Statistics
in Medicine

Research Article

Received 17 June 2013, Accepted 20 October 2013 Published online 6 December 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6046

Optimal linear combinations of multiple diagnostic biomarkers based on Youden index

Jingjing Yin and Lili Tian*†

In practice, usually multiple biomarkers are measured on the same subject for disease diagnosis. Combining these biomarkers into a single score could improve diagnostic accuracy. Many researchers have addressed the problem of finding the optimal linear combination based on maximizing the area under *ROC* curve (*AUC*). Actually, such combined score might have less than optimal property at the diagnostic threshold. In this paper, we propose the idea of using Youden index as an objective function for searching the optimal linear combination. The combined score directly achieves the maximum overall correct classification rate at the diagnostic threshold corresponding to Youden index; in other words, it is the optimal linear combination score for making the disease diagnosis. We present both empirical and numerical searching methods for the optimal linear combination. We carry out extensive simulation study to investigate the performance of the proposed methods. Additionally, we empirically compare the optimal overall classification rates between the proposed combination based on Youden index and the traditional one based on *AUC* and demonstrate a significant gain in diagnostic accuracy for the proposed combination. In the end, we apply the proposed methods to a real data set.

Keywords: ROC analysis; Youden index; linear combination; diagnostic accuracy

1. Introduction

The receiver operating characteristic (*ROC*) curve is a very useful tool in diagnostics for the purpose of evaluating the discriminatory ability of biomarkers or diagnostic tests. For a continuous-scaled marker, the *ROC* curve graphically depicts the marker's diagnostic ability for all threshold values in a unit square by plotting proportion of true positives (sensitivity) versus proportion of false positives (1− specificity). Extensive statistical research has been performed in this field. For excellent reviews of statistical methods involving *ROC* curves; see Shapiro [1], Zhou *et al.* [2], Pepe [3], and Zou *et al.* [4]. The area under the *ROC* curve (*AUC*) is the most popular overall diagnostic accuracy index, and it has been extensively used by many researchers for biomarker evaluation and selection. Greater *AUC* value indicates greater discriminatory ability of a biomarker over all threshold values. There exist many research papers on inference methods for *AUC*, for example, Delong *et al.* [5], Wieand *et al.* [6], Obuchowski and Lieber [7], Tian [8], and Qin and Hotilovac [9].

In practice, it is very common that several biomarkers are measured on the same subjects in order to provide as much information as possible for more accurate disease diagnosis. Combining these biomarkers into a single score usually gives better diagnostic performance than each biomarker alone. There are many ways to combine multiple biomarkers, and the easiest and most popular way is to combine them linearly. The linear combination methods using *AUC* as objective function have been explored extensively. For example, Su and Liu [10] presented analytical form for the coefficient vector of linear combination under multivariate normal assumption. Pepe and Thompson [11] empirically performed

Department of Biostatistics, State University of New York at Buffalo, 717 Kimball Tower, 3435 Main St. Bldg. 26, Buffalo, NY 14214-3000, U.S.A.

**Correspondence to: Lili Tian, Department of Biostatistics, State University of New York at Buffalo, 717 Kimball Tower, 3435 Main St. Bldg. 26, Buffalo, NY 14214-3000 U.S.A.*

†*E-mail: ltian@buffalo.edu*

a grid search for the optimal linear combination in the p-dimensional space without any parametric assumption. However, researchers following Pepe and Thompson's [11] work pointed out that their method becomes computationally challenging when the number of biomarkers is larger than 3, and hence some more efficient searching approaches are proposed afterwards. For instance, Liu *et al.* [12] performed a min–max approach that only combines the minimum and maximum of all biomarkers. Recently, Kang *et al.* [13] proposed a stepwise combination method to maximize volume under the *ROC* surface (*VUS*) for multiple biomarkers of three ordinal disease stages.

For the purpose of making diagnosis, that is, classifying a subject as either diseased or healthy, a threshold value (or cut-off point) is needed. The diagnostic accuracy is often measured by the overall correct classification rate (i.e., sensitivity plus specificity) at the threshold. The combined score based on maximization of AUC does not necessarily have the best diagnostic ability at the optimal threshold as *AUC* is a global summary measure of diagnostic accuracy across all possible thresholds. Therefore, to generate an optimal combined marker for making diagnosis, which achieves the largest possible overall correct classification rate, naturally the overall correct classification rate should be used as the objective function.

Very few research articles addressed the issue of combining biomarkers using other summary statistics of *ROC* curve as objective functions besides *AUC*. Among them, Gao *et al.* [14] numerically searched for optimum linear combination of multivariate normal distributed biomarkers that maximizes sensitivity at a fixed specificity. Liu *et al.* [15] obtained best linear combination based on maximizing sensitivity over a low or high range of specificity.

Youden index (J), another popular summary index of *ROC* curve besides *AUC*, is defined as $J = \max_c\{Sensitivity(c) + Specificity(c) - 1\}$ where c stands for a threshold value. The c value that corresponds to J is referred as the optimal cut-point [16]. For diagnostic threshold selection for a particular biomarker, there exist several approaches out of which Youden index is the most popular one in practice because the concept of Youden index ties nicely into the *ROC* framework [17]. In medical and biological sciences, Youden index has been extensively used to select a threshold for making diagnoses and classification. Youden index is important because not only it offers the optimal cut-off point but it is also a direct measure of the maximum overall correct classification rate a marker can achieve. Combining markers using Youden index as an objective function will lead to a combined score with the largest possible overall correct classification rate at the diagnostic threshold among all possible linear combinations. Hence, such combined biomarker has the best accuracy for making diagnosis.

