Published in final edited form as:

Biom J. 2014 March ; 56(2): 270–286. doi:10.1002/bimj.201200012.

Confidence interval estimation of the difference between two sensitivities to the early disease stage

Tuochuan Dong¹, Le Kang¹, Alan Hutson¹, Chengjie Xiong², and Lili Tian^{1,*}

¹Department of Biostatistics, University at Buffalo, Buffalo, NY 14214, USA

²Division of Biostatistics, Washington University at St. Louis, St. Louis, MO 63110, USA

Abstract

Although most of the statistical methods for diagnostic studies focus on disease processes with binary disease status, many diseases can be naturally classified into three ordinal diagnostic categories, that is normal, early stage, and fully diseased. For such diseases, the volume under the ROC surface (VUS) is the most commonly used index of diagnostic accuracy. Because the early disease stage is most likely the optimal time window for therapeutic intervention, the sensitivity to the early diseased stage has been suggested as another diagnostic measure. For the purpose of comparing the diagnostic abilities on early disease detection between two markers, it is of interest to estimate the confidence interval of the difference between sensitivities to the early diseased stage. In this paper, we present both parametric and non-parametric methods for this purpose. An extensive simulation study is carried out for a variety of settings for the purpose of evaluating and comparing the performance of the proposed methods. A real example of Alzheimer's disease (AD) is analyzed using the proposed approaches.

Keywords

Box-Cox transformation; Diagnostic studies; Disease with three ordinal stages; Generalized inference; Sensitivity to the early diseased stage

1 Introduction

In diagnostic studies, the probability of correctly identifying a healthy subject as negative is called specificity and the probability of correctly diagnosing a diseased subject as positive is called sensitivity. For a marker measured on continuous scale, the ROC (receiver operating characteristic) curve is a plot of the sensitivity versus (1 – specificity) over all possible threshold values. ROC is a popular way to graphically display the discriminatory accuracy of a marker for distinguishing between two populations. To evaluate the diagnostic accuracy of a marker, researchers usually summarize the information of a ROC curve by indexes,

Conflict of interest

The authors have declared no conflict of interest.

such as area under curve (AUC), partial area under curve (PAUC) or sensitivity at a given specificity.

In reality, a disease process might be classified into three stages rather than two: the normal healthy stage without even the earliest subtle disease symptoms, the early stage of the disease, and the stage of fully blown development of the disease. One of the examples for such a case is the diagnosis of Alzheimer's disease (AD). The AD process begins with a presymptomatic or preclinical stage during which either no or very few clinical symptoms can be detected to subtle decrements in performance. As individuals approach the threshold of dementia, they begin to exhibit subtle signs of functional and cognitive deterioration, suggestive of early-stage clinical dementia. Hence, mild cognitive impairment (MCI) and/or early-stage Alzheimer's disease is a transitional stage between the cognitive changes of normal aging and the more serious problems. More background information can be found from Katzman et al. (1988), Crystall et al. (1988), Morris et al. (1991), and Xiong et al. (2006). To be specific, let X , Y , and Z denote the results of a diagnostic test or measurements of a biomarker and F_x , F_y , and F_z denote the corresponding cumulative distribution functions for the subjects at healthy stage, early disease stage, and the fully diseased stage, respectively. Assume the results are measured on a continuous scale and that higher values indicate greater severity of the disease. Let c_1 and c_3 denote threshold values ($c_1 < c_3$) for classifying a subject into the nondiseased ($< c_1$) group, early-stage group (between c_1 and c_3) and the fully diseased stage group ($> c_3$). Define $P_1 = F_x(c_1)$, $P_3 = 1 - F_z(c_3)$, and define

$$P_2 = P(c_1 < Y < c_3) = F_y(c_3) - F_y(c_1) = F_y[F_z^{-1}(1 - P_3)] - F_y[F_x^{-1}(P_1)]. \quad (1)$$

Obviously, P_1 is the correct classification rate for the healthy group (i.e., specificity), P_3 is the correct classification rate for the fully diseased group (i.e., sensitivity to the fully diseased stage) and P_2 is the correct classification rate for the early diseased group, that is the probability that a randomly selected subject from the early diseased stage group has a test result between c_1 and c_3 (i.e., sensitivity to early disease stage). As a function of P_1 and P_3 , $P_2 = P_2(P_1, P_3)$ defines a surface in the three-dimensional space (P_1, P_2, P_3) , that is the ROC surface. For such diseases, the volume under ROC surface (VUS) and the partial volume under surface (PVUS) have been widely used as a summary measure of the diagnostic accuracy, parallel to AUC and PAUC for diseases with binary status.

Most of the existing statistical methods in diagnostic studies were developed for the diseases with binary status. Some of the existing statistical methods for diseases with binary status have been extended to diseases with three ordinal stages; for example Mossman (1999), Dreiseitl (2000), Heckerling (2001), Nakas and Yiannoutsos (2004), Xiong et al. (2006), and Tian et al. (2010).

For disease processes with three ordinal stages, the early stage is usually considered the most critical for intervention and therefore it is crucial to correctly identify patients in this stage. For example, no pharmaceutical treatments to-date are effective for the late stage of AD and most of the existing medicines are only devoted to slow down the development of the disease. Hence the transitional stage (i.e., the early stage) of the AD process is of paramount importance to detect not only because it is most likely the best time window to

treat the symptoms of the disease but also because this is the time for family caregivers to discuss care options and family future planning while the patient can still take part in making decisions. Furthermore, early diagnosis may also lead to early treatment with some of the new drugs that have proven effective in slowing the progression of memory loss and improving the quality of life for many patients. However, because of the subtle clinical symptoms in the early stage of many complex disease processes such as AD and the fact that no single test can determine whether a person has the disease, diagnosing someone in the early stage of a disease is often challenging. More information can be found in Katzman et al. (1988), Crystall et al. (1988), Morris et al. (1991), and Xiong et al. (2006). Overall speaking, for a specific biomarker or a diagnostic test, its ability to correctly identify the patients in the early stage of the disease, that is P_2 , given the specificity P_1 and the sensitivity for the fully diseased stage P_3 , is a key measure of diagnostic accuracy for early disease stage diagnosis. Most recently, Dong et al. (2012) proposed both parametric and nonparametric approaches for the confidence interval estimation of P_2 given specificity and sensitivity for the fully diseased stage.

When there exist multiple biomarkers, investigators often need to compare the diagnostic accuracy between biomarkers. For disease processes with two stages, such comparison is frequently addressed by comparing the resulting paired AUCs or PAUCs. Additionally, there also exist approaches for comparing two sensitivities at a fixed level of specificity. For example, Greenhouse and Mantel (1950) and Linnet (1987) compared two sensitivities at a fixed level of specificity in a nonparametric approach, and Zhou and Qin (2006) developed several methods for estimating the confidence intervals for the difference of two sensitivities at a fixed level of specificity. For disease processes with three ordinal stages, the comparison of the diagnostic accuracy between two biomarkers can be addressed by comparing the resulting paired VUSs or PVUSs. For example, Xiong et al. (2007) developed a large sample approach for comparing several VUSs for normally distributed data, and Tian et al. (2010) proposed an approach based on generalized inference for confidence interval estimation of the difference between paired VUSs and PVUSs. However, for disease processes with three ordinal stages, the problem of comparing sensitivities to the early disease stage between two biomarkers or tests has not been addressed. Therefore, in this paper, we will present parametric and nonparametric confidence intervals of the difference between the P_2 s for two biomarkers given specificity P_1 and sensitivity to the fully diseased stage P_3 .

This paper is organized as follows. In Section 2, we propose parametric confidence interval estimations for the difference of the P_2 s. In Section 3, nonparametric confidence intervals are presented. In Section 4, we conduct extensive simulation studies to assess the finite sample performance of the proposed confidence intervals. In Section 5, we analyze the data from an Alzheimer's Disease study. Section 6 presents a discussion.

2 Parametric approaches

Under the assumption that two markers have a joint bivariate normal distribution, we propose a method for confidence interval estimation for the difference between two paired sensitivities to the early-stage disease based on generalized inference. When normality is not

satisfied, confidence intervals based on generalized inference will be obtained using Box-Cox transformed data.

