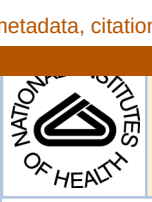
View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Carolina Digital Repository



NIH Public Access

Author Manuscript

Med Decis Making. Author manuscript; available in PMC 2011 July 1.

Published in final edited form as:

Med Decis Making. 2010 ; 30(4): 499–508. doi:10.1177/0272989X09353452.

Bivariate Random Effects Meta-analysis of Diagnostic Studies Using Generalized Linear Mixed Models

HAITAO CHU,
Department of Biostatistics and Lineberger Comprehensive Cancer Center, Univerity of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA

HONGFEI GUO, and
Division of Biostatistics and Clinical and Translational Science Institute, University of Minnesota, Minneapolis, MN 55414

YIJIE ZHOU
Merck Research Laboratories, Merck & Co., Inc., Rahway, NJ 07065

HAITAO CHU: hchu@bios.unc.edu; HONGFEI GUO: hfguo@umn.edu; YIJIE ZHOU: yijie_zhou@merck.com

Abstract

Bivariate random effect models are currently one of the main methods recommended to synthesize diagnostic test accuracy studies. However, only the logit-transformation on sensitivity and specificity has been previously considered in the literature. In this paper, we consider a bivariate generalized linear mixed model to jointly model the sensitivities and specificities, and discuss the estimation of the summary receiver operating characteristic curve (ROC) and the area under the ROC curve (AUC). As the special cases of this model, we discuss the commonly used logit, probit and complementary log-log transformations. To evaluate the impact of misspecification of the link functions on the estimation, we present two case studies and a set of simulation studies. Our study suggests that point estimation of the median sensitivity and specificity, and AUC is relatively robust to the misspecification of the link functions. However, the misspecification of link functions has a noticeable impact on the standard error estimation and the 95% confidence interval coverage, which emphasizes the importance of choosing an appropriate link function to make statistical inference.

Keywords

meta-analysis; bivariate random effect models; sensitivity; specificity; receiver operating characteristic curve; area under the ROC curve

1. Introduction

Accurate diagnosis of a disease condition such as tumor mutation is often the first step toward its control and prevention. Performance of a diagnostic test is often measured by paired indices, such as sensitivity and specificity, positive and negative predictive values, or positive and negative diagnostic likelihood ratios [1,2]. Sensitivity and specificity are often regarded as intrinsic properties of a diagnostic test. Sensitivity (Se), also referred to as the true positive fraction (TPF), is defined as the conditional probability of testing positive in diseased subjects, i.e., Pr(T = 1| D = 1) where T and D denote the binary test and disease

Correspondence to: HAITAO CHU, hchu@bios.unc.edu.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

status, respectively. Specificity (Sp), also known as the true negative fraction (TNF), is defined as the conditional probability of test negative in non-diseased subjects, i.e., i.e., Pr(T = 0| D = 0).

The rapid growth of evidence-based medicine has led to a dramatic increase in attention to evidence-based diagnosis by meta-analysis of diagnostic test accuracy studies [3]. Meta-analysis allows us to summarize the results from similar diagnostic test accuracy studies quantitatively. In situations where studies compare a diagnostic test with its gold standard, numerous methods are available to take the heterogeneity between studies into account [4–14]. Such heterogeneity arises between studies due to the differences in disease prevalence, study design as well as laboratory and other errors. Because of this heterogeneity, random effects models including the hierarchical summary receiver operating characteristic model [4] and bivariate random effects meta-analysis on sensitivities and specificities [6,8,10], which are identical in some situations, have been recommended [11,12,15]. Furthermore, Riley and others [16–18] suggested that bivariate random-effects meta-analysis offers numerous advantages over separate univariate meta-analysis through extensive simulations. Chu et al. [19] also discussed trivariate nonlinear random-effects models on jointly modeling the disease prevalence, sensitivities and specificities, and an alternative parameterization on jointly modeling the test prevalence and the predictive values. When the diagnostic test itself and the reference test are both imperfect, Walter [20] and Chu et al. [21] discussed the latent class random effects models for a meta-analysis of two diagnostic tests. Sutton et al. [22] discussed the integration of meta-analysis and economic decision modeling for evaluating diagnostic tests. Walter discussed the properties of the summary receiver operating characteristic curve for diagnostic test data [23] and the partial area under the summary ROC curve.