This paper aims to propose several empirical and numerical (i.e., derivation based) methods for finding the optimal linear combination of multiple biomarkers based on maximizing Youden index. The rest of the paper is organized as follows. In Section 2, we present notations and background. In Section 3, we discuss two distribution-free empirical searching methods. Section 4 presents two derivation-based methods: a parametric method under multivariate normality and a non-parametric kernel smoothing method. Section 5 presents simulation results about the performance of the proposed linear combination methods. Especially, we empirically compared the optimal overall classification rates between the proposed combination based on Youden index and the traditional one based on *AUC* and demonstrated a significant gain in diagnostic accuracy for the proposed approach. In Section 6, we used a real data set for illustrating the proposed methods. Section 7 contains a summary and discussion.

2. Preliminary

Assume that there are n_1 and n_2 subjects for diseased and non-diseased groups, respectively, and p markers on continuous scale are measured for the same subject. Let $\boldsymbol{Y}_1$ be the p-dimensional vector of the biomarker measurements in the diseased group and $\boldsymbol{Y}_{11}, \boldsymbol{Y}_{12}, \ldots, \boldsymbol{Y}_{1n_1}$ be a random sample of size n_1, and

$$\boldsymbol{Y}_{1i} = \left(Y_{1i1}, Y_{1i2}, \ldots, Y_{1ip}\right)^T,\ i = 1, 2, \ldots, n_1,$$

where Y_{1ik} $(k = 1, 2, \ldots, p)$ denote the measurement of the k^{th} biomarker on the i^{th} individual in the diseased group. Similarly, let $\boldsymbol{Y}_2$ be the p-dimensional vector of the biomarker measurements in the non-diseased group and $\boldsymbol{Y}_{21}, \boldsymbol{Y}_{22}, \ldots, \boldsymbol{Y}_{2n_2}$ be a random sample of size n_2, and

$$\boldsymbol{Y}_{2j} = \left(Y_{2j1}, Y_{2j2}, \ldots, Y_{2jp}\right)^T,\ j = 1, 2, \ldots, n_2,$$

where Y_{2jk} $(k = 1, 2, \ldots, p)$ denote the measurement of the k^{th} biomarker on the j^{th} individual in the non-diseased group. Let $\boldsymbol{w} = \left(w_1, w_2, \ldots, w_p\right)^T$ be the vector of combination coefficients. The combined scores are defined as

$$Z_g = \boldsymbol{w}^T \boldsymbol{Y}_g \quad \text{for } g = 1, 2 \tag{1}$$

for diseased and healthy groups, respectively.

For the k^{th} biomarker, let $F_{Y_{1k}}(.)$ and $F_{Y_{2k}}(.)$ denote the CDFs of measurements for the diseased and healthy populations, respectively. Given a threshold value c, sensitivity (P_{1k}) and specificity (P_{2k}) are

$$P_{1k}(c) = 1 - F_{Y_{1k}}(c), \; P_{2k}(c) = F_{Y_{2k}}(c).$$

The corresponding estimates are

$$\hat{P}_{1k}(c) = 1 - \hat{F}_{Y_{1k}}(c), \; \hat{P}_{2k}(c) = \hat{F}_{Y_{2k}}(c)$$

where $\hat{F}_{Y_{1k}}(.)$ and $\hat{F}_{Y_{2k}}(.)$ are the estimated CDFs of diseased and healthy samples, respectively.

In practice, the cut-off value c is usually unknown and is selected by optimizing an objective function, such as Youden index. The Youden index (J_k) for the k^{th} marker is defined as $J_k = \max_c\{P_{1k}(c) + P_{2k}(c) - 1\}$ [16]. The optimal threshold value for the k^{th} marker determined by Youden index is denoted as c_k, and it is obtained as

$$\begin{aligned} c_k &= \{c : \max_c \left(P_{1k}(c) + P_{2k}(c) - 1\right)\}; \\ &= \left\{c : \max_c \left(F_{Y_{2k}}(c) - F_{Y_{1k}}(c)\right)\right\}. \end{aligned} \tag{2}$$

The sensitivity and specificity at the optimal cut-point via Youden index are

$$P_{1k}(c_k) = 1 - F_{Y_{1k}}(c_k); P_{2k}(c_k) = F_{Y_{2k}}(c_k).$$

The estimated Youden index for the k^{th} marker is

$$\hat{J}_k = \hat{P}_{1k}(\hat{c}_k) + \hat{P}_{2k}(\hat{c}_k) - 1 = \hat{F}_{Y_{2k}}(\hat{c}_k) - \hat{F}_{Y_{1k}}(\hat{c}_k) \tag{3}$$

where

$$\hat{c}_k = \left\{c : \max_c \left(\hat{F}_{Y_{2k}}(c) - \hat{F}_{Y_{1k}}(c)\right)\right\}. \tag{4}$$

For the combined score Z_g in (1), denote the corresponding CDFs as $F_{Z_g}(.)$ for $g = 1, 2$, and the optimal cut-point associated with Youden index as $c_{\boldsymbol{w}}$. The estimated Youden index of the combined score is

$$\hat{J}_{\boldsymbol{w}} = \hat{F}_{Z_2}(\hat{c}_{\boldsymbol{w}}) - \hat{F}_{Z_1}(\hat{c}_{\boldsymbol{w}}) \tag{5}$$

where

$$\hat{c}_{\boldsymbol{w}} = \left\{c : \max_c \left(\hat{F}_{Z_2}(c) - \hat{F}_{Z_1}(c)\right)\right\}. \tag{6}$$

3. Empirical searching methods

In this section, we propose two distribution-free searching methods: a min–max type of method in Section 3.1 followed by a stepwise approach in Section 3.2.