2.1 Under the bivariate normal assumptions

Under the bivariate normality assumption, we employ the generalized inference approach to calculate the confidence interval for the difference of P_2 s for two diagnostic tests. The generalized variables and generalized pivots were introduced by Tsui and Weerahandi (1989) and Weerahandi (1993); see the book by Weerahandi (2003) for a detailed discussion. The concepts of generalized confidence interval and generalized p -value have been successfully applied to a variety of practical settings where standard exact solutions do not exist for confidence intervals and hypothesis testing. It has been shown that generalized inference approaches generally have good performance, even at small sample sizes (see e.g., Weerahandi, 1995; Weerahandi and Berger, 1999; Krishnamoorthy and Lu, 2003; Tian and Cappelleri, 2004; Tian, 2008; Li et al., 2008a, 2008b).

Consider the case with two biomarkers (1 and 2) that are measured on each subject. Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ denote the measurements for the non-diseased, early stage and diseased subjects, respectively, and assume all of them follow bivariate normal distributions. Specifically,

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(\boldsymbol{\mu}_x, \Sigma_x),$$

where X_1 is the test result for marker 1 and X_2 for marker 2 and

$$\boldsymbol{\mu}_x = \begin{pmatrix} \mu_{x1} \\ \mu_{x2} \end{pmatrix} \text{ and } \Sigma_x = \begin{pmatrix} \sigma_{x11}^2 & \sigma_{x12} \\ \sigma_{x21} & \sigma_{x22}^2 \end{pmatrix}.$$

For the early disease stage and fully diseased stage, $\mathbf{Y}$, $\boldsymbol{\mu}_y$, Σ_y , and $\mathbf{Z}$, $\boldsymbol{\mu}_z$, Σ_z are defined similarly.

Under the bivariate normality assumption, the sensitivity to the early disease stage for marker 1 can be written as:

$$P_2^{(1)} = \Phi \left[\frac{\mu_{z1} + \Phi^{-1}(1 - P_3)\sigma_{z11} - \mu_{y1}}{\sigma_{y11}} \right] - \Phi \left[\frac{\mu_{x1} + \Phi^{-1}(P_1)\sigma_{x11} - \mu_{y1}}{\sigma_{y11}} \right],$$

and similarly for marker 2:

$$P_2^{(2)} = \Phi \left[\frac{\mu_{z2} + \Phi^{-1}(1 - P_3)\sigma_{z22} - \mu_{y2}}{\sigma_{y22}} \right] - \Phi \left[\frac{\mu_{x2} + \Phi^{-1}(P_1)\sigma_{x22} - \mu_{y2}}{\sigma_{y22}} \right],$$

where Φ is the cumulative distribution function for the standard normal. The difference between $P_2^{(1)}$ and $P_2^{(2)}$ is the parameter of interest:

$$\Delta = P_2^{(1)} - P_2^{(2)}. \quad (2)$$

Let $\bar{\mathbf{X}}$, $\bar{\mathbf{Y}}$, and $\bar{\mathbf{Z}}$ denote the sample mean vectors, S_x , S_y , and S_z denote the corresponding sample covariance matrices, and n_1 , n_2 , and n_3 denote the sample sizes for nondiseased, early stage, and diseased groups, respectively. Define

$$\mathbf{T}_x = \frac{n_1 - 1}{n_1} S_x, \mathbf{T}_y = \frac{n_2 - 1}{n_2} S_y, \mathbf{T}_z = \frac{n_3 - 1}{n_3} S_z.$$

Let $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, $\mathbf{t}_x$, $\mathbf{t}_y$, and $\mathbf{t}_z$ denote the corresponding observed values of the above statistics. Then according to Lin et al. (2007), the generalized pivots for the multivariate normal means are:

$$\mathbf{R}_{\mu_x} = \bar{\mathbf{x}} - (\mathbf{t}_x^{1/2} \mathbf{W}_{x1}^{-1} \mathbf{t}_x^{1/2})^{1/2} \mathbf{U}_x,$$

$$\mathbf{R}_{\mu_y} = \bar{\mathbf{y}} - (\mathbf{t}_y^{1/2} \mathbf{W}_{y1}^{-1} \mathbf{t}_y^{1/2})^{1/2} \mathbf{U}_y,$$

$$\mathbf{R}_{\mu_z} = \bar{\mathbf{z}} - (\mathbf{t}_z^{1/2} \mathbf{W}_{z1}^{-1} \mathbf{t}_z^{1/2})^{1/2} \mathbf{U}_z$$

and the generalized pivots for the covariance matrices are:

$$\mathbf{R}_{\Sigma_x} = n_1 \mathbf{t}_x^{1/2} \mathbf{W}_{x2}^{-1} \mathbf{t}_x^{1/2}, \mathbf{R}_{\Sigma_y} = n_2 \mathbf{t}_y^{1/2} \mathbf{W}_{y2}^{-1} \mathbf{t}_y^{1/2}, \mathbf{R}_{\Sigma_z} = n_3 \mathbf{t}_z^{1/2} \mathbf{W}_{z2}^{-1} \mathbf{t}_z^{1/2}, \quad (3)$$

where $\mathbf{U}_x, \mathbf{U}_y, \mathbf{U}_z \sim N_2(\mathbf{0}, \mathbf{I}_2)$, $\mathbf{W}_{x1}, \mathbf{W}_{x2} \sim \text{Wishart}_2(n_1 - 1, \mathbf{I}_2)$, $\mathbf{W}_{y1}, \mathbf{W}_{y2} \sim \text{Wishart}_2(n_2 - 1, \mathbf{I}_2)$ and $\mathbf{W}_{z1}, \mathbf{W}_{z2} \sim \text{Wishart}_2(n_3 - 1, \mathbf{I}_2)$ with $\mathbf{I}_2$ being the 2×2 identity matrix. Furthermore,

$$\mathbf{R}_{\mu_x} = \begin{pmatrix} R_{\mu_{x1}} \\ R_{\mu_{x2}} \end{pmatrix}, \mathbf{R}_{\mu_y} = \begin{pmatrix} R_{\mu_{y1}} \\ R_{\mu_{y2}} \end{pmatrix}, \mathbf{R}_{\mu_z} = \begin{pmatrix} R_{\mu_{z1}} \\ R_{\mu_{z2}} \end{pmatrix},$$

and

$$\mathbf{R}_{\Sigma_x} = \begin{pmatrix} R_{\sigma_{x11}} & R_{\sigma_{x12}} \\ R_{\sigma_{x21}} & R_{\sigma_{x22}} \end{pmatrix}, \mathbf{R}_{\Sigma_y} = \begin{pmatrix} R_{\sigma_{y11}} & R_{\sigma_{y12}} \\ R_{\sigma_{y21}} & R_{\sigma_{y22}} \end{pmatrix}, \mathbf{R}_{\Sigma_z} = \begin{pmatrix} R_{\sigma_{z11}} & R_{\sigma_{z12}} \\ R_{\sigma_{z21}} & R_{\sigma_{z22}} \end{pmatrix}.$$

Finally, the generalized pivotal quantity for $P_2^{(1)}$ can be obtained as:

$$R_{P_2^{(1)}} = \Phi \left[\frac{R_{\mu_{z1}} + \Phi^{-1}(1 - P_3) \sqrt{R_{\sigma_{z11}}} - R_{\mu_{y1}}}{\sqrt{R_{\sigma_{y11}}}} \right] - \Phi \left[\frac{R_{\mu_{x1}} + \Phi^{-1}(P_1) \sqrt{R_{\sigma_{x11}}} - R_{\mu_{y1}}}{\sqrt{R_{\sigma_{y11}}}} \right]. \quad (4)$$

Similarly, the generalized pivotal quantity for $P_2^{(2)}$ is

$$R_{P_2^{(2)}} = \Phi \left[\frac{R_{\mu_{z2}} + \Phi^{-1}(1 - P_3) \sqrt{R_{\sigma_{z22}}} - R_{\mu_{y2}}}{\sqrt{R_{\sigma_{y22}}}} \right] - \Phi \left[\frac{R_{\mu_{x2}} + \Phi^{-1}(P_1) \sqrt{R_{\sigma_{x22}}} - R_{\mu_{y2}}}{\sqrt{R_{\sigma_{y22}}}} \right]. \quad (5)$$

Then the generalized pivotal quantity for the difference is

$$R_{\Delta} = R_{P_2^{(1)}} - R_{P_2^{(2)}}. \quad (6)$$

One can easily check that R is a bona fide generalized pivot given the following: 1) the distributions of R is independent of any unknown parameters; and 2) the observed value of R equals to Δ as defined in (2) for given $\mathbf{x}, \bar{\mathbf{y}}, \mathbf{z},$ and $\mathbf{t}_x, \mathbf{t}_y, \mathbf{t}_z$.

