Communications for Statistical Applications and Methods
2022, Vol. 29, No. 5, 533–545

https://doi.org/10.29220/CSAM.2022.29.5.533
Print ISSN 2287-7843 / Online ISSN 2383-4757

Estimating the AUC of the MROC curve in the presence of measurement errors

Siva G[a], Vishnu Vardhan R[1,a], Asha Kamath[b]

[a]Department of Statistics, Pondicherry University, India;
[b]Department of Data Science, MAHE, Manipal, India

Abstract

Collection of data on several variables, especially in the field of medicine, results in the problem of measurement errors. The presence of such measurement errors may influence the outcomes or estimates of the parameter in the model. In classification scenario, the presence of measurement errors will affect the intrinsic cum summary measures of Receiver Operating Characteristic (ROC) curve. In the context of ROC curve, only a few researchers have attempted to study the problem of measurement errors in estimating the area under their respective ROC curves in the framework of univariate setup. In this paper, we work on the estimation of area under the multivariate ROC curve in the presence of measurement errors. The proposed work is supported with a real dataset and simulation studies. Results show that the proposed bias-corrected estimator helps in correcting the AUC with minimum bias and minimum mean square error.

Keywords: multivariate ROC curve, area under the curve, measurement errors, minimax approach

1. Introduction

In many statistical applications, we are interested in classifying subjects into one of two or more groups that have better percentage of correct classification. The Receiver Operating Characteristics (ROC) curve is one such classification technique that plays an important role in evaluating the diagnostic test performance and classifies the individuals into one of the predefined groups. It is a graphical plot between a false positive rate (1 − Specificity, $1 - S_p$ or FPR) and a true positive rate (Sensitivity, S_n or TPR) for all possible values of a variable of interest.

The ROC curve can be estimated under both parametric and non-parametric approaches. The most commonly used parametric ROC model is the *Bi-Normal ROC curve* (Egan, 1975), which is based on the assumption that both populations underlie normal distribution. Considering practical situations, the multivariate versions of the ROC curve were proposed by Su and Liu (1993); Schisterman *et al.* (2004); Yuan and Ghosh (2008); Yin and Tian (2014); and Sameera *et al.* (2016). The area under the curve (AUC) is an important summary measure that tells how well a classifier separates the individuals. Bamber (1975) stated the AUC of a ROC curve as "$\Delta = P(Y > X)$, which measure how well a biomarker distinguishes between the two groups, assuming both groups are normal where X and Y are the biomarker values on controls and cases, respectively."

In general, more care should be taken while collecting information on several markers, particularly in the field of medicine. Some markers may be susceptible to substantial measurement errors (ME),

[1] Corresponding author: Department of Statistics, Pondicuerry University, 605014 Puducherry, Kalapet, Chinna Kalapet, India. E-mail: vrstatsguru@gmail.com

Published 30 September 2022 / journal homepage: http://csam.or.kr

which may be attributed to instruments used in the laboratory, the lack of knowledge of the technicians, biological variability, temporal changes in subjects, etc. Shear *et al.* (1987) measured systolic and diastolic blood pressure on children, which were used as forecasters of future hypertension. It is well known that measurements are recorded by a person using laboratory instruments. To obtain the recordings of systolic and diastolic blood pressure, laboratory instruments are used by the technicians. However, such recordings may be susceptible to measurement errors if either the instruments or the technicians report faulty values. For more examples on measurement errors, readers can look at Begg and Greene (1983), Begg and McNeil (1988), and Berbaum *et al.*, (1989).

In the context of ROC analysis, the existence of such measurement errors leads to inconsistent estimates and also masks the true information. Coffin and Sukhatme (1996, 1997) came out with a bias-corrected estimator for AUC (Δ) in both parametric and non-parametric cases and the asymptotic distribution of these estimators were developed by Kim and Gleser (2000). Faraggi (2000) and Raiser (2000) focused on explaining the effect of measurement errors on the confidence intervals of Δ and derived a new confidence interval expression for Δ, which had the actual coverage. Using the reliability approach, Dunn (1989) and Schisterman *et al.* (2001) have also developed the corrected estimate for Δ based on the assumption that replicated observations are available on the study objects. Tosteson *et al.* (2005) and Perkins *et al.* (2009) proposed methodologies for correcting the AUC of a ROC curve using the replicate measures and the maximum likelihood methods with the assumption the existence of normal measurement errors.

These research works, which are discussed above, mostly focused on the univariate structure by developing procedures to correct the AUC of a ROC curve. However, we come across multiple markers in real situations. The existing procedures will not address the problem of measurement errors in estimating the AUC for such data. Hence, in this work, we have attempted to derive the approximate bias-corrected estimator for AUC under a multivariate normality assumption in the presence of measurement errors. Monte-Carlo simulation studies are performed at different sample sizes and estimated the AUC's along with the corresponding mean square error (MSE) values. The results are discussed in Section 3, which depicts how the bias and the MSE of the estimators are influenced by measurement errors. In addition, a real data set is also considered to demonstrate the effect or impact of measurement errors in estimating the AUC of a multivariate ROC (MROC) curve.

2. Methodology

In medicine, many markers are sensitive to different aspects of a disease. Thus, the validity of a single biomarker to judge an individual's health status may be reduced. For example, let us consider the dataset from The British United Provident Association (BUPA) on liver disorders that consists of 345 samples with seven markers, such as mean corpuscular volume (mcv), alkaline phosphotase (alkphos), alamine aminotransferase (sgpt), aspartate aminotransferase (sgot), gamma-glutamyl transpeptidase (gammagt), and the number of half-pint equivalents of alcoholic beverages drunk per day (drinks). Each marker has the disease status with or without the liver disorder. This scenario creates a necessity to develop a model that can accommodate more than one biomarker for better classification. Hence, the work presented in this paper is based on the Multivariate Receiver Operating Characteristic (MROC) model proposed by Sameera *et al.* (2016).