Using empirical estimates for the CDFs of diseased and healthy populations in (3) and (5), the empirical estimate of Youden index (J) for the k^{th} marker is

$$\begin{aligned} \hat{J}_k &= \hat{F}_{Y_{2k}}(\hat{c}_k) - \hat{F}_{Y_{1k}}(\hat{c}_k) \\ &= \frac{\sum_{j=1}^{n_2} I\left(y_{2jk} \leq \hat{c}_k\right)}{n_2} - \frac{\sum_{i=1}^{n_1} I(y_{1ik} \leq \hat{c}_k)}{n_1} \end{aligned} \tag{7}$$

where $\hat{c}_k$ is the empirical cut-off value for the k^{th} biomarker defined in (4) for $k = 1, 2 \ldots, p$. Similarly, for the combined score Z_g $(g = 1, 2)$ in (1), the empirical estimate of Youden index $(J_{\boldsymbol{w}})$ is

$$\begin{aligned} \hat{J}_{\boldsymbol{w}} &= \hat{F}_{Z_2}(\hat{c}_{\boldsymbol{w}}) - \hat{F}_{Z_1}(\hat{c}_{\boldsymbol{w}}) \\ &= \frac{\sum_{j=1}^{n_2} I\left(\boldsymbol{w}^T \boldsymbol{y}_{2j} \leqslant \hat{c}_{\boldsymbol{w}}\right)}{n_2} - \frac{\sum_{i=1}^{n_1} I\left(\boldsymbol{w}^T \boldsymbol{y}_{1i} \leqslant \hat{c}_{\boldsymbol{w}}\right)}{n_1} \end{aligned} \tag{8}$$

where $\hat{c}_{\boldsymbol{w}}$ is the empirical cut-off value defined in (6).

The goal is to obtain the linear combination coefficient vector $\boldsymbol{w}$ that maximizes $\hat{J}_{\boldsymbol{w}}$, that is, Youden index of the combined marker. Because $\hat{J}_{\boldsymbol{w}}$ involves the indicator function $I(.)$, it is not a smooth function of $\boldsymbol{w}$. Therefore, instead of analytical approach, an extensive grid search in p-dimensional space is required for such optimization problem. A similar grid-search method has been proposed by Pepe and Thompson [11] for the purpose of searching for the optimal linear combination that maximizes *AUC*. Such searching methods are generally computationally formidable when $p > 3$. Hence, some alternative methods are needed. In the following, we discuss two searching methods with computational ease.

3.1. *The min–max approach*

Liu *et al.* [12] proposed a distribution-free min–max approach that linearly combines only the minimum and maximum values of the p biomarker measurements to maximize *AUC* based on the reasonings as follows. For all $1 \leqslant k \leqslant p$, at the cut-point c, the sensitivity satisfies

$$Pr(\min(\boldsymbol{Y}_{1i}) > c) \leqslant Pr(Y_{1ik} > c) \leqslant Pr(\max(\boldsymbol{Y}_{1i}) > c)$$

for all $i = 1, 2 \ldots, n_1$; and specificity satisfies

$$Pr(\max(\boldsymbol{Y}_{2j}) \leqslant c) \leqslant Pr(Y_{2jk} \leqslant c) \leqslant Pr(\min(\boldsymbol{Y}_{2j}) \leqslant c)$$

for all $j = 1, 2 \ldots, n_2$. These results indicate that a compromise between the maximum and the minimum (e.g., a linear combination of the maximum and the minimum) of biomarker values might be a plausible combination to achieve best sensitivity and specificity. Liu *et al.* [12] claimed that for the purpose of maximizing *AUC*, such combination is expected to perform well. Obviously, the previously stated reasonings directly apply to the proposed optimization criterion, that is, Youden index, which is entirely a simple function of sensitivity and specificity.

Following the same vein as Liu *et al.* [12], the min–max approach can be extended to use Youden index as the objective function instead of *AUC* as follows:

$$\hat{J}_w = \frac{\sum_{j=1}^{n_2} I\left(y_{2j,\max} + w y_{2j,\min} \leqslant \hat{c}_w\right)}{n_2} - \frac{\sum_{i=1}^{n_1} I\left(y_{1i,\max} + w y_{1i,\min} \leqslant \hat{c}_w\right)}{n_1} \tag{9}$$

