

Published in final edited form as:

Biom J. 2011 May ; 53(3): 464–476. doi:10.1002/bimj.201000083.

ROC curve inference for best linear combination of two biomarkers subject to limits of detection

Neil J. Perkins^{1,*}, Enrique F. Schisterman¹, and Albert Vexler²

¹Eunice Kennedy Shriver National Institute of Child Health and Human Development, 6100 Executive Blvd, rm 7b03, Rockville, MD 20852, USA

²Department of Biostatistics, The State University of New York at Buffalo, 246 Farber Hall, 3435 Main St., Buffalo, NY 14214, USA

Abstract

The receiver operating characteristic (ROC) curve is a tool commonly used to evaluate biomarker utility in clinical diagnosis of disease. Often, multiple biomarkers are developed to evaluate the discrimination for the same outcome. Levels of multiple biomarkers can be combined via best linear combination (BLC) such that their overall discriminatory ability is greater than any of them individually. Biomarker measurements frequently have undetectable levels below a detection limit sometimes denoted as limit of detection (LOD). Ignoring observations below the LOD or substituting some replacement value as a method of correction has been shown to lead to negatively biased estimates of the area under the ROC curve for some distributions of single biomarkers. In this paper, we develop asymptotically unbiased estimators, via the maximum likelihood technique, of the area under the ROC curve of BLC of two bivariate normally distributed biomarkers affected by LODs. We also propose confidence intervals for this area under curve. Point and confidence interval estimates are scrutinized by simulation study, recording bias and root mean square error and coverage probability, respectively. An example using polychlorinated biphenyl (PCB) levels to classify women with and without endometriosis illustrates the potential benefits of our methods.

Keywords

Area under the curve; Best linear combination; Left censoring; Limit of detection; ROC

1 Introduction

The use of biomarkers to assist medical decision making, the diagnosis and prognosis of individuals with a given disease, is increasingly common in both clinical settings and epidemiological research. This has spurred an increase in exploration for and development

© 2011 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

*Corresponding author: perkinson@mail.nih.gov, Phone: +1-301-435-6928, Fax: +1-301-402-2084.

Conflict of interest

The authors have declared no conflict of interest.

Supporting Information for this article is available from the author or on the WWW under <http://dx.doi.org/10.1002/bimj.201000083>

of new biomarkers. All biomarkers, established and emerging, are limited by the sensitivity of the measurement instrument which can lead to censored observations, often more frequently in emerging biomarkers as laboratory methods are underdeveloped. This censoring process is due to a limit of detection (LOD) or the inability of an instrument to reliably measure samples below (or in some cases above) some threshold. Polychlorinated biphenyl (PCB) levels fit this scenario well as they have been linked with several adverse outcomes and their measurements are affected by LODs. A common statistical tool used to evaluate the utility of a potential biomarker such as PCBs is the receiver operating characteristic (ROC) curve.

Consider populations of diseased and non-diseased people with levels of a specific biomarker denoted by independent random variables X and Y , respectively, with cumulative distribution functions $F(x)$ and $G(y)$. Suppose, the biomarker is utilized as an indicator of disease status where a level above some cut point, c , indicates a positive test for the disease and a level below c corresponds to a negative test. The sensitivity (the true positive test rate) and specificity (the true negative test rate) of the biomarker for a given c are $q(c) = 1 - F(c)$ and $p(c) = G(c)$, respectively. The ROC curve is then a mapping of $\{1 - p(c), q(c)\}$ across all possible c . Proposed uses for the ROC curve (Zhou et al., 2002; Pepe, 2003) include assessing discriminatory ability over all c , over a specific range of $q(c)$ or $p(c)$ and the maximum ability to differentiate between the populations. The area under the ROC curve, denoted here by AUC, is the most commonly used summary measure (Zhou et al., 2002; Pepe, 2003) and has been shown to be $P(X > Y)$ for continuously measured biomarkers. AUC tends to range from 0.5 to 1 with larger values indicating greater separation between diseased and non-diseased biomarker levels. As a result, given multiple markers for the same disease, a researcher would be inclined to choose the one with the highest AUC to aid in decision making. Another approach would be to utilize a linear combination of these multiple biomarkers in lieu of selecting one. The attractiveness of a linear combination lies in being able to better discriminate, achieve a higher AUC, than if using any single biomarker alone.

Now consider the case where two biomarkers, $\vec{X} = (X_1, X_2)^T$ and $\vec{Y} = (Y_1, Y_2)^T$, measured in individuals with and without a disease, respectively, are independent bivariate normally distributed. Clearly, each biomarker considered individually would be normally distributed and AUC's for each could be calculated and contrasted. The alternative described previously would be to use a linear combination, say $U = \beta^T \vec{X} = \beta_1 X_1 + \beta_2 X_2$ and $V = \beta^T \vec{Y} = \beta_1 Y_1 + \beta_2 Y_2$, as a composite "biomarker" for decision making. Conveniently, U and V are also normally distributed with an ROC curve that is now a function of the choice of β^T and the corresponding AUC can be denoted by AUC_β .

Linear combinations for binary outcomes are often estimated using logistic regression for a given set of covariates. However, given this scenario of having several biomarkers following a multivariate normal distribution, Su and Liu (1993) developed a best linear combination (BLC), $\vec{\beta}_0^T$, leading to AUC_0 , that is "best" with respect to maximizing AUC_β over all real β^T directly rather than maximizing the logistic regression model which, can be concordant or discordant depending on the scenario (Pepe et al., 2006). This would allow for better

discrimination than from any individual biomarker. When normal parameter values are unknown, we can use random samples to calculate estimates $\hat{\beta}_0^T$ and $A\hat{U}C_0$ via maximum likelihood techniques.