However, in situations where studies compare a diagnostic test with its gold standard reference test, to our knowledge, only logit transformation has been used for the bivariate random effects meta-analysis of sensitivity and specificity parameters (i.e., Se_i and Sp_i) in practice. The other transformations such as the probit and complementary log-log have not been utilized in this setting. It is conceivable that some transformations may provide a better goodness of fit than others for a particular meta-analysis, and in return may provide a better statistical inference for the parameters of interest. For example, complementary log-log models are frequently used when the probability of an event is very small or very large and thus be more applicable for a diagnostic test with very high sensitivity and specificity. Furthermore, unlike the logit and probit transformation, which are symmetrical, the complementary log-log transformation is asymmetrical. This property implies that the bivariate normal distribution assumption on (Se_i, $1 - Sp_i$), (Se_i, Sp_i), ($1 - Se_i$, Sp_i), or ($1 - Se_i$, $1 - Sp_i$) in the transformed scale will provide the same goodness of fit and inference if we use logit or probit transformation, but will generally provide different goodness of fit if we use the complementary log-log transformation.

In this article, we focus on situations where the reference test can be considered as a gold standard, and consider a bivariate generalized linear mixed effects model for meta-analysis of diagnostic accuracy studies with logit, probit and complementary log-log transformation as special cases. Specifically, in Section 2, we present the generalized bivariate random effects model in this setting, and discuss the estimation of parameters and the summary receiver operating characteristic curve (ROC). Furthermore, we discuss the estimation of the area under the ROC curve (AUC) and the impact of misspecification of link functions on parameter estimation, which has not been discussed in the literature. In Section 3, we reanalyze two real data sets as illustrating examples. We present a simulation study in Section 4, and a brief discussion in Section 5.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2. Bivariate Random Effects Meta-regression Model Using Generalized Linear Mixed Model

First, we discuss statistical methods focusing on the setting where each study presents the number of true positive, true negative, false positive and false negative subjects without any study-level or individual-level covariates. In the i^{th} diagnostic studies from a meta-analysis, let n_{i11}, n_{i00}, n_{i01}, and n_{i10} be the number of true positive, true negative, false positive and false negative subjects, respectively. Furthermore, let $n_{i1+} = n_{i11} + n_{i10}$ and $n_{i0+} = n_{i01} + n_{i00}$ be the number of diseased and non-diseased individuals. Conditional on the number of diseased and non-diseased patients in each study, the bivariate random-effects meta-analysis model first assumes that n_{i01} and n_{i11} are binomially distributed as Bin(n_{i0+}, $1 - Sp_i$) and Bin(n_{i1+}, Se_i) respectively, where Sp_i and Se_i are the specificity and sensitivity parameters for the i^{th} diagnostic studies. Although it is common in practice to transform the specificity and sensitivity parameters Sp_i and Se_i with the logit transformation, other transformations such as the probit and complementary log-log can be used as well. In this article, we consider a bivariate generalized linear mixed effects model as a general framework for the meta-analysis of diagnostic tests when a gold standard reference test is available. When using some transformations such as the complementary log-log transformation, the bivariate normal distributional assumption of (Se_i, $1 - Sp_i$), (Se_i, Sp_i), ($1 - Se_i$, Sp_i) or ($1 - Se_i$, $1 - Sp_i$) in the transformed scale will generally provide different goodness of fit. To simplify our discussion, we focus on a bivariate generalized linear mixed effects model assuming bivariate normal distributional assumption of (Se_i, $1 - Sp_i$), which is specified as follows,

$$g(\mathrm{Se}_i)=\mu_i, \quad g(1-Sp_i)=\nu_i, \quad \text{and} \quad (\mu_i,\nu_i)^T \sim N(\boldsymbol{\mu}, \sum) \tag{1}$$

where $g()$ is a monotone link function such as commonly used logit, probit and complementary log-log transformation, the mean vector $\boldsymbol{\mu}=(\mu_0, \nu_0)^T$, and the variance-covariance matrix $\sum = \begin{pmatrix} \sigma_\mu^2 & \rho\sigma_\mu\sigma_\nu \\ \rho\sigma_\mu\sigma_\nu & \sigma_\nu^2 \end{pmatrix}$. The class of models in equation (1) can be further extended to allow different transformations for sensitivity and specificity parameters, e.g., $g_1(Se_i) = \mu_i$, $g_2(1 - Sp_i) = \nu_i$. For simplicity and ease of discussion, we focus on using the same transformation function for both sensitivity and specificity in this article. Based on this model, the median sensitivity and specificity for the population is $Se_M = g^{-1}(\mu_0)$ and $Sp_M = 1 - g^{-1}(\nu_0)$. The mean sensitivity and specificity for the population can be estimated as $Se_E = \int_{-\infty}^{+\infty} g^{-1}(\mu_0+x) f_\mu(x)dx$ and $Sp_E = \int_{-\infty}^{+\infty} g^{-1}(\nu_0+x) f_\nu(x)dx$ where $f_\mu(x)$ and $f_\nu(x)$ are normal density functions with mean 0 and standard deviations of σ_μ and σ_ν, respectively.