2.1. The MROC curve

Let $X \sim \text{MVN}_p(\mu_H, \Sigma_H)$ and $Y \sim \text{MVN}_p(\mu_D, \Sigma_D)$, where μ_H, μ_D, Σ_H and Σ_D are mean vectors and covariance matrices of Healthy (H) and Diseased (D) populations, respectively. The expressions for

the False Positive Rate (FPR) and the True Positive Rate (TPR) are defined as,

$$\text{FPR} = P(S > c \mid H) = 1 - \Phi\left(\frac{c - b'\mu_H}{\sqrt{b'\Sigma_H b}}\right) \tag{2.1}$$

and

$$\text{TPR} = P(S > c \mid D) = \Phi\left(\frac{b'\mu_D - c}{\sqrt{b'\Sigma_D b}}\right), \tag{2.2}$$

where $S = \{X, Y\}$, which is a data matrix that contains information on both H and D populations, 'c' is the cutoff and the vector 'b' contains 'p' coefficients, which are obtained through equation (2.3) and is the minimax procedure (Anderson and Bahadur, 1962),

$$b = [t\Sigma_D + (1-t)\Sigma_H]^{-1}(\mu_D - \mu_H) \tag{2.3}$$

where 't' is constant and is determined by the trial and error method in the interval 0 and 1. Using equation (2.1), we obtain $c = b'\mu_H + \sqrt{b'\Sigma_H b}\,\Phi^{-1}(1 - \text{FPR})$, and the MROC expression is

$$\text{ROC}(c) = \Phi\left[\frac{b'(\mu_D - \mu_H) - \sqrt{b'\Sigma_H b}\,\Phi^{-1}(1 - \text{FPR})}{\sqrt{b'\Sigma_D b}}\right]. \tag{2.4}$$

Using probabilistic notations, the AUC expression for the MROC will be

$$\Delta = \Phi\left[\frac{b'(\mu_D - \mu_H)}{\sqrt{b'(\Sigma_D + \Sigma_H)b}}\right]. \tag{2.5}$$

Under the MROC curve framework, the linear combination that acts as a classifier is

$$U = b'Z,$$

where $b' = [b_1, b_2, \ldots, b_k]$. Here, if $U > c$ then it is classified to the D population.
The maximum likelihood estimator of Δ is

$$\hat{\Delta} = \Phi\left[\frac{\hat{b}'(\hat{\mu}_D - \hat{\mu}_H)}{\sqrt{\hat{b}'\left(\hat{\Sigma}_D + \hat{\Sigma}_H\right)\hat{b}}}\right], \tag{2.6}$$

where $\hat{b} = [t\,\hat{\Sigma}_D + (1-t)\,\hat{\Sigma}_H]^{-1}(\hat{\mu}_D - \hat{\mu}_H)$ and $\hat{\Sigma}_H$, $\hat{\Sigma}_D$ are the estimated sample covariance matrices of the H and D populations. Using the Taylor series expansion, it can be easily shown that $E(\hat{\Delta}) = \Delta + O(n^{*^{-1}})$; where $n^* = \min(m, n)$, and m and n are the number of samples in the H and D populations, respectively.

2.2. Proposed bias-corrected approximation

If the observations are measured with error, then the 'true' values of the X and Y will be masked and the parameter estimates may not be reliable to proceed further in order to classify the individuals. In a notational manner, we define the affected variables as,

$$A_i = X_i + \Sigma_\epsilon, \quad i = 1, 2, \ldots, m,$$
$$B_j = Y_j + \Sigma_\eta, \quad j = 1, 2, \ldots, n,$$

where $\epsilon \sim N(0, b'\Sigma_\epsilon b)$ and $\eta \sim N(0, b'\Sigma_\eta b)$. The assumption of equal covariances of Σ_ϵ and Σ_η is not necessary, but in many instances, it is justified that the variations leading to measurement errors are due to laboratory errors, and do not depend on the individual's risk or the true variable value. The AUC of a MROC curve in the presence of a measurement error will be

$$\Delta^* = \Phi\left[\frac{b'(\mu_B - \mu_A)}{\sqrt{b'(\Sigma_B + \Sigma_A)b}}\right] \tag{2.7}$$

and the likelihood estimator for Δ^* is

$$\hat{\Delta}^* = \Phi\left[\frac{\hat{b}'\left(\bar{B} - \bar{A}\right)}{\sqrt{\hat{b}'\left(\hat{\Sigma}_B + \hat{\Sigma}_A\right)\hat{b}}}\right], \tag{2.8}$$

where $\hat{\Sigma}_A, \hat{\Sigma}_B$ are the estimated sample covariance matrices of the populations A and B. Here $\bar{B}$ and $\bar{A}$ are the estimated sample mean vectors of the B and A populations, respectively. In order to correct the bias in $\hat{\Delta}^*$, let $\boldsymbol{\delta} = \boldsymbol{\epsilon} - \boldsymbol{\eta} \sim N(0, b'\Sigma_\epsilon b + b'\Sigma_\eta b)$ and the unbiased estimate of $\hat{\Delta}^*$ can be written in the following form (Coffin and Sukhatme, 1996),

$$\begin{aligned} E\left(\hat{\Delta}^*\right) &\simeq P(Y > X + \delta) \\ &\simeq \iint [1 - G_Y(s+t)] f_X(s) f_\delta(t) dx ds \\ &\simeq \Delta - \frac{E(\delta\delta')}{2} \int g'_Y(s) f_X(s) ds \\ E\left[\hat{\Delta}^*\right] &\simeq \Delta + (-\Omega), \end{aligned}$$

where $\Omega = E(\delta\delta')/2 \int g'_Y(s) f_X(s) ds$, $G_Y(\cdot)$ is the cdf of Y and $f_\delta(\cdot)$ is the density function of the distribution of δ. In this case, $g_Y(\cdot)$ and $f_X(\cdot)$ denote the multivariate normal densities of the D and H populations, respectively. Thus, the Ω component reduces to

$$\begin{aligned} \Omega &= \frac{E(\delta\delta')}{2} \int g'_Y(s) f_X(s) ds \\ &= \frac{1}{2} \frac{b'(\Sigma_\epsilon + \Sigma_\eta)}{\sqrt{2\pi} K} \left(\frac{b'(\mu_D - \mu_H)}{\sqrt{K}}\right) \exp\left\{-\frac{1}{2}\left(\frac{b'(\mu_D - \mu_H)}{\sqrt{K}}\right)^2\right\}, \end{aligned} \tag{2.9}$$

where $K = b'(\Sigma_D + \Sigma_H)b$.

