



Published in final edited form as:

Stat Methods Med Res. 2016 August ; 25(4): 1359–1380. doi:10.1177/0962280213481053.

Linear combination methods to improve diagnostic/prognostic accuracy on future observations

Le Kang¹, Aiyi Liu², and Lili Tian³

¹Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD, USA

²Biostatistics and Bioinformatics Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA

³Department of Biostatistics, State University of New York at Buffalo, Buffalo, NY, USA

Abstract

Multiple diagnostic tests or biomarkers can be combined to improve diagnostic accuracy. The problem of finding the optimal linear combinations of biomarkers to maximise the area under the receiver operating characteristic curve has been extensively addressed in the literature. The purpose of this article is threefold: (1) to provide an extensive review of the existing methods for biomarker combination; (2) to propose a new combination method, namely, the nonparametric stepwise approach; (3) to use leave-one-pair-out cross-validation method, instead of re-substitution method, which is overoptimistic and hence might lead to wrong conclusion, to empirically evaluate and compare the performance of different linear combination methods in yielding the largest area under receiver operating characteristic curve. A data set of Duchenne muscular dystrophy was analysed to illustrate the applications of the discussed combination methods.

Keywords

Multiple biomarkers; receiver operating characteristic curve; area under the receiver operating characteristic curve; linear combination; diagnostic/prognostic accuracy

1 Introduction

In diagnostic study, multiple tests are often performed on the same individual to provide clinicians as much information as possible as it is becoming increasingly clear that one single diagnostic test or biomarker is not sufficient to make accurate disease diagnosis or prognosis.¹ It is therefore of critical importance to combine the information available in an optimal way to improve the diagnostic/prognostic accuracy.²

We consider the cases when the diagnostic outcome is binary, i.e., non-diseased and diseased. Let S_1 and S_2 denote the scores resulting from a diagnostic test or biomarker and F_1 and F_2 be the corresponding cumulative distribution functions for the non-diseased and

diseased subjects, respectively. Assume that the results of a diagnostic test are measured on a continuous scale and higher values indicate greater likelihood of having the disease. For a threshold value c , let $F_1(c)$ and $1 - F_2(c)$ be the true classification rates for non-diseased and diseased category, also known as the specificity and sensitivity of the marker, respectively. For all possible $c \in \mathbb{R}$, a plot of $\{1 - F_1(c), 1 - F_2(c)\}$ produces the receiver operating characteristic (ROC) curve of the marker,³ the most widely used graphical and statistical tool for assessing a diagnostic test's ability to distinguish between two disease populations. The area under this curve (AUC), the most commonly used diagnostic accuracy measure, is then given by

$$\text{AUC} = \int_c [1 - F_2(c)] d[1 - F_1(c)] = \int_0^1 [1 - F_2(F_1^{-1}(1 - t))] dt.$$

One can show that the AUC is mathematically equivalent to the probability $P(S_1 < S_2)$.⁴ Note that the unbiased nonparametric Mann-Whitney U statistic of the AUC is given by

$$U = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(S_{1i} < S_{2j}), \quad (1.1)$$

where n_1 and n_2 are the sample sizes for non-diseased and diseased subjects, respectively, and $I(\cdot)$ stands for the indicator function. Under the normality assumption $S_d \sim N(\mu_d, \sigma_d^2)$, $d = 1, 2$, the AUC can be further expressed as

$$\text{AUC} = \Phi \left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right), \quad (1.2)$$

where $\Phi(\cdot)$ is the standard normal distribution function, see Pepe⁵ for details. For a useless biomarker (e.g., when normally distributed S_1 and S_2 have the identical means), the AUC is 0.5.

When several diagnostic tests and biomarkers are available, one can combine them in a linear fashion into a composite score that achieves better diagnostic/prognostic accuracy. An optimal linear combination of biomarkers is defined as the one for which the composite score would achieve the maximum AUC over all possible linear combinations. Many articles have addressed the problem of finding the optimal linear combination to maximise the AUC. For instance, Su and Liu⁶ extended Fisher's discriminant function and derived an optimal linear combination that maximises the AUC when the markers in the non-diseased and diseased category follow multivariate normal distributions. Without assumptions on the distributions of the markers, Pepe and Thompson⁷ considered an empirical search of the optimal linear combination that maximises the Mann-Whitney statistic of AUC, although this approach is computationally formidable when the number of biomarkers is large.⁸ Liu et al.⁹ developed a semi-linear min-max combination approach which only involves searching for a single coefficient that maximises the Mann-Whitney U statistic of AUC and thus is

computationally efficient. However, as stated by the authors, when not all markers are measured on the same scale the feasibility of this combination method might be an issue. Recently, Jin and Lu¹⁰ proved that, if the data satisfy a logistic regression model, the coefficient from a fitted logistic regression with binary diagnostic outcomes is the optimal linear combination in the sense that it provides the highest sensitivity uniformly over the entire range of specificity and therefore yields the largest AUC among all possible linear combinations. All of the above-mentioned methods have pros and cons so that there is no clear winner. In this paper, we propose a nonparametric stepwise approach for the same purpose. This new approach is flexible and easy to implement, and performs better than other methods under certain scenarios. More details will be given in Section 4.

So far, the comparison of performance among the existing combination methods has been done^{6,7,9,10} using re-substitution method for AUC estimation which consists of the following steps: (1) linear combination coefficients were first estimated from a particular data set; (2) a composite score was then calculated by linearly combining multiple diagnostic tests using the estimated coefficients; (3) and finally the AUC was estimated based on the combined score. The investigators usually concluded the superiority of a certain linear combination method if the associated AUC is the largest. However, as pointed out by a few researchers,^{11–13} the estimated AUC using the composite score by re-substitution method usually is overoptimistic for estimating the diagnostic/prognostic accuracy on future observations. In other words, a linear combination rule might perform well in yielding the largest AUC on observed dataset used for obtaining the estimated combination coefficients; however, it might not be the optimal rule on different dataset, e.g., future observations. Therefore, using re-substitution method to estimate AUC for the purpose of comparing between combination methods might be misleading. This is a common phenomenon between training set and validation set in the discipline of machine learning.¹⁴ On the other hand, cross-validation method is considered as the simplest and most widely used method for estimating the prediction error.^{15–17} Recently, Huang et al.¹¹ proposed several methods to adjust for the upward bias from estimating the AUC associated with the estimated coefficients by re-substitution. Among the investigated methods including bootstrap and sigmoid function smoothing, the leave-one-pair-out cross-validation (LOPO CV) approach is especially advocated because it produces nearly unbiased AUC estimate associated with the estimated combination coefficients. They also mentioned an approximated cross-validation to reduce the computing cost.

To our knowledge, no work has been done in evaluating and comparing the performance of linear combination methods, i.e., the methods by Su and Liu,⁶ logistic regression and Liu et al.,⁹ using LOPO CV estimate of AUC. Therefore, in this article, besides presenting an extensive review of the existing combination methods and proposing a new stepwise combination method, we are also concerned with comparing the performance of the combination methods using LOPO CV estimate of AUC instead of re-substitution method. More details will be given in Section 2.

The rest of our article is organised as follows. In Section 2, a LOPO CV approach to estimating the diagnostic/prognostic accuracy of a linear combination rule based on AUC for future observations was discussed. An overview of the historic developments of linear

combination methods was briefly given in Section 3. In Section 4, a new nonparametric stepwise approach was introduced. An extensive simulation study was presented in Section 5 for comparing the performance of different linear combination methods in maximising the diagnostic/prognostic accuracy for future subjects based on AUC. In Section 6, existing approaches as well as the proposed approach were applied to a real data set of 125 females on Duchenne muscular dystrophy (DMD) from Carnegie Mellon University Statlib Datasets Archive to combine four markers to increase the diagnostic/prognostic accuracy of screening females as potential DMD carriers. A broader discussion on deriving linear combinations of diagnostic tests and biomarkers to improve the diagnostic/prognostic accuracy is presented in Section 7.

2 A LOPO CV approach to estimating AUC of linear combination rules

2.1 Notations and preliminary

Suppose we have p diagnostic tests or biomarkers available on each individual. The diagnostic category is denoted as $D = d$, where $d = 1, 2$, representing non-diseased and diseased subjects, respectively. Let

$$\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip}), i = 1, 2, \dots, n_1,$$

be the p -dimensional observed scores from a random sample of size n_1 in the non-diseased category, and

$$\mathbf{Y}_j = (Y_{j1}, Y_{j2}, \dots, Y_{jp}), j = 1, 2, \dots, n_2,$$

be the p -dimensional observed scores from a random sample of size n_2 in the diseased category. The data is often stacked together in a matrix form

$$\begin{pmatrix} [\mathbf{X}_i]_{n_1 \times p} & \mathbf{1}_{n_1} \\ [\mathbf{Y}_j]_{n_2 \times p} & \mathbf{2}_{n_2} \end{pmatrix},$$

where the first p columns form the matrix of observed scores concatenated from \mathbf{X}_i and \mathbf{Y}_j by row and the last column indicates the diagnostic category. Here, we are not focusing on high dimensional data, restricting $p \ll \min(n_1, n_2)$.

