5	93.9	-0.1	1.5	1.5	94.0	Se	-0.2	1.6	1.5	93.9	-0.1	1.5	1.5	94.0
		Sp	-0.1	0.9	0.8	94.4	-0.1	0.8	0.8	94.4	Sp	-0.1	0.9	0.8	94.4	-0.1	0.8	0.8	94.4
	0.3	LR+	0.23	3.20	3.10	93.8	0.30	3.21	3.11	93.9	LR+	0.23	3.20	3.10	93.8	0.30	3.21	3.11	93.9
		LR-	0.00	0.02	0.02	93.9	0.00	0.02	0.02	94.0	LR-	0.00	0.02	0.02	93.9	0.00	0.02	0.02	94.0
50	-0.3	Se	-0.1	1.6	1.5	94.1	-0.1	1.6	1.5	93.7	Se	-0.1	1.6	1.5	94.1	-0.1	1.6	1.5	93.7
		Sp	-0.1	0.8	0.8	94.7	-0.1	0.8	0.8	94.5	Sp	-0.1	0.8	0.8	94.7	-0.1	0.8	0.8	94.5
	-0.6	LR+	0.23	3.13	3.03	94.0	0.30	3.11	3.04	93.8	LR+	0.23	3.13	3.03	94.0	0.30	3.11	3.04	93.8
		LR-	0.00	0.02	0.02	93.9	0.00	0.02	0.02	93.5	LR-	0.00	0.02	0.02	93.9	0.00	0.02	0.02	93.5
50	-0.6	Se	-0.1	1.6	1.5	94.0	-0.1	1.5	1.5	94.7	Se	-0.1	1.6	1.5	94.0	-0.1	1.5	1.5	94.7
		Sp	-0.1	0.8	0.8	94.5	-0.1	0.8	0.8	94.7	Sp	-0.1	0.8	0.8	94.5	-0.1	0.8	0.8	94.7
	-0.9	LR+	0.22	3.04	2.95	93.8	0.23	2.99	2.95	93.5	LR+	0.22	3.04	2.95	93.8	0.23	2.99	2.95	93.5
		LR-	0.00	0.02	0.02	93.8	0.00	0.02	0.02	94.7	LR-	0.00	0.02	0.02	93.8	0.00	0.02	0.02	94.7

SL method<sup>†</sup>: standard maximum likelihood method based on the BGLMM

CL method\*: proposed composite likelihood method

**Table 3**

Estimates of bias (BIAS), standard errors (SE), model-based standard errors (MBSE) and coverage probabilities (CP) of estimates in 5,000 simulations, with study size = 50 for each simulation. The data are generated from models  $\theta_2 \sim \theta_4$ , respectively. True values of model parameters are: sensitivity = 70.0, specificity = 80.0, LR+ = 3.50, and LR- = 0.38. All entries for sensitivity and specificity are multiplied by 100.

$\rho$	SL method <sup>†</sup>					CL method*				
	BIAS	SE	MBSE	CP(%)	CP(%)	BIAS	SE	MBSE	CP(%)	CP(%)
$\theta_2$	0.0	Se	-0.2	4.0	3.9	93.8	-0.2	4.0	3.9	94.1
		Sp	-0.3	3.1	3.0	94.2	-0.3	3.1	3.1	94.8
		LR+	0.02	0.59	0.57	93.6	0.01	0.57	0.57	93.9
		LR-	0.01	0.05	0.05	94.0	0.01	0.05	0.05	94.7
-0.3		Se	-0.3	4.1	3.9	93.7	-0.2	4.0	3.9	94.0
		Sp	-0.3	3.1	3.0	94.0	-0.4	3.1	3.1	94.5
		LR+	0.01	0.53	0.52	93.5	-0.01	0.52	0.52	93.1
		LR-	0.01	0.05	0.05	93.9	0.01	0.05	0.05	94.5
-0.6		Se	-0.2	4.0	4.0	94.0	-0.4	4.0	4.0	94.0
		Sp	-0.3	3.1	3.1	94.2	-0.2	3.1	3.0	93.9
		LR+	0.02	0.46	0.46	93.5	0.01	0.47	0.46	93.8
		LR-	0.00	0.04	0.04	94.1	0.01	0.04	0.04	94.5
$\theta_3$	0.0	Se	-0.9	6.3	6.1	90.9	-0.7	6.3	6.4	93.3
		Sp	-1.2	5.7	5.4	89.0	-1.1	5.6	5.7	93.9
		LR+	0.02	1.15	1.03	86.5	0.06	1.14	0.78	78.7
		LR-	0.02	0.09	0.08	92.5	0.02	0.09	0.08	91.5
-0.3		Se	-0.7	6.1	4.4	62.6	-0.7	6.4	6.4	93.6
		Sp	-1.1	6.0	3.8	61.0	-1.1	5.6	5.7	93.6
		LR+	0.07	1.18	0.66	55.9	0.02	1.04	0.72	81.8
		LR-	0.01	0.08	0.06	63.2	0.01	0.08	0.07	91.5
-0.6		Se	-0.8	6.5	3.6	49.4	-0.7	6.4	6.4	93.5
		Sp	-1.0	5.9	3.0	49.5	-1.0	5.7	5.7	93.2