Computing algorithm—Given a dataset consisting of $X_{i1}, \dots, X_{in_1}$ from the nondiseased group, $Y_{i1}, \dots, Y_{in_2}$ from the early diseased stage subjects, and $Z_{i1}, \dots, Z_{in_3}$ from the fully diseased subjects where $i = 1, 2$ representing the results for marker 1 and 2, respectively, the confidence interval for Δ in (2) using generalized inference approach can be obtained by the following steps:

1. Calculate $\mathbf{x}, \bar{\mathbf{y}},$ and $\mathbf{z}, \mathbf{t}_x, \mathbf{t}_y, \mathbf{t}_z$ as defined above.
2. Generate $\mathbf{U}_x, \mathbf{U}_y, \mathbf{U}_z$ from standard bivariate normal distributions $N_2(0, \mathbf{I}_2)$, and $\mathbf{W}_{x1}, \mathbf{W}_{x2} \sim \text{Wishart}_2(n_1 - 1, \mathbf{I}_2)$, $\mathbf{W}_{y1}, \mathbf{W}_{y2} \sim \text{Wishart}_2(n_2 - 1, \mathbf{I}_2)$, and $\mathbf{W}_{z1}, \mathbf{W}_{z2} \sim \text{Wishart}_2(n_3 - 1, \mathbf{I}_2)$.
3. Calculate R as in (6).
4. Repeat Steps 2–3 for a total of B times to obtain a set of $R_{\Delta}^b, b = 1, 2, \dots, B$. In this paper, we use $B = 2000$. The $(1 - \alpha)100\%$ level generalized Inference **GI** interval for Δ is:

$$(R_{\Delta, \alpha/2}^b, R_{\Delta, 1-\alpha/2}^b)$$

where $R_{\Delta, \alpha/2}^b, R_{\Delta, 1-\alpha/2}^b$ are the $100(\alpha/2)$ th and $100(1 - \alpha/2)$ th percentile of R_{Δ}^b .

2.2 Without the bivariate normal assumptions

When the bivariate normal assumptions are not satisfied, the above introduced generalized inference approach can not be directly applied to the data. For such data, due to the fact that P_2 is invariant under monotonic transformations, Box-Cox transformation can be applied first to transform the data into bivariate normal distribution, and then the normality-based methods proposed in Section 2.1 for confidence interval estimation can be used. This type of

approaches have been found useful in the context of ROC analysis for a wide variety of scenarios (e.g., Zou et al., 1997; Zou and Hall, 2000; Faraggi and Reiser, 2002; Fluss et al., 2005; Schisterman et al., 2004, 2006; Molodianovitch et al., 2006; Dong et al., 2012).

Specifically, for the i th ($i = 1, 2$) marker, the Box-Cox transformation is defined as:

$$X_i^{(\lambda_i)} = \begin{cases} \frac{X_i^{\lambda_i} - 1}{\lambda_i} & \lambda_i \neq 0 \\ \log(X_i) & \lambda_i = 0. \end{cases}$$

It is assumed that $(X_1^{(\lambda_1)}, X_2^{(\lambda_2)}) \sim N_2(\mu_x, \Sigma_x)$, and

$$\mu_x = \begin{pmatrix} \mu_{x1} \\ \mu_{x2} \end{pmatrix} \text{ and } \Sigma_x = \begin{pmatrix} \sigma_{x11}^2 & \rho_x \sigma_{x11} & \sigma_{x12} \\ \rho_x \sigma_{x11} & \sigma_{x12} & \sigma_{x22}^2 \end{pmatrix}.$$

$Y_i^{(\lambda_i)}$ and $Z_i^{(\lambda_i)}$ ($i = 1, 2$) are defined similarly. The log-likelihood function can be written as:

$$\begin{aligned} & -\frac{n_1}{2} \log(1 - \rho_x^2) - \frac{1}{2(1 - \rho_x^2)} \sum_{i=1}^{n_1} \left(\left(\frac{x_{1i}^{(\lambda_1)} - \mu_{x1}}{\sqrt{\sigma_{x11}}} \right)^2 \right. \\ & \quad - 2\rho_x \left(\frac{x_{1i}^{(\lambda_1)} - \mu_{x1}}{\sqrt{\sigma_{x11}}} \right) \left(\frac{x_{2i}^{(\lambda_2)} - \mu_{x2}}{\sqrt{\sigma_{x22}}} \right) \\ & \quad \left. + \left(\frac{x_{2i}^{(\lambda_2)} - \mu_{x2}}{\sqrt{\sigma_{x22}}} \right)^2 \right) + (\lambda_1 - 1) \sum_{i=1}^{n_1} \log(x_{1i}) \\ & \quad + (\lambda_2 - 1) \sum_{i=1}^{n_1} \log(x_{2i}) \\ & \quad - n_1 \log \left(2\pi \sqrt{\sigma_{x11} \sigma_{x22}} \right) \\ & -\frac{n_2}{2} \log(1 - \rho_y^2) - \frac{1}{2(1 - \rho_y^2)} \sum_{i=1}^{n_2} \left(\left(\frac{y_{1i}^{(\lambda_1)} - \mu_{y1}}{\sqrt{\sigma_{y11}}} \right)^2 \right. \\ & \quad - 2\rho_y \left(\frac{y_{1i}^{(\lambda_1)} - \mu_{y1}}{\sqrt{\sigma_{y11}}} \right) \left(\frac{y_{2i}^{(\lambda_2)} - \mu_{y2}}{\sqrt{\sigma_{y22}}} \right) \\ & \quad \left. + \left(\frac{y_{2i}^{(\lambda_2)} - \mu_{y2}}{\sqrt{\sigma_{y22}}} \right)^2 \right) + (\lambda_1 - 1) \sum_{i=1}^{n_2} \log(y_{1i}) \\ & \quad + (\lambda_2 - 1) \sum_{i=1}^{n_2} \log(y_{2i}) \\ & \quad - n_2 \log \left(2\pi \sqrt{\sigma_{y11} \sigma_{y22}} \right) \end{aligned}$$

$$\begin{aligned}
& -\frac{n_3}{2} \log(1 - \rho_z^2) - \frac{1}{2(1 - \rho_z^2)} \sum_{i=1}^{n_3} \left(\left(\frac{z_{1i}^{(\lambda_1)} - \mu_{z_1}}{\sqrt{\sigma_{z_{11}}}} \right)^2 \right. \\
& \quad - 2\rho_z \left(\frac{z_{1i}^{(\lambda_1)} - \mu_{z_1}}{\sqrt{\sigma_{z_{11}}}} \right) \left(\frac{z_{2i}^{(\lambda_2)} - \mu_{z_2}}{\sqrt{\sigma_{z_{22}}}} \right) \\
& \quad \left. + \left(\frac{z_{2i}^{(\lambda_2)} - \mu_{z_2}}{\sqrt{\sigma_{z_{22}}}} \right)^2 \right) + (\lambda_1 - 1) \sum_{i=1}^{n_3} \log(z_{1i}) \\
& \quad + (\lambda_2 - 1) \sum_{i=1}^{n_3} \log(z_{2i}) \\
& \quad - n_3 \log \left(2\pi \sqrt{\sigma_{z_{11}} \sigma_{z_{22}}} \right). \tag{7}
\end{aligned}$$

The maximum likelihood estimates λ_1, λ_2 can be obtained by maximizing the above function. After applying the Box-Cox transformation, the generalized inference approach proposed in Section 2.1 is applied to the transformed data. This confidence interval is called **BCGI** (Box-Cox generalized inference) interval.