One common complication in biomarker evaluation is that for a variety of reasons, random samples X and Y of biomarkers are often evaluated with non-detects or missing data below some LOD, quantified as d , effectively censoring the data below d . Omitting these values and proceeding with a complete case analysis has been shown to lead to biased $A\hat{U}C$ for univariate normal and gamma distributed biomarkers (Perkins et al., 2007). Categorizing the missing biomarker levels as ties at the lowest score, the standard non-parametric ROC curve and empirical AUC yield unbiased estimates of the biomarkers' effectiveness given the measurement limitations. However, when underlying discriminatory ability is of interest, parametric methods can be used to estimate the ROC curve below the LOD and thus an AUC for a latent variable that might be measured completely. Generally, substituting a replacement value such as 0 , $d/2$, $d/\sqrt{2}$ and d for unobservable data have been shown as a simple method to lessen biased parametric estimation (e.g. mean, variance and potential AUC) but the magnitude and direction of the remaining bias is highly dependent on the value chosen as well as the parameter being estimated (Hornung and Reed, 1990). It has been shown (Haas and Scheff, 1990; Lyles et al., 2001; Singh and Nocerino, 2002; Lynn, 2001; Perkins et al., 2007) that all of these methods lead to biased estimation of mean and standard deviation parameters, regression coefficients, odds ratio and AUC for a single biomarker following common distributions. In the logistic regression framework which has similarities here, Lynn (2001) demonstrated that the bias of these replacement values lessens the closer they are to the expected value below d , which performed similarly to more sophisticated methods of estimation.

Lyles et al. (2001) proposed a more thoughtful solution of constructing a likelihood function for two censored bivariate normally distributed random variables. Based on this likelihood and under the assumption of normality, subsequent maximum likelihood estimators (MLEs) are efficient and asymptotically unbiased point estimators.

In this paper, we propose in Section 2 to obtain MLEs for normal distribution parameters using Lyles et al.'s (2001) bivariate normal developments for parameter estimates based on a sample with left censoring in order to construct $\hat{\beta}_0^T$ and $A\hat{U}C_0$ to estimate that underlying potential of the BLC of two biomarkers. In Section 3, we extend Lyles et al. (2001) and consider the asymptotic distributions of parameter estimates by constructing the Fisher Information matrix for this case. Also in Section 3, these developments are subsequently used in finding the asymptotic distributions of point estimates $\hat{\beta}_0^T$ and $A\hat{U}C_0$, leading to accompanying confidence intervals (CIs). This procedure allows for the estimation of the BLC's underlying discriminatory ability or the potential that could be realized if it were possible to eliminate the LOD and censoring in the tail. This would be especially useful when researchers are exploring biomarkers for a given outcome using more cost-effective but less-sensitive assays, with the idea of further measurements or re-measurement being conducted on a narrowed set of promising biomarkers using a more sensitive and costly,

“gold standard” assay. When a more sensitive assay is lacking this would help identify biomarkers with potential that are worth additional resources in refining a measurement process. However, the discriminatory ability of the BLC of these biomarkers as measured with LOD is better estimated using the AUC for traditional empirical ROC curve, denoted $A\tilde{U}C_0$ here, which appropriately accounts for the censored values by essentially treating them as ties because they are indiscernible from one another. In Section 4, simulation is used to assess $\hat{\beta}_0^T$ and point estimators $A\hat{U}C_0$ and $A\tilde{U}C_0$ as well as CIs for AUC_0 . An example in Section 5 using levels of the environmental toxicants PCBs to classify women with endometriosis is used to illustrate empirical and maximum likelihood techniques. We end with a brief discussion of issues surrounding estimation based on two biomarkers affected by LODs.

2 Methods

Suppose that pairs of biomarkers' levels $X = (X_1, X_2)^T$ and $Y = (Y_1, Y_2)^T$ for cases and controls, respectively, are independent and have bivariate normal distributions $f(X; \mu_X, \Sigma_X)$ and $g(Y; \mu_Y, \Sigma_Y)$, respectively. These distributions can be written in a matrix form or explicitly expanded

$$g(\vec{W}; \vec{\mu}, \Sigma) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(\vec{W} - \vec{\mu})^T \Sigma^{-1}(\vec{W} - \vec{\mu}) \right\} \\ = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)}(\eta_1^2 - 2\rho\eta_1\eta_2 + \eta_2^2) \right\},$$

where $\eta_l = (w_l - \mu_l)/\sigma_l$, $l = 1, 2$, $\vec{\mu} = (\mu_1, \mu_2)^T$ is the mean vector and the covariance matrix Σ consists of variances, $\Sigma_{ll} = \sigma_l^2$ and covariance terms, $\Sigma_{12} = \Sigma_{21} = \rho\sigma_1\sigma_2$. For ease of development, we will exclusively work with the latter, explicit form. Under this assumption Su and Liu (1993) showed that the formulae for coefficients leading to the BLC's ROC curve is

$$\vec{\beta}_0^T \propto (\vec{\mu}_X - \vec{\mu}_Y)^T (\Sigma_X + \Sigma_Y)^{-1}, \quad (1)$$

which in turn leads to,

$$AUC_0 = P(U > V) = \Phi \left(\sqrt{(\vec{\mu}_X - \vec{\mu}_Y)^T (\Sigma_X + \Sigma_Y)^{-1} (\vec{\mu}_X - \vec{\mu}_Y)} \right). \quad (2)$$

Now suppose that the biomarker levels are measured with fixed LODs $d = (d_1, d_2)^T$. Let the measured observations, Z_X and Z_Y , be the componentwise transformation of X and Y , respectively, such that

$$Z_{X_l} = \begin{cases} X_l; & X_l \geq d_l, \\ N/A; & X_l < d_l, \end{cases} \quad \text{and} \quad Z_{Y_l} = \begin{cases} Y_l; & Y_l \geq d_l, \\ N/A; & Y_l < d_l, \end{cases}$$

where for a fixed d_l , $l = 1, 2$, the l -th biomarker level is either quantified or not. Without loss of generality we assumed that both diseased and non-diseased biomarker measurements were affected by the same point of censoring, say $d_{Xl} = d_{Yl} = d_l$.

Lyles et al. (2001) considered the case of two censored bivariate normally distributed random variables and developed the likelihood

$$\begin{aligned} L(\vec{\mu}, \Sigma; \vec{Z}_w) = & \prod_{\vec{Q}=(1,1)} \frac{1}{\sigma_1 \sigma_2 \sqrt{1-\rho^2}} \phi(\eta_{1j}) \phi(\eta_{2|1j}^*) \\ & \times \prod_{\vec{Q}=(1,0)} \frac{1}{\sigma_1} \phi(\eta_{1j}) \Phi(\eta_{d2|1j}^*) \\ & \times \prod_{\vec{Q}=(0,1)} \frac{1}{\sigma_2} \phi(\eta_{2j}) \Phi(\eta_{d1|2j}^*) \\ & \times \prod_{\vec{Q}=(0,0)} \Phi(\eta_{d1j}) \Phi(\eta_{d2|1j}^*), \end{aligned} \quad (3)$$