Using the unbiased estimates $\bar{Y}$ and $\bar{X}$ of μ_D and μ_H in equation (2.9), gives us the expression of Ω as,

$$\hat{\Omega} = \frac{b'\left(\hat{\Sigma}_\epsilon + \hat{\Sigma}_\eta\right)}{2\sqrt{2\pi}(\tau)} \left(\frac{b'\left(\bar{Y} - \bar{X}\right)}{\sqrt{\tau}}\right) \exp\left\{-\frac{1}{2}\left(\frac{b'\left(\bar{Y} - \bar{X}\right)}{\sqrt{\tau}}\right)^2\right\},$$

where $\tau = b'[(\hat{\Sigma}_D + \hat{\Sigma}_H) - (\hat{\Sigma}_\epsilon + \hat{\Sigma}_\eta)]b$, then the bias-corrected estimator of Δ is

$$\tilde{\Delta} = \hat{\Delta}^* + \hat{\Omega}, \tag{2.10}$$

Table 1: $X_1 \sim MVN(\boldsymbol{\mu}_H, \Sigma_H), X_2 \sim MVN(\boldsymbol{\mu}_D, \Sigma_D)$

Sets	μ_H	μ_D	Σ_H	Σ_D
A	$\begin{pmatrix}17.9\\16.2\end{pmatrix}$	$\begin{pmatrix}20.8\\18.0\end{pmatrix}$	$\begin{pmatrix}2 & 1\\1 & 4\end{pmatrix}$	$\begin{pmatrix}2 & 1\\1 & 6\end{pmatrix}$
B	$\begin{pmatrix}20.8\\18.0\end{pmatrix}$	$\begin{pmatrix}20.8\\18.0\end{pmatrix}$	$\begin{pmatrix}2 & 1\\1 & 4\end{pmatrix}$	$\begin{pmatrix}2 & 1\\1 & 4\end{pmatrix}$
C	$\begin{pmatrix}17.9\\16.2\end{pmatrix}$	$\begin{pmatrix}20.8\\18.0\end{pmatrix}$	$\begin{pmatrix}2 & 1\\1 & 4\end{pmatrix}$	$\begin{pmatrix}2 & 1\\1 & 4\end{pmatrix}$
D	$\begin{pmatrix}20.8\\18.0\end{pmatrix}$	$\begin{pmatrix}20.8\\18.0\end{pmatrix}$	$\begin{pmatrix}2 & 1\\1 & 4\end{pmatrix}$	$\begin{pmatrix}2 & 1\\1 & 6\end{pmatrix}$

which is the corrected estimator for Δ in the presence of measurement errors.

To demonstrate the proposed methodology and to showcase the behaviour of Σ_ϵ and Σ_η, simulations for different cases were carried out and are presented in the next section. The three different cases are:

1. both Σ_ϵ and Σ_η have common variances, i.e., $\Sigma_\epsilon = \Sigma_\eta = (\sigma^2)$,
2. both Σ_ϵ and Σ_η have different variances, i.e., $\Sigma_\epsilon = \Sigma_\eta = (\sigma_i^2);\ i = 1, 2, \ldots, n$,
3. and some of the markers does not get influenced by the measurement errors. For instance, let there be four markers M_1, M_2, M_3, and M_4, and of these, if M_2 does not get influenced with measurement error, then the covariance matrix is given by (assuming equal variances for the remaining markers),

$$\Sigma_\epsilon = \Sigma_\eta = \begin{bmatrix}\sigma^2 & 0 & 0 & 0\\ & 0 & 0 & 0\\ & & \sigma^2 & 0\\ & & & \sigma^2\end{bmatrix}.$$

3. Simulation study I

For each case of the error covariance matrices, Monte-Carlo simulations were performed to assess the behavior of the proposed bias-corrected estimator under the structure of a bivariate normal distribution. In Table 1, four sets are considered, and sets A & C are defined to explain the case of better classification with unequal and equal covariance matrices, respectively. Sets B & D are defined with equal mean vectors; and equal and unequal covariance matrices, respectively, which are considered in order to explain the worst-case scenarios. In each population, random samples of size $n = \{25, 50, 100, 200, 500\}$ were generated using the parameter values defined in the four sets.

- Case (i) Both $\Sigma_\epsilon = \Sigma_\eta = (\sigma^2)$

In this case, it is assumed that both the error covariance matrices have same variances.

$$\text{i.e.,} \quad \Sigma_\epsilon = \Sigma_\eta = \begin{pmatrix}1.3 & 0.0\\0.0 & 1.3\end{pmatrix}.$$

Table 2 contains the estimated values of the AUC, with and without ME, and also bias-corrected AUC. The discrepancy in the AUC is observed through the bias and MSE.