where $y_{1i,\max} = \max(\boldsymbol{y}_{1i})$ and $y_{1i,\min} = \min(\boldsymbol{y}_{1i})$ for diseased group, $y_{2j,\max} = \max(\boldsymbol{y}_{2j})$ and $y_{2j,\min} = \min(\boldsymbol{y}_{2j})$ for non-diseased group, and $\hat{c}_w$ is the empirical optimal cut-off value for such combined score. Hence, the formidable p-dimensional grid search reduces to an empirical search of a single coefficient w. Setting w to be 201 equally spaced values in $[-1, 1]$, the value of w within $[-1, 1]$, which gives largest Youden index in (9), is obtained by an empirical search over the 201 values: For each given value of w, first the optimal cut-off value $\hat{c}_w$ is empirically searched by maximizing $\hat{J}_w$; then, the value of $\hat{J}_w$ is calculated at the selected optimal cut-point for each value of w; finally, the optimal value of w is selected if it gives the largest corresponding $\hat{J}_w$. Recognizing the fact that multiplying any constant does not affect the Youden index of the combined score because *ROC* curve is invariant to any monotone transformation, we can divide the linear combination by w, that is, $\frac{1}{w} \times \max + \min = r \times \max + \min$. By setting $r\ \left(= \frac{1}{w}\right)$ to be 201 equally spaced values in $[-1, 1]$, the values of w outside $[-1, 1]$ are explored as well. Therefore, an extensive empirical search of all possible values of coefficient w from $-\infty$ to ∞ is performed. Finally, after obtaining the optimal values of w within and outside $[-1, 1]$, respectively, we compare their corresponding $\hat{J}_w$ values to decide on which w value is the optimal combination coefficient across the real line $\mathbb{R}$.

3.2. The stepwise approach

The primary drawback of the min–max approach is that this method only uses the minimum and maximum of the marker values, so it may not be the best way to achieve the optimal linear combination. Kang *et al.* [13] proposed a stepwise combination approach using volume under surface (VUS) as an objective function for biomakers with three ordinal disease stages as another efficient non-parametric method, which combines all markers in a stepwise manner. They proposed two types of stepwise methods, that is, step-up and step-down methods, and recommended the use of step-down procedure. Following the same vein, we will extend the step-down approach to find the optimal linear combination using Youden index as the objective function in the following steps:

(1) Calculate the empirical estimate of Youden index for each of p biomarkers following (7);
(2) Order these p biomarkers based on the values of the empirical Youden index estimates from largest to smallest. Denote the corresponding values for diseased group as $\boldsymbol{y}_{1i} = \left(y_{1i,(p)}, y_{1i,(p-1)}, \ldots, y_{1i,(1)}\right)^T$ for $i = 1, 2, \ldots, n_1$ and for non-diseased group as $\boldsymbol{y}_{2j} = \left(y_{2j,(p)}, y_{2j,(p-1)}, \ldots, y_{2j,(1)}\right)^T$ for $j = 1, 2, \ldots, n_2$.
(3) Combine the first two biomarkers empirically using objective function

$$\hat{J}_{w_{p-1}} = \frac{\sum_{j=1}^{n_2} I\left(y_{2j,(p)} + w_{p-1} y_{2j,(p-1)} \leq \hat{c}_{w_{p-1}}\right)}{n_2} - \frac{\sum_{i=1}^{n_1} I\left(y_{1i,(p)} + w_{p-1} y_{1i,(p-1)} \leq \hat{c}_{w_{p-1}}\right)}{n_1} \tag{10}$$

which is evaluated on 201 equally spaced values of w_{p-1} in $[-1, 1]$ as follows. For each given value of w_{p-1}, first the optimal cut-off value $\hat{c}_{w_{p-1}}$ is empirically searched by maximizing $\hat{J}_{w_{p-1}}$; then, the value of $\hat{J}_{w_{p-1}}$ is calculated at the selected optimal cut-point for each w_{p-1}.
(4) Similar to step 3, combine the first two biomarkers empirically as

$$\hat{J}_{r_{p-1}} = \frac{\sum_{j=1}^{n_2} I\left(r_{p-1} y_{2j,(p)} + y_{2j,(p-1)} \leq \hat{c}_{r_{p-1}}\right)}{n_2} - \frac{\sum_{i=1}^{n_1} I\left(r_{p-1} y_{1i,(p)} + y_{1i,(p-1)} \leq \hat{c}_{r_{p-1}}\right)}{n_1}, \tag{11}$$

and $\hat{J}_{r_{p-1}}$ is evaluated on 201 equally spaced values of r_{p-1} in $[-1, 1]$.
(5) Decide on the coefficient $\left(w_{p-1} \text{ or } r_{p-1}\right)$, which gives the largest Youden index $\left(\hat{J}_{w_{p-1}} \text{ or } \hat{J}_{r_{p-1}}\right)$ value as the optimal combination coefficient for the first two markers out of the 201 values of $\hat{J}_{w_{p-1}}$ and the other 201 values of $\hat{J}_{r_{p-1}}$.
(6) Having derived the univariate combined score of the first two biomarkers in step 5, combine it with the third marker, that is, the $(p-2)^{th}$ ordered marker, using the same procedure in steps 3–5. Proceed in the same way until all biomarkers are included in the linear combination.

4. Derivation-based numerical searching methods

In this section, we propose two derivation-based methods for numerically searching for the optimal linear combination using Youden index as an objective function. We discuss a parametric method under multivariate normality in Section 4.1 followed by a non-parametric approach based on kernel smoothing technique in Section 4.2.