The summary receiver operating characteristic (ROC) curve can be obtained through a characterization of the estimated bivariate normal distribution in (1) by a line. A straightforward choice may be the regression line of $g(Se_i)$ on $g(1 - Sp_i)$. Please refer to Arends et al. [13] for other potential choices. Based on the bivariate normality assumption of $(\mu_i, \nu_i)^T$, the expected sensitivity for a chosen specificity in the transformed scale is given by

$$g(\mathrm{Se})=\mu_0+\rho\sigma_\mu/\sigma_\nu\,[g(1-\mathrm{Sp})-\nu_0]=(\mu_0-\rho\nu_0\sigma_\mu/\sigma_\nu)+\rho\sigma_\mu/\sigma_\nu\,[g(1-\mathrm{Sp})] \tag{2}$$

Let $\varphi()$ be a standard Gaussian density function. The expected sensitivity for a given specificity is given by

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$E(Se|Sp)=\int_{-\infty}^{+\infty} g^{-1}\left\{(\mu_0-\rho\nu_0\sigma_\mu/\sigma_\nu)+\rho\sigma_\mu/\sigma_\nu\left[\mathrm{g}(1-\mathrm{Sp})\right]+x\right\}\frac{1}{\sqrt{\sigma_\mu^2(1-\rho^2)}}\varphi\left(x/\sqrt{\sigma_\mu^2(1-\rho^2)}\right)dx, \tag{3}$$

which may be approximated by the median sensitivity for a given specificity as

$$M(Se|Sp)=g^{-1}\left\{(\mu_0-\rho\nu_0\sigma_\mu/\sigma_\nu)+\rho\sigma_\mu/\sigma_\nu\left[\mathrm{g}(1-\mathrm{Sp})\right]\right\}. \tag{4}$$

Thus, the expected area under the summary operating characteristic (ROC) curve (AUC) can be estimated as

$$\mathrm{AUC}_E=\int_0^1\int_{-\infty}^{+\infty} g^{-1}\left\{(\mu_0-\rho\nu_0\sigma_\mu/\sigma_\nu)+\rho\sigma_\mu/\sigma_\nu\left[\mathrm{g}(1-\mathrm{Sp})\right]+x\right\}\frac{1}{\sqrt{\sigma_\mu^2(1-\rho^2)}}\varphi\left(x/\sqrt{\sigma_\mu^2(1-\rho^2)}\right)dx d\mathrm{Sp}, \tag{5}$$

which can be approximated by integration of the summary ROC based on the median sensitivity for a given specificity as

$$\mathrm{AUC}_M=\int_0^1 g^{-1}\left\{(\mu_0-\rho\nu_0\sigma_\mu/\sigma_\nu)+\rho\sigma_\mu/\sigma_\nu\left[\mathrm{g}(1-\mathrm{Sp})\right]\right\}dSp. \tag{6}$$

To select a link function that can give a better goodness of fit, we used the Akaike's Information Criterion (AIC) as the guideline [24]. The smaller value of AIC, the better goodness-of-fit. The bivariate generalized linear mixed effects model can be fitted using commonly used statistical software such as SAS, SPLUS/R and STATA. We implement it through the SAS NLMIXED procedure (SAS Institute Inc., Cary, NC), which uses an adaptive Gaussian quadrature to approximate the likelihood integrated over the random effects by dual quasi-Newton optimization techniques [25]. Furthermore, the NLMIXED built-in delta method is used to compute the population estimates of the back-transformed parameters of interest including the median sensitivity and specificity, the area under the summary ROC curve by numerical integration with trapezoidal rule with 1,000 equal space subintervals, and their confidence intervals based on a normal approximation. In this paper, we will focus on inference about the median sensitivity, median specificity and AUC_M. Besides computational efficiency, we focus on the medians instead of the means because the distributions of these parameters are generally skewed in this context.

3. Two Data Examples

To illustrate the bivariate generalized linear mixed effects model discussed in this article, we apply them to two meta-analysis data sets as follows.

3.1 Example 1: Diagnostic accuracy of FDG-PET for malignant focal pulmonary lesions

Gould et al. [26] presented 40 studies estimating the diagnostic accuracy of positron emission tomography (PET) with the glucose analog 18-fluorodexoxyglucose (FDG) of pulmonary lesions to identify malignant focal pulmonary nodules and mass lesions. FDG-PET is a noninvasive functional imaging test capitalized on the observation that malignant

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

cells have increased rates of glucose metabolism. Among the 40 studies, six studies did not report specificity and three studies examined FDG imaging with a modified gamma camera. To illustrate and compare different models on sensitivity and specificity for FDG-PET, we will exclude these nine studies. Table 1 shows the frequencies of the FDG-PET outcomes based on the final diagnosis of malignant or benign pulmonary nodules or masses, i.e., the number of true positives, false negatives, false positives, and true negatives subjects, for these 31 studies.