The problem of interest is to obtain a vector combination coefficient \mathbf{c} such that the univariate composite scores $S_{1i} = \mathbf{X}_i \mathbf{c}$ and $S_{2j} = \mathbf{Y}_j \mathbf{c}$ for non-diseased and diseased category, respectively, have the largest overall discriminating ability of classifying subjects into their corresponding diagnostic category, in this case, yielding the largest AUC. Denote the estimated combination coefficient from the observed data $[\mathbf{X}_i]_{n_1 \times p}$ and $[\mathbf{Y}_j]_{n_2 \times p}$ to be $\hat{\mathbf{c}}$, the literature^{6,7,9,10} often show the optimality of their linear combination methods by presenting the estimated AUC associated with $\hat{\mathbf{c}}$, $\widehat{\text{AUC}} = (\mathbf{X}_i \hat{\mathbf{c}} < \mathbf{Y}_j \hat{\mathbf{c}})$, which can be simply estimated by re-substitution as follows,

$$\widehat{AUC} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(\mathbf{X}_i \hat{\mathbf{c}} < \mathbf{Y}_j \hat{\mathbf{c}}),$$

without any distributional assumption of \mathbf{X}_i and \mathbf{Y}_j , or

$$\widehat{AUC} = \Phi \left(\frac{\hat{\mathbf{c}}'(\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)}{\sqrt{\hat{\mathbf{c}}'(\hat{\boldsymbol{\Sigma}}_1 + \hat{\boldsymbol{\Sigma}}_2)\hat{\mathbf{c}}}} \right),$$

where $\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2$ are the sample means and $\hat{\boldsymbol{\Sigma}}_1, \hat{\boldsymbol{\Sigma}}_2$ are the sample variance-covariance matrices of observed \mathbf{X}_i and \mathbf{Y}_j , respectively, given that \mathbf{X}_i and \mathbf{Y}_j follow a multivariate normal distribution $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$.

Unfortunately, these estimates of AUC for overall diagnostic/prognostic accuracy are overoptimistic, especially for small sample size problems; for example, see Huang et al.¹¹ and Efron.¹² When the composite score is calculated based on estimated combination coefficients from a given dataset, it tends to discriminate subjects better than it should be if applying the same combination rule to another random dataset from the same populations. In an extreme case, one may argue that with certain nonlinear combinations, it is possible to achieve perfect discriminations (i.e., non-diseased and diseased are fully separated) on the given dataset. However, it is impossible to perform perfect discriminations on another independent dataset.

Note that our ultimate goal is to estimate the AUC of the composite score for future observations ($\mathbf{X}_{ind}, \mathbf{Y}_{ind}$), independent of the observed data used for obtaining $\hat{\mathbf{c}}$ as follows,

$$\widehat{AUC} = (\mathbf{X}_{ind} \hat{\mathbf{c}} < \mathbf{Y}_{ind} \hat{\mathbf{c}}).$$

This in fact serves as a better assessment of discriminatory or prognosis accuracy of certain combination rule for the purpose of improving diagnostic/prognostic accuracy on future observations. Hence, re-substitution procedure is not appropriate for such purpose.

2.2 An LOPO CV procedure solution

The cross-validation^{12,14,15} is the simplest and most widely used technique for assessing how the results of a statistical analysis would generalise to an independent data set and how accurately a predictive model will perform in practice. First developed by Quenouille¹⁶ in the form of “leave-one-out” procedure (Jackknife), it was used to estimate the bias of an estimator. Lachenbruch¹⁷ also discussed an almost unbiased method of obtaining confidence intervals for misclassification rate in discriminant analysis based on $n_1 - 1$ and n_2 observations or n_1 and $n_2 - 1$ observations. Similarly for our context, Huang et al.¹¹ discussed a LOPO CV procedure to estimate the AUC associated with linear combination $\hat{\mathbf{c}}$ for future observations ($\mathbf{X}_{ind}, \mathbf{Y}_{ind}$). The so-called LOPO CV estimate of AUC of the combination rule is as follows,

$$\widehat{\text{AUC}}^{\text{CV}} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(\mathbf{X}_i \hat{\mathbf{c}}^{(-ij)} < \mathbf{Y}_j \hat{\mathbf{c}}^{(-ij)}), \quad (2.1)$$

where $\hat{\mathbf{c}}^{(-ij)}$ is the linear combination coefficient calculated from the observed data $[\mathbf{X}_{(-i)}]_{(n_1-1) \times p}$ and $[\mathbf{Y}_{(-j)}]_{(n_2-1) \times p}$, where $[\mathbf{X}_{(-i)}]_{(n_1-1) \times p}$ is obtained by removing the i th row from $[\mathbf{X}_i]_{n_1 \times p}$ and $[\mathbf{Y}_{(-j)}]_{(n_2-1) \times p}$ is obtained by removing the j th row from $[\mathbf{Y}_j]_{n_2 \times p}$. The authors pointed out that alternative 5-fold cross-validation and 10-fold cross-validation can be applied instead of LOPO CV to gain computational efficiency. Besides the cross-validation methods, they also discussed several methods including bootstrap for estimation of AUC associated with the estimated coefficients. In conclusion, they recommended the use of LOPO CV for the AUC estimate associated with the combination rule for future observations as it has been shown to be nearly unbiased in their simulation studies.

In this article, our main concern is the point estimation of the AUC associated with $\hat{\mathbf{c}}$ for future observations. Therefore, we employ the LOPO CV method in evaluating and comparing the performance of different linear combination methods in improving diagnostic/prognostic accuracy based on AUC for future observations. Here, we only consider the LOPO CV approach for AUC nonparametric Mann-Whitney U statistic-based estimator as in equation (1.1). This is because if $[\mathbf{X}_{(-i)}]_{(n_1-1) \times p}$ and $[\mathbf{Y}_{(-j)}]_{(n_2-1) \times p}$ are already used for estimating linear combination coefficient $\hat{\mathbf{c}}^{(-ij)}$, it is awkward to estimate the variance-covariance matrices Σ_1 and Σ_2 based on the only remaining pair of $\mathbf{X}_{(i)}$ and $\mathbf{Y}_{(j)}$, should we want to use parametric AUC estimator as in equation (1.2). Furthermore, the LOPO CV approach for nonparametric AUC estimator, which is nearly unbiased,¹¹ fully serves the purpose of this article.

3 The existing approaches

3.1 Su and Liu's approach

Again, assume that \mathbf{X}_i and \mathbf{Y}_j follow a multivariate normal distribution $N_p(\boldsymbol{\mu}_1, \Sigma_1)$ and $N_p(\boldsymbol{\mu}_2, \Sigma_2)$ for the non-diseased and diseased subjects, respectively. Anderson and Bahadur¹⁸ first discussed classification problems for two multivariate normal distributions with different covariance matrices. Realizing that the result can be extended to AUC, Su and Liu⁶ derived the best linear combination that maximises the AUC with

$$\mathbf{c} \propto (\Sigma_1 + \Sigma_2)^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1).$$

Because the AUC is invariant to scalar transformation, apart from a constant coefficient,

$$\mathbf{c} = (\Sigma_1 + \Sigma_2)^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1). \quad (2.2)$$

Su and Liu⁶ proved that the AUC of this combination \mathbf{c} is maximised among all possible linear combinations under the multivariate normality assumption. In practice, the mean

vector and variance-covariance matrix for each diagnostic category can be estimated from the data. The estimates can then be substituted into equation (2.2) for calculating the combination coefficient $\hat{\mathbf{c}} = (\hat{\Sigma}_1 + \hat{\Sigma}_2)^{-1} (\hat{\mu}_2 - \hat{\mu}_1)$. The estimated coefficient $\hat{\mathbf{c}}$, together with the estimated means and variance-covariances can be re-substituted into equation (1.2) to estimate the AUC associated with the combination $\hat{\mathbf{c}}$ as follows,

$$\begin{aligned} \widehat{\text{AUC}} &= \Phi \left(\frac{\hat{\mathbf{c}}' (\hat{\mu}_2 - \hat{\mu}_1)}{\sqrt{\hat{\mathbf{c}}' (\hat{\Sigma}_1 + \hat{\Sigma}_2) \hat{\mathbf{c}}}} \right) \\ &= \Phi \left(\frac{(\hat{\mu}_2 - \hat{\mu}_1)' (\hat{\Sigma}_1 + \hat{\Sigma}_2)^{-1} (\hat{\mu}_2 - \hat{\mu}_1)}{\sqrt{(\hat{\mu}_2 - \hat{\mu}_1)' (\hat{\Sigma}_1 + \hat{\Sigma}_2)^{-1} (\hat{\mu}_2 - \hat{\mu}_1)}} \right) \\ &= \Phi \left(\sqrt{(\hat{\mu}_2 - \hat{\mu}_1)' (\hat{\Sigma}_1 + \hat{\Sigma}_2)^{-1} (\hat{\mu}_2 - \hat{\mu}_1)} \right). \end{aligned} \quad (2.3)$$

It is already mentioned above that this estimate is overoptimistic for AUC of the combination rule on future observations. Furthermore, note that Su and Liu's approach was developed under the assumption of normality. Hence, without normality assumption, especially when sample sizes n_1 and n_2 are not large enough, the asymptotic result for this combination approach may not hold, and thus the linear combination from this approach may not be optimal.