where $j = 1, \dots, n$, $\eta_{d1j}^* = (w_{2j} - \mu_2 - \rho\sigma_2/\sigma_1(w_{1j} - \mu_1))/(\sigma_2 \sqrt{1-\rho^2})$, \vec{Q} is a vector of indicator functions with $Q_l = 1$ if $w_l \geq d_l$ and $Q_l = 0$ otherwise, and ϕ and Φ are the univariate standard normal *pdf* and *cdf*, respectively. Given random samples, \mathbf{Z}_X and \mathbf{Z}_Y , of two biomarkers levels measured with LODs in n_X and n_Y individuals, respectively, we can maximize Eq. (3) in order to generate MLEs for underlying normal parameters, $\vec{\theta} = (\vec{\mu}, \vec{\Sigma}) = (\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$, for cases and for controls. Substituting these estimators for the appropriate parameters in Eqs. (1) and (2), the MLE's $\hat{\beta}_0$ and \hat{AUC}_0 are formed based on samples with multiple biomarkers affected by LODs.

3 Asymptotic results

Previously, the MLE \hat{AUC} , for a single biomarker affected by an LOD (Perkins et al., 2007), was developed along with $1-\alpha$ level CI formed using the asymptotic properties of \hat{AUC} .

Applying those developments to $\hat{AUC}_0 = AUC_0(\hat{\theta}_X, \hat{\theta}_Y)$ results in

$\sqrt{n_X + n_Y}(\hat{AUC}_0 - AUC_0) \sim N(0, \sigma_A^2)$ where \sim denotes the asymptotic distribution. Again similar to the univariate case, the variance σ_A^2 is obtained by the standard delta method

$$\begin{aligned} \sigma_A^2 = & \lambda^{-1} \left(\frac{\partial AUC_0}{\partial \vec{\theta}_X} \right)^T \left(\lim_{n_X \rightarrow \infty} n_X \text{Cov}(\hat{\theta}_X) \right) \left(\frac{\partial AUC_0}{\partial \vec{\theta}_X} \right) \\ & + (1-\lambda)^{-1} \left(\frac{\partial AUC_0}{\partial \vec{\theta}_Y} \right)^T \left(\lim_{n_Y \rightarrow \infty} n_Y \text{Cov}(\hat{\theta}_Y) \right) \left(\frac{\partial AUC_0}{\partial \vec{\theta}_Y} \right) \end{aligned} \quad (4)$$

with $\lambda = n_X/(n_X+n_Y) \in (0, 1)$ as $n_X \rightarrow \infty$ and $n_Y \rightarrow \infty$ and $(\text{AUC}_0/\theta)_{ij}^{\rightarrow} = \text{AUC}_0/\theta_{ij}^{\rightarrow}$ being the ij -th element of $(\text{AUC}_0/\theta)^{\rightarrow}$. The covariance matrices of MLE's of unknown vectors of parameters are evaluated by the inverse of the Fisher information matrices, say I_X, I_Y , i.e.

$\text{Cov}(\vec{\theta}_X) = [I_X]^{-1}$ and $\text{Cov}(\vec{\theta}_Y) = [I_Y]^{-1}$. The asymptotic properties of $\hat{\beta}_0$ mirror those for

AUC_0 with distribution $\sqrt{n_X+n_Y}(\hat{\beta}_0^T - \vec{\beta}_0^T) \sim N_p(\vec{0}, \Sigma_{\beta})$, where $\vec{0} = (0, 0)^T$ and the standard delta method is used to obtain the 2×2 covariance matrix

$$\Sigma_{\beta} = \lambda^{-1} \left(\frac{\partial \vec{\beta}_0^T}{\partial \vec{\theta}_X} \right)^T \left(\lim_{n_X \rightarrow \infty} n_X \text{Cov}(\hat{\theta}_X) \right) \left(\frac{\partial \vec{\beta}_0^T}{\partial \vec{\theta}_X} \right) + (1-\lambda)^{-1} \left(\frac{\partial \vec{\beta}_0^T}{\partial \vec{\theta}_Y} \right)^T \left(\lim_{n_Y \rightarrow \infty} n_Y \text{Cov}(\hat{\theta}_Y) \right) \left(\frac{\partial \vec{\beta}_0^T}{\partial \vec{\theta}_Y} \right) \quad (5)$$

with $(\partial \vec{\beta}_0^T / \partial \vec{\theta})_{ij} = \partial \vec{\beta}_{0j}^T / \partial \theta_i$ being the ij -th element of $\partial \vec{\beta}_0^T / \partial \vec{\theta}$. To calculate CI's for the elements of $\hat{\beta}_0$ and AUC_0 , we must find Σ_{β} and σ_A^2 , respectively, which consist of the covariance of the MLE's $\hat{\theta}_X$ and $\hat{\theta}_Y$ and the partial derivatives of $\vec{\beta}_0^T$ and AUC_0 with respect to each parameter. Here we are considering the case where two biomarkers' levels are bivariate normally distributed with unknown parameters $\vec{\mu}_Y$ and Σ_Y . The covariances, $\text{Cov}(\hat{\theta}_X)$ and $\text{Cov}(\hat{\theta}_Y)$, can be determined by

$$\text{Cov}(\vec{\theta}) = \text{Cov}(\hat{\mu}_1, \hat{\sigma}_1, \hat{\mu}_2, \hat{\sigma}_2, \hat{\rho}) = [I]^{-1} = \begin{bmatrix} I_{11} & \cdots & I_{15} \\ \vdots & \ddots & \vdots \\ I_{51} & \cdots & I_{55} \end{bmatrix} \quad (6)$$

of bivariate normal distributions with LODs, where

$$I_{ab} = \lim_{n \rightarrow \infty} \frac{-1}{n} \text{E} \left[\frac{\partial^2 \log(L(\vec{\theta}; \vec{Z}))}{\partial \theta_a \partial \theta_b} \right] \quad (a, b=1, \dots, 5).$$

The details of Eq. (6) and the partial derivatives of $\vec{\beta}_0^T$ and AUC_0 for Eqs. (4) and (5) are found in the Supporting Information.

When all or a portion of the parameters are unknown, the MLE's $\hat{\mu}_X, \hat{\Sigma}_X, \hat{\mu}_Y$ and $\hat{\Sigma}_Y$ are substituted for the appropriate parameters and generate approximate variances Σ_{β} and σ_A^2 for $\hat{\beta}_0^T$ and the AUC_0 , respectively. We can then use this distributional information to assess the variability in our BLC coefficients, $\hat{\beta}_0^T$, and to approximate α -level CI's for AUC_0 by $\text{AUC}_0 \pm z_{\alpha/2} \sqrt{\sigma_A^2 / \sqrt{n_X+n_Y}}$.