4.1. A parametric approach under multivariate normality

Assume multivariate normality for $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, that is, $\boldsymbol{Y}_g \sim MVN_p\left(\boldsymbol{\eta}_g, \boldsymbol{\Sigma}_g\right), g = 1, 2$ for diseased and healthy groups, respectively. Given the combination vector $\boldsymbol{w}$, the combined score $Z_g = \boldsymbol{w}^T \boldsymbol{Y}_g$ follows a univariate normal distribution, that is, $Z_g \sim N\left(\mu_g, \sigma_g^2\right)$, where

$$\mu_g = \boldsymbol{w}^T \boldsymbol{\eta}_g \tag{12}$$

and

$$\sigma_g^2 = \boldsymbol{w}^T \boldsymbol{\Sigma}_g \boldsymbol{w}. \tag{13}$$

Schisterman and Perkins [18] presented results for Youden index and optimal cut-off point for a single marker under normality. Based on their results, we can easily derive Youden index and the optimal cut-off point for the combined marker in the following.

Given combination vector $\boldsymbol{w}$, when the combined marker has different variances between diseased and non-diseased groups, that is, $\sigma_1^2 \neq \sigma_2^2$ (this scenario occurs when $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$), the optimal cut-point for the combined score is

$$c_w = \frac{\mu_2\left(b^2-1\right) - a + b\sqrt{a^2 + \left(b^2-1\right)\sigma_2^2 \ln\left(b^2\right)}}{b^2-1}, \tag{14}$$

and Youden index is

$$J_w = \Phi\left(\frac{\mu_1 - c_w}{\sigma_1}\right) + \Phi\left(\frac{c_w - \mu_2}{\sigma_2}\right) - 1 \tag{15}$$

where

$$a = \mu_1 - \mu_2 = \boldsymbol{w}^T\left(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\right),\ b = \frac{\sigma_1}{\sigma_2} = \frac{\sqrt{\boldsymbol{w}^T\boldsymbol{\Sigma}_1\boldsymbol{w}}}{\sqrt{\boldsymbol{w}^T\boldsymbol{\Sigma}_2\boldsymbol{w}}}$$

and $\Phi(.)$ denotes the standard normal cumulative distribution function. When $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (i.e., $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$),

$$c_w = \frac{\mu_1 + \mu_2}{2}$$

and

$$J_w = 2\Phi\left(\frac{\mu_1 - \mu_2}{2\sqrt{\sigma^2}}\right) - 1.$$

An estimate of Youden index $\hat{J}_w$ can be obtained by plugging estimates $\hat{\boldsymbol{\eta}}_g$'s and $\hat{\boldsymbol{\Sigma}}_g$'s into (14) and (15). Obviously, $\hat{J}_w$ is an explicit continuous function of combination vector $\boldsymbol{w}$ under multivariate normality and is differentiable with respect to $\boldsymbol{w}$. Although writing out the analytical solution is tedious, $\hat{J}_w$ can be numerically optimized with respect to vector $\boldsymbol{w}$ via some quasi-Newton algorithms, for example, R package optim() or nlm(). Thus, the optimal combination vector $\boldsymbol{w}$ can be obtained.

4.2. A non-parametric approach based on kernel smoothing (KS)

For the scenarios without multivariate normality assumption, similar derivation-based method for $\boldsymbol{w}$ can also be developed.

The non-parametric estimator of Youden index for the combined score, that is, $\hat{J}_{\boldsymbol{w}}$ in (8), contains some indicator functions $I(.)$, and hence it is not differentiable directly. In order to obtain a smoothed and differentiable version of $\hat{J}_{\boldsymbol{w}}$ with respect to $\boldsymbol{w}$, Gaussian kernel is applied to smooth the empirical distribution functions of diseased and healthy groups as follows:

$$\hat{F}_{Z_g}^{KS}(t) = \frac{1}{n_g}\sum_{i=1}^{n_g}\Phi\left(\frac{t - z_{gi}}{h_{Z_g}}\right),\ g = 1, 2 \tag{16}$$

where the bandwidths

$$h_{Z_g} = 0.9\min\left\{SD(z_g), IQR(z_g)/1.34\right\} n_g^{-0.2},$$

and $SD(z_g)$ and $IQR(z_g)$ are the standard deviation and the inter quartile range of the combined score for disease and healthy groups, respectively. Such bandwidth is recommended by Silverman [19] and utilized by many researchers in diagnostic studies [20–22]. By kernel smoothing, we obtain an estimate of Youden index as

$$\hat{J}_{\boldsymbol{w}}^{KS} = \hat{F}_{Z_2}^{KS}\left(\hat{c}_{\boldsymbol{w}}^{KS}\right) - \hat{F}_{Z_1}^{KS}\left(\hat{c}_{\boldsymbol{w}}^{KS}\right), \tag{17}$$

where $\hat{c}_{\boldsymbol{w}}^{KS}$ denote the optimal cut-point via Youden index for the combined score obtained using kernel smoothing technique. Note that unlike the parametric method in Section 4.1, where $\hat{c}_{\boldsymbol{w}}$ is an explicit function of coefficient vector $\boldsymbol{w}$, $\hat{c}_{\boldsymbol{w}}^{KS}$ is evaluated simultaneously with $\boldsymbol{w}$ through some derivation-based algorithms. Therefore, the unknown parameter vector to be optimized is $\boldsymbol{w}^* = \left(c_{\boldsymbol{w}}^{KS}, \boldsymbol{w}^T\right)^T$.

After kernel smoothing, $\hat{J}_{\boldsymbol{w}}^{KS}$ is an explicit continuous function of $\boldsymbol{w}^*$; thus, it is differentiable with respect to $\boldsymbol{w}$ and $c_{\boldsymbol{w}}^{KS}$. Therefore, similar to the parametric method in Section 4.1, we could numerically optimize the $\hat{J}_{\boldsymbol{w}}^{KS}$ function with respect to $\boldsymbol{w}^*$ by some quasi-Newton algorithm, for example, R package optim() or nlm().