3.2 Pepe and Thompson's approach

Pepe and Thompson⁷ considered maximising the AUC without normality assumptions on the distribution of \mathbf{X}_j and \mathbf{Y}_j . For simplicity, they addressed the issue of finding optimal linear combinations with only $p=2$, i.e., $\mathbf{X}_j = (X_{j1}, X_{j2})$, $i = 1, 2, \dots, n_1$, and $\mathbf{Y}_j = (Y_{j1}, Y_{j2})$, $j = 1, 2, \dots, n_2$. Such setting avoids the potential computational difficulties, which we will revisit later. In this scenario, the vector combination coefficient is $\mathbf{c} = (c_1, c_2)'$. Due to the fact that the AUC is invariant to scalar transformation, finding the combination coefficient $\mathbf{c} = (c_1, c_2)'$ which maximises the AUC is equivalent to finding $\mathbf{c} = (1, \alpha)'$, where $\alpha \in (-\infty, \infty)$.

It is straightforward to show that the Mann-Whitney U statistic of the AUC associated with combination coefficient $\mathbf{c} = (1, \alpha)'$ is

$$U(\alpha) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(X_{i1} + \alpha X_{i2} < Y_{j1} + \alpha Y_{j2}). \quad (2.4)$$

As Pepe and Thompson⁷ pointed out, one might choose α such that $U(\alpha)$ is maximised. Since $U(\alpha)$ is not a continuous function of α , a search rather than a derivative-based method is needed for the maximisation procedure. It means that general-purpose optimization

algorithms such as conjugate-gradient or Newton-type methods are not appropriate for this maximisation.

To implement the maximisation by searching α , $U(\alpha)$ is evaluated for 201 equally spaced values of $\alpha \in [-1, 1]$. For $\alpha < -1$ and $\alpha > 1$,

$$U(\alpha) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(\gamma X_{i1} + X_{i2} < \gamma Y_{j1} + Y_{j2}), \text{ where } \gamma = \frac{1}{\alpha} \in [-1, 1], \text{ thus } U(\alpha) \text{ is}$$

evaluated for another 201 equally spaced values of $\gamma = \frac{1}{\alpha} \in [-1, 1]$. The optimal combination coefficient is $\hat{\mathbf{c}} = (1, \hat{\alpha})'$ or $\hat{\mathbf{c}} = (\hat{\gamma}, 1)'$ that maximises the $U(\alpha)$. See Pepe and Thompson⁷ for details.

Eventually, when $p > 2$ markers are involved, we need to search $p - 1$ coefficients $\{\alpha_2, \dots, \alpha_p\}$ using the same scheme such that

$$U(\boldsymbol{\alpha}) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(X_{i1} + \alpha_2 X_{i2} + \dots + \alpha_p X_{ip} < Y_{j1} + \alpha_2 Y_{j2} + \dots + \alpha_p Y_{jp})$$

is maximised. The idea is straightforward at the first glance, however, when the number of markers is large, i.e., $p \geq 3$, this approach is computationally formidable.⁸

3.3 Liu et al.'s approach (Min–Max)

To address the computational difficulty from Pepe and Thompson,⁷ Liu et al.⁹ proposed a nonparametric min-max combination approach that linearly combines only the minimum and maximum values of p markers to maximise the Mann-Whitney U statistic of AUC, i.e.,

$$U(\alpha) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(X_{i,\max} + \alpha X_{i,\min} < Y_{j,\max} + \alpha Y_{j,\min}) \quad (2.5)$$

where

$$X_{i,\max} = \max_{1 \leq k \leq p} X_{ik}, \quad X_{i,\min} = \min_{1 \leq k \leq p} X_{ik};$$

and

$$Y_{j,\max} = \max_{1 \leq k \leq p} Y_{jk}, \quad Y_{j,\min} = \min_{1 \leq k \leq p} Y_{jk}.$$

The searching of α is exactly the same as in Pepe and Thompson.⁷ However, such a combination only involves searching for a single coefficient and thus is computationally efficient. They showed under certain circumstances, the proposed min-max combination may yield larger AUC than empirical search of combination coefficient by Pepe and Thompson.⁷

Although this procedure is easy to implement, it comes with a few possible drawbacks. For instance, the authors commented that when the markers are measured with different units/scales, the measurements need to be standardised first before proceeding combination using this min-max combination approach. Also, since this approach just uses the minimum and maximum values of p markers, it is not clear whether the information contained in the data is fully utilised. Another difficulty of interpretation of the estimated linear combination coefficient lies in the fact that the minimum and maximum of p markers may come from different markers for different subjects.

3.4 Logistic regression approach

Walker and Duncan¹⁹ proposed logistic regression as a way of modelling the probability of an event given several independent variables. Richards et al.²⁰ presented a method for multiple test combinations that is based on a modified Bayes formula analogous to logistic regression. The logistic regression yields a linear combination of markers that intuitively discriminates non-diseased subjects from the diseased. Let \mathbf{M}_p denote a vector of p - variate observed scores from either non-diseased or diseased category. The logistic regression approach produces an intercept β_0 and a vector coefficient \mathbf{c} , i.e.,

$$\log \frac{P(D=2)}{P(D=1)} = \beta_0 + \mathbf{M}_p \mathbf{c}.$$

Of course, the vector coefficient $\hat{\mathbf{c}}$ is chosen to maximise the logistic likelihood function rather than to maximise the AUC when the method was proposed.

Recently, Jin and Lu¹⁰ proved that under the condition of generalised linear models, the coefficient from a fitted logistic regression with binary diagnostic outcomes is the optimal linear combination in the sense that it provides the highest sensitivity uniformly over the entire range of specificity and therefore yields the largest AUC among all possible linear combinations. The solution is quite appealing, although checking the assumptions underlying the generalised linear models, e.g., the correct specification of the link and variance functions, is not easy in reality.

Efron²¹ and Ruiz-Velasco²² pointed out the logistic regression is generally less efficient than the normal discriminant analysis when the normal assumption is met and thus is less efficient than Su and Liu's method based on multivariate normality. On the other hand, Cox and Snell²³ suggested the logistic regression will be more robust because estimation of the best linear combination needs no assumption of the joint distribution of the multiple biomarkers. What can be expected is that when the normality assumption is met, the logistic regression approach would not perform as well as Su and Liu's method. It is interesting to explore the performance of linear combinations from the logistic regression approach with non-normal data.

4 The proposed method

In this section, a new nonparametric approach for linearly combining markers to maximise the AUC will be discussed. This distribution-free stepwise approach aims to find the optimal combination empirically by maximising the Mann-Whitney statistic of the AUC at each step.

As in Section 3.2, the empirical estimate of AUC of the combination $\hat{\mathbf{c}} = (1, \boldsymbol{\alpha})' = (1, \alpha_2, \dots, \alpha_p)'$ is

$$U(\boldsymbol{\alpha}) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(X_{i1} + \alpha_2 X_{i2} + \dots + \alpha_p X_{ip} < Y_{j1} + \alpha_2 Y_{j2} + \dots + \alpha_p Y_{jp}).$$

When the number of markers $p \geq 3$, the empirical search for \mathbf{c} is computationally inaccessible. The nonparametric min-max procedure by Liu et al.⁹ provides alternative solution, but it still comes with aforementioned drawbacks.

To overcome all the shortcomings of the current existing combination methods, we develop a distribution-free approach that combines all the markers in a stepwise fashion as follows.

1. Estimate the AUC for each of p markers based on the Mann-Whitney statistic.
2. Assign the order from 1 to p for each marker based on their estimated AUC from largest to smallest.
3. Combine the first two markers with first two largest AUC using empirical search for combination coefficients by Pepe and Thompson.⁷
4. Having derived the combined score obtained in step 3, combine it with the marker with the third largest estimated AUC.
5. Proceed in this fashion until the marker with smallest estimated AUC is included in the linear combination.

We also consider another stepwise combination approach proceeding from the marker with the smallest estimated AUC to the marker with the largest estimated AUC. The reason that we choose these two stepwise methods for investigation is rooted in order restricted inference,²⁴ where it is argued that any other stepwise method selecting different proceeding orders would perform somewhere in between. The advantages of our stepwise approach are (1) it is distribution-free and therefore it is robust; (2) it is easy to implement and thus it offers a relief from the computational burden in the empirical search of combination coefficients in $p - 1$ dimensional spaces; (3) interpretation of estimated linear combination coefficient is relatively easy. Similar stepwise combination method has been proposed for the biomarkers with three ordinal diagnostic categories by Kang et al.²⁵; however, no work has been done for biomarkers with binary disease status.