4 Evaluation

Utilizing the R programming language, we simulated $B = 2000$ data sets of biomarker values via the `mvtnorm` package ($n_Y = n_X = 50, 100, 200$) for non-diseased and diseased individuals from various bivariate normal distributions with $\theta_Y = (\vec{\mu}_Y, \Sigma_Y) = (\mu_{Y1}, \sigma_{Y1}, \mu_{Y2}, \sigma_{Y2}, \rho_Y) = (0, 1, 0, 1, 0)$ and $\theta_X = (\vec{\mu}_X, \Sigma_X) = (\mu_{X1}, \sigma_{X1}, \mu_{X2}, \sigma_{X2}, \rho_X), (\sigma_{X1}^2, \sigma_{X2}^2) = (1, 1), (0.5, 0.5), (1, 0.5)$, mean $\vec{\mu}_X$ corresponding to $AUC_0 = 0.6, 0.7, 0.8, 0.9$ and $\rho_X = 0, 0.2, 0.5, 0.8$. For each simulated group of samples, $\hat{\vec{\mu}}_X, \hat{\Sigma}_X, \hat{\vec{\mu}}_Y$ and $\hat{\Sigma}_Y$ were calculated by maximizing Eq. (3) via the `optim` function with `method = "L-BFGS-B"` allowing for bounded results necessary for the variance/covariance terms. Using Eqs. (1) and (2), these estimators then lead to $\hat{\beta}_0^T$ and subsequently \hat{AUC}_0 . Additionally, we calculated the empirical \tilde{AUC}_0 using the same $\hat{\beta}_0^T$ and replacing missings with the often used replacement values $a_l = d_l, l = 1, 2$, for each biomarker although any value $0 < a < d$ will result in the same \tilde{AUC}_0 for these positive $\hat{\beta}_0^T$. Using methods laid out in Section 3 and bootstrapped quantiles (1000 resamplings using the `boot` function), 95% CI's for AUC_0 accompany \hat{AUC}_0 and \tilde{AUC}_0 , respectively. We calculated these estimators first with no LOD, no missingness, to appreciate the baseline properties of the estimators under these various conditions and then applied increasing levels of censoring, where values of d were chosen to be the 20, 40, 60 and 80th quantiles of marginal distributions of the non-diseased population.

The $\hat{\beta}_0^T$ themselves were assessed and, while bias increased slightly as missingness increased, the proportionality of the coefficients' bias remained consistent with the proportion of the true coefficients, e.g. $\hat{\beta}_{01}/(\hat{\beta}_{01} + \hat{\beta}_{02}) \approx \beta_{01}/(\beta_{01} + \beta_{02})$. This is an excellent result as Eq. (1) only requires proportionality in the coefficients. As expected, the estimated relative bias, $(\hat{AUC}_0 - AUC_0)/AUC_0$, and root mean squared error (RMSE) of \hat{AUC}_0 decrease as sample sizes increase. The ranges of relative bias were 0.0027 to 0.1463, 0.0019 to 0.0880, 0.0003 to 0.0446 and were 0.0245 to 0.1504, 0.0174 to 0.0973, 0.0121 to 0.0618 for RMSE, respectively for $n_Y = n_X = 50, 100, 200$. Figure 1 depicts the relative bias and RMSE for $(\sigma_{X1}^2, \sigma_{X2}^2) = (1, 1)$; the $(\sigma_{X1}^2, \sigma_{X2}^2) = (0.5, 0.5)$ and $(1, 0.5)$ cases display similar relations in direction and magnitude regarding increasing sample size and percent missing. Relative bias in Fig. 1 clearly increases as ρ_X and percent missing increase and decreases as a function of sample size and AUC_0 . Changes in relative bias are slight for all levels of ρ_X and from 0 to 40% missing, increasing moderately at 60% and only substantially from 60 to 80% missing. Generally, \hat{AUC}_0 generated from only $n_X = n_Y = 100$, with over half of those values missing, has bias that is less than two percent of AUC_0 , which is generally equivalent to the relative bias of \hat{AUC}_0 based on a full data set. The second row of plots in Fig. 1 show that RMSE of \hat{AUC}_0 has generally the same relationship with sample size, correlation, percent missing and level of discrimination as with relative bias. Again, \hat{AUC}_0 calculated using only $n_X = n_Y = 100$, with over half of those values missing, has relative RMSE around one percent, which is comparable to that of \hat{AUC}_0 based on all the data.

The empirical estimator \tilde{AUC}_0 , is obviously unbiased for the discriminatory ability of the BLC of the biomarkers as measured with ties for missing values below the LOD. However, Fig. 2 displays a plotting of pairs of relative bias of \hat{AUC}_0 and \tilde{AUC}_0 with regard to the underlying latent AUC_0 for the various scenarios with $n_x=n_y=50$. This depiction shows,

relatively, how consistently \tilde{AUC}_0 can estimate AUC_0 with the $\hat{\beta}_0^T$ and the current measurements. Toward this end, scenarios of relative bias of \tilde{AUC}_0 close to the horizontal dashed line indicating consistent estimates range from higher levels of AUC_0 , (triangles=0.8, cross=0.9) and for 0, 20 and 40% missingness and also for lower AUC_0 with 40 and 60% missingness. In the latter instance, \tilde{AUC}_0 displayed better consistency for AUC_0 than \hat{AUC}_0 for these small sample sizes. Figure 2 also shows the potential bias of using \hat{AUC}_0 in lieu of \tilde{AUC}_0 when measurement cannot be improved.

The coverage probability of 95% CI's that accompanied \tilde{AUC}_0 were nominal or near nominal coverage for all AUC_0 ranging from 0.890 to 0.964, 0.913 to 0.965 and 0.933 to 0.960 for $n_x=n_y=50, 100$ and 200, respectively. No discernable patterns exist regarding $\rho_X=0, 0.2, 0.5, 0.8$ or the percent missing due to the LOD. It should be noted that unlike Reiser and Faraggi's (1997) CI for AUC using complete data, our CI's do not take into account the bounded nature of the AUC and coverage may be adversely affected near these bounds. However, our simulation did not reflect any difficulties for the sample sizes and levels of discrimination shown. Figure 3 displays the width of MLE and bootstrapped CI's, open and solid points, respectively, versus their coverage probability for our smallest sample size, $n_x=n_y=50$. Generally, CI's increase in width to reflect uncertainty from increasing missingness, where larger point size corresponds to a larger proportion missing. Additional plots for the larger sample sizes displaying increased coverage and tighter widths are included in the Supporting Information.