5. Simulation studies

This section contains two parts: (i) comparing the performance of the proposed combination methods based on Youden index, presented in Section 5.1, and (ii) investigating the gain in maximum total correct classification rate for the combined score using Youden index as objective function compared with *AUC*, presented in Section 5.2. Throughout this section, first, the optimal linear combinations are obtained for each method, and then the empirical estimates of Youden index in (8) of the corresponding optimal linear combination are calculated and compared.

5.1. Comparison of the performance of the proposed combination methods

We carried out simulation studies to assess the performance of the proposed combination methods, that is, derivation-based multivariate normal approach (*MVN*), derivation-based kernel-smoothing non-parametric approach (KS), min–max non-parametric approach (*MMX*), and stepwise non-parametric approach with downward direction (*SWD*).

We considered a wide range of joint distributions of four diagnostic biomarkers: (i) multivariate normal distributions with equal variance; (ii) multivariate normal distributions with unequal variance; (iii) multivariate exponential distribution generated using a method proposed by Proschan and Sullo [23]; and (iv) a joint distribution with different marginal distributions for each biomarker generated using normal copula with exchangeable correlation.

For each joint distribution, we chose two sets of mean parameters. For each simulation setting, we generated $10,000$ Monte Carlo random samples from the underlying joint distribution at sample sizes $(n_1, n_2) = (10, 20), (30, 30), (50, 30), (50, 50)$ for diseased group and non-diseased group. In Tables I–IV, we present the mean value of Youden index as well as the probability of obtaining the largest Youden index out of $10,000$ simulations for each combination method.

5.1.1. Multivariate normal distributions: equal variances for diseased group and healthy group. Assume diseased and healthy samples are from multivariate normal distributions with mean vector $\boldsymbol{\eta}_2 = (0, 0, 0, 0)^T$ for healthy group. The mean vector for diseased group is set as follows: (i) $\boldsymbol{\eta}_1 = (0.2, 0.5, 1.0, 0.7)^T$ and (ii) $\boldsymbol{\eta}_1 = (0.4, 1.0, 1.5, 1.2)^T$. The variance matrices of diseased and healthy populations are set to be equal: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = (1-\gamma)\boldsymbol{I}_{4\times4} + \gamma \boldsymbol{J}_{4\times4}$ where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{J}$ is a matrix of all 1s. Correlation parameter γ is set to be 0.3, 0.5, and 0.7 to present scenarios with low, medium, and high correlation between markers, respectively. Table I presents the simulation results.

The results in Table I indicate that the derivation-based kernel-smoothing non-parametric approach dominates all other three methods for all cases and yields the largest Youden index for a majority of time. The stepwise-down non-parametric approach and the parametric method under multivariate normality perform similarly.

5.1.2. Multivariate normal distributions: unequal variances for diseased group and non-diseased group. The assumption for the mean vectors for diseased and healthy groups is the same as in Section 5.1.1, that is, $\boldsymbol{\eta}_2 = (0, 0, 0, 0)^T$ for healthy group, and (i) $\boldsymbol{\eta}_1 = (0.2, 0.5, 1.0, 0.7)^T$ and (ii) $\boldsymbol{\eta}_1 = (0.4, 1.0, 1.5, 1.2)^T$ for diseased group. However, we assume unequal variance matrices for diseased and healthy groups, that is, $\boldsymbol{\Sigma}_1 = 0.3\boldsymbol{I}_{4\times4} + 0.7\boldsymbol{J}_{4\times4}$ and $\boldsymbol{\Sigma}_2 = 0.7\boldsymbol{I}_{4\times4} + 0.3\boldsymbol{J}_{4\times4}$. Tables II presents the simulation results.

Table II shows that when diseased and healthy samples are normally distributed with unequal variances, the derivation-based kernel-smoothing non-parametric approach and stepwise-down non-parametric approach perform comparably well among the four proposed combinations with the kernel-smoothing one being slightly better.

Table I. Mean Youden index and probability of obtaining largest Youden index for different combination methods under multivariate normal distributions with equal variance.