3 Nonparametric confidence interval estimation

The parametric approaches presented above rely on the bivariate normal assumptions and/or involve solving for the MLEs of the transformation parameters. Furthermore, not all the data can be transformed to normal distribution through Box-Cox transformation. Hence, it is worthwhile to explore the non-parametric approaches for confidence interval estimation of $\Delta = P_2^{(1)} - P_2^{(2)}$.

First assume the cumulative distributions for the nondiseased group ($F_X^{(1)}$ for marker 1, $F_X^{(2)}$ for marker 2) and the fully diseased group ($F_Z^{(1)}$ for marker 1, $F_Z^{(2)}$ for marker 2) are known. For the j th marker and i th subject in the early-stage group, define

$A_i^{(j)} = I_{[[F_X^{(j)}]^{-1}(P_1) \leq Y_{1i} \leq [F_Z^{(j)}]^{-1}(1-P_3)]}$. Therefore $A_i^{(j)}$'s ($i = 1, 2, \dots, n_2$) are Bernoulli random variables with the probability of success as $P_2^{(j)} = P[[F_X^{(j)}]^{-1}(P_1) \leq Y_j \leq [F_Z^{(j)}]^{-1}(1-P_3)]$ where $[F_X^{(j)}]^{-1}(P_1)$ and $[F_Z^{(j)}]^{-1}(1-P_3)$ stand for the quantiles for $X^{(j)}$ at P_1 and for $Z^{(j)}$ at P_2 respectively. For the j th marker, the maximum likelihood estimator (MLE) of $P_2^{(j)}$ is:

$$\bar{P}_2^{(j)} = \sum_{i=1}^{n_2} A_i^{(j)} / n_2. \tag{8}$$

Agresti and Coull (1998) proposed a better estimation for binomial proportions:

$$\bar{P}_2^{AC(j)} = \frac{\sum_{i=1}^{n_2} I_{[[F_X^{(j)}]^{-1}(P_1) \leq Y_{1i} \leq [F_Z^{(j)}]^{-1}(1-P_3)]} + z_{1-\alpha/2}^2/2}{n_2 + z_{1-\alpha/2}^2}. \tag{9}$$

Given $\alpha = 0.05$, $z_{1-\alpha/2}$ approximately equals to 2. Hence the above estimator can be interpreted as the adjusted Wald estimator by adding two successes and two failures to Bernoulli observations. Agresti and Coull (1998) recommended this method because their simulation study has shown that the confidence interval derived by this estimator has good coverage accuracy even for small sample sizes. Later on, Zhou and Qin (2006) proposed a refined version of Agresti and Coull's estimator for paired samples:

$$\bar{P}_2^{ZQ(j)} = \frac{\sum_{i=1}^{n_2} I_{[[F_X^{(j)}]^{-1}(P_1) \leq Y_{1i} \leq [F_Z^{(j)}]^{-1}(1-P_3)]} + 1}{n_2 + 2}. \quad (10)$$

In reality, the true distributions $F_X^{(j)}$ and $F_Z^{(j)}$ are unknown, therefore the sample quantiles $[\hat{F}_X^{(j)}]^{-1}(P_1)$ and $[\hat{F}_Z^{(j)}]^{-1}(1-P_3)$ are used instead. Hence, we have

$$\hat{P}_2^{(j)} = \frac{\sum_{i=1}^{n_2} I_{[[\hat{F}_X^{(j)}]^{-1}(P_1) \leq Y_{1i} \leq [\hat{F}_Z^{(j)}]^{-1}(1-P_3)]}}{n_2},$$

$$\hat{P}_2^{AC(j)} = \frac{\sum_{i=1}^{n_2} I_{[[\hat{F}_X^{(j)}]^{-1}(P_1) \leq Y_{1i} \leq [\hat{F}_Z^{(j)}]^{-1}(1-P_3)]} + z_{1-\alpha/2}^2/2}{n_2 + z_{1-\alpha/2}^2},$$

$$\hat{P}_2^{ZQ(j)} = \frac{\sum_{i=1}^{n_2} I_{[[\hat{F}_X^{(j)}]^{-1}(P_1) \leq Y_{1i} \leq [\hat{F}_Z^{(j)}]^{-1}(1-P_3)]} + 1}{n_2 + 2}.$$

In Zhou and Qin (2006), $\hat{P}_2^{ZQ(j)}$ was used in estimating the confidence intervals for the difference of two sensitivities at a fixed level of specificity for diseases processes with binary status, and their simulation study indicated it leads to better coverage properties than $\hat{P}_2^{AC(j)}$ does.

Finally, we obtain the three estimators for $\Delta = P_2^{(1)} - P_2^{(2)}$:

$$\hat{\Delta} = \hat{P}_2^{(1)} - \hat{P}_2^{(2)},$$

$$\hat{\Delta}^{AC} = \hat{P}_2^{AC(1)} - \hat{P}_2^{AC(2)},$$

and

$$\hat{\Delta}^{ZQ} = \hat{P}_2^{ZQ(1)} - \hat{P}_2^{ZQ(2)}. \quad (11)$$

The variance of three estimators can be obtained via simulations using bootstrap samples. Below is the algorithm for $\hat{ZQ}$.

Computing algorithm

Given a dataset consisting of $X_{i1}, \dots, X_{in1}$ from the nondiseased group, $Y_{i1}, \dots, Y_{in2}$ from the early-stage group, and $Z_{i1}, \dots, Z_{in3}$ from the fully diseased group, where $i = 1, 2$ represent marker 1 and 2 respectively, the confidence interval for $\hat{ZQ}$ can be obtained as follows.

1. Calculate $\hat{ZQ}$.

For $b = 1$ to 500 bootstrap iterations, do the following:

2. Draw a resample of n_1 pairs, (X_{1j}^b, X_{2j}^b) with replacements from the nondiseased sample (X_{1j}, X_{2j}) where $j = 1, \dots, n_1$. Similarly, draw a resample of size n_2 pairs from the early-stage group and a resample of n_3 pairs from the fully diseased group.
3. Calculate the bootstrap version of $\hat{P}_2^{ZQ(1),b}$, $\hat{P}_2^{ZQ(2),b}$, and $\hat{ZQ},b$.
4. Repeat Steps 2–3 for a total of $B = 500$ times to obtain the set of bootstrap replications $\hat{P}_2^{ZQ(1),b}$, $\hat{P}_2^{ZQ(2),b}$, $\hat{ZQ},b$, $b = 1, 2, \dots, B$.

The first proposed bootstrap variance estimator is defined as:

$$\hat{Var}^{boot1}(\hat{\Delta}^{ZQ}) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\Delta}^{ZQ,b} - \overline{\hat{\Delta}^{ZQ}} \right)^2, \quad (12)$$

where $\overline{\hat{\Delta}^{ZQ}} = (1/B) \sum_{b=1}^B \hat{\Delta}^{ZQ,b}$.