$\rho$	SL method <sup>†</sup>						CL method <sup>*</sup>					
	BIAS	SE	MBSE	CP(%)	BIAS	SE	MBSE	CP(%)	BIAS	SE	MBSE	CP(%)
$\rho_4$ (0.0, -0.3)	LR+	0.04	1.00	0.51	45.7	0.03	0.98	0.67	81.4			
	LR-	0.01	0.07	0.05	49.7	0.01	0.07	0.06	92.2			
	Se	-0.3	4.6	3.1	81.0	0.0	3.2	3.2	94.1			
	Sp	-0.2	3.4	2.4	81.7	-0.1	2.5	2.4	93.7			
	LR+	0.04	0.64	0.44	82.0	0.03	0.45	0.44	93.6			
	LR-	0.01	0.06	0.04	80.9	0.00	0.04	0.04	93.8			
(0.0, -0.6)	Se	-0.3	4.6	3.1	81.1	-0.1	3.2	3.2	94.3			
	Sp	-0.2	3.3	2.4	83.4	-0.1	2.5	2.4	94.2			
	LR+	0.04	0.58	0.42	83.8	0.02	0.42	0.42	94.1			
	LR-	0.00	0.06	0.04	81.0	0.00	0.04	0.04	94.7			
(-0.3, -0.6)	Se	-0.3	4.6	3.1	81.1	-0.1	3.2	3.2	94.0			
	Sp	-0.2	3.4	2.4	81.8	-0.1	2.5	2.4	94.0			
	LR+	0.04	0.57	0.39	82.1	0.02	0.40	0.40	93.6			
	LR-	0.00	0.05	0.04	80.7	0.00	0.04	0.04	94.0			

SL method<sup>†</sup>: standard maximum likelihood method based on the BGLMM

CL method<sup>\*</sup>: proposed composite likelihood method

**Table 4**

Summary of 5,000 simulations with data generated from  $\theta_5$ : bias (BIAS), standard errors (SE), model-based standard errors (MBSE), coverage probabilities (CP) and relative efficiency (RE) of estimates of regression coefficients. True values of coefficients are:  $\beta_{10} = 1.71$ ,  $\beta_{11} = -1.27$ ,  $\beta_{12} = 0.00$ ,  $\beta_{20} = 1.91$ ,  $\beta_{21} = 1.26$ , and  $\beta_{22} = 0.00$ .

m	$\rho$	SL method <sup>†</sup>						CL method*					
		BIAS	SE	MBSE	CP(%)	RE(%)	BIAS	SE	MBSE	CP(%)	RE%		
30	0.0	$\beta_{10}$	0.00	0.32	0.30	92.1	100.0	0.00	0.37	0.33	92.6	78.8	
		$\beta_{11}$	0.00	0.45	0.41	91.4	100.0	-0.01	0.51	0.46	91.2	76.8	
		$\beta_{12}$	0.00	0.38	0.36	92.2	100.0	0.00	0.44	0.40	91.6	75.8	
		$\beta_{20}$	0.00	0.30	0.28	92.2	100.0	0.00	0.31	0.28	92.1	99.3	
		$\beta_{21}$	0.01	0.46	0.43	91.9	100.0	0.00	0.46	0.43	92.4	101.3	
		$\beta_{22}$	0.00	0.40	0.37	92.0	100.0	0.01	0.39	0.37	92.3	102.0	
-0.3	$\beta_{10}$	0.00	0.32	0.30	91.8	100.0	0.00	0.37	0.33	91.9	77.9		
	$\beta_{11}$	0.00	0.45	0.41	91.4	100.0	0.00	0.50	0.46	91.3	80.5		
	$\beta_{12}$	0.00	0.38	0.36	92.1	100.0	-0.01	0.44	0.40	91.7	74.7		
	$\beta_{20}$	0.00	0.31	0.28	92.1	100.0	0.00	0.31	0.28	92.3	100.0		
	$\beta_{21}$	0.01	0.46	0.42	91.8	100.0	0.00	0.46	0.43	92.2	103.1		
	$\beta_{22}$	0.00	0.40	0.37	92.5	100.0	0.00	0.40	0.37	92.5	99.5		
-0.6	$\beta_{10}$	0.00	0.32	0.29	91.0	100.0	0.00	0.37	0.33	92.1	75.3		
	$\beta_{11}$	0.00	0.44	0.41	90.4	100.0	-0.02	0.50	0.46	91.6	78.2		
	$\beta_{12}$	0.00	0.38	0.35	91.3	100.0	0.00	0.44	0.40	91.8	73.6		
	$\beta_{20}$	0.00	0.31	0.28	90.5	100.0	0.00	0.31	0.28	92.0	98.7		
	$\beta_{21}$	0.01	0.46	0.42	90.7	100.0	0.00	0.46	0.43	91.8	100.9		
	$\beta_{22}$	0.00	0.39	0.37	91.3	100.0	0.00	0.40	0.37	92.6	97.0		
50	0.0	$\beta_{10}$	0.00	0.24	0.23	93.5	100.0	0.00	0.28	0.26	93.2	78.1	
		$\beta_{11}$	-0.01	0.33	0.32	93.6	100.0	-0.01	0.38	0.36	93.0	77.5	
		$\beta_{12}$	0.01	0.29	0.28	93.4	100.0	-0.01	0.33	0.31	93.1	76.3	