The results of set A & C (better classification) show that the estimated AUCs are downward, whereas the worst-case scenario (set B & D) shows that, the AUC's are overestimated at smaller

Table 2: The Bias and MSE of the estimated and bias-corrected estimator of Δ for case (i)

Sets	Δ	n	$\hat{\Delta}$ (95% CI)	Bias	MSE	$\tilde{\Delta}$ (95% CI)	Bias	MSE
A	0.92423	25	0.89002 (0.85802, 0.92202)	−0.03420	0.00312	0.93697 (0.90137, 0.97257)	0.01274	0.00025
		50	0.88943 (0.87713, 0.90173)	−0.03479	0.00281	0.93153 (0.91003, 0.95303)	0.00730	0.00008
		100	0.87569 (0.84358, 0.90780)	−0.04854	0.00240	0.92880 (0.89635, 0.96125)	0.00457	0.00004
		200	0.87047 (0.83806, 0.90288)	−0.05376	0.00212	0.92590 (0.88060, 0.97120)	0.00167	0.00004
		500	0.86817 (0.82586, 0.91048)	−0.05605	0.00203	0.92317 (0.88330, 0.96304)	−0.00106	0.00002
B	0.53510	25	0.60483 (0.58263, 0.62703)	0.06973	0.00914	0.56306 (0.53319, 0.59293)	0.02796	0.00645
		50	0.57328 (0.54115, 0.60541)	0.03818	0.00363	0.54836 (0.50606, 0.59066)	0.01326	0.00241
		100	0.54849 (0.52717, 0.56981)	0.01339	0.00119	0.54176 (0.51855, 0.56497)	0.00666	0.00058
		200	0.53479 (0.51583, 0.55375)	−0.00031	0.00040	0.53803 (0.50236, 0.57370)	0.00293	0.00010
		500	0.52244 (0.48984, 0.55504)	−0.01266	0.00015	0.53561 (0.48983, 0.58139)	0.00051	0.00001
C	0.91739	25	0.89497 (0.86559, 0.92435)	−0.02242	0.00626	0.92990 (0.89708, 0.96272)	0.01251	0.00012
		50	0.88766 (0.86817, 0.90715)	−0.02973	0.00409	0.92613 (0.89764, 0.95463)	0.00874	0.00005
		100	0.87080 (0.84513, 0.89647)	−0.0466	0.00397	0.92390 (0.89052, 0.95728)	0.00651	0.00004
		200	0.87304 (0.83850, 0.90758)	−0.04435	0.00358	0.92241 (0.89394, 0.95089)	0.00502	0.00001
		500	0.86734 (0.84389, 0.89079)	−0.05005	0.00346	(0.91192 (0.87238, 0.95147)	−0.00547	0.00001
D	0.54013	25	0.59050 (0.55598, 0.62502	0.05037	0.01066	0.56918 (0.54354, 0.59482)	0.02905	0.00653
		50	0.58075 (0.54052, 0.62098)	0.04062	0.00948	(0.55336 (0.53350, 0.57322)	0.01323	0.00418
		100	0.54189 (0.49611, 0.58767)	0.00176	0.00782	(0.54412 (0.52217, 0.56607)	0.00399	0.00040
		200	0.52177 (0.48752, 0.55602)	−0.01836	0.00091	(0.54256, (0.52022, 0.56491)	0.00243	0.00028
		500	0.51717 (0.47696, 0.55738)	−0.02296	0.00034	0.54019 (0.51081, 0.56957)	0.00006	0.00001

sample sizes. This will provide an observation that at smaller sample sizes, the corrected AUC is overestimated; and with large samples, true estimates of the corrected AUC are witnessed. This refers to the phenomenon that with moderate and large sample sizes, the bias-corrected estimator of AUC ($\tilde{\Delta}$) has a smaller bias and is closer to the true AUC values (Δ). The same scenario is observed in all three cases.

The graphical representations of the ROC plots for the true MROC model and the MROC models (errors in the data) at various sample sizes are shown in Figure 1. These graphs depict that the actual performance of the classifier is affected due to the ME. Once, we correct the MROC model, then the accuracy curve will be restored using bias-corrected estimator.

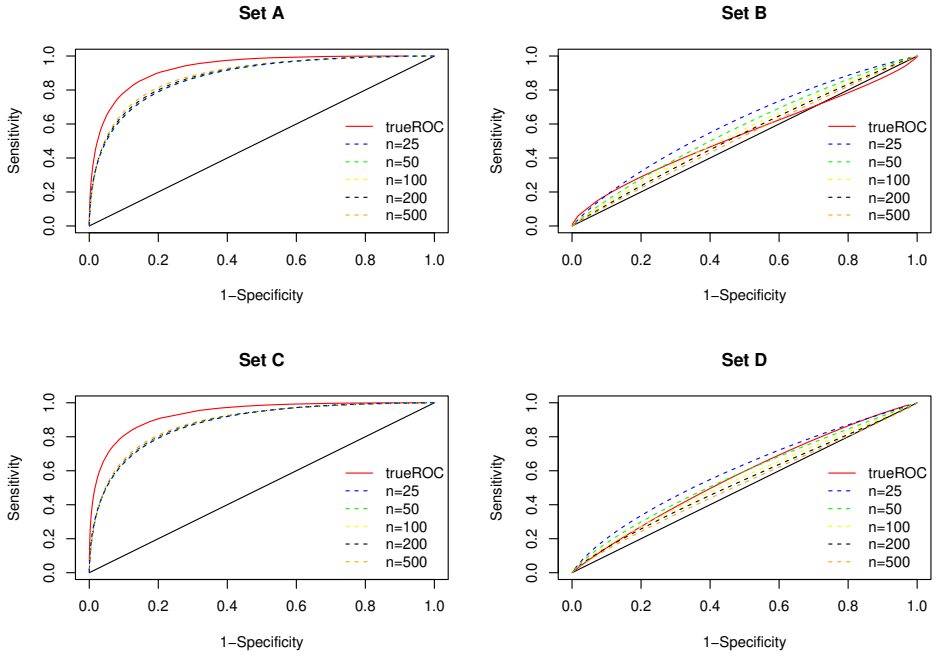

Figure 1: *The true MROC curve and the estimated MROC curves at various sample sizes under case (i).*

- Case (ii) Σ_ϵ and Σ_η have different variances

In this case, the bivariate error covariance matrices are considered with the assumption that the markers possess different variances.

$$\text{i.e.,}\ \Sigma_\epsilon = \Sigma_\eta = \begin{pmatrix} 2.0 & 0.0 \\ 0.0 & 2.8 \end{pmatrix}.$$

In this case as well, and in consideration of the error covariance matrices, four datasets are generated at various sample sizes and the AUC is obtained, before and after correction, as shown in Table 3. The results show that the bias-corrected estimator has less bias and MSE compared with the estimated AUCs and is closer to the true AUC measure. The true MROC curve and the estimated MROC curves in the presence of measurement errors for these simulations at various sample sizes are presented in Figure 2.