The second proposed variance estimator is:

$$\hat{Var}^{boot2}(\hat{\Delta}^{ZQ}) = \hat{Var} \left(\hat{P}_2^{ZQ(1)} \right) + \hat{Var} \left(\hat{P}_2^{ZQ(2)} \right) - 2\hat{Cov} \left(\hat{P}_2^{ZQ(1)}, \hat{P}_2^{ZQ(2)} \right), \quad (13)$$

where

$$\hat{Var} \left(\hat{P}_2^{ZQ(1)} \right) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{P}_2^{ZQ(1),b} - \overline{\hat{P}_2^{ZQ(1)}} \right)^2,$$

$$\hat{Var} \left(\hat{P}_2^{ZQ(2)} \right) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{P}_2^{ZQ(2),b} - \overline{\hat{P}_2^{ZQ(2)}} \right)^2,$$

$$\hat{Cov}\left(\hat{P}_2^{ZQ(1)}, \hat{P}_2^{ZQ(2)}\right) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{P}_2^{ZQ(1),b} - \bar{\hat{P}}_2^{ZQ(1)} \right) \left(\hat{P}_2^{ZQ(2),b} - \bar{\hat{P}}_2^{ZQ(2)} \right)$$

$$\text{with } \bar{\hat{P}}_2^{ZQ(1)} = (1/B) \sum_{b=1}^B \hat{P}_2^{ZQ(1),b} \text{ and } \bar{\hat{P}}_2^{ZQ(2)} = (1/B) \sum_{b=1}^B \hat{P}_2^{ZQ(2),b}.$$

There are several ways to construct confidence intervals for $\hat{P}_2$ using bootstrap samples. The $(1 - \alpha)100\%$ level bootstrap percentile confidence interval (BTP) for $\hat{P}_2$ is:

$$(Q_{\alpha/2}(\hat{\Delta}^{ZQ,b}), Q_{1-\alpha/2}(\hat{\Delta}^{ZQ,b}))$$

where $Q_{\alpha}(\hat{\Delta}^{ZQ,b})$ is the α th percentile of $\hat{\Delta}^{ZQ,b}$. Furthermore, using the two estimates for $\hat{P}_2$, namely, $\hat{P}_2^{ZQ}$ and $\bar{\hat{P}}_2^{ZQ}$, and two variance estimators, namely, $\hat{Var}^{boot1}(\hat{\Delta}^{ZQ})$ and $\hat{Var}^{boot2}(\hat{\Delta}^{ZQ})$, we can compute the $(1 - \alpha)100\%$ level confidence intervals in four difference ways, namely

$$\text{theBTBInterval: } \left(\hat{\Delta}^{ZQ} - z_{1-\alpha/2} \sqrt{\hat{Var}^{boot1}(\hat{\Delta}^{ZQ})}, \hat{\Delta}^{ZQ} + z_{1-\alpha/2} \sqrt{\hat{Var}^{boot1}(\hat{\Delta}^{ZQ})} \right),$$

$$\text{theBTIInterval: } \left(\bar{\hat{\Delta}}^{ZQ} - z_{1-\alpha/2} \sqrt{\hat{Var}^{boot1}(\hat{\Delta}^{ZQ})}, \bar{\hat{\Delta}}^{ZQ} + z_{1-\alpha/2} \sqrt{\hat{Var}^{boot1}(\hat{\Delta}^{ZQ})} \right),$$

$$\text{theBTICInterval: } \left(\hat{\Delta}^{ZQ} - z_{1-\alpha/2} \sqrt{\hat{Var}^{boot2}(\hat{\Delta}^{ZQ})}, \hat{\Delta}^{ZQ} + z_{1-\alpha/2} \sqrt{\hat{Var}^{boot2}(\hat{\Delta}^{ZQ})} \right),$$

and

$$\text{theBTIICInterval: } \left(\bar{\hat{\Delta}}^{ZQ} - z_{1-\alpha/2} \sqrt{\hat{Var}^{boot2}(\hat{\Delta}^{ZQ})}, \bar{\hat{\Delta}}^{ZQ} + z_{1-\alpha/2} \sqrt{\hat{Var}^{boot2}(\hat{\Delta}^{ZQ})} \right).$$

4 Simulation studies

Simulation studies were carried out to assess the performances of the proposed confidence interval estimations for $\hat{P}_2$ under bivariate normal and bivariate exponential distributions. The sample sizes for the healthy group, the early diseased group and the fully diseased group were set to be (10, 10, 10), (20, 20, 20), (30, 30, 30), (50, 50, 50), and (50, 30, 30). While both the specificity and sensitivity to the fully disease were fixed at 80%, the parameters of the distributions were determined to set $(P_2^{(1)}, P_2^{(2)})$ as (50%, 50%), (70%, 70%), (90%, 90%), (50%, 90%), and (70%, 90%), in order to evaluate the performance of the confidence intervals either when the P_2 s are the same or different from each other. For each parameter

setting, 2000 random samples were generated. Besides, considering 60% would be the minimum reasonable value for the specificity and the sensitivity to the fully disease, we have also conducted simulations for the following combinations of (P_1, P_3) : (60%, 60%), (60%, 80%), and (80%, 60%). The results are very similar to the above (80%, 80%) setting, so they are omitted in the paper but available upon request.

Remark

Our simulation study showed that out of the three estimators for $\hat{Q}$, $\hat{Q}$ works the best in terms of coverage probabilities. Therefore, only simulation results using $\hat{Q}$ are presented.

Table 1 presents the estimated coverage probabilities and lengths under bivariate normality. The covariance matrices $(\Sigma_1, \Sigma_2 \text{ and } \Sigma_3)$ are set as follows:

$$\Sigma_1 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1.1 & 0.5 \\ 0.5 & 1.1 \end{pmatrix}, \Sigma_3 = \begin{pmatrix} 1.2 & 0.5 \\ 0.5 & 1.2 \end{pmatrix}.$$

To save space, the simulation results using other settings for covariance matrices $(\Sigma_1, \Sigma_2, \text{ and } \Sigma_3)$ with different correlations between two markers are presented in Supplementary Material. The simulations results indicate that the **GI** method can generally achieve the nominal level, except that the coverage probability tends to be a slightly conservative as the sample sizes are small as (10, 10, 10). The performance of the nonparametric methods *BTP*, *BTIB*, *BTIIB*, *BTIC*, and *BTIIC* varies with different settings of $P_2^{(1)}$ and $P_2^{(2)}$. Specifically, we note the following: (1) as $P_2^{(1)}=0.7, P_2^{(2)}=0.9$, *BTP*, *BTIIB*, and *BTIIC* provide reasonable coverage while *BTIB* and *BTIC* can be liberal for certain scenarios; (2) as $P_2^{(1)}=0.7, P_2^{(2)}=0.7$, and $P_2^{(1)}=0.9, P_2^{(2)}=0.9$, all of these nonparametric methods tends to be conservative; (3) as $P_2^{(1)}=0.5, P_2^{(2)}=0.5$, *BTIB*, and *BTIC* provide reasonable coverage while *BTP* and *BTIIB* tend to be conservative; (4) as $P_2^{(1)}=0.5, P_2^{(2)}=0.9$, *BTP* provides reasonable coverage while *BTIB*, *BTIIB*, *BTIC*, *BTIIC* tend to be liberal. Overall speaking, out of the nonparametric methods, *BTP* method is the most preferable one, followed by *BTIB*, *BTIC*. Furthermore, the length of *GI* confidence intervals are shorter than that by non-parametric methods except for a few cases with small sample sizes.

Table 2 presents the estimated coverage probabilities and lengths under bivariate exponential distributions with the correlation coefficient for multinormal copulas being 0.5. To save space, the simulation results using settings with correlation coefficient as 0.2 and 0.8 for multinormal copulas are presented in Supplementary Material. The bivariate random numbers are generated for each of the three groups with the marginal distributions as exponential distributions. The parameter λ for these exponential distributions are determined by the given $(P_2^{(1)}, P_2^{(2)})$. The *GI* method was applied to Box-Cox transformed data (*BCGI*) and simulation results show *BCGI* performs reasonably except that it tends to be conservative for small sample sizes. The performance of the nonparametric methods are similar to that for normal data in Table 1. Specifically, *BTP* method has the best performance followed by *BTIB*, *BTIC*.

In summary, the *GI* and *BCGI* approaches work the best for the scenarios considered. Otherwise, the non-parametric method *BTP* is a good option.

Remark

Extensive additional simulation study has been performed for the following settings: (1) different P_1 and P_3 ; (2) unequal sample sizes for three groups; (3) unequal means for two biomarkers for equal P_2 cases. The conclusions are similar. The results are not presented here, but available upon request.