5 Simulation studies

Simulations were conducted to compare the performance of different combination methods in improving the diagnostic/prognostic accuracy on future observations based on AUC. Overall, we compare the performance of five approaches, namely, Su and Liu's method (SULIU), Liu et al.'s min-max approach (MIN-MAX), logistic regression approach (LOGISTIC), the stepwise method proceeding from marker with largest estimated AUC to marker with smallest estimated AUC (SW1) and the stepwise method proceeding from marker with smallest estimated AUC to marker with largest estimated AUC (SW2).

The performance of all the above approaches to obtaining the largest AUC was investigated through extensive simulation studies. Eight different settings of the joint distributions of four markers ($p = 4$) were considered. For each setting, observations were generated from the underlying distribution with different sample sizes. Each combination method was applied and the estimated combination coefficients $\hat{\mathbf{c}}$ were obtained. The AUC of the combination rule was estimated from both re-substitution (Re-SUB) and leave-one-pair-out cross-validation (LOPO) for comparison purpose, although only the estimate from LOPO is considered to be unbiased and thus accurate. For each setting, 1000 Monte Carlo samples were generated to calculate the mean AUC of the combination rule and its standard error (SE). For each method, the empirical probability of yielding the largest AUC among different approaches in various simulation settings was also reported. The results are summarized in Tables 1–4.

5.1 Multivariate normal distributions with equal variance

Data from multivariate normal distributions with different mean vectors and equal variance matrices for non-diseased and diseased category were generated with the following 2 settings

$$\begin{aligned} \text{A: } \boldsymbol{\mu}_1 &= \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{pmatrix}, \boldsymbol{\mu}_2 = \begin{pmatrix} 0.6 \\ 0.8 \\ 1.0 \\ 1.2 \end{pmatrix}, \text{ and } \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{pmatrix}; \\ \text{B: } \boldsymbol{\mu}_1 &= \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{pmatrix}, \boldsymbol{\mu}_2 = \begin{pmatrix} 1.1 \\ 1.4 \\ 1.7 \\ 2.0 \end{pmatrix}, \text{ and } \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{pmatrix}. \end{aligned}$$

Under these 2 settings, the stepwise method proceeding from marker with largest estimated AUC to marker with smallest estimated AUC (SW1) produces the largest AUC estimated from re-substitution (Re-SUB) on average. However, AUC estimates from Re-SUB are biased. If we look at the AUC of the associated linear combination estimated by LOPO CV, SW2 performs much better in the sense that it produces a combination rule that would have larger discriminatory ability on future observations. This is reflected in Table 1 where SW2 produces the largest mean AUCs by LOPO and is most likely to have the largest AUCs for each Monte Carlo sample among different combination methods, except under mean configuration B with sample sizes (50, 50) in which case SULIU surpasses SW2 marginally.

LOGISTIC produces slightly larger mean AUCs than SW1 by LOPO, but it is less likely to have the largest AUCs for each Monte Carlo sample. MIN-MAX, on the other hand, has certain chance to obtain the largest AUCs among different methods, although the chance diminishes as sample sizes are getting large, under both mean configurations A and B. Thus, in terms of the linear combinations on the future observations to improve AUC under multivariate normality with equal variance assumption, SW2 method outperforms other methods. When the sample sizes are large (≥ 50), SULIU based on asymptotic formula could be better.

5.2 Multivariate normal distributions with unequal variance

Now we consider multivariate normal distributions with different mean vectors and unequal variances matrices for non-diseased and diseased category. The mean settings A and B are the same as in Section 5.1, with variance matrices set as follows,

$$\Sigma_1 = \begin{pmatrix} 1 & 0.3 & 0.3 & 0.3 \\ 0.3 & 1 & 0.3 & 0.3 \\ 0.3 & 0.3 & 1 & 0.3 \\ 0.3 & 0.3 & 0.3 & 1 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1 & 0.7 & 0.7 & 0.7 \\ 0.7 & 1 & 0.7 & 0.7 \\ 0.7 & 0.7 & 1 & 0.7 \\ 0.7 & 0.7 & 0.7 & 1 \end{pmatrix}.$$

For these settings, it is interesting to observe that MIN-MAX is far more superior to other methods in yielding the largest AUC estimated from LOPO under mean configuration A, while SW2 and SULIU have better performance under mean configuration B, as reflected in Table 2. The results with mean configuration B are somewhat similar to what we have observed in Table 1. The big difference between Tables 1 and 2 under mean configuration A suggests MIN-MAX would filter out the best linear combination when non-diseased and diseased populations are not far apart and the variances of two populations are not the same, which are common situations in practice. Notice that SULIU always performs slightly better than LOGISTIC approach, possibly due to the fact that the normality of the data with unequal variances is utilised. SW1 is always inferior to SW2. We could also see that in general when sample sizes increase, the AUCs of the associated linear combination estimated by Re-SUB are getting closer to the ones estimated by LOPO. Our recommendation for linear combinations with unequal-variance multivariate normal data is to use MIN-MAX approach if two population means are relatively close or use SW2 and/or SULIU if two population means are far apart.

5.3 Multivariate log-normal distributions with unequal variance

In this section, we would like to investigate the performances of different combination methods, assuming that the p -dimensional markers follow multivariate log-normal distributions, that is, the log-transformed markers are multivariate normally distributed. Data were first generated from the multivariate normal settings in Section 5.2 and then exponentiated to get the multivariate lognormal observations.

From Table 3, it is clearly shown that MIN-MAX, under either mean configuration A or B, is dominant in obtaining the largest AUC for each Monte Carlo sample. It suggests that for highly skewed multivariate data, MIN-MAX would most likely produce a composite score

that has the best discriminatory ability on future observations. SW2 proceeding from marker with smallest estimated AUC to marker with largest estimated AUC is inferior to MIN-MAX but superior to the remaining methods. SULIU based on multivariate normality has the worst performance, which is expected.

5.4 Multivariate normal/chi-squared/exponential/gamma distributions via normal copula

We further investigate the performances of different combination methods with p -dimensional ($p = 4$) scores assuming that the first score follows a normal, the second follows a chi-squared, an exponential for the third and a gamma distribution for the last, respectively, and coupled together via a normal copula^{26,27} with exchangeable correlation $\rho = 0.3$ and 0.7 in non-diseased and diseased category, respectively. The marginal distributions under configurations A and B for p markers for non-diseased and diseased subjects, respectively, was chosen as follows,

$$A: \begin{pmatrix} N(0.1, 1) \\ \chi_{0.1}^2 \\ \text{Exp}(0.1) \\ \Gamma(0.1, 1) \end{pmatrix}, \begin{pmatrix} N(0.6, 1) \\ \chi_{0.8}^2 \\ \text{Exp}(1.0) \\ \Gamma(1.2, 1) \end{pmatrix}; B: \begin{pmatrix} N(0.1, 1) \\ \chi_{0.1}^2 \\ \text{Exp}(0.1) \\ \Gamma(0.1, 1) \end{pmatrix}, \begin{pmatrix} N(1.1, 1) \\ \chi_{1.4}^2 \\ \text{Exp}(1.7) \\ \Gamma(2.0, 1) \end{pmatrix}$$

Under these 2 settings, the mean vectors were exactly the same as in Section 5.1. From Table 4, we can see obviously MIN-MAX seldom works. SULIU approach only has small chance in obtaining the largest AUC for the combined marker. Our proposed SW2 outperforms the other methods in terms of the mean AUC of the combined marker estimated from LOPO for the future observations and the probability in yielding the largest AUC among all the methods investigated, when the sample sizes are small to moderate. When the sample sizes are getting large (≥ 50), LOGISTIC has the best performance with marginal configuration A, and SW1 has the best performance with marginal configuration B, among all the investigated approaches.

In summary, considering the overall performance of all the five methods for the four scenarios presented in Tables 1–4, the proposed SW2 approach generally performs well using the LOPO CV method for the scenarios with or without normality assumption, although there is no clear winner. The SULIU method was developed under multivariate normality and hence it works asymptotically when population means are far apart and sample sizes are large enough to guarantee the asymptotic normality. The MIN-MAX method, on the other hand, excels in different simulation scenarios involving highly skewed multivariate data and multivariate normal data with relatively close population means and unequal variance matrices, although it still suffers from the aforementioned drawbacks including the difficulties in interpretation of linear combination coefficient. The LOGISTIC approach has slight chance to produce the best linear combinations when the multivariate normality assumption is not satisfied. The proposed SW2 outperforms the other methods in simulation settings including equal-variance multivariate normal data, unequal-variance multivariate normal data with population means far apart and normal alike data, i.e., multivariate normal-copula data. The re-substitution method, on the other hand, obviously

favours the SW1 method. This discrepancy clearly indicates the necessity of advocating the use of LOPO CV method in evaluating linear combination methods.

6 Analysis of DMD data: An example

In this section, the proposed stepwise methods (SW1/SW2) as well as SULIU, LOGISTIC and MIN-MAX approaches are applied to a real data set of 125 females on DMD available from Carnegie Mellon University Statlib Datasets Archive at <http://lib.stat.cmu.edu/datasets/biomed.desc> to combine four markers to increase the diagnostic accuracy of screening females as potential DMD carriers.