m	p	SL method <sup>†</sup>						CL method <sup>*</sup>					
		BIAS	SE	MBSE	CP(%)	RE(%)	RE%	BIAS	SE	MBSE	CP(%)	RE%	
-0.3	$\beta_{20}$	-0.01	0.23	0.22	93.6	100.0	94.1	0.00	0.24	0.22	93.3	94.1	
	$\beta_{21}$	0.01	0.35	0.33	93.6	100.0	98.3	0.00	0.35	0.33	93.3	98.3	
	$\beta_{22}$	0.00	0.30	0.29	93.2	100.0	102.7	-0.01	0.30	0.29	94.3	102.7	
-0.6	$\beta_{10}$	0.00	0.24	0.23	93.5	100.0	79.8	0.00	0.27	0.26	93.7	79.8	
	$\beta_{11}$	-0.01	0.33	0.32	93.6	100.0	79.2	0.00	0.37	0.36	93.6	79.2	
	$\beta_{12}$	0.00	0.29	0.28	93.5	100.0	79.1	-0.01	0.33	0.31	93.5	79.1	
-0.6	$\beta_{20}$	0.00	0.23	0.22	93.8	100.0	99.1	0.00	0.23	0.22	93.8	99.1	
	$\beta_{21}$	0.01	0.35	0.33	93.4	100.0	100.0	0.00	0.35	0.33	93.0	100.0	
	$\beta_{22}$	0.00	0.30	0.29	93.5	100.0	100.0	0.00	0.30	0.29	93.5	100.0	
-0.6	$\beta_{10}$	0.00	0.24	0.23	93.3	100.0	76.8	0.00	0.28	0.26	93.5	76.8	
	$\beta_{11}$	0.00	0.33	0.32	93.5	100.0	76.6	-0.01	0.38	0.36	93.2	76.6	
	$\beta_{12}$	0.00	0.29	0.28	93.4	100.0	77.1	-0.01	0.33	0.31	93.3	77.1	
-0.6	$\beta_{20}$	0.00	0.23	0.22	93.6	100.0	99.1	0.01	0.23	0.22	93.2	99.1	
	$\beta_{21}$	0.01	0.35	0.33	93.2	100.0	101.7	0.01	0.35	0.33	93.6	101.7	
	$\beta_{22}$	0.00	0.30	0.28	93.5	100.0	99.3	0.00	0.30	0.29	93.6	99.3	

SL method<sup>†</sup>: standard maximum likelihood method based on the BGLMM

CL method<sup>\*</sup>: proposed composite likelihood method

**Table 5**

Model selection using the CL-AIC and CL-BIC when analyzing the data in Xing et al. (2011).

Model	$d_s^*$	-2logCL	CL-AIC	CL-BIC
baseline	12	1050	1074	1106
+I(Regional)*I(US)	14	1050	1078	1115
+I(Regional)*I(CT)	14	1048	1076	1113
+I(Regional)*I(PET)	14	1050	1078	1114
+I(Regional)*I(US)+I(Regional)*I(CT)	16	1048	1080	1122
+I(Regional)*I(US)+I(Regional)*I(PET)	16	1050	1082	1124
+I(Regional)*I(CT)+I(Regional)*I(PET)	16	1048	1080	1122
+I(Regional)*I(US)+I(Regional)*I(CT)+I(Regional)*I(PET)	18	1048	1084	1131

baseline: meta-regression model with study-level covariates of I(Regional) + I(US) + I(CT) + I(PET).