5 An example

In this section, we apply the proposed confidence interval estimators for $\Delta = P_2^{(1)} - P_2^{(2)}$ to a dataset from a study on Alzheimer's disease, based on a longitudinal cohort from the Washington University Alzheimer's Disease Research Center (WU ADRC). Alzheimer's disease (AD) is one of the most common degenerative dementias for aged people, and it is becoming even more prevalent, thus resulting in a major healthcare crisis. So far, the earliest and the most reliable evidence of AD is cognitive and behavioral deficits. However, by the time these neuropsychology deficits are fully developed, the disease probably has been present for many years. On the other hand, most of the medicines for AD only have limited efficacy at a later stage and hence the early-stage AD is likely the better time window for treatment interventions than the fully developed AD. Therefore a major challenge is to identify the individuals in the early phase of AD. In this data set, the severity of dementia was staged by the clinical dementia rating (CDR) according to published rules (Morris, 1993). A global CDR is derived from individual ratings in multiple domains by an experienced clinician such that CDR 0 indicated no dementia, and CDR 0.5, 1, 2, and 3 represent very mild, mild, moderate, and severe dementia, respectively. Because neuropsychometric data from people with more advanced dementia (i.e., CDR 2 or 3) are rarely available, we will only focus on subjects with CDR 0, 0.5, and 1. In this age-matched sample, there are 45 subjects in nondemented group, 44 in the very mildly demented group, and 29 in the mildly demented group. Besides clinical evaluation, participants also completed standard psychometric tests. In addition, based on two factor analysis, Kanne et al. (1998) and Rubin et al. (1998) described several different factors which play significant roles at different stages of dementia including global factor, temporal factor, parietal factor and frontal factor and therefore these composite factor scores were also computed from the database. The descriptive statistics of these factors are presented in Table 3.

Furthermore, the sample covariance matrices for these four composite factors are as follows:

$$\begin{pmatrix} 0.7885 & 1.5743 & 0.8276 & 1.1509 \\ & 5.0558 & 1.2074 & 2.141 \\ & & 1.6779 & 0.7801 \\ & & & 3.156 \end{pmatrix} \text{ for non-demented group (CDR0),}$$

$$\begin{pmatrix} 2.9649 & 4.7466 & 2.9853 & 2.8862 \\ & 9.5968 & 4.4839 & 4.3004 \\ & & 4.207 & 2.5113 \\ & & & 4.8947 \end{pmatrix} \text{ for very mildly demented group(CDR0.5), and}$$

$$\begin{pmatrix} 2.8863 & 4.258 & 2.7024 & 2.4275 \\ & 10.7552 & 2.6894 & 3.2999 \\ & & 6.4956 & 1.5315 \\ & & & 4.2711 \end{pmatrix} \text{ for mildly demented group(CDR1).}$$

The goodness-of-fit tests showed that only the two pairs (temporal factor, frontal factor) and (parietal factor, frontal factor) satisfy the bivariate normality for original data. No other pairs satisfy the bivariate normality for either the original or Box-Cox transformed data.

Therefore, for these two pairs, that is (temporal factor, frontal factor) and (parietal factor, frontal factor), the generalized inference confidence intervals (*GI*) were computed for the differences; for the other pairs, only the nonparametric confidence intervals using the *BTP* method were computed. Results are presented in Table 4. From this table, we can see that global factor and temporal factor have similar sensitivity to the early stage, and so are parietal factor and frontal factor. Furthermore, both the global factor and temporal factor have better sensitivity to the early stage than parietal factor. Overall speaking, the biomarkers with the best identification accuracy for the early disease stage are global factor and temporal factor. Hence for the purpose of early disease identification, global factor and temporal factor are better biomarkers to use compared to parietal factor and frontal factor.

6 Discussion

For the disease processes with three ordinal disease stages, the early stage is usually considered the most important stage for intervention. Hence, sensitivity to the early stage given specificity and sensitivity to fully diseased stage is an important index of diagnostic accuracy. In practice, multiple biomarkers or diagnostic tests are usually measured on the same individuals to search for the best test for the early diagnosis of disease. This article provides inference methods for confidence interval estimation of the difference in sensitivity to early stage between two paired markers. We proposed several parametric and nonparametric approaches for this purpose and compared their performance.

Based on the simulation studies, the following recommendations are made. First of all, if bivariate normality is satisfied for either original or transformed data, the generalized inference (*GI*) or Box-Cox transformed generalized inference (*BCGI*) approaches are recommended. They are easy to use, and have good coverage probability. Secondly, when the bivariate normality assumption is not met, the nonparametric methods work when sample sizes are larger than 10, though sometimes maybe a little conservative or liberal. It appears the *BTP* method is preferable compared to other nonparametric methods. All of the proposed approaches are simulation based. The *GI* approach based on normality is an easy-to-use approach while the *BCGI* approach involves solving equations and checking the

assumptions. The nonparametric approaches, especially the *BTP* method, are easy to implement except that for *BTIB*, *BTIB*, *BTIC*, and *BTIC* methods the variances are computed through bootstrap samples. An R program is available upon request from ltian@buffalo.edu.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Dr. Chengjie Xiong was partly supported by National Institute on Aging grant P50AG05681, P01 AG03991, R01 AG029672, and R01 AG034119.

References

- Agresti A, Coull BA. Approximate is better than “exact” for interval estimation of Binomial proportions. *The American Statistician*. 1998; 52:119–126.
- Crystal H, Dickson D, Fuld P, Masur D, Scott R, Mehler M, Masdeu J, Kawas C, Aronson M, Wolfson L. Clinico-pathologic studies in dementia: nondemented subjects with pathologically confirmed Alzheimers disease. *Neurology*. 1988; 38:1682–1687. [PubMed: 3185902]
- Dong T, Tian L, Hutson A, Xiong C. Parametric and non-parametric confidence intervals of the probability of identifying early dis-ease stage given sensitivity to full disease and specificity with three ordinal diagnostic groups. *Statistics in Medicine*. 2011; 30:3532–3545. [PubMed: 22139763]
- Dreiseitl S, Ohno-Machado L, Binder M. Comparing three-class diagnostic tests by three-way ROC analysis. *Medical Decision Making*. 2000; 20:323–331. [PubMed: 10929855]
- Faraggi D, Reiser B. Estimation of the area under the ROC curve. *Statistics in Medicine*. 2002; 21:3093–3106. [PubMed: 12369084]
- Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cut-off point. *Biometrical Journal*. 2005; 47:458–472. [PubMed: 16161804]
- Greenhouse SW, Mantel N. The evaluation of diagnostic test. *Bimetrics*. 1950; 6:399–412.
- Heckerling PS. Parametric three-way receiver operating characteristic surface analysis using Mathematica. *Medical Decision Making*. 2001; 21:409–417. [PubMed: 11575490]
- Kanne SM, Balota DA, Storandt M, McKeel DW Jr, Morris JC. Relating anatomy to function in Alzheimers disease: neuropsychological profiles predict regional neuropathology 5 years later. *Neurology*. 1998; 50:979–985. [PubMed: 9566382]
- Katzman R, Terry R, DeTeresa R, Brown T, Davies P, Fuld P, Renbing X, Peck A. Clinical pathological and neurochemical changes in dementia: a subgroup with preserved mental status and numerous neocortical plaques. *Annals of Neurology*. 1988; 23:138–144. [PubMed: 2897823]
- Krishnamoorthy K, Lu Y. Inference on the common means of several normal populations based on the generalized variable method. *Biometrics*. 2003; 59:237–247. [PubMed: 12926708]
- Li C, Liao C, Liu J. On the exact interval estimation for the difference in paired areas under the ROC curves. *Statistics in Medicine*. 2008a; 27:224–242. [PubMed: 17139702]
- Li C, Liao C, Liu J. A non-inferiority test for diagnostic accuracy based on the paired partial areas under ROC curves. *Statistics in Medicine*. 2008b; 27:1762–1776. [PubMed: 17968858]
- Lin SH, Lee JC, Wang RS. Generalized inferences on the common mean vector of several multivariate normal populations. *Journal of Statistical Planning and Inference*. 2007; 137:2240–2249.
- Linnet K. Comparison of quantitative diagnostic tests: type I error, power, and sample size. *Statistics in Medicine*. 1987; 6:147–158. [PubMed: 3589244]
- Molodianovitch K, Faraggi D, Reiser B, Molodianovitch K, Faraggi D, Reiser B. Comparing the areas under two correlated ROC curves: parametric and non-parametric approaches. *Biometrical Journal*. 2006; 5:745–757. [PubMed: 17094340]