This data first discussed by Cox et al.²⁸ was gathered as part of a program to develop an effective method for screening female DMD carriers. DMD is a recessive genetic disorder passed from a mother carrier to her children, which often results in muscle degeneration, difficulty walking, breathing and even death. It is the most severe of the human dystrophies and usually appears in male children before age 5. Progressive proximal muscle weakness of the legs and pelvis associated with a loss of muscle mass is observed first. This weakness spreads to the arms, neck, and other areas and eventually leads to paralysis. As there is no effective treatment at present, it is of paramount importance to diagnose potential DMD carriers. Carriers generally have no physical symptoms but they tend to have elevated levels of certain serum enzymes. Blood samples were taken on two sets of subjects, 87 non-diseased and 38 carriers. Four different variables $M_1 - M_4$ were measured in each blood sample. For some of the subjects who had blood drawn at several different times, the average was taken. The data was then processed by a log transformation.

The empirical estimates of the AUC for these four measurements are 0.9012, 0.7494, 0.8161 and 0.8626, respectively. The stepwise method proceeding from marker with largest estimated AUC to marker with smallest estimated AUC (SW1) provides the following combination

$$M_1+4.6598 \times M_2+0.6724 \times M_3+1.7241 \times M_4$$

with an estimated AUC from LOPO to be 0.9448 for the combined marker. The stepwise method proceeding from marker with smallest estimated AUC to marker with largest estimated AUC (SW2) provides the following combination

$$M_1+1.9231 \times M_2+0.7500 \times M_3+1.1731 \times M_4$$

with an estimated AUC to be 0.9422 for the combined marker. The SULIU approach provides the following combination

$$M_1+3.6559 \times M_2+0.3522 \times M_3+1.4125 \times M_4$$

with an estimated AUC to be 0.9480; the LOGISTIC approach provides the combination

$$M_1 + 2.7052 \times M_2 + 0.3150 \times M_3 + 0.2507 \times M_4$$

with an estimated AUC to be 0.9457 and the MIN-MAX approach provides the combination

$$\max\{M_1, M_2, M_3, M_4\} + 0.2420 \times \min\{M_1, M_2, M_3, M_4\}$$

with an estimated AUC from LOPO to be 0.9140 for the combined marker. Each of the combinations provides a linearly combined score that yields a larger AUC than any of the original measurement. Since the data was pre-processed by a log transformation, it was more normally distributed. In this case, SULIU approach provides the best linear combination that has the largest discriminatory ability on future observations.

7 Discussions

In this article, we reviewed the historic developments for linear combination methods that maximise the most important diagnostic accuracy index for binary outcomes, namely, the area under the ROC curve (AUC), and proposed nonparametric stepwise methods for the same purpose. A potential problem of overestimating the AUC of the linear combination rule for future observations caused by re-substitution method in the past literature was discussed and addressed. Simulation studies were conducted to empirically compare the performances of different linear combination methods in yielding the largest AUC estimated from LOPO CV on future subjects.

Note that throughout the article, all the markers for each of the non-diseased and diseased subjects are completely observed and assumed to be accurate. Perkins et al.^{29,30} discussed ROC curve inference for biomarkers subject to limits of detection and measurement errors and proposed best linear combination of two biomarkers subject to limits of detection. Chang³¹ proposed to maximise an ROC-type measure via linear combination of markers when the gold standard is continuous rather than dichotomous. In fact, there are many types of optimal combination methods that maximise certain objective functions. In this article, we only focus on assessing linear combination methods designed to maximise the AUC.

As a referee pointed out, when the sample sizes are large enough $p \ll \min(n_1, n_2)$, the re-substitution method should be about the same as the LOPO CV method. The similar phenomenon was observed in Chen et al.³² with infinite training and independent testing data. However, in reality, the number of observations is usually not order of magnitude greater than the number of variables. So it is of critical importance to understand or at least know how to assess the performance of different linear combination methods with finite datasets.

The proposed approach is a stepwise approach which is distribution-free in nature and hence is robust with non-normal data. The computing effort and cost in obtaining the combination coefficient is significantly less than the empirical search in $p - 1$ dimensional spaces. Our simulations show the stepwise method proceeding from marker with smallest estimated AUC to marker with largest estimated AUC (SW2) is more likely to produce the optimal

linear combination that yields the largest AUC for future observations than the stepwise method proceeding from marker with largest estimated AUC to marker with smallest estimated AUC (SW1). This result, in conjunction with the estimated AUCs from re-substitution, confirms the fact that the linear combinations which work best for the current dataset might not be optimal for the future observations. Hence the LOPO CV method should always be used to assess diagnostic/prognostic accuracy of the linear combination rule performed on future observations.

We argue the fact that SW2 method is consistently better than SW1 on future observations is caused by the over-fitting/over-training issue. With small sample sizes and large number of markers, SW1 tends to put too much weight on the seemingly best markers because of its empirical top-down search strategy, which is problematic. It is also arguable that all other stepwise methods choosing different proceeding orders are inferior to SW2. We have run additional simulations that compare between stepwise methods with random proceeding orders and SW1/SW2 methods. For the same simulation configurations as in Section 5, it appears that SW2 is most likely to produce the largest AUC for future observations. These results are available upon request.

For data that is highly skewed from multivariate normal, say, multivariate log-normal, MIN-MAX approach by Liu et al.⁹ has superior performance in producing a composite score that has the largest estimated AUC. Also, with unequal-variance multivariate normal data, when two means for non-diseased and diseased populations are close to each other, MIN-MAX may be better than other methods. It is interesting to explore if adding some other order statistics, e.g., median, will improve the combinations additionally while maintaining its computational efficiency in the future research.

Last but not least, Su and Liu's method⁶ has certain chance to produce the best linear combination that yields the largest AUC for normal data only. Without normality assumption, large sample sizes are needed to guarantee the asymptotic of the formula to hold. The logistic regression approach in general has little chance to produce the best linear combination that yields the largest AUC for either normal or non-normal data. All the aforementioned discussion does not suggest Su and Liu's method⁶ and the logistic regression approach are useless. It is so simply because in reality we seldom have dataset that completely satisfy the assumption underlying Su and Liu's method⁶ and the logistic regression approach. Therefore, we recommend using Su and Liu's method⁶ and the logistic regression approach with caution in combining the diagnostic tests and biomarkers to improve the diagnostic/prognostic accuracy. Our simulation studies, although designed to cover a wide variety of parameter settings, only explored limited situations. A small-scale pilot simulation study can provide useful information on which linear combination method would outperform under certain scenarios.

Acknowledgement

Le Kang was supported in part by an appointment to the Research Participation Program at the Center for Devices and Radiological Health administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the US Food and Drug Administration. Aiyi Liu was supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development of the National Institutes of Health. The opinions expressed are solely those of the

authors and not necessarily those of the Editors. The authors thank the referees for helpful discussion and comments. A supplementary R code to implement the methods described in this article is in the Appendix.

References

1. Sidransky D. Emerging molecular markers of cancer. *Nature Rev Cancer*. 2002; 2:210–219. [PubMed: 11990857]
2. Etzioni R, Kooperberg C, Pepe M, et al. Combining biomarkers to detect disease with application to prostate cancer. *Biostatistics*. 2003; 4:523–538. [PubMed: 14557109]
3. Swets, JA.; Pickett, RM. Evaluation of diagnostic systems: Methods from signal detection theory. New York: Academic Press; 1982.
4. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Mathematical Psychol*. 1975; 12:387–415.
5. Pepe, MS. The statistical evaluation of medical tests for classification and prediction. Oxford: Oxford Statistical Science Series. Oxford University Press; 2003.
6. Su JQ, Liu JS. Linear combinations of multiple diagnostic markers. *J Am Stat Assoc*. 1993; 88:1350–1355.
7. Pepe MS, Thompson ML. Combining diagnostic test results to increase accuracy. *Biostatistics*. 2000; 1:123–140. [PubMed: 12933515]
8. 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]
9. Liu C, Liu A, Halabi S. A min-max combination of biomarkers to improve diagnostic accuracy. *Stat Med*. 2011; 30:2005–2014. [PubMed: 21472763]
10. Jin H, Lu Y. The optimal linear combination of multiple predictors under the generalized linear models. *Stat Probabil Lett*. 2009; 79:2321–2327.
11. Huang X, Qin GS, Fang YX. Optimal combinations of diagnostic tests based on AUC. *Biometrics*. 2011; 67:568–576. [PubMed: 20560934]
12. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am Stat Assoc*. 1983; 78:316–331.
13. Copas JB, Corbett P. Overestimation of the receiver operating characteristic curve for logistic regression. *Biometrika*. 2002; 89:315–331.
14. Alpaydin, E. Introduction to machine learning. (Adaptive Computation and Machine Learning). Cambridge, MA: MIT Press; 2004.
15. Friedman, J.; Hastie, T.; Tibshirani, R. The elements of statistical learning. New York: Springer Series in Statistics; 2009.
16. Quenouille MH. Notes on bias in estimation. *Biometrika*. 1956; 43:353–360.
17. Lachenbruch PA. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*. 1967; 23:639–645. [PubMed: 6080201]
18. Anderson TW, Bahadur RR. Classification into two multivariate normal distributions with different covariance matrices. *Ann Math Stat*. 1962; 33:420–431.
19. Walker SH, Duncan DB. Estimation of the probability of an event as a function of several independent variables. *Biometrika*. 1967; 54:167–179. [PubMed: 6049533]
20. Richards RJ, Hammitt JK, Tsevat J. Finding the optimal multiple-test strategy using a method analogous to logistic regression. *Med Decision Making*. 1996; 16:367–375.
21. Efron B. The efficiency of logistic regression compared to normal discriminant analysis. *J Am Stat Assoc*. 1975; 70:892–898.
22. Ruiz-Velasco S. Asymptotic efficiency of logistic regression relative to linear discriminant analysis. *Biometrika*. 1991; 78:235–243.
23. Cox, DR.; Snell, EJ. Analysis of binary data. 2nd ed. London: Chapman & Hall; 1989.
24. Robertson, T.; Wright, FT.; Dykstra, RL. Order restricted statistical inference. New York: John Wiley & Sons; 1988.