We fitted the bivariate generalized linear mixed effects models as described in Section 2 on the data of 31 studies on the diagnostic accuracy of FDG-PET of pulmonary nodules and mass lesions. We assumed a bivariate normal distribution of $(Se_i, 1 - Sp_i)$ on the transformed scale using the logit, probit, and complementary log-log transformation. Since the complementary log-log transformation is asymmetrical, we also fitted the bivariate generalized linear mixed effects models for the pairs of (Se_i, Sp_i), $(1 - Se_i, 1 - Sp_i)$ and $(1 - Se_i, Sp_i)$ using the complementary log-log transformation. Table 2 presents the parameter estimates and their standard errors including the median sensitivity and specificity, and the area under the summary operating characteristics curve (AUC), and the goodness of fit measurement Akaike's Information Criterion (AIC) resulting from the bivariate random effects meta-analysis. From the Table 2, it is clear that the results of median sensitivity and median specificity of different transformations are very similar. The median sensitivity estimates of FDG-PET range from 0.974 to 0.976 and the median specificity estimates of FDG-PET range from 0.780 to 0.787. The AUC_M estimates for the logit, probit and complementary log-log transformations are very similar. The AIC indicates that the model with the best goodness of fit is modeling pair of $(1 - Se_i, 1 - Sp_i)$ using the complementary log-log transformation link function. Comparing the best-fitted complementary log-log transformation to the frequently used logit transformation, the AIC difference is 232.6 – 234.2 = −1.6, which suggests improvement in goodness-of-fit. Arguably, one may want to add some extra penalty when comparing complementary log-log transformation to the other two transformations to account for the fact that a best-out-of-four complementary log-log transformation is used in the comparison. Figure 1a plots the summary receiver operating characteristic curves (ROC) and the boundary of 95% prediction regions, which has a probability of 95% to include the "true" sensitivity and specificity of a future study, based on the logit, probit and best fitted complementary log-log (C-log-log) transformations. For this case study, the ROC curves and the boundaries from three models are very similar.

3.2 Example 2: Diagnostic accuracy of semi-quantitative or quantitative catheter segment culture for intravascular device-related bloodstream infection

To identify the most accurate methods for diagnosis of intravascular devices (IVD)-related bloodstream infection, Safdar et al. [27] studied 8 diagnostic methods and presented 51 studies in a meta-analysis. Safdar et al. found out the most accurate catheter segment culture test was quantitative culture followed by semi-quantitative culture from analyzing 14 studies of quantitative catheter segment culture and 19 studies of semi-quantitative catheter segment culture. To illustrate our methods, we will analyze these 33 studies of semi-quantitative or quantitative catheter segment culture for the diagnosis of IVD-related bloodstream infection. Table 3 shows the frequencies of the catheter segment culture test outcomes based on the final diagnosis of bloodstream infection for these 33 studies where study 1–19 were semi-quantitative catheter segment culture test outcomes and study 20–33 were quantitative catheter segment culture test outcomes. This data example has a larger sample size (mean sample size of 256) comparing to the first data example (mean sample size of 48).

We fitted the bivariate generalized linear mixed effects models as described in Section 2 on the data of 33 studies on the diagnosis of IVD-related bloodstream infection. Similar to Section 3.1, we assumed a bivariate normal distribution of $(Se_i, 1 - Sp_i)$ on the transformed

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

scale using the logit, probit and complementary log-log transformation. Since the complementary log-log transformation is asymmetrical, we also fitted the bivariate generalized linear mixed effects models for pairs of (Se_i, Sp_i), ($1 - Se_i$, $1 - Sp_i$) and ($1 - Se_i$, Sp_i) using the complementary log-log transformation. Since there is no statistically significant difference of sensitivities and specificities between semi-quantitative and quantitative catheter segment culture, and the estimates are very close (not presented), we pool semi-quantitative and quantitative catheter segment culture together in this analysis. Table 4 presents the parameter estimates and their standard errors including the median sensitivity and specificity, and AUC, and AIC resulting from the bivariate random effects meta-analysis. In this data example, the estimates of median sensitivity and median specificity of different transformations are also similar. The median sensitivity estimates range from 0.851 to 0.863 and the median specificity estimates range 0.858 to 0.873. The AUC_M estimates are also very similar with different transformations. The AIC indicated the best goodness of fit model is modeling pair of ($1-Se_i$, Sp_i) using the complementary log-log transformation among the six models we studied. Comparing the best-fitted complementary log-log transformation to the frequently used logit transformation, the AIC difference is 413.6 – 418.3 = − 4.7, suggesting significant improvement of goodness-of-fit. Figure 1b plots the summary receiver operating characteristic curves (ROC) and the boundaries of 95% prediction region based on the logit, probit and best fitted complementary log-log (C-log-log) transformations. For this case study, the ROC curves from three models are very similar, while the boundaries of 95% prediction region are noticeably different, potentially suggesting the importance of selecting an appropriate link function for prediction.