- Morris JC, McKeel DW, Storandt M, Rubin EH, Price L, Grant EA, Ball MJ, Berg L. Very mild Alzheimers disease: informant-based clinical, psychometric and pathologic distinction from normal aging. *Neurology*. 1991; 41:469–478. [PubMed: 2011242]
- Morris JC. The clinical dementia rating (CDR): current version and scoring rules. *Neurology*. 1993; 43:2412–2414. [PubMed: 8232972]
- Mossman D. Three-way ROCs. *Medical Decision Making*. 1999; 19:78–89. [PubMed: 9917023]
- Nakas C, Yiannoutsos C. Ordered multiple-class ROC analysis with continuous measurement. *Statistics in Medicine*. 2004; 23:3437–3449. [PubMed: 15505886]
- Rubin EH, Storandt M, Miller JP, Kinscherf DA, Grant EA, Morris JC, Berg L. A prospective study of cognitive function and onset of dementia in cognitively healthy elders. *Archives of neurology*. 1998; 55:395–401. [PubMed: 9520014]
- Schisterman E, Faraggi D, Reiser B. Adjusting the generalized ROC curve for covariates. *Statistics in Medicine*. 2004; 23:3319–3331. [PubMed: 15490426]
- Schisterman E, Reiser B, Faraggi D. ROC analysis for markers with mass at zero. *Statistics in Medicine*. 2006; 25:623–638. [PubMed: 16345033]
- Tian L. Confidence intervals for $P(Y1 > Y2)$ with normal outcomes in linear models. *Statistics in Medicine*. 2008; 27:4221–4237. [PubMed: 18407578]
- Tian L, Cappelleri JC. A new approach for interval estimation and hypothesis testing of a certain intraclass correlation coefficient: the generalized variable method. *Statistics in Medicine*. 2004; 23:2125–2135. [PubMed: 15211607]
- Tian L, Xiong C, Lai C, Vexler A. Exact confidence interval estimation for the difference in diagnostic accuracy with three ordinal diagnostic groups. *Journal of Statistical Planning and Inference*. 2010; 141:549–558. [PubMed: 23538945]
- Tsui KW, Weerahandi S. Generalized p -values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of American Statistical Association*. 1989; 84:602–607.
- Weerahandi S. Generalized confidence intervals. *Journal of American Statistical Association*. 1993; 88:899–905.
- Weerahandi S. ANOVA under unequal error variances. *Biometrics*. 1995; 51:589–599.
- Weerahandi, S. *Exact Statistical Methods for Data Analysis*. New York, NY: Springer; 2003.
- Weerahandi S, Berger VW. Exact inference for growth curves with intraclass correlation structure. *Biometrics*. 1999; 55:921–924. [PubMed: 11315029]
- Xiong CJ, Gerald VB, Philip M, John CM. Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Statistics in Medicine*. 2006; 25:1251–1273. [PubMed: 16345029]
- Xiong C, van Belle G, Miller JP, Yan Y, Gao F, Feng S, Yu K, Morris JC. A parametric comparison of diagnostic accuracy with three ordinal diagnostic groups. *Biometrical Journal*. 2007; 49:682–693. [PubMed: 17763377]
- Zhou XH, Qin GS, Hsu YS. New confidence intervals for the difference between two sensitivities at a fixed level of specificity. *Statistics in Medicine*. 2006; 25:3487–3502. [PubMed: 16345124]
- Zou KH, Hall WJ. Two transformation models for estimating an ROC curve derived from continuous data. *Journal of Applied Statistics*. 2000; 27:621–631.
- Zou KH, Hall WJ, Shapiro DE. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine*. 1997; 16:2143–2156. [PubMed: 9330425]

Table 1

Simulation results for the difference of P_2 s under bivariate normality (based on 2000 simulations) with

$$\Sigma_1 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1.1 & 0.5 \\ 0.5 & 1.1 \end{pmatrix}, \Sigma_3 = \begin{pmatrix} 1.2 & 0.5 \\ 0.5 & 1.2 \end{pmatrix}.$$

Sample sizes	Coverage probability					Length				
	BTP	BTIB	BTIC	BTIC	BTIC	BTP	BTIB	BTIC	BTIC	GI
$\mu_1 = (0, 0)', \mu_2 = (2.5, 2.5)', \mu_3 = (3.69, 3.69)', P_2^{(1)} = 0.5, P_2^{(2)} = 0.5$										
(10, 10, 10)	0.9920	0.9580	0.9860	0.9585	0.9860	0.9820	0.9213	0.9375	0.9311	0.8946
(20, 20, 20)	0.9835	0.9625	0.9795	0.9600	0.9835	0.9620	0.7714	0.7846	0.7839	0.6033
(30, 30, 30)	0.9860	0.9665	0.9800	0.9660	0.9780	0.9630	0.6636	0.6681	0.6757	0.4875
(50, 50, 50)	0.9835	0.9630	0.9805	0.9660	0.9785	0.9545	0.5333	0.5410	0.5422	0.3760
(50, 30, 30)	0.9800	0.9595	0.9760	0.9615	0.9770	0.9580	0.6509	0.6644	0.6640	0.4773
$\mu_1 = (0, 0)', \mu_2 = (2.5, 2.5)', \mu_3 = (4.31, 4.31)', P_2^{(1)} = 0.7, P_2^{(2)} = 0.7$										
(10, 10, 10)	0.9935	0.9745	0.9905	0.9775	0.9935	0.9820	0.8263	0.8207	0.8174	0.8677
(20, 20, 20)	0.9940	0.9755	0.9910	0.9685	0.9855	0.9690	0.6736	0.6736	0.6744	0.5633
(30, 30, 30)	0.9870	0.9740	0.9830	0.9685	0.9870	0.9625	0.5698	0.5717	0.5694	0.4451
(50, 50, 50)	0.9805	0.9600	0.9805	0.9610	0.9805	0.9545	0.4478	0.4507	0.4544	0.3343
(50, 30, 30)	0.9830	0.9755	0.9845	0.9745	0.9885	0.9570	0.5539	0.5564	0.5583	0.4322
$\mu_1 = (0, 0)', \mu_2 = (2.5, 2.5)', \mu_3 = (5.51, 5.51)', P_2^{(1)} = 0.9, P_2^{(2)} = 0.9$										
(10, 10, 10)	0.9995	0.9895	0.9985	0.9960	0.9995	0.9695	0.5382	0.5308	0.5346	0.7121
(20, 20, 20)	0.9970	0.9910	0.9980	0.9895	0.9970	0.9660	0.4071	0.4106	0.4077	0.3902
(30, 30, 30)	0.9955	0.9875	0.9960	0.9895	0.9970	0.9630	0.3334	0.3283	0.3348	0.2855
(50, 50, 50)	0.9865	0.9850	0.9905	0.9765	0.9870	0.9515	0.2550	0.2530	0.2524	0.2020
(50, 30, 30)	0.9895	0.9850	0.9965	0.9865	0.9925	0.9620	0.3135	0.3100	0.3128	0.2679
$\mu_1 = (0, 0)', \mu_2 = (2.5, 2.5)', \mu_3 = (3.69, 5.51)', P_2^{(1)} = 0.5, P_2^{(2)} = 0.9$										
(10, 10, 10)	0.9530	0.8960	0.9125	0.8985	0.9240	0.9790	0.7824	0.8093	0.8057	0.8454
(20, 20, 20)	0.9615	0.9090	0.9280	0.9130	0.9290	0.9590	0.6511	0.6599	0.6629	0.5620

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Sample sizes	Coverage probability						Length					
	BTP	BTIB	BTIB	BTIC	BTIC	BTIC	BTIB	BTIB	BTIB	BTIC	BTIC	GI
(30, 30, 30)	0.9575	0.9210	0.9365	0.9255	0.9430	0.9550	0.5668	0.5668	0.5675	0.5675	0.5675	0.4509
(50, 50, 50)	0.9585	0.9205	0.9360	0.9330	0.9475	0.4436	0.4505	0.4505	0.4543	0.4543	0.4543	0.3445
(50, 30, 30)	0.9560	0.9240	0.9405	0.9325	0.9415	0.5427	0.5594	0.5594	0.5580	0.5580	0.5580	0.4399
$\mu_1 = (0, 0)', \mu_2 = (2.5, 2.5)', \mu_3 = (4.31, 5.51)', P_2^{(1)} = 0.7, P_2^{(2)} = 0.9$												
(10, 10, 10)	0.9470	0.9305	0.9525	0.9280	0.9450	0.7012	0.7160	0.7160	0.7164	0.7164	0.7164	0.8205
(20, 20, 20)	0.9595	0.9450	0.9650	0.9375	0.9575	0.5678	0.5757	0.5757	0.5777	0.5777	0.5777	0.5123
(30, 30, 30)	0.9640	0.9435	0.9560	0.9425	0.9595	0.4782	0.4910	0.4910	0.4883	0.4883	0.4883	0.4002
(50, 50, 50)	0.9575	0.9340	0.9485	0.9315	0.9475	0.3791	0.3862	0.3862	0.3794	0.3794	0.3794	0.3005
(50, 30, 30)	0.9525	0.9340	0.9465	0.9360	0.9515	0.4603	0.4693	0.4693	0.4706	0.4706	0.4706	0.3876

Table 2

Simulation results for the difference of P_2 s under bivariate exponential distributions (based on 2000 simulations) with bivariate normal copular with correlation coefficient 0.5.