25. Kang L, Xiong C, Crane P, et al. Linear combinations of biomarkers to improve diagnostic accuracy with three ordinal diagnostic categories. *Stat Med.* 2012; 32(4):631–643. [PubMed: 22865796]
26. Nelsen, RB. An introduction to copulas. New York: Springer; 1999.
27. Kojadinovic I, Yan J. Modeling multivariate distributions with continuous margins using the copula R package. *J Stat Software.* 2010; 34:1–20.
28. Cox, LH.; Johnson, MM.; Kafadar, K. Exposition of statistical graphics technology; Proceedings of the American Statistical Association, Computation Section; 1982. p. 55-56.
29. Perkins NJ, Schisterman EF, Vexler A. Generalized ROC curve inference for a biomarker subject to a limit of detection and measurement error. *Stat Med.* 2009; 28(13):1841–1860. [PubMed: 19340817]
30. Perkins NJ, Schisterman EF, Vexler A. ROC curve inference for best linear combination of two biomarkers subject to limits of detection. *Biomet J.* 2011; 53(3):464–476.
31. Chang YI. Maximizing an ROC-type measure via linear combination of markers when the gold reference is continuous. *Stat Med.* 2012
32. Chen W, Wagner RF, Yousef WA, et al. Comparison of classifier performance estimators: a simulation study. *Med Imag.* 2009 72630X-11.

Appendix: Relevant R Code

```
#####
## Mann-Whitney U stat for AUC: continuous data
#####
nonp.auc <- function(u,v) {
  n1 = length(u)
  n2 = length(v)
  return(sum(sapply(u,function(x) sum(x<v)))/n1/n2)
#####
## Su and Liu's method
#####
suliu <- function(new.1,new.2) {
  a = var(new.1)+var(new.2)
  b = colMeans(new.2) - colMeans(new.1)
  est.coef = as.numeric(solve(a)%*%b)
  check.sign = nonp.auc(new.1%*%est.coef,new.2%*%est.coef)
  if(check.sign>=0.5) return(list(coef=est.coef,
    auc.combined=check.sign)) else return(list(coef=-est.coef,
    auc.combined=1-check.sign))
}
#####
## logistic regression approach
#####
logistic <- function(new.1,new.2) {
  n1 = nrow(new.1)
  n2 = nrow(new.2)
  dat.lr <- data.frame(cbind(response = rep(c(0,1),times=c(n1,n2)),
    rbind(new.1,new.2)))
```

```

obj.lr <- try(glm(response
~.,data=dat.lr,family=binomial(link="logit")),silent=T)
if (any(class(obj.lr)== "try-error")) return(list(coef=rep(0,ncol(new.1)),
auc.combined=0.5)) else est.coef=as.numeric(obj.lr$coef[-1])
check.sign = nonp.auc(new.1%%est.coef,new.2%%est.coef)
if(check.sign>=0.5) return(list(coef=est.coef,
auc.combined=check.sign)) else return(list(coef=-est.coef,
auc.combined=1-check.sign))
}
#####
##### data.1, data.2 must be of two-column
#####
nonpar.combine2.auc <- function(alpha,rate,data.1,data.2) {
n1 = nrow(data.1)
n2 = nrow(data.2)
new.1 = data.1%%c(alpha,rate)
new.2 = data.2%%c(alpha,rate)
nonp.auc(new.1,new.2)
}
nonpar.combine2.coef <- function(new.1,new.2,evalnum=201) {
rate=seq(-1,1,length=evalnum)
alpha=rev(rate)[-1]
auc.rate_x = sapply(rate, nonpar.combine2.auc, alpha=1,data.1=new.1,data.
2=new.2)
auc.alpha_x = sapply(alpha, nonpar.combine2.auc,
rate=1,data.1=new.1,data.2=new.2)
auc.0 = c(auc.rate_x,auc.alpha_x)
amax.idx = which.max(auc.0)
if (amax.idx<=evalnum) return(c(alpha=1,rate=rate[amax.idx],
auc.max=auc.0[amax.idx]))
if (amax.idx> evalnum) return(c(alpha=alpha[amax.idx-evalnum],rate=1,
auc.max=auc.0[amax.idx]))
}
nonp.auc.check <- function(health,middle) {
auc.i=numeric(ncol(health))
for (i in 1:ncol(health)) {
new.1=health[,i]
new.2=middle[,i]
auc.i[i]=nonp.auc(new.1,new.2)}
auc.i
}
#####
### Step-wise method
#####

```

```

step.coef <- function(new.1,new.2,design='step-down') {
n1 = nrow(new.1)
n2 = nrow(new.2)
VARnum = ncol(new.1)
combcoef = matrix(0,nrow=VARnum-1,ncol=2)
if (design=='step-down') {
auc.order = sort(nonp.auc.check(health=new.1,middle=new.2),
  index.return=T,decreasing=T)$ix) else {
auc.order = sort(nonp.auc.check(health=new.1,middle=new.2),
  index.return=T,decreasing=F)$ix)
combmarker.1=new.1[,auc.order[1]]
combmarker.2=new.2[,auc.order[1]]
nal.coef = 1
for (i in 2:VARnum) {
combmarker.1 = cbind(combmarker.1,new.1[,auc.order[i]])
combmarker.2 = cbind(combmarker.2,new.2[,auc.order[i]])
temp.info = nonpar.combine2.coef(combmarker.1,combmarker.2)
combcoef[i-1,] = temp.info[1:2]
nal.coef = c(nal.coef*combcoef[i-1,1],combcoef[i-1,2])
combmarker.1 = combmarker.1%%temp.info[1:2]
combmarker.2 = combmarker.2%%temp.info[1:2]
}
nal.coef = nal.coef[sort(auc.order,index.return=T)$ix]
check.sign = nonp.auc(new.1%%nal.coef,new.2%%nal.coef)
if(check.sign<=0.5) nal.coef=-nal.coef
return(list(coef=as.numeric(nal.coef),
  auc.combined=as.numeric(temp.info[3]),
  check=(max(check.sign,1-check.sign))==temp.info[3])
))
}
#####
##### Min-Max method
#####
liu.coef <- function(data.1,data.2) {
max_min.1 = cbind(apply(data.1,1,max),apply(data.1,1,min))
max_min.2 = cbind(apply(data.2,1,max),apply(data.2,1,min))
est.coef = nonpar.combine2.coef(max_min.1,max_min.2)[1:2]
check.sign = nonp.auc(max_min.1%%est.coef,max_min.2%%est.coef)
if(check.sign>=0.5) return(list(coef=est.coef,
  auc.combined=check.sign)) else return(list(coef=-est.coef,
  auc.combined=1-check.sign))
}

```

Table 1
 Mean area under receiver operating characteristic curve (AUC) ± SE and probability of obtaining the largest AUC (beneath in parentheses).