5 Example

To illustrate our method, we use PCBs, environmental toxicants, as potential indicators of endometriosis (Louis et al., 2005). Endometriosis is a gynecological disease exclusive to species that menstruate such as humans and other primates, occurring predominantly in women of reproductive age. Data from experimental studies in animals and observational human studies suggest an association between dioxin and PCBs and endometriosis. In our data, PCBs 153 and 180 were measured in 28 women with and 51 women without endometriosis. The biomarkers were measured jointly and were expected to be correlated. However, the sensitivity of the measurement process differed for each biomarker with PCB 153 having an LOD of $d_{153}=0.2$, resulting in 64% of the cases and 74% of the controls having unobservable levels and PCB 180 having an LOD of $d_{180}=0.034$, resulting in no missing cases but 11% of the controls having unobservable levels (Whitcomb et al., 2005).

Analyzing the biomarkers univariately, empirical methods led to $\tilde{AUC}_{153} = 0.564$ and $\tilde{AUC}_{180} = 0.609$. As measured, PCB 180 and 153 appear to have some discriminatory ability for women with endometriosis, greater for PCB 180. To investigate what potential might lie below the LODs, we can employ parametric techniques. Univariate normal distributions were assessed on the log transformed biomarkers similar to Lyles et al. (2001)

via q - q plots. Figure 4 displays q - q plots where both biomarkers fit log normal distributions well, as evidenced by diagonal points corresponding to the points above the LODs. The horizontal points are essentially quantile place holders for the observations below the LOD. Assuming that the data we don't see follow the data we do see, univariate normal likelihoods led to the MLE's $\hat{AUC}_{153} = 0.511$ and $\hat{AUC}_{180} = 0.630$ for PCB 153 and PCB 180, respectively (Perkins et al., 2007). Again, PCB 180 displays a moderate ability to differentiate but the potential of PCB 153 seems to have been diminished. The result $\hat{AUC}_{153} > \hat{AUC}_{180}$ is due to a large "hook" in the ROC curve (Pesce and Metz, 2007) resulting from disparate variances. This behavior is rarely reflective of actual etiology but rather is likely to be a result of highly variable parameter estimates given the small sample sizes and large amount of missings. Regardless, PCB 153 might be discarded as potentially lacking discriminatory ability for endometriosis.

However, these two biomarkers are closely linked biologically and are suspected as being highly correlated as a result of being in a mixture of environmental exposures. Using a variety of methods, we considered the joint distribution of PCB 153 and 180 to investigate whether PCB 153 might have some potential as a contributor in a BLC. First, Eq. (3) was maximized using log transformed PCB levels, MLEs for distribution parameters were

calculated as well as $\hat{\beta}_0^T = \{\beta_{153}, \beta_{180}\} = \{-1.73, 1.65\}$ via Eq. (1). Using Eq. (2), we estimated that the potential from this BLC if measurements could be improved to be $\hat{AUC}_0 = 0.747$ with 95% CI (0.580, 0.915). This result shows that the BLC of PCBs 153 and 180 could be far superior to either biomarker alone and is potentially a good discriminator of women with and without endometriosis if the measurement sensitivity of PCB 153 could be improved. However, the sample sizes here have led to a significant but wide CI which does not rule out $AUC_0 < 0.630$.

Table 1 compares \hat{AUC}_0 to several alternative estimators of the potential AUC_0 . Naïve methods to estimate potential AUC_0 in Table 1 shows the results of using replacement values (e.g. imputing $a_I = d_I/2$ for values below the LODs) and standard parametric methods to obtain naïve BLC, $\hat{\beta}_{OR}^T$. Imputing values in this fashion clearly violates the assumption of bivariate normality. However, we generated such estimates for the sake of comparison since standard methodology is often applied to data sets with imputed values. While these results are a slight improvement over the univariate biomarkers, it is vastly less than the proposed \hat{AUC}_0 . These naïve cases also point to widely varying linear combinations, as in column 3 of Table 1.

The second set of estimates in Table 1 begins with this BLC's current diagnostic ability, where \hat{AUC}_0 was estimated empirically using various replacement values (d , $d/2$ and $a_I = E[X_I | X_I < d_I]$, the expected values below the LODs) and the mles $\hat{\beta}_0^T$ estimated previously. These estimates vary but all are less than \hat{AUC}_{180} and even \hat{AUC}_{153} . Alternatively, we could use these replacement values in conjunction with simple logistic regression as a means to identify the "best" coefficients to estimate \hat{AUC}_0 as measured. Lynn (2001) showed that in some cases these naïve methods can provide logistic regression coefficient estimates comparable to those using a full maximum likelihood approach for logistic regression

accounting for an LOD. These “logistic” $A\tilde{U}C_0$ in Table 1 are on par with those using the $\frac{\hat{\Delta}T}{\hat{\beta}_0}$, “mle,” likely due to the logistic coefficients having the same signs and not grossly different proportionality. However, they do shift some of the focus to PCB 180 as evidenced by the slightly larger proportion of the linear combination, making the subsequent $A\tilde{U}C_0$ better than $A\tilde{U}C_{153}$. The ordering $A\tilde{U}C_{153} < A\tilde{U}C_0 < A\tilde{U}C_{180}$ is due to the difference in signs for the coefficients and in percent below the LOD for the two biomarkers. If both coefficients were positive, the order of PCB 180 would be preserved when ties from PCB 153 are introduced to a BLC and thus $A\tilde{U}C_0 > A\tilde{U}C_{180}$. As measured, the BLC of PCBs 153 and 180 would not be of use beyond what PCB 180 can already achieve, however, Fig. 2 shows us that we need only to reduce the missingness in PCB 153 to between 20 and 40% to realize most of the difference between the potential latent and as measured effectiveness.

6 Discussion

While multiple biomarkers are often available or obtainable for the same outcome, investigators often ignore all but one and in doing so essentially throw out potential ability to discriminate beyond the biomarker they have chosen. For this reason and where biomarkers can be assumed to be normally distributed, the proposed BLC should be used in lieu of a single marker, especially in cases where multiple biomarkers come at no additional cost, as is the case with multiplex assays.