- Case (iii) Some of the markers are measured correctly

This case is quite different, because it is considered that one of the markers of the bivariate error covariance matrices possesses a measurement error.

$$\text{i.e.,}\ \Sigma_\epsilon = \Sigma_\eta = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Table 4, shows the results of the estimated and the bias-corrected estimator of AUCs compared with the true AUC values, which show the MSE of the corrected estimator being smaller than the estimated

Table 3: The Bias and MSE of the estimated and bias-corrected estimator of Δ for case (ii)

Sets	Δ	n	$\hat{\Delta}$ (95% CI)	Bias	MSE	$\tilde{\Delta}$ (95% CI)	Bias	MSE
A	0.92423	25	0.87353 (0.83901, 0.90805)	−0.05069	0.00529	0.94926 (0.92888, 0.96963)	0.02503	0.00049
		50	0.86818 (0.83818, 0.89817)	−0.05605	0.00433	0.94046 (0.90405, 0.97687)	0.01623	0.00035
		100	0.86032 (0.84320, 0.87745)	−0.06390	0.00352	0.93528 (0.90264, 0.96793)	0.01105	0.00028
		200	0.85940 (0.82668, 0.89213)	−0.06483	0.00309	0.93023 (0.89664, 0.96382)	0.00600	0.00011
		500	0.84210 (0.81741, 0.86679)	−0.08213	0.00212	0.92683 (0.90754, 0.94611)	0.00260	0.00010
B	0.53510	25	0.59206 (0.56714, 0.61697)	0.05696	0.00496	0.57094 (0.55052, 0.59137)	0.03584	0.00173
		50	0.58826 (0.55361, 0.62291)	0.05316	0.00295	0.56245 (0.54086, 0.58404)	0.02735	0.00043
		100	0.54983 (0.52790, 0.57176)	0.01473	0.00070	0.55204 (0.52217, 0.58192)	0.01694	0.00021
		200	0.53296 (0.51473, 0.55119)	−0.00214	0.00053	0.53639 (0.49942, 0.57335)	0.00129	0.00001
		500	0.52050 (0.48540, 0.55560)	−0.01460	0.00055	0.53401 (0.50394, 0.56408)	−0.00109	0.00001
C	0.91739	25	0.86526 (0.84328, 0.88725)	−0.05213	0.00633	0.93421 (0.90606, 0.96237)	0.01682	0.00040
		50	0.85923 (0.84550, 0.87296)	−0.05816	0.00467	0.92941 (0.89489, 0.96392)	0.01202	0.00028
		100	0.85854 (0.84813, 0.86894)	−0.05886	0.00493	0.92225 (0.88653, 0.95797)	0.00486	0.00020
		200	0.84928 (0.81161, 0.88695)	−0.06811	0.00530	0.92085 (0.89322, 0.94848)	0.00346	0.00002
		500	0.84749 (0.82754, 0.86745)	−0.06990	0.00451	0.91952 (0.88849, 0.95055)	0.00213	0.00001
D	0.54013	25	0.58934 (0.55650, 0.62218)	0.04921	0.00549	0.56426 (0.52346, 0.60506)	0.02413	0.00032
		50	0.57383 (0.53023, 0.61743)	0.03370	0.00531	0.55892 (0.53249, 0.58536)	0.01879	0.00024
		100	0.55490 (0.52821, 0.58160)	0.01477	0.00067	0.55424 (0.53557, 0.57291)	0.01411	0.00024
		200	0.53903 (0.50402, 0.57404)	−0.00110	0.00048	0.54935 (0.51690, 0.58180)	0.00922	0.00002
		500	0.53323 (0.49676, 0.56970)	−0.00690	0.00034	0.54053 (0.51069, 0.57038)	0.00040	0.00000

AUCs. From the results, it is understood that the involvement of measurement errors in any one of the markers may influence the outcome of the MROC model. The graphical representations of the ROC plots for the true MROC model and the estimated MROC models at various sample sizes are shown in Figure 3.

The idea behind considering three cases, that is equal, unequal and the absence of error variance for one variable, is to observe the impact on the AUC. It was noticed that under equal and unequal variance cases, the AUC was impacted significantly and attained a lower value than the true AUC. However, pertaining to case (iii), the impact was less with a lower AUC value and the same was reflected in the MSE. Therefore, if the measurement error is not observed in one variable, then the deviation from the true AUC will not be that significant. The MSE values observed at different sample