Sample size	Mean config	SULIU	LOGISTIC	SW1	SW2	MIN-MAX		
(20,20)	A	Re-SUB	.831 ± .002 (0.171)	.832 ± .002 (0.195)	.836 ± .002 (0.470)	.822 ± .002 (0.087)	.783 ± .002 (0.078)	
		LOPO	.760 ± .003 (0.165)	.758 ± .003 (0.102)	.750 ± .003 (0.126)	.770 ± .003 (0.362)	.737 ± .003 (0.245)	
	B	Re-SUB	.945 ± .001 (0.116)	.948 ± .001 (0.256)	.950 ± .001 (0.464)	.943 ± .001 (0.107)	.917 ± .001 (0.057)	
		LOPO	.915 ± .002 (0.271)	.910 ± .001 (0.111)	.906 ± .002 (0.127)	.916 ± .001 (0.310)	.894 ± .002 (0.179)	
	(20,30)	A	Re-SUB	.826 ± .002 (0.151)	.827 ± .002 (0.196)	.830 ± .002 (0.506)	.817 ± .002 (0.071)	.777 ± .002 (0.076)
			LOPO	.767 ± .002 (0.151)	.767 ± .002 (0.140)	.758 ± .003 (0.140)	.774 ± .002 (0.345)	.741 ± .003 (0.223)
(30,50)	B	Re-SUB	.939 ± .001 (0.130)	.941 ± .001 (0.242)	.943 ± .001 (0.470)	.936 ± .001 (0.100)	.911 ± .001 (0.059)	
		LOPO	.913 ± .001 (0.258)	.910 ± .001 (0.131)	.906 ± .001 (0.119)	.913 ± .001 (0.324)	.892 ± .002 (0.168)	
	A	Re-SUB	.816 ± .001 (0.161)	.816 ± .002 (0.174)	.818 ± .001 (0.568)	.808 ± .002 (0.056)	.771 ± .002 (0.041)	
		LOPO	.779 ± .002 (0.150)	.779 ± .002 (0.138)	.774 ± .002 (0.215)	.781 ± .002 (0.366)	.748 ± .002 (0.132)	
	B	Re-SUB	.935 ± .001 (0.116)	.936 ± .001 (0.214)	.938 ± .001 (0.571)	.932 ± .001 (0.068)	.908 ± .001 (0.030)	
		LOPO	.919 ± .001 (0.271)	.918 ± .001 (0.120)	.915 ± .001 (0.194)	.918 ± .001 (0.323)	.897 ± .001 (0.094)	
(50,50)	A	Re-SUB	.815 ± .001 (0.165)	.816 ± .001 (0.198)	.817 ± .001 (0.554)	.808 ± .001 (0.055)	.772 ± .001 (0.028)	
		LOPO	.788 ± .002 (0.184)	.788 ± .002 (0.141)	.784 ± .002 (0.235)	.788 ± .001 (0.337)	.756 ± .002 (0.103)	

Sample size	Mean config	SULIU	LOGISTIC	SW1	SW2	MIN-MAX
B	Re-SUB	.933 ± .001 (0.120)	.934 ± .001 (0.226)	.935 ± .001 (0.585)	.930 ± .001 (0.048)	.905 ± .001 (0.021)
	LOPO	.921 ± .001 (0.307)	.920 ± .001 (0.141)	.918 ± .001 (0.205)	.920 ± .001 (0.294)	.897 ± .001 (0.053)

Simulation setting 1: normal data with equal variance

$$A; \mu_1 = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{pmatrix}, \mu_2 = \begin{pmatrix} 0.6 \\ 0.8 \\ 1.0 \\ 1.2 \end{pmatrix}; B; \mu_1 = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{pmatrix}, \mu_2 = \begin{pmatrix} 1.1 \\ 1.4 \\ 1.7 \\ 2.0 \end{pmatrix}$$

$$\Sigma_1 = \Sigma_2 = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{pmatrix}$$

Re-SUB: AUC of the associated linear combination estimated by re-substitution (biased); LOPO: AUC of the associated linear combination estimated by leave-one-pair-out cross-validation (unbiased); SULIU: Su and Liu's method; LOGISTIC: linear combination coefficient from logistic regression approach; SW1: stepwise method proceeding from marker with largest estimated AUC to marker with smallest estimated AUC; SW2: stepwise method proceeding from marker with smallest estimated AUC to marker with largest estimated AUC; MIN-MAX: Liu et al.'s min-max approach.

Table 2
 Mean area under receiver operating characteristic curve (AUC) ± SE and probability of obtaining the largest AUC (beneath in parentheses).

Sample Size	Mean config	SULIU	LOGISTIC	SW1	SW2	MIN-MAX	
(20,20)	A	Re-SUB	.837 ± .002 (0.171)	.840 ± .002 (0.291)	.826 ± .002 (0.044)	.832 ± .002 (0.374)	
		LOPO	.766 ± .003 (0.111)	.755 ± .003 (0.080)	.773 ± .003 (0.193)	.797 ± .003 (0.525)	
	B	Re-SUB	.943 ± .001 (0.120)	.945 ± .001 (0.231)	.939 ± .001 (0.432)	.919 ± .001 (0.114)	
		LOPO	.911 ± .002 (0.216)	.906 ± .002 (0.123)	.913 ± .001 (0.304)	.896 ± .002 (0.253)	
	(20,30)	A	Re-SUB	.831 ± .002 (0.124)	.835 ± .002 (0.303)	.821 ± .002 (0.030)	.825 ± .002 (0.380)
			LOPO	.769 ± .002 (0.098)	.766 ± .002 (0.075)	.776 ± .002 (0.207)	.796 ± .002 (0.523)
B		Re-SUB	.939 ± .001 (0.109)	.941 ± .001 (0.238)	.943 ± .001 (0.450)	.916 ± .001 (0.113)	
		LOPO	.911 ± .001 (0.221)	.908 ± .001 (0.119)	.912 ± .001 (0.289)	.898 ± .001 (0.249)	
(30,50)		A	Re-SUB	.825 ± .001 (0.130)	.825 ± .001 (0.301)	.816 ± .001 (0.020)	.823 ± .001 (0.401)
			LOPO	.785 ± .002 (0.096)	.783 ± .002 (0.080)	.787 ± .002 (0.162)	.806 ± .002 (0.551)
	B	Re-SUB	.936 ± .001 (0.131)	.937 ± .001 (0.217)	.938 ± .001 (0.518)	.910 ± .001 (0.088)	
		LOPO	.918 ± .001 (0.275)	.916 ± .001 (0.112)	.913 ± .001 (0.153)	.899 ± .001 (0.164)	
	A	Re-SUB	.813 ± .001 (0.097)	.813 ± .001 (0.150)	.805 ± .001 (0.265)	.815 ± .001 (0.482)	
		LOPO	.786 ± .002 (0.111)	.785 ± .002 (0.073)	.781 ± .002 (0.104)	.803 ± .001 (0.582)	

Sample Size	Mean config	SULIU	LOGISTIC	SW1	SW2	MIN-MAX
B	Re-SUB	.934 ± .001 (0.114)	.934 ± .001 (0.215)	.935 ± .001 (0.544)	.930 ± .001 (0.051)	.910 ± .001 (0.076)
	LOPO	.922 ± .001 (0.271)	.921 ± .001 (0.123)	.919 ± .001 (0.203)	.920 ± .001 (0.261)	.903 ± .001 (0.142)

Simulation setting 2: normal data with unequal variance

$$A; \mu_1 = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{pmatrix}, \mu_2 = \begin{pmatrix} 0.6 \\ 0.8 \\ 1.0 \\ 1.2 \end{pmatrix}; B; \mu_1 = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{pmatrix}, \mu_2 = \begin{pmatrix} 1.1 \\ 1.4 \\ 1.7 \\ 2.0 \end{pmatrix}$$

$$\Sigma_1 = \begin{pmatrix} 1 & 0.3 & 0.3 & 0.3 \\ 0.3 & 1 & 0.3 & 0.3 \\ 0.3 & 0.3 & 1 & 0.3 \\ 0.3 & 0.3 & 0.3 & 1 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1 & 0.7 & 0.7 & 0.7 \\ 0.7 & 1 & 0.7 & 0.7 \\ 0.7 & 0.7 & 1 & 0.7 \\ 0.7 & 0.7 & 0.7 & 1 \end{pmatrix}$$

Re-SUB: AUC of the associated linear combination estimated by re-substitution (biased); LOPO: AUC of the associated linear combination estimated by leave-one-pair-out cross-validation (unbiased); SULIU: Su and Liu's method; LOGISTIC: linear combination coefficient from logistic regression approach; SW1: stepwise method proceeding from marker with largest estimated AUC to marker with smallest estimated AUC; SW2: stepwise method proceeding from marker with smallest estimated AUC to marker with largest estimated AUC; MIN-MAX: Liu et al.'s min-max approach.

Table 3

Mean area under receiver operating characteristic curve (AUC) ± SE and probability of obtaining the largest AUC (beneath in parentheses).