Using two biomarkers measured with LODs can result in one or both measured with censored values for an individual. The methods developed here allow us to properly account for the missing values while simultaneously taking advantage of the benefits in discriminatory ability realized by a BLC. While replacement has been shown previously to be useful in certain situations as an ad hoc approach for the AUC in the univariate case, we’ve shown here that standard parametric methods with replacement values can yield estimates of the potential AUC_0 that vary greatly from the MLE (Perkins et al., 2007). For this reason, replacing missing values by a constant is not recommended when estimating, even naively, the potential of a BLC of biomarkers.

The method proposed here is related to the approaches for handling censored covariates in logistic regression setting. Lynn (2001) discusses an approach for estimating the linear combination of subject-to-LOD covariates for prediction of a binary outcome. Due to the invariance of the ROC curve with respect to order-preserving transformations and because of the assumed multivariate normality of the covariates, the potential AUC_0 for Lynn’s linear combination can be computed using the standard formulation for binormal AUC. The novelty of the approach developed here is that the proposed BLC targets maximization of the AUC_0 rather than maximization of the likelihood under the logistic regression model. Depending on the parameters of the distributions the results can be either similar or substantially different (Pepe et al., 2006).

While the likelihood ratio has been demonstrated to achieve the highest possible AUC for a given set of biomarkers, in the case of two bivariate normally distributed biomarkers the absolute difference in estimated AUC can be minimal. However, this is not always the case as likelihood ratio ROC curves are assuredly “proper” or completely concave while

conventional “binormal” ROC curves and those based on linear combinations of normal biomarkers can have minimal to severe non-concave portions or “hooks” (e.g. $\sigma_x^2 \gg \sigma_y^2$, $(\beta^T \Sigma_x \beta) \gg (\beta^T \Sigma_y \beta)$ or $b \ll 1$ from Pesce and Metz (2007)). The linear combination here has the benefit of being simple to interpret and can incorporate the empirical estimate of the biomarkers’ effectiveness as currently measured where the likelihood ratio does not. More importantly though is that in the context of the latent discriminatory ability, both would work similarly in identifying biomarkers with potential that could be measured more precisely and more extensively in the future. While, Section 3 considers the asymptotic properties of $A\hat{U}C_0$ for the BLC of two multivariate normal biomarkers, Section 4 shows small sample performance that is nearly unbiased and achieves nominal coverage probability for sample sizes as little as 100 with 20, 40 and 60% missing values due to LODs. The need for further investigation regarding magnitude and potential causes of differences between the likelihood- ratio and conventional “binormal” ROC curves in the context here remains.

As with all parametric estimation, this BLC of two biomarkers corrected for LODs is dependent on distributional assumptions specifically that the biomarkers or their transforms are multivariate normally distributed. The literature (Molodianovitch et al., 2006; Perkins et al., 2007) has shown MLEs of $A\hat{U}C$ for the univariate case are robust to minor deviations from the normal assumption and unfavorable results when assumptions are grossly violated. Estimates of the AUC based on replicates of a single biomarker, a special case of the multivariate normal developments here, have shown favorable robustness by generating biomarker levels from gamma distributions (Perkins et al., 2009). While the bias was on average less than one percent of the AUC being estimated, caution and diligence should still be taken when employing this BLC resulting in $A\hat{U}C_0$.

A biomarker may provide differential discrimination based on some associated factor. The inclusion of this covariate information can of course be incorporated by conducting a stratified analysis for each level of such a factor. A better approach might be to directly include the covariate information in the likelihood function.

In our example we applied several methods to empirically estimate the as measured discriminatory ability, $A\tilde{U}C_0$, by using simple replacement values in standard logistic regression or with $\hat{\beta}_0^T$. Lynn (2001) showed that in some cases, naïve replacement could yield similar logistic regression coefficient estimates to those estimated using maximum likelihood accounting for the LOD properly. We found a similar concordance here in $A\tilde{U}C_0$ by virtue of the naïve logistic coefficients and our $\hat{\beta}_0^T$ yielded similarly proportional estimated coefficients. However, instances where this is not the case surely exist and care should be taken to investigate potential discordance before applying either BLC in practice.

These methods for estimating AUC_0 while properly accounting for missing data censored below LODs will provide an asymptotically unbiased view of the potential discriminatory ability of the BLC of a set of two biomarkers. We must note again that the $A\hat{U}C_0$ in this scenario reflect potential discriminatory ability and cannot be realized in practice until observations can be measured and thus allow for differentiation below the LODs. This is

often the case where a less sensitive, but cheaper assay could be conducted to narrow down potential biomarkers before conducting more expensive but more precisely measured assays. We have also demonstrated the usefulness of the empirical estimate of what is possible “as measured” and how the majority of two biomarkers’ latent discriminatory ability might be realized by improving measurement, reducing missingness to between 20 and 40% rather than to zero. Estimating this potential is increasingly important as it allows researchers to focus limited resources to improve the measurement process in biomarkers that display promise and subsequently lead to improved diagnostic care.

Acknowledgments

The authors thank the Editor and reviewer for their valuable comments and suggestions that have added greatly to this manuscript. This research was supported by the Intramural Research Program of the Epidemiology branch of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH.