4. Simulation Studies

To study the impact of misspecification of link functions on the estimation of sensitivity, specificity, and the area under the ROC curve (AUC), we performed three sets of simulations with 5000 replicates each. For each replicate, we simulated 40 meta-studies with 100 cases and 100 non-cases per study. In these three sets of simulations, we assumed a bivariate normal distribution for (Se_i, $1 - Sp_i$) in the logit scale, the probit scale, and the complementary log-log scale, respectively, with medians of (Se_i, $1 - Sp_i$) = (0.8, 0.1), a positive correlation coefficient $\rho = 0.5$, and standard deviations (σ_μ, σ_v) = (1.0, 1.0). It corresponds to an expected sensitivity and specificity of (0.761, 0.866), (0.724, 0.818) and (0.736, 0.812) for the logit, probit and complementatary log-log transformations, respectively. The 2.5 and 97.5 percentiles of the distribution correspond to (0.360, 0.966) and (0.015, 0.441) for (Se_i, $1 - Sp_i$) by the logit transformation; (0.249, 0.999) and (0.001, 0.751) by the probit transformation; and (0.203, 1.000) and (0.015, 0.527) by the complementary log-log transformation. Figure 2 presents the true summary ROC curves that are based on the joint distributions of (Se_i, $1 - Sp_i$) for the three link functions considered. By equation (6), the true AUC_M is 0.8990, 0.9076, and 0.9332 for the logit, the probit and the complementary log-log transformations, respectively, with the complementary log-log transformation achieving the largest AUC_M. For each set of simulations, we fit the bivariate generalized linear mixed effect model as described in Section 2 with all the three link functions, and we estimated the back-transformed median sensitivity and median specificity as well as the AUC_M under each model. In addition, AIC was calculated to select among the three random effect models. The 5000 replicates for each set of simulations provides a reasonably small standard error of 0.0031 for the estimation of 95% confidence interval coverage probability [28]. The results were averaged across the 5000 replicates.

Table 5 presents the empirical probabilities of selecting among the three candidate link functions using AIC based on the 5000 replicates. It shows that the probability of the correct selection is 0.59, 0.74, and 0.85 when the true link function is logit, probit, and complementary log-log, respectively. Furthermore, Table 6 presents the empirical

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

probabilities of pairwise incorrect selection using AIC based on the 5000 replicates. For example, it shows that the probability of selecting a probit or a complementary log-log link is 0.34 and 0.19 respectively if the true link function is logit. Based on the limited simulations, it seems that misspecification resulting from AIC-based model selection is more likely to occur when the logit link function is true than when the complementary log-log link function is true. Table 7 presents the average estimates of AUC_M, median sensitivity and median specificity together with their standard errors and the 95% confidence interval coverage probabilities across the 5000 replicates, when using the three link functions. In summary, the estimated AUC_M, and the median sensitivity and the median specificity are nearly unbiased upon misspecification of link functions. It suggests that point estimation of the three quantities is approximately robust to the choice of the link functions. However, the misspecification of link functions has a noticeable impact on the standard error estimation and the 95% confidence interval coverage. Although the confidence interval coverage probabilities are slightly lower than the expected 95% even if the link function is correctly specified, they generally perform well and range from 0.925 to 0.942. However, if the link function is misspecified, a very low coverage probability of 0.67 is observed for median sensitivity if we incorrectly specified a logit link function when the true link function is complementary log-log. It emphasizes the importance to carefully choose an appropriate link function to make statistical inference. On the other hand, when the complementary log-log function is fit to data generated from a logit or probit function, coverage probabilities of the 95% confidence interval for AUC_M are still 0.941 and 0.958, respectively. It suggests that the complementary log-log transformation may be more flexible and robust to misspecification than the logit and probit transformations.

5. Discussion

Performance of a diagnostic test is often measured by paired indices, e.g. sensitivity and specificity, rather than one single summary statistic. Sensitivity and specificity are often jointly modeled in the meta-analysis using random effects models to synthesize the diagnostic test across similar studies. In this article we proposed a bivariate generalized linear mixed effects model for meta-analysis of diagnostic accuracy studies using a general link function including logit transformation as a special case. We fitted the models using the dual quasi-Newton optimization techniques with SAS NLMIXED procedure and provided methods to estimate the median sensitivity, median specificity, to construct summary operating characteristic curve (SROC) and to estimate the area under the SROC.