Sample size	Mean config	SULIU	LOGISTIC	SW1	SW2	MIN-MAX	
(20,20)	A	Re-SUB	.801 ± .002 (0.040)	.821 ± .002 (0.154)	.836 ± .002 (0.372)	.818 ± .002 (0.372)	.828 ± .002 (0.372)
		LOPO	.721 ± .003 (0.071)	.741 ± .003 (0.071)	.741 ± .003 (0.115)	.756 ± .003 (0.222)	.790 ± .002 (0.522)
	B	Re-SUB	.897 ± .002 (0.010)	.933 ± .001 (0.231)	.940 ± .001 (0.453)	.932 ± .001 (0.118)	.917 ± .001 (0.187)
		LOPO	.850 ± .002 (0.052)	.884 ± .002 (0.132)	.885 ± .002 (0.115)	.898 ± .002 (0.115)	.893 ± .002 (0.386)
(20,30)	A	Re-SUB	.795 ± .002 (0.020)	.815 ± .002 (0.169)	.828 ± .002 (0.323)	.812 ± .002 (0.024)	.827 ± .002 (0.464)
		LOPO	.727 ± .003 (0.065)	.744 ± .003 (0.078)	.747 ± .003 (0.089)	.758 ± .003 (0.165)	.798 ± .002 (0.603)
	B	Re-SUB	.894 ± .001 (0.011)	.931 ± .001 (0.227)	.936 ± .001 (0.464)	.929 ± .001 (0.098)	.914 ± .001 (0.201)
		LOPO	.859 ± .002 (0.068)	.891 ± .002 (0.137)	.892 ± .002 (0.123)	.900 ± .001 (0.293)	.895 ± .001 (0.379)
(30,50)	A	Re-SUB	.784 ± .002 (0.018)	.802 ± .002 (0.121)	.814 ± .001 (0.311)	.800 ± .002 (0.015)	.821 ± .002 (0.536)
		LOPO	.739 ± .002 (0.044)	.756 ± .002 (0.070)	.760 ± .002 (0.105)	.764 ± .002 (0.142)	.804 ± .002 (0.639)
	B	Re-SUB	.896 ± .001 (0.007)	.927 ± .001 (0.201)	.931 ± .001 (0.536)	.925 ± .001 (0.042)	.912 ± .001 (0.214)
		LOPO	.876 ± .001 (0.054)	.902 ± .001 (0.149)	.903 ± .001 (0.156)	.907 ± .001 (0.296)	.901 ± .001 (0.345)
A	Re-SUB	.779 ± .001 (0.018)	.794 ± .002 (0.080)	.807 ± .001 (0.317)	.795 ± .001 (0.006)	.817 ± .001 (0.579)	
	LOPO	.747 ± .002 (0.026)	.762 ± .002 (0.067)	.767 ± .002 (0.125)	.770 ± .002 (0.139)	.804 ± .001 (0.643)	

Sample size	Mean config	SULIU	LOGISTIC	SW1	SW2	MIN-MAX
B	Re-SUB	.892 ± .001 (0.006)	.920 ± .001 (0.134)	.925 ± .001 (0.597)	.920 ± .001 (0.046)	.909 ± .001 (0.217)
	LOPO	.876 ± .001 (0.041)	.903 ± .001 (0.146)	.904 ± .001 (0.179)	.906 ± .001 (0.302)	.901 ± .001 (0.332)

Simulation setting 3: log-normal data with unequal variance

$$\begin{aligned}
 A; \mu_1 = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{pmatrix}, \mu_2 = \begin{pmatrix} 0.6 \\ 0.8 \\ 1.0 \\ 1.2 \end{pmatrix}; B; \mu_1 = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \\ 0.1 \end{pmatrix}, \mu_2 = \begin{pmatrix} 1.1 \\ 1.4 \\ 1.7 \\ 2.0 \end{pmatrix} \\
 \Sigma_1 = \begin{pmatrix} 1 & 0.3 & 0.3 & 0.3 \\ 0.3 & 1 & 0.3 & 0.3 \\ 0.3 & 0.3 & 1 & 0.3 \\ 0.3 & 0.3 & 0.3 & 1 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1 & 0.7 & 0.7 & 0.7 \\ 0.7 & 1 & 0.7 & 0.7 \\ 0.7 & 0.7 & 1 & 0.7 \\ 0.7 & 0.7 & 0.7 & 1 \end{pmatrix}
 \end{aligned}$$

Re-SUB: AUC of the associated linear combination estimated by re-substitution (biased); LOPO: AUC of the associated linear combination estimated by leave-one-pair-out cross-validation (unbiased); SULIU: Su and Liu's method; LOGISTIC: linear combination coefficient from logistic regression approach; SW1: stepwise method proceeding from marker with largest estimated AUC to marker with smallest estimated AUC; SW2: stepwise method proceeding from marker with smallest estimated AUC to marker with largest estimated AUC; MIN-MAX: Liu et al.'s min-max approach.

Table 4

Mean area under receiver operating characteristic curve (AUC) ± SE and probability of obtaining the largest AUC (beneath in parentheses).

Sample size	Mean config	SULIU	LOGISTIC	SW1	SW2	MIN-MAX		
(20,20)	A	Re-SUB	.947 ± .001 (0.037)	.971 ± .001 (0.359)	.973 ± .001 (0.394)	.968 ± .001 (0.208)	.828 ± .002 (0.002)	
		LOPO	.907 ± .002 (0.147)	.930 ± .001 (0.279)	.921 ± .002 (0.183)	.935 ± .001 (0.380)	.777 ± .003 (0.011)	
	B	Re-SUB	.975 ± .001 (0.066)	.992 ± .000 (0.327)	.993 ± .000 (0.348)	.991 ± .000 (0.243)	.926 ± .001 (0.016)	
		LOPO	.950 ± .001 (0.192)	.960 ± .001 (0.260)	.952 ± .001 (0.120)	.969 ± .001 (0.401)	.892 ± .002 (0.027)	
	(20,30)	A	Re-SUB	.944 ± .001 (0.016)	.970 ± .001 (0.423)	.970 ± .001 (0.403)	.965 ± .001 (0.159)	.821 ± .002 (0.000)
			LOPO	.913 ± .001 (0.134)	.935 ± .001 (0.304)	.931 ± .001 (0.203)	.936 ± .001 (0.357)	.778 ± .003 (0.001)
(30,50)	B	Re-SUB	.975 ± .001 (0.054)	.991 ± .000 (0.347)	.992 ± .000 (0.367)	.990 ± .000 (0.227)	.921 ± .001 (0.005)	
		LOPO	.958 ± .001 (0.191)	.965 ± .001 (0.220)	.962 ± .001 (0.144)	.973 ± .001 (0.434)	.893 ± .002 (0.012)	
	A	Re-SUB	.947 ± .001 (0.018)	.966 ± .001 (0.386)	.967 ± .001 (0.477)	.962 ± .001 (0.119)	.813 ± .001 (0.000)	
		LOPO	.928 ± .001 (0.093)	.945 ± .001 (0.364)	.943 ± .001 (0.253)	.943 ± .001 (0.290)	.783 ± .002 (0.000)	
	B	Re-SUB	.974 ± .001 (0.011)	.989 ± .000 (0.361)	.990 ± .000 (0.485)	.988 ± .000 (0.143)	.915 ± .001 (0.000)	
		LOPO	.964 ± .001 (0.112)	.975 ± .001 (0.279)	.975 ± .001 (0.258)	.977 ± .001 (0.350)	.894 ± .001 (0.001)	
(50,50)	A	Re-SUB	.947 ± .001 (0.023)	.962 ± .001 (0.348)	.964 ± .001 (0.528)	.958 ± .001 (0.100)	.810 ± .001 (0.000)	
		LOPO	.933 ± .001 (0.075)	.948 ± .001 (0.379)	.947 ± .001 (0.319)	.945 ± .001 (0.227)	.785 ± .002 (0.000)	

Sample size	Mean config	SULIU	LOGISTIC	SW1	SW2	MIN-MAX
B	Re-SUB	.974 ± .000 (0.011)	.987 ± .000 (0.280)	.989 ± .000 (0.618)	.986 ± .000 (0.091)	.914 ± .001 (0.000)
	LOPO	.966 ± .001 (0.110)	.978 ± .000 (0.287)	.979 ± .000 (0.331)	.979 ± .000 (0.271)	.897 ± .001 (0.000)

Simulation setting 4: normal- χ^2 -exponential-gamma copula data

$$\begin{matrix}
 \text{A:} \\
 \left(\begin{matrix} N(0.1, 1) \\ \chi_{0.1}^2 \\ \text{Exp}(0.1) \\ \Gamma(0.1, 1) \end{matrix} \right) , \left(\begin{matrix} N(0.6, 1) \\ \chi_{0.8}^2 \\ \text{Exp}(1.0) \\ \Gamma(1.2, 1) \end{matrix} \right) \text{B:} \\
 \left(\begin{matrix} N(0.1, 1) \\ \chi_{0.1}^2 \\ \text{Exp}(0.1) \\ \Gamma(0.1, 1) \end{matrix} \right) , \left(\begin{matrix} N(1.1, 1) \\ \chi_{1.4}^2 \\ \text{Exp}(1.7) \\ \Gamma(2.0, 1) \end{matrix} \right)
 \end{matrix}$$

Re-SUB: AUC of the associated linear combination estimated by re-substitution (biased); LOPO: AUC of the associated linear combination estimated by leave-one-pair-out cross-validation (unbiased); SULIU: Su and Liu's method; LOGISTIC: linear combination coefficient from logistic regression approach; SW1: stepwise method proceeding from marker with largest estimated AUC to marker with smallest estimated AUC; SW2: stepwise method proceeding from marker with smallest estimated AUC to marker with largest estimated AUC; MIN-MAX: Liu et al.'s min-max approach.