References

- Haas CH, Scheff PA. Estimation of averages in truncated samples. *Environmental Science & Technology*. 1990; 24:912–919.
- Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. *Applied Occupational and Environmental Hygiene*. 1990; 5:46–51.
- Louis GM, Weiner JM, Whitcomb BW, Sperrazza R, Schisterman EF, Lobdell DT, Crickard K, Greizerstein H, Kostyniak PJ. Environmental PCB exposure and risk of endometriosis. *Human Reproduction*. 2005; 20:279–285. [PubMed: 15513976]
- Lyles RH, Williams JK, Chuachoowong R. Correlating two viral load assays with known detection limits. *Biometrics*. 2001; 57:1238–1244. [PubMed: 11764265]
- Lynn HS. Maximum likelihood inference for left-censored HIV RNA data. *Statistics in Medicine*. 2001; 20:33–45. [PubMed: 11135346]
- Molodianovitch K, Faraggi D, Reiser B. Comparing the areas under two correlated ROC curves: parametric and non-parametric approaches. *Biometrical Journal*. 2006; 48:745–757. [PubMed: 17094340]
- Pepe, MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press; New York: 2003.
- Pepe MS, Cai T, Longton G. Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometrics*. 2006; 62:221–229. [PubMed: 16542249]
- Perkins NJ, Schisterman EF, Vexler A. Receiver operating characteristic curve inference from a sample with a limit of detection. *American Journal of Epidemiology*. 2007; 165:325–333. [PubMed: 17110640]
- Perkins NJ, Schisterman EF, Vexler A. Generalized ROC curve inference for a biomarker subject to a limit of detection and measurement error. *Statistics in Medicine*. 2009; 28:1841–1860. [PubMed: 19340817]
- Pesce LL, Metz CE. Reliable and computationally efficient maximum-likelihood estimation of “proper” binormal ROC curves. *Academic Radiology*. 2007; 14:814–829. [PubMed: 17574132]
- Reiser B, Faraggi D. Confidence intervals for the generalized ROC criterion. *Biometrics*. 1997; 53:644–652. [PubMed: 9235120]
- Singh A, Nocerino J. Robust estimation of mean and variance using environmental data sets with below detection limit observations. *Chemometrics and Intelligent Laboratory Systems*. 2002; 60:69–86.
- Su JQ, Liu JS. Linear combinations of multiple diagnostic markers. *Journal of the American Statistical Association*. 1993; 88:1350–1355.

- Whitcomb BW, Schisterman EF, Buck GM, Weiner JM, Greizerstein H, Kostyniak PJ. Relative concentrations of organochlorines in adipose tissue and serum among reproductive age women. *Environmental Toxicology and Pharmacology*. 2005; 19:203–213. [PubMed: 21783478]
- Zhou, XH.; Obuchowski, NA.; McClish, DK. *Statistical Methods in Diagnostic Medicine*. Wiley; New York: 2002.

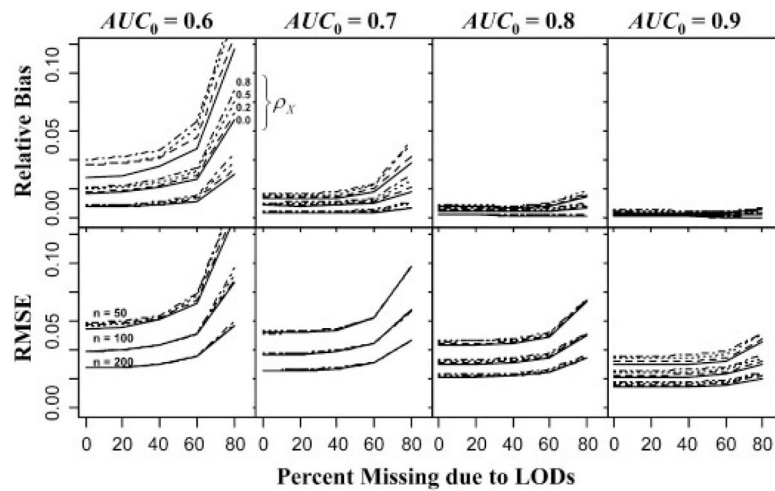


Figure 1.

Relative bias and RMSE of maximum likelihood estimates of various levels of AUC_0 based on simulated bivariate normally distributed data of various sample sizes and correlations as a function of the percent of the non-diseased population missing due to LODs. Measurements of diseased were also affected by the LODs but to a degree that lessens as AUC_0 increases.

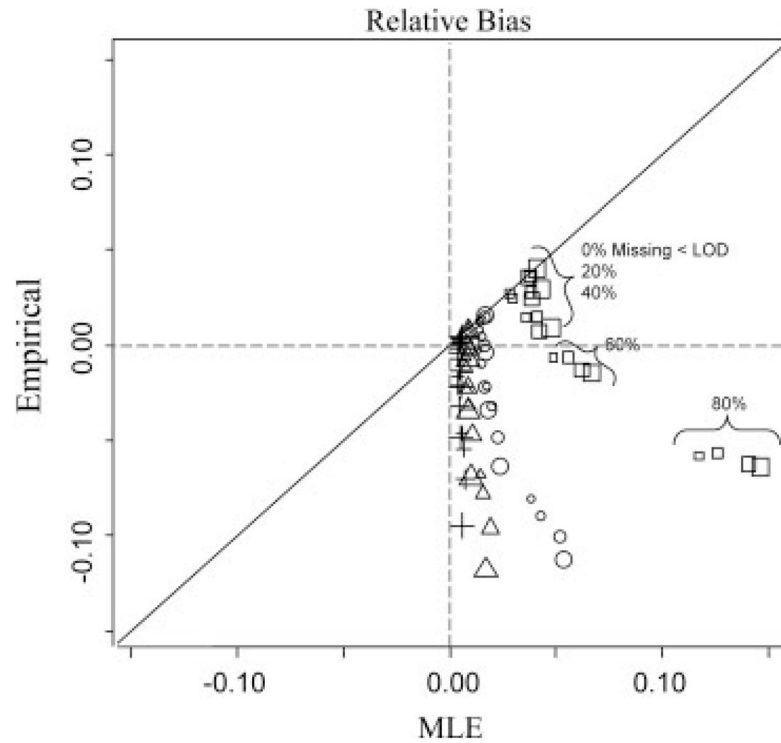


Figure 2. Relative bias of maximum likelihood versus empirical estimators of AUC_0 based on the estimated BLC of simulated bivariate normally distributed data of various correlations (point size) and missingness due to LODs from samples of 50 health and 50 diseased. Squares, circles, triangles and crosses indicate estimators with true $AUC_0=0.6, 0.7, 0.8$ and 0.9 , respectively.