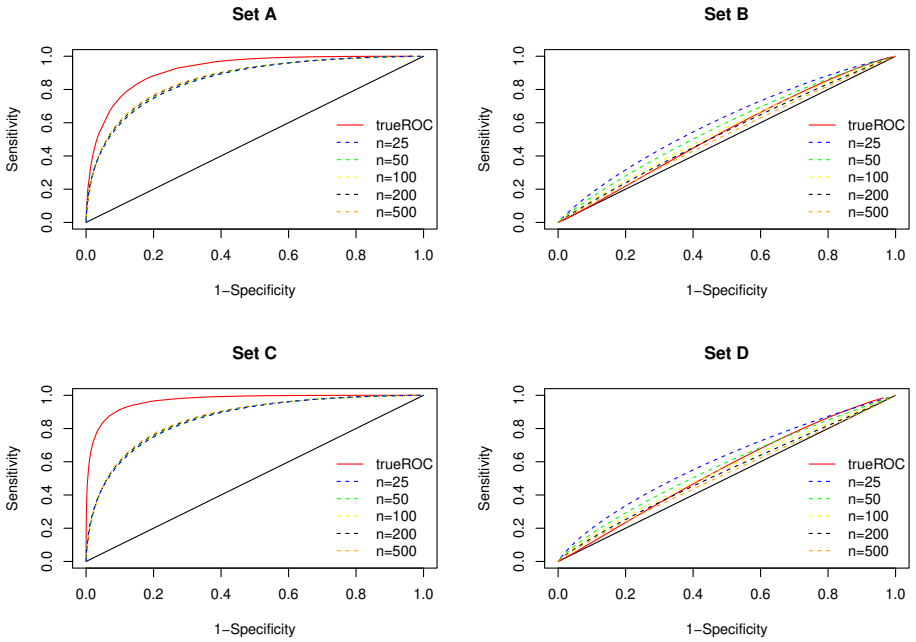

Figure 2: *The true MROC curve and the estimated MROC curves at various sample sizes under case (ii).*

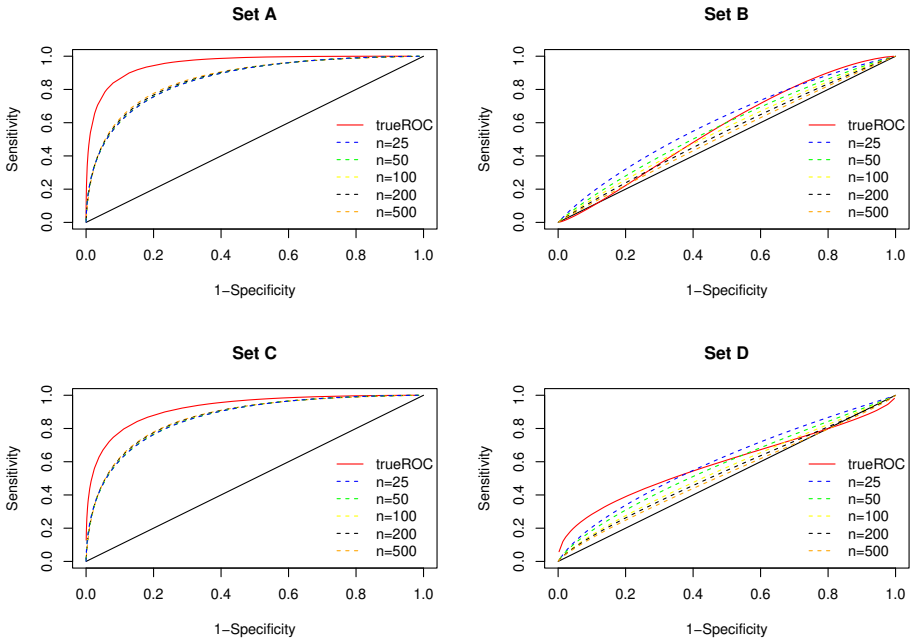

Figure 3: *The true MROC curve and the estimated MROC curves at various sample sizes under case (iii).*

Table 4: The Bias and MSE of the estimated and bias-corrected estimator of Δ for case (iii)

Sets	Δ	n	$\hat{\Delta}$ (95% CI)	Bias	MSE	$\tilde{\Delta}$ (95% CI)	Bias	MSE
A	0.92423	25	0.88812 (0.86489, 0.91135)	−0.03610	0.00618	0.94036 (0.90499, 0.97573)	0.01613	6.3441E-05
		50	0.87646 (0.85460, 0.89832)	−0.04777	0.00545	0.93913 (0.90010, 0.97816)	0.01490	5.3107E-05
		100	0.87434 (0.83115, 0.91753)	−0.04989	0.00508	0.92535 (0.88588, 0.96482)	0.00112	4.8620E-06
		200	0.87110 (0.83725, 0.90495)	−0.05313	0.00438	0.92513 (0.90138, 0.94888)	0.00091	4.1177E-07
		500	0.87016 (0.83251, 0.90781)	−0.05407	0.00402	0.92487 (0.88050, 0.96924)	0.00064	3.7878E-07
B	0.53510	25	0.59041 (0.56411, 0.61671)	0.05531	0.00543	0.54987 (0.50897, 0.59077)	0.01477	5.3684E-03
		50	0.58717 (0.54573, 0.62861)	0.05207	0.00396	0.54925 (0.51356, 0.58494)	0.01415	4.5187E-04
		100	0.56820 (0.53201, 0.60439)	0.03310	0.00034	0.53926 (0.51154, 0.56698)	0.00416	1.7718E-04
		200	0.53102 (0.50315, 0.55889)	−0.00408	0.00034	0.53537 (0.49715, 0.57359)	0.00027	1.1523E-05
		500	0.51230 (0.47108, 0.55352)	−0.02280	0.00029	0.53353 (0.50126, 0.56580)	−0.00157	1.2883E-04
C	0.91739	25	0.87930 (0.85113, 0.90747)	−0.03809	0.00559	0.93482 (0.91168, 0.95796)	0.01743	6.9813E-06
		50	0.87847 (0.84002, 0.91692)	−0.03892	0.00490	0.93204 (0.90058, 0.96350)	0.01465	4.3485E-06
		100	0.86513 (0.82578, 0.90448)	−0.05226	0.00410	0.92860 (0.88347, 0.97373)	0.01121	4.0867E-06
		200	0.85684 (0.81538, 0.89830)	−0.06055	0.00375	0.92369 (0.88675, 0.96063)	0.00630	1.5410E-06
		500	0.84321 (0.81202, 0.87440)	−0.07418	0.00302	0.91265 (0.87645, 0.94885)	−0.00474	1.0254E-06
D	0.54013	25	0.60943 (0.56631, 0.65255)	0.06930	0.00571	0.57233 (0.52868, 0.61598)	0.03220	5.8768E-03
		50	0.58236 (0.54632, 0.61840)	0.04223	0.00502	0.56507 (0.53418, 0.59596)	0.02494	5.5041E-04
		100	0.56915 (0.53051, 0.60779)	0.02902	0.00041	0.54398 (0.50491, 0.58305)	0.00385	2.1186E-06
		200	0.54549 (0.51695, 0.57403)	0.00536	0.00037	0.54136 (0.50652, 0.57620)	0.00123	1.5418E-06
		500	0.52201 (0.48773, 0.55629)	−0.01812	0.00030	0.54053 0.49951 0.58155	0.00040	1.1863E-06