To our knowledge, only logit transformation has been used for the bivariate random effects meta-analysis on the sensitivity and specificity in the literature. Our contribution in this article has been to extend the transformation of the sensitivity and specificity to a general link function including logit transformation as the special case and compare the performance of parameter estimation and the goodness of fit for the proposed link functions of the bivariate generalized linear mixed effects model. We discussed three link functions, the commonly used logit transformation and two additional link functions of probit transformation and complementary log-log transformation. We proposed to select a link function that can give a better goodness of fit using the AIC. Our data examples illustrated different link function provided different goodness of fit based on the AICs. Furthermore, since the complementary log-log transformation is asymmetrical, it is more flexible than the logit and probit transformations. Specifically, modeling pairs of (Se_i, Sp_i), (Se_i, $1 - Sp_i$), ($1 - Se_i$, Sp_i) or ($1 - Se_i$, $1 - Sp_i$) using complementary log-log transformation may provide different goodness of fit. Our two data examples illustrated a better goodness of fit based on one of the pairs using complementary log-log transformation.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

We evaluated the impact of the misspecification of the link functions on the parameter estimation of the bivariate generalized linear mixed models through a simulation study. Our simulation study indicated that the point estimation of the median sensitivity and specificity, and the AUC were robust to the misspecification of the link functions. But both the standard errors and the 95% confidence interval coverage probabilities were not robust to the misspecification of link functions. We observed a low coverage probability of 67% for the median sensitivity if we incorrectly specify a logit link when the true link function is complementary log-log transformation. However, approximately 95% coverage probability can still be obtained for AUC_M if the complementary log-log link is incorrectly specified when the true link is logit and probit. We also examined the performance of AIC on selecting a candidate link function through a simulation study. Our simulations indicated that the AIC method performed relatively well.

The bivariate generalized linear mixed models we proposed in this article do not include the study-level or individual level covariates. Generalization of extending our models to include such covariates is straightforward through the SAS NLMIXED procedure. An alternative method is to fit those models using Bayesian approaches that can be easily fitted by some free downloadable software such as WinBUGS.

Acknowledgments

Dr. Haitao Chu was supported in part by the Lineberger Cancer Center Core Grant CA16086 from the U.S. National Cancer Institute.

Reference List

1. Zhou, XH.; Obuchowski, NA.; McClish, DK. Statistical methods in diagnostic medicine. New York: John Wiley & Sons; 2002.
2. Pepe, MS. The statistical evaluation of medical tests for classification and prediction. Oxford: Oxford University Press; 2003.
3. Egger, M.; Smith, GD.; Altman, DG. Systematic reviews in health care: meta-analysis in context. 2. BMJ Publishing Group; 2001.
4. Rutter CA, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Statistics in Medicine 2001;20(19):2865–84. [PubMed: 11568945]
5. Song FJ, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. Int J Epidemiol 2002;31(1):88–95. [PubMed: 11914301]
6. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in Medicine 2002;21:589–624. [PubMed: 11836738]
7. Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. Journal of Clinical Epidemiology 2004;57(9):925–32. [PubMed: 15504635]
8. 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):982–90. [PubMed: 16168343]
9. Mallett S, Deeks JJ, Halligan S, Hopewell S, Cornelius V, Altman DG. Systematic reviews of diagnostic tests in cancer: review of methods and reporting. BMJ 2006;333(7565):413. [PubMed: 16849365]
10. 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):1331–2. [PubMed: 17098577]
11. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JAC. A unification of models for meta-analysis of diagnostic accuracy studies. Biostat 2007;8(2):239–51.
12. Zwinderman AH, Bossuyt PM. We should not pool diagnostic likelihood ratios in systematic reviews. Statistics in Medicine. 2007 Ref Type: In Press.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

13. Arends LR, Hamza TH, van Houwelingen JC, Heijenbrok-Kal MH, Hunink MGM, Stijnen T. Bivariate Random Effects Meta-Analysis of ROC Curves. Med Decis Making 2008;28(5):621–38. [PubMed: 18591542]
14. Walter SD. The partial area under the summary ROC curve. Statistics in Medicine 2005;24(13): 2025–40. [PubMed: 15900606]
15. Chu H, Guo H. Letter to the editor: a unification of models for meta-analysis of diagnostic accuracy studies. Biostat 2009;10(1):201–3.
16. Riley R, Abrams K, Sutton A, Lambert P, Thompson J. Bivariate random-effects meta-analysis and the estimation of between-study correlation. BMC Medical Research Methodology 2007;7(1):3. [PubMed: 17222330]
17. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. Biostat 2008;9(1):172–86.
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. Chu H, Nie L, Cole SR, Poole C. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: Alternative parameterizations and model selection. Statistics in Medicine 2009;28(18):2384–99. [PubMed: 19499551]
20. Walter SD, Irwig L, Glasziou PP. Meta-analysis of diagnostic tests with imperfect reference standards. Journal of Clinical Epidemiology 1999;52(10):943–51. [PubMed: 10513757]
21. Chu H, Chen S, Louis TA. Random Effects Models in a Meta-Analysis of the Accuracy of Two Diagnostic Tests without a Gold Standard. Journal of the American Statistical Association 2009;104:512–23. [PubMed: 19562044]
22. Sutton AJ, Cooper NJ, Goodacre S, Stevenson M. Integration of Meta-analysis and Economic Decision Modeling for Evaluating Diagnostic Tests. Med Decis Making 2008;28(5):650–67. [PubMed: 18753686]
23. Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. Statistics in Medicine 2002;21(9):1237–56. [PubMed: 12111876]
24. Burnham, KP.; Anderson, DR. Model Selection and Inference: A Practical Information-Theoretic Approach. New York: Springer-Verlag; 1998.
25. Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. Journal of Computational and Graphical Statistics 1995;4(1):12–35.
26. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions - A meta-analysis. Jama-Journal of the American Medical Association 2001;285(7):914–24.
27. Safdar N, Fine JP, Maki DG. Meta-Analysis: Methods for Diagnosing Intravascular Device-Related Bloodstream Infection. Ann Intern Med 2005;142(6):451–66. [PubMed: 15767623]
28. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. Statistics in Medicine 2006;25(24):4279–92. [PubMed: 16947139]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