Normal means $(\eta_{11}, \eta_{12}, \eta_{13}, \eta_{14})$	Sample sizes (n_1, n_2)	Mean Youden index				Probability of largest Youden index			
		MVN	*KS*	*MMX*	*SWD*	*MVN*	*KS*	*MMX*	*SWD*
$\Sigma_1 = \Sigma_2 = 0.7I_{4\times4} + 0.3J_{4\times4}$ (low correlation)									
(0.2, 0.5, 1.0, 0.7)	(10, 20)	0.6350	0.6815	0.5718	0.6885	0.1108	0.3850	0.0950	0.4092
	(30, 30)	0.5505	0.5915	0.4848	0.5932	0.0858	0.4425	0.0542	0.4175
	(50, 30)	0.5284	0.5735	0.4650	0.5662	0.0583	0.5458	0.0600	0.3358
	(50, 50)	0.5013	0.5348	0.4374	0.5330	0.0642	0.4883	0.0442	0.4033
(0.4, 1.0, 1.5, 1.2)	(10, 20)	0.6382	0.6920	0.5460	0.6655	0.1450	0.4475	0.0800	0.3275
	(30, 30)	0.5402	0.5862	0.4630	0.5663	0.0858	0.5342	0.0383	0.3417
	(50, 30)	0.5381	0.5738	0.4426	0.5608	0.0875	0.5025	0.0200	0.3900
	(50, 50)	0.5093	0.5407	0.4209	0.5293	0.1025	0.4575	0.0150	0.4250
$\Sigma_1 = \Sigma_2 = 0.5I_{4\times4} + 0.5J_{4\times4}$ (medium correlation)									
(0.2, 0.5, 1.0, 0.7)	(10, 20)	0.6742	0.7188	0.5365	0.6518	0.2129	0.5471	0.0488	0.1912
	(30, 30)	0.5785	0.6195	0.4502	0.5630	0.2175	0.6475	0.0242	0.1108
	(50, 30)	0.5748	0.6058	0.4379	0.5591	0.2117	0.6217	0.0000	0.1667
	(50, 50)	0.5461	0.5816	0.4206	0.5301	0.1650	0.7150	0.0050	0.1150
(0.4, 1.0, 1.5, 1.2)	(10, 20)	0.7835	0.8158	0.7200	0.8075	0.1567	0.4167	0.1017	0.3250
	(30, 30)	0.7060	0.7393	0.6415	0.7270	0.0933	0.5558	0.0600	0.2908
	(50, 30)	0.7064	0.7401	0.6364	0.7287	0.0683	0.6458	0.0125	0.2733
	(50, 50)	0.6878	0.7139	0.6208	0.7070	0.0767	0.5875	0.0233	0.3125
$\Sigma_1 = \Sigma_2 = 0.3I_{4\times4} + 0.7J_{4\times4}$ (large correlation)									
(0.2, 0.5, 1.0, 0.7)	(10, 20)	0.7888	0.8068	0.6970	0.7958	0.2150	0.3600	0.0717	0.3533
	(30, 30)	0.7087	0.7287	0.6170	0.7202	0.1588	0.4479	0.0362	0.3571
	(50, 30)	0.6947	0.7186	0.6071	0.7155	0.0967	0.4517	0.0200	0.4317
	(50, 50)	0.6752	0.6923	0.5864	0.6882	0.1446	0.4396	0.0179	0.3979
(0.4, 1.0, 1.5, 1.2)	(10, 20)	0.8110	0.8338	0.6908	0.7790	0.2221	0.4871	0.0729	0.2179
	(30, 30)	0.7400	0.7632	0.6213	0.7117	0.2217	0.5833	0.0167	0.1783
	(50, 30)	0.7321	0.7506	0.6074	0.7227	0.1950	0.5625	0.0050	0.2375
	(50, 50)	0.7120	0.7293	0.5742	0.6826	0.1904	0.6504	0.0188	0.1404

MVN, derivation-based multivariate normal approach; *KS*, derivation-based kernel-smoothing non-parametric method; *MMX*, min–max non-parametric approach; *SWD*, stepwise non-parametric approach with downward direction.

Table II. Simulation results for mean Youden index and probability of obtaining largest Youden index for different combination methods under multivariate normal distributions ($\Sigma_1 = 0.3I_{4\times4} + 0.7J_{4\times4}$ and $\Sigma_2 = 0.7I_{4\times4} + 0.3J_{4\times4}$).

Normal means $(\eta_{11}, \eta_{12}, \eta_{13}, \eta_{14})$	Sample sizes (n_1, n_2)	Mean Youden index				Probability of largest Youden index			
		MVN	*KS*	*MMX*	*SWD*	*MVN*	*KS*	*MMX*	*SWD*
(0.2, 0.5, 1.0, 0.7)	(10, 20)	0.6561	0.7043	0.5920	0.6726	0.1728	0.4285	0.1472	0.2515
	(30, 30)	0.5533	0.5945	0.5095	0.5735	0.1762	0.4075	0.1115	0.3048
	(50, 30)	0.5333	0.5685	0.4879	0.5575	0.1620	0.3893	0.0843	0.3643
	(50, 50)	0.5140	0.5469	0.4721	0.5325	0.1422	0.4778	0.0915	0.2885
(0.4, 1.0, 1.5, 1.2)	(10, 20)	0.7906	0.8184	0.6899	0.8088	0.1948	0.3792	0.0755	0.3505
	(30, 30)	0.7041	0.7289	0.6024	0.7170	0.1557	0.4517	0.0427	0.3500
	(50, 30)	0.6920	0.7150	0.5927	0.7107	0.1037	0.4467	0.0330	0.4167
	(50, 50)	0.6752	0.6972	0.5717	0.6897	0.1085	0.4555	0.0215	0.4145

MVN, derivation-based multivariate normal approach; *KS*, derivation-based kernel-smoothing non-parametric method; *MMX*, min–max non-parametric approach; *SWD*, stepwise non-parametric approach with downward direction.