sizes in case (iii) were slightly lower compared to the MSE values in case (i) and (ii), respectively. Parallel to this, the influence of the sample size was also observed and resulted in the overestimated values of corrected AUCs at smaller sample sizes, while the corrected AUCs were closer to the true AUC at moderate (n = 100) and large samples. This scenario was noticed in all three cases.

4. Simulation study II (with the real dataset)

In this section, the *Vertebral Column dataset* (Guilherme and Ajalmar, 2011), which consists of 310 samples and three categories, namely Normal, Disk Hernia and Spondylolisthesis, is considered. The MROC curve analysis is confined to binary classification, so for illustration purposes, the same dataset

Table 5: Error covariance matrices for both the populations of DH and SP datasets

(i) Disk Hernia dataset

$$\begin{pmatrix} 16.4 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ & 23.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ & & 19.2 & 0.0 & 0.0 & 0.0 \\ & & & 18.9 & 0.0 & 0.0 \\ & & & & 21.9 & 0.0 \\ & & & & & 26.5 \end{pmatrix}$$

(ii) Spondylolisthesis dataset

$$\begin{pmatrix} 34.6 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ & 36.4 & 0.0 & 0.0 & 0.0 & 0.0 \\ & & 38.9 & 0.0 & 0.0 & 0.0 \\ & & & 38.7 & 0.0 & 0.0 \\ & & & & 36.9 & 0.0 \\ & & & & & 31.9 \end{pmatrix}$$

Table 6: Bias and MSE of the estimated and the bias-corrected estimator of AUC of the Disk Hernia dataset

Dataset	Δ	$\hat{\Delta}$	Bias	MSE	$\tilde{\Delta}$	Bias	MSE
Disk Hernia	**0.896123**	0.868275	−0.027849	0.000776	**0.897044**	0.000921	0.000001
Spondylolisthesis	**0.947690**	0.938546	−0.009144	0.000084	**0.947007**	−0.000683	0.000000

is subdivided into two datasets, such as the Disk Hernia dataset and Spondylolisthesis dataset, respectively.

1. **Disk Hernia (DH)** contains a total of 160 samples, and of these, 100 samples belong to normal subjects and the rest are disk hernia subjects.

2. **Spondylolisthesis (SP)** has 250 samples, and of these, 100 are normal subjects and the remaining 150 samples belong to Spondylolisthesis subjects.

Each dataset has six attributes, namely, pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius, and the grade of spondylolisthesis. In order to show the working nature of the methodology, error covariance matrices are considered for both the DH and SP population datasets (Table 5). Random samples are generated using these error covariances matrices and are added to the actual observations in the data. This gives us a situation where the variables in the dataset are affected due to measurement error.

The proposed bias-corrected estimator, true AUC and related measures were estimated and presented in Table 6. Here Δ denotes the AUC of the original dataset and $\hat{\Delta}$ is the AUC obtained after the inclusion of error observations. This clearly indicates that when the variables are affected by measurement error, then the true AUC may not be observed because the bias depicts the difference from the information that got masked due the presence of measurement error. Additionally, $\tilde{\Delta}$ is the AUC after the correction. Therefore, the bias-corrected estimator has helped in retaining the true AUC value. In the Disk Hernia dataset, the Δ (0.896123) and $\tilde{\Delta}$ (0.897044) are very close and the same is true for MSE (0.000001). Hence, the proposed estimator helps in obtaining the true estimate of AUC, even if the data is influenced by measurement error. Similar kinds of results are obtained with respect to the SP dataset. In this case as well, the Δ (0.947690) and $\tilde{\Delta}$ (0.947007) are very close and the MSE is almost equal to zero.

These results show that by adding error observations to the variables, the accuracy was affected. In such situations, using the bias-corrected estimator can achieve the true accuracy values, which have less bias and MSE when compared to the estimated AUC. With the support of bias and MSE values, there is a clear need to correct the bias to estimate the AUC in the presence of measurement errors in the data. In Figure 4, the MROC plots are drawn for the two datsets, both before and after the correction of AUC. In the MROC curves for both datasets, it is clear that the performance of the classifier got affected in the presence of measurement errors. To demonstrate the application of the proposed methodology mentioned above, random samples have been generated using the error

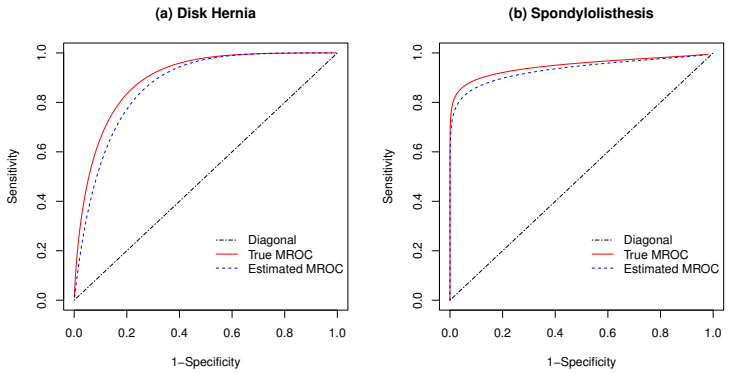

Figure 4: *True and estimated MROC curves for (a) the Disk Hernia dataset and (b) the Spondylolisthesis dataset.*

covariance matrices, which is a limitation because getting a real dataset with observed measurement errors was difficult. However, our attempt in generating random samples through error covariance matrices showed that the proposed methodology works well if the data contains measurement errors. Hence, we treated this exercise as a simulation study using a real dataset.