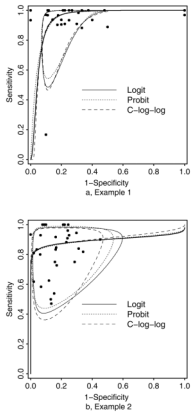


Figure 1.
Summary receiver operating characteristic curves and the boundaries of 95% prediction region in the conventional ROC space based on the logit, probit and the best-fitted complementary log-log bivariate generalized linear mixed models for two case studies.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

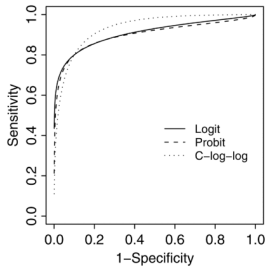


Figure 2.
The summary receiver operating characteristic curves in the conventional ROC space based on the true parameters for the simulation studies.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Example 1: Data from a meta-analysis of studies on the accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions [26].

Study	True Positive	False Negative	False Positive	True Negative
1	2	10	1	9
2	12	0	0	7
3	19	2	2	8
4	33	0	2	16
5	29	2	2	3
6	59	2	3	24
7	44	3	3	12
8	22	0	2	7
9	57	2	5	23
10	34	0	4	15
11	18	0	2	4
12	33	0	2	15
13	30	3	2	7
14	26	0	7	19
15	29	0	6	10
16	82	0	12	13
17	17	0	0	9
18	30	2	2	12
19	40	4	3	7
20	12	0	0	7
21	59	1	9	20
22	14	1	1	3
23	15	2	2	4
24	28	3	4	19
25	14	1	0	2
26	8	1	1	1
27	24	0	1	2
28	64	2	3	27
29	43	0	3	9
30	91	3	3	0
31	37	0	4	0

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

First Data Example: a meta-analysis of studies on the accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions

	Logit (Se & 1−Sp)	Probit (Se & 1−Sp)	Complementary log-log			
			Se & 1−Sp	Se & Sp	1−Se & 1−Sp	1−Se & Sp
AIC	234.2	236.7	239.9	240.6	**232.6**	233.4
μ_0	3.682 (0.381)	1.969 (0.185)	1.298 (0.128)	1.300 (0.128)	**−3.656 (0.352)**	−3.663 (0.353)
ν_0	−1.298 (0.156)	−0.783 (0.090)	−1.430 (0.143)	0.414 (0.078)	**−1.432 (0.143)**	0.415 (0.078)
σ_μ	1.477 (0.341)	0.738 (0.175)	0.536 (0.139)	0.538 (0.140)	**1.338 (0.297)**	1.345 (0.298)
σ_ν	0.474 (0.205)	0.266 (0.115)	0.435 (0.187)	0.226 (0.101)	**0.430 (0.187)**	0.224 (0.102)
ρ	0.662 (0.477)	0.663 (0.462)	0.603 (0.461)	−0.609 (0.469)	**−0.664 (0482)**	0.679 (0.495)
Median Se	0.975 (0.009)	0.976 (0.011)	0.974 (0.012)	0.975 (0.012)	**0.975 (0.009)**	0.975 (0.009)
Median Sp	0.786 (0.027)	0.783 (0.026)	0.787 (0.027)	0.780 (0.026)	**0.788 (0.027)**	0.780 (0.026)
AUC_M	0.940 (0.061)	0.948 (0.058)	0.953 (0.045)	0.961 (0.044)	**0.936 (0.063)**	0.947 (0.077)

Med Decis Making. Author manuscript; available in PMC 2011 July 1.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Example 2: Data from a meta-analysis of studies on semi-quantitative (the first nineteen studies) or quantitative (the last fourteen studies) catheter segment culture for diagnosis of intravascular device-related bloodstream infection [27].