Sample sizes	Coverage probability					Length				
	BTP	BTIB	BTIB	BTIC	BTIC	BCGI	BTP	BTIB	BTIC	BCGI
$\lambda_1 = (20, 20)'; \lambda_2 = (1, 1)'; \lambda_3 = (0.26, 0.26)'; P_2^{(1)} = 0.5, P_2^{(2)} = 0.5$										
(10, 10, 10)	0.9920	0.9565	0.9860	0.9525	0.9855	0.9775	0.9516	0.9770	0.9709	0.9008
(20, 20, 20)	0.9860	0.9540	0.9770	0.9485	0.9740	0.9705	0.8367	0.8573	0.8558	0.6550
(30, 30, 30)	0.9865	0.9580	0.9805	0.9630	0.9815	0.9650	0.7451	0.7528	0.7550	0.5400
(50, 50, 50)	0.9795	0.9565	0.9810	0.9545	0.9735	0.9600	0.6140	0.6211	0.6205	0.4232
(50, 30, 30)	0.9860	0.9605	0.9780	0.9555	0.9770	0.9575	0.7386	0.7533	0.7453	0.5374
$\lambda_1 = (20, 20)'; \lambda_2 = (1, 1)'; \lambda_3 = (0.15, 0.15)'; P_2^{(1)} = 0.7, P_2^{(2)} = 0.7$										
(10, 10, 10)	0.9965	0.9565	0.9870	0.9610	0.9890	0.9795	0.9027	0.9190	0.9096	0.9565
(20, 20, 20)	0.9920	0.9680	0.9890	0.9715	0.9920	0.9735	0.7978	0.8011	0.8111	0.6678
(30, 30, 30)	0.9875	0.9700	0.9890	0.9690	0.9870	0.9620	0.7061	0.7135	0.7128	0.5397
(50, 50, 50)	0.9865	0.9450	0.9745	0.9685	0.9840	0.9650	0.5846	0.5890	0.5851	0.4119
(50, 30, 30)	0.9885	0.9660	0.9835	0.9695	0.9880	0.9640	0.7064	0.7111	0.7076	0.5389
$\lambda_1 = (20, 20)'; \lambda_2 = (1, 1)'; \lambda_3 = (0.06, 0.06)'; P_2^{(1)} = 0.9, P_2^{(2)} = 0.9$										
(10, 10, 10)	0.9985	0.9880	0.9965	0.9915	0.9975	0.9805	0.7133	0.7009	0.6977	0.9302
(20, 20, 20)	0.9980	0.9885	0.9960	0.9900	0.9980	0.9730	0.5709	0.5593	0.5617	0.5209
(30, 30, 30)	0.9925	0.9895	0.9965	0.9840	0.9915	0.9710	0.4721	0.4646	0.4640	0.3700
(50, 50, 50)	0.9845	0.9870	0.9945	0.9875	0.9925	0.9685	0.3465	0.3450	0.3457	0.2450
(50, 30, 30)	0.9925	0.9885	0.9935	0.9865	0.9925	0.9710	0.4587	0.4554	0.4587	0.3700
$\lambda_1 = (20, 20)'; \lambda_2 = (1, 1)'; \lambda_3 = (0.26, 0.06)'; P_2^{(1)} = 0.5, P_2^{(2)} = 0.9$										
(10, 10, 10)	0.9085	0.8645	0.8820	0.8745	0.8920	0.9675	0.8468	0.8869	0.8845	0.9233
(20, 20, 20)	0.9450	0.9125	0.9300	0.9175	0.9295	0.9545	0.7406	0.7553	0.7672	0.6356
(30, 30, 30)	0.9545	0.9280	0.9420	0.9295	0.9395	0.9490	0.6419	0.6545	0.6585	0.5098

Sample sizes	Coverage probability						Length					
	BTP	BTIB	BTIB	BTIC	BTIC	BTIC	BTP	BTIB	BTIB	BTIC	BTIC	BCGI
(50, 50, 50)	0.9650	0.9490	0.9550	0.9420	0.9535	0.9480	0.5255	0.5404	0.5404	0.5287	0.5287	0.3937
(50, 30, 30)	0.9535	0.9345	0.9470	0.9235	0.9425	0.9485	0.6414	0.6679	0.6679	0.6593	0.6593	0.5055
$\lambda_1 = (20, 20); \lambda_2 = (1, 1); \lambda_3 = (0.15, 0.06); P_2^{(1)} = 0.7, P_2^{(2)} = 0.9$												
(10, 10, 10)	0.9365	0.9150	0.9385	0.9225	0.9420	0.9785	0.8267	0.8245	0.8245	0.8302	0.8302	0.9470
(20, 20, 20)	0.9485	0.9290	0.9535	0.9425	0.9605	0.9705	0.7017	0.7157	0.7157	0.7127	0.7127	0.6197
(30, 30, 30)	0.9590	0.9385	0.9580	0.9410	0.9560	0.9555	0.6080	0.6242	0.6242	0.6247	0.6247	0.4874
(50, 50, 50)	0.9690	0.9470	0.9595	0.9410	0.9545	0.9520	0.4925	0.5020	0.5020	0.4984	0.4984	0.3642
(50, 30, 30)	0.9675	0.9390	0.9590	0.9450	0.9565	0.9595	0.6117	0.6228	0.6228	0.6145	0.6145	0.4855

Note: λ_1^* contains the parameters (inverse of the mean) used to generate marginal exponential distributions for markers 1 and 2 for nondiseased group, λ_2^* early-stage group, and λ_3^* for diseased group.

Table 3

Means (standard deviations) of composite factors from the WU ADRC sample.

Variable	CDR 0 (<i>n</i> =45)	CDR 1 (<i>n</i> =44)	CDR 3 (<i>n</i> =29)
Global factor	0.569 (0.888)	−1.622 (1.722)	−4.199 (1.699)
Temporal factor	4.085 (2.249)	−0.986 (3.315)	−5.855 (3.223)
Parietal factor	1.803 (1.295)	−0.241 (2.051)	−2.377 (2.549)
Frontal factor	2.866 (1.777)	0.373 (2.212)	−2.682 (2.067)

Table 4

Estimated 95% confidence intervals for the pairwise differences $P_2^{(1)} - P_2^{(2)}$ between cognitive factors in the first column and those in the first row from WU ADRC sample (P_1 and P_3 are set at 0.8).

Var	Method	Temporal factor		Parietal factor		Frontal factor	
		Lower	upper	Lower	upper	Lower	upper
Global factor	BTP	-0.1522	0.1957	0.1087	0.6522	0	0.5652
	GI/BCGI	*	*	*	*	*	*
Temporal factor	BTP			0.0870	0.6739	0.0000	0.5870
	GI/BCGI			*	*	-0.0125	0.3926
Parietal factor	BTP					-0.4565	0.3043
	GI/BCGI					-0.4657	0.0290

The * indicate the cases where *GI* or *BCGI* can not be applied due to the fact that either the original or the transformed data do not satisfy normality.

Note:

Lower: Lower bound of the confidence interval.

Upper: Upper bound of the confidence interval.