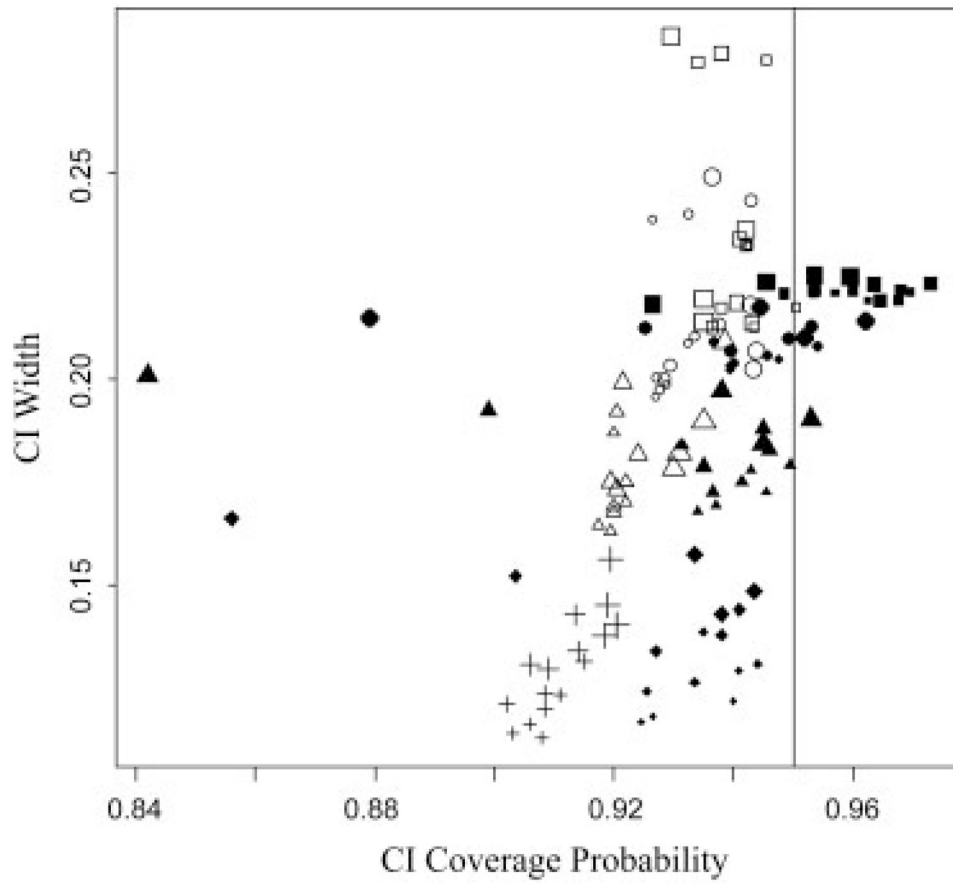


Figure 3. Confidence interval (CI) width and coverage probability of 95% CI's estimated via maximum likelihood (open points) or bootstrapped percentiles (solid points) that accompany ML and empirical estimates of the AUC_0 , respectively. True $AUC_0=0.6, 0.7, 0.8$ and 0.9 are identified by squares, circles, triangles and crosses/diamonds, respectively.

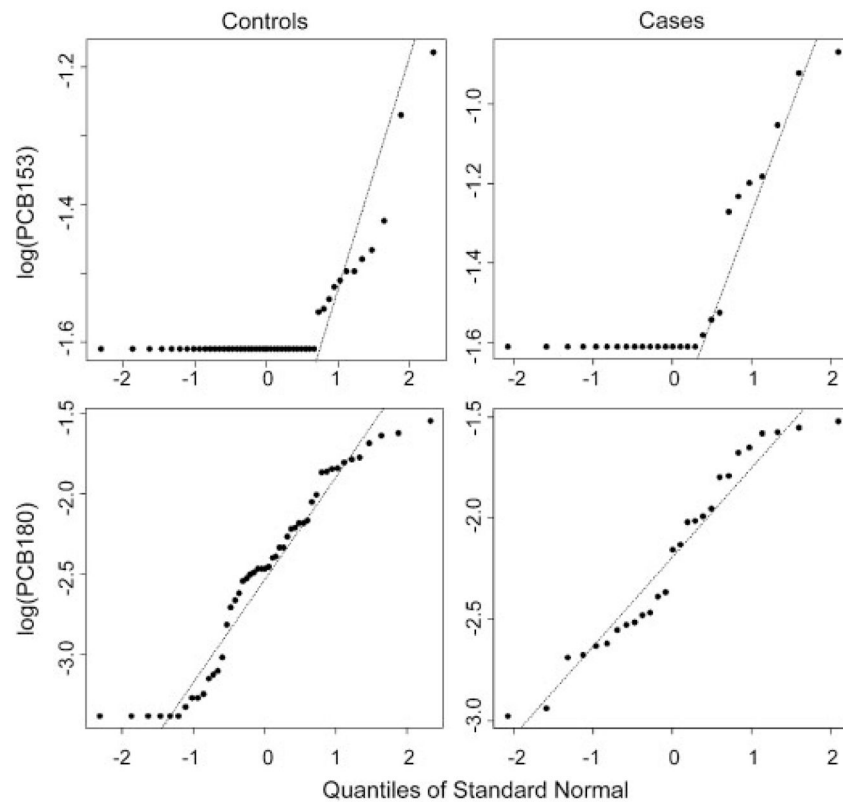


Figure 4. q–q normal plots of the log transformed biomarker levels of PCBs 153 and 180 for women with endometriosis (cases) and without (controls). Horizontal points correspond to observations censored below an LOD.

Table 1

Estimates of linear combinations of log PCBs 153 and 180 measured with LODs intending to maximize the AUC_0 for women with and without endometriosis using various methods.

| Type of AUC_0 | Estimation method ^{a)} | Estimates | |
|-----------------|---------------------------------|------------------------------|---------|
| | | $(\beta_{153}, \beta_{180})$ | AUC_0 |
| Potential | mle | (-1.733, 1.651) | 0.747 |
| | naïve mle (d) | (0.927, 0.204) | 0.639 |
| | naïve mle ($d/2$) | (0.014, 0.390) | 0.639 |
| | naïve mle ($E[x x<d]$) | (0.138, 0.346) | 0.629 |
| As measured | mle/emp (d) | (-1.733, 1.651) | 0.557 |
| | mle/emp ($d/2$) | (-1.733, 1.651) | 0.534 |
| | mle/emp ($E[x x<d]$) | (-1.733, 1.651) | 0.532 |
| | logistic/emp (d) | (-0.322, 0.472) | 0.581 |
| | logistic/emp ($d/2$) | (-0.348, 0.557) | 0.553 |
| | logistic/emp ($E[x x<d]$) | (-0.195, 0.449) | 0.575 |
| | naïve mle/emp (d) | (0.927, 0.204) | 0.608 |
| | naïve mle/emp ($d/2$) | (0.014, 0.390) | 0.607 |

The AUC_0 estimated are for the biomarker as measured with missing values below the LODs and for the potential of the latent biomarkers if they could be measured completely.

^{a)} Method for estimating $(\beta_{153}, \beta_{180})$ /method for estimating AUC_0 (replacement value used).