Study	True Positive	False Negative	False Positive	True Negative
1	12	0	29	289
2	10	2	14	72
3	17	1	36	85
4	13	0	18	67
5	4	0	21	225
6	15	2	122	403
7	45	5	28	34
8	18	4	69	133
9	5	0	11	34
10	8	9	15	96
11	5	0	7	63
12	11	2	122	610
13	5	1	6	145
14	7	5	25	342
15	10	1	93	296
16	5	5	41	271
17	5	0	15	53
18	55	13	19	913
19	6	2	12	30
20	42	26	19	913
21	5	3	5	37
22	13	0	11	125
23	20	0	24	287
24	7	6	13	72
25	48	2	15	47
26	11	1	14	72
27	15	5	32	170
28	68	13	5	11
29	13	1	5	72
30	8	3	66	323
31	13	1	98	293
32	14	1	0	155
33	8	2	4	60

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 4

Second Data Example: a meta-analysis of studies on semi-quantitative or quantitative catheter segment culture for diagnosis of intravascular device-related bloodstream infection.

	Logit (Se & 1−Sp)	Probit (Se & 1−Sp)	Complementary log-log			
			Se & 1−Sp	Se & Sp	1−Se & 1−Sp	1−Se & Sp
AIC	418.3	415.9	420.5	413.8	420.0	**413.6**
μ_0	1.829 (0.222)	1.068 (0.119)	0.645 (0.098)	0.645 (0.099)	−1.916 (0.199)	**−1.912 (0.199)**
ν_0	−1.906 (0.169)	−1.103 (0.088)	−1.999 (0.157)	0.669 (0.070)	−1.999 (0.157)	**0.669 (0.070)**
σ_μ	0.876 (0.213)	0.485 (0.115)	0.410 (0.098)	0.417 (0.010)	0.764 (0.182)	**0.763 (0.181)**
σ_ν	0.923 (0.132)	0.482 (0.068)	0.855 (0.122)	0.384 (0.054)	0.854 (0.122)	**0.383 (0.054)**
ρ	0.212 (0.225)	0.204 (0.226)	0.202 (0.223)	−0.190 (0.225)	−0.227 (0.224)	**0.220 (0.231)**
Median Se	0.862 (0.026)	0.857 (0.027)	0.851 (0.028)	0.851 (0.028)	0.863 (0.025)	**0.863 (0.025)**
Median Sp	0.871 (0.019)	0.864 (0.019)	0.873 (0.019)	0.858 (0.020)	0.873 (0.019)	**0.858 (0.020)**
AUC_M	0.897 (0.034)	0.898 (0.039)	0.886 (0.040)	0.905 (0.046)	0.892 (0.030)	**0.908 (0.034)**

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 5

The empirical probability of selecting a candidate among the three candidate link functions using AIC based on simulation studies with 5000 replicates. The bolded cells represent the probability of identifying the correct model.

Selected Random Effects Model	True Random Effects Model		
	Logit	Probit	Complementary log-log
Logit	**0.590**	0.177	0.028
Probit	0.279	**0.740**	0.120
Complementary log-log	0.131	0.083	**0.852**

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 6

The empirical probability of pairwise incorrect selection using AIC based on simulation studies with 5000 replicates.

Selected Random Effects Model	True Random Effects Model		
	Logit	Probit	Complementary log-log
Logit	—	0.188	0.069
Probit	0.344	—	0.142
Complementary log-log	0.189	0.092	—

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 7

The estimation and coverage performance of each link function based on simulation studies with 5000 replicates. The bolded cells represent the correctly chosen model.

Fitted Random Effects Model		True Random Effects Model: Logit			True Random Effects Model: Probit			True Random Effects Model: Complementary log-log		
		AUC (0.8990)	Se (0.80)	Sp (0.90)	AUC (0.9076)	Se (0.80)	Sp (0.90)	AUC (0.9332)	Se (0.80)	Sp (0.90)
Logit	Mean	**0.8950**	**0.7986**	**0.8989**	0.9070	0.8143	0.9091	0.9343	0.8513	0.8961
	Standard Error	**0.0206**	**0.0259**	**0.0151**	0.0247	0.0433	0.0254	0.0169	0.0409	0.0169
	95% CICP*	**0.9245**	**0.936**	**0.9421**	0.9002	0.8803	0.8684	0.8736	0.6659	0.9505
Probit	Mean	0.9002	0.7913	0.8939	**0.9026**	**0.7972**	**0.8972**	0.9344	0.8327	0.8901
	Standard Error	0.0222	0.0262	0.0159	**0.0273**	**0.0443**	**0.0286**	0.0184	0.0435	0.0180
	95% CICP*	0.8862	0.9343	0.9441	**0.9285**	**0.9309**	**0.9307**	0.8785	0.8151	0.9452
Complementary log-log	Mean	0.8910	0.7774	0.9010	0.8903	0.7585	0.9149	**0.9259**	**0.7980**	**0.8989**
	Standard Error	0.0270	0.0278	0.0145	0.0363	0.0517	0.0225	**0.0232**	**0.0507**	**0.0159**
	95% CICP*	0.9412	0.8881	0.9325	0.9576	0.8986	0.8051	**0.9393**	**0.9293**	**0.9415**

*95% CICP = 95% confidence interval coverage probability based on normal assumption.