5. Summary

In this paper, discussions were focused on the influence of measurement errors that affect the performance of the MROC curve. When the data possesses measurement errors, the AUC will be biased downward. To address this, a bias-corrected estimator is derived and shown to work with simulated and real datasets, even if measurement errors are present in the data, without any loss of information to determine the true AUC value. To mimic the existence of measurement error, the error covariance matrices were added to the original DH and SP datasets and the measures of the MROC, bias-corrected estimator ($\tilde{\Delta}$) were computed. In each instance, the results were similar. The bias of the corrected estimator $\tilde{\Delta}$ was much smaller than the bias of $\hat{\Delta}$, and thus, the bias in estimating Δ was reduced without inflating the mean squared error. Therefore, by using the bias-corrected estimator, it is possible to quantify the deviation of the true accuracy. The results reveal that the proposed estimate works well in retaining the true accuracy of the classifier.

References

Anderson TW and Bahadur RR (1962). Classification into two multivariate normal distributions with different covariance matrices, *The Annals of Mathematical Statistics*, **33**, 420–431.

Bamber D (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph, *Journal of Mathematical Psychology*, **12**, 387–415.

Begg CB and Greenes RA (1983). Assessment of diagnostic tests when disease verification is subject to selection bias, *Biometrics*, **39**, 207–215.

Begg CB and McNeil BJ (1988). Assessment of radiologic tests: control of bias and other design considerations, *Radiology*, **167**, 565–569.

Berbaum KS, Dorfman DD, and Franken Jr EA (1989). Measuring observer performance by ROC analysis: indications and complications, *Investigative Radiology*, **24**, 228–233.

Coffin M and Sukhatme S (1996). A parametric approach to measurement errors in receiver operating

characteristic studies, In *Lifetime Data: Models in Reliability and Survival Analysis*, 71–75.

Coffin M and Sukhatme S (1997). Receiver operating characteristic studies and measurement errors, *Biometrics*, **53**, 823–837.

Dunn G (1989). *Design and Analysis of Reliability Studies: The Statistical Evaluation of Measurement Errors*, Edward Arnold Publishers, New York.

Egan JP (1975). *Signal Detection Theory and ROC Analysis*, Academic Press, New York.

Faraggi D (2000). The effect of random measurement error on receiver operating characteristic (ROC) curves, *Statistics in Medicine*, **19**, 61–70.

Fuller WA (2009). Measurement Error Models, John Wiley & Sons, New York.

Guilherme AB and Ajalmar Rago RN (2011). Department of Teleinformatics Engineering, Federal University of Ceara, Fortaleza, Ceara, Brazil, *UCI Machine Learning Repository*, Available from: https://archive.ics.uci.edu/ml/datasets/Vertebral+Column

Hand DJ (2010). Evaluating diagnostic tests: the area under the ROC curve and the balance of errors, *Statistics in Medicine*, **29**, 1502–1510.

Kim J and Gleser LJ (2000). SIMEX approaches to measurement error in ROC studies, *Communications in Statistics-Theory and Methods*, **29**, 2473–2491.

Perkins NJ, Schisterman EF, and Vexler A (2009). Generalized ROC curve inference for a biomarker subject to a limit of detection and measurement error, *Statistics in Medicine*, **28**, 1841–1860.

Reiser B (2000). Measuring the effectiveness of diagnostic markers in the presence of measurement error through the use of ROC curves, *Statistics in Medicine*, **19**, 2115–2129.

Sameera G, Vardhan RV, and Sarma KVS (2016). Binary classification using multivariate receiver operating characteristic curve for continuous data, *Journal of Biopharmaceutical Statistics*, **26**, 421–431.

Schisterman EF (1999). *Lipid peroxidation and cardiovascular disease: an ROC approach (Doctoral dissertation*, State University of New York at Buffalo ProQuest Dissertations Publishing, New York.

Schisterman EF, Faraggi D, Reiser B, and Trevisan M (2001). Statistical inference for the area under the receiver operating characteristic curve in the presence of random measurement error, *American Journal of Epidemiology*, **154**, 174–179.

Schisterman EF, Faraggi D, and Reiser B (2004). Adjusting the generalized ROC curve for covariates, *Statistics in Medicine*, **23**, 3319–3331.

Shear L, Burke GL, Freedman DS, Webber LS, and Berenson GS (1987). Designation of children with high blood pressure—considerations on percentile cut points and subsequent high blood pressure: the Bogalusa Heart Study, *American Journal of Epidemiology*, **125**, 73–84.

Su JQ and Liu JS (1993). Linear combinations of multiple diagnostic markers, *Journal of the American Statistical Association*, **88**, 1350–1355.

Tosteson TD, Buonaccorsi JP, Demidenko E, and Wells WA (2005). Measurement error and confidence intervals for ROC curves, *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, **47**, 409–416.

Vexler A, Schisterman EF, and Liu A (2008). Estimation of ROC curves based on stably distributed biomarkers subject to measurement error and pooling mixtures, *Statistics in Medicine*, **27**, 280–296.

Yin J and Tian L (2014). Optimal linear combinations of multiple diagnostic biomarkers based on Youden index, *Statistics in Medicine*, **33**, 1426–1440.

Yuan Z and Ghosh D (2008). Combining multiple biomarker models in logistic regression, *Biometrics*, **64**, 431–439.

Received January 1, 2022; Revised March 11, 2022; Accepted May 19, 2022