5.1.3. Multivariate exponential distributions. We generated the multivariate exponential random samples using a method proposed by Proschan and Sullo [23]. Assume that healthy sample is from multivariate exponential distribution with mean vector $\eta_2 = (0.2, 0.2, 0.2, 0.2)^T$ and diseased sample from multivariate exponential distributions with mean vector: (i) $\eta_1 = (0.30, 0.35, 0.40, 0.30)^T$ and (ii) $\eta_1 =$

$(0.35, 0.45, 0.45, 0.35)^T$. The correlations between any two measurements are set as 0.25 for healthy group, and for diseased group, the correlations vary between pairs (i.e., $(\rho_{12}, \rho_{13}, \rho_{14}, \rho_{23}, \rho_{24}, \rho_{34}) = (0.1927, 0.2070, 0.1767, 0.2294, 0.1927, 0.2070)$ for mean setting 1) and $(\rho_{12}, \rho_{13}, \rho_{14}, \rho_{23}, \rho_{24}, \rho_{34}) = (0.2153, 0.2153, 0.1867, 0.2542, 0.2153, 0.2153)$ for mean setting 2)). Table III presents the simulation results.

For extremely skewed distributions such as the multivariate exponential distribution, stepwise-down non-parametric approach dominates all other three methods for all the scenarios considered and generally yields the largest Youden index. One possible interpretation for the stepwise-down approach to be superior than the derivation-based kernel-smoothing method is that the kernel function used in this paper is normal kernel, and hence the method based on kernel smoothing may not perform well if the underlying distribution is far from normal.

5.1.4. Multivariate chi-squared/normal/gamma/exponential distributions via normal copula. To investigate the scenarios when markers follow different marginal distributions, we generated multivariate chi-squared/normal/gamma/exponential distributions with normal copula of exchangeable correlation ($\rho = 0.3$ for healthy group and $\rho = 0.7$ for diseased group). Assume that biomarkers from healthy sample are marginally distributed as $\chi^2_{0.1}$, $N(0.1, 1)$, $\Gamma(0.1, 1)$ and $Exp(0.1)$. We consider two marginal settings for diseased samples: (i) $\chi^2_{0.1}$, $N(0.3, 1)$, $\Gamma(0.4, 1)$ and $Exp(0.1)$ and (ii) $\chi^2_{0.1}$, $N(0.6, 1)$, $\Gamma(0.8, 1)$ and $Exp(0.1)$. Table IV lists the simulation results.

For multivariate distribution with different marginal distributions of each biomarker, stepwise-down non-parametric approach dominates all other three methods for almost all cases and generally yields

Table III. Mean Youden index and probability of obtaining largest Youden index for different combination methods under multivariate exponential distributions.

Exponential means	Sample sizes	Mean Youden index				Probability of largest Youden index			
$(\eta_{11}, \eta_{12}, \eta_{13}, \eta_{14})$	(n_1, n_2)	*MVN*	*KS*	*MMX*	*SWD*	*MVN*	*KS*	*MMX*	*SWD*
(0.30, 0.35, 0.4, 0.30)	(10, 20)	0.6923	0.7578	0.6290	0.7682	0.1048	0.3422	0.0805	0.4724
	(30, 30)	0.6099	0.6712	0.5505	0.6911	0.0351	0.2816	0.0519	0.6314
	(50, 30)	0.6025	0.6535	0.5362	0.6799	0.0201	0.1730	0.0316	0.7753
	(50, 50)	0.5900	0.6437	0.5164	0.6661	0.0181	0.1974	0.0172	0.7673
(0.35, 0.45, 0.45, 0.35)	(10, 20)	0.8138	0.8490	0.6992	0.8639	0.1678	0.3060	0.0368	0.4894
	(30, 30)	0.7399	0.7829	0.6243	0.8011	0.0577	0.2752	0.0110	0.6561
	(50, 30)	0.7324	0.7682	0.6129	0.7896	0.0323	0.1865	0.0079	0.7732
	(50, 50)	0.7232	0.7617	0.5952	0.7802	0.0267	0.2033	0.0018	0.7683

MVN, derivation-based multivariate normal approach; *KS*, derivation-based kernel-smoothing non-parametric method; *MMX*, min–max non-parametric approach; *SWD*, stepwise non-parametric approach with downward direction.

Table IV. Mean Youden index and probability of obtaining largest Youden index for different combination methods under multivariate chi-squared/normal/gamma/exponential distributions with normal copula.

Parameter	Sample sizes	Mean Youden index				Probability of largest Youden index			
	(n_1, n_2)	*MVN*	*KS*	*MMX*	*SWD*	*MVN*	*KS*	*MMX*	*SWD*
$N(0.3, 1)$	(10, 20)	0.5590	0.6873	0.4311	0.7197	0.0641	0.3478	0.0191	0.5690
$\Gamma(0.4, 1)$	(30, 30)	0.4704	0.6156	0.3203	0.6553	0.0229	0.2797	0.0018	0.6955
	(50, 30)	0.4651	0.6095	0.2985	0.6519	0.0125	0.2056	0.0000	0.7819
	(50, 50)	0.4533	0.5979	0.2704	0.6295	0.0152	0.2553	0.0000	0.7295
$N(0.6, 1)$	(10, 20)	0.7710	0.8515	0.5019	0.8450	0.1366	0.4435	0.0041	0.4159
$\Gamma(0.8, 1)$	(30, 30)	0.7295	0.8071	0.4046	0.8069	0.0760	0.4559	0.0000	0.4681
	(50, 30)	0.7290	0.8013	0.3890	0.8081	0.0519	0.3920	0.0000	0.5561
	(50, 50)	0.7196	0.7911	0.3653	0.7928	0.0547	0.4390	0.0000	0.5062

MVN: derivation-based multivariate normal approach;
KS: derivation-based kernel-smoothing non-parametric method;
MMX: min–max non-parametric approach;
SWD: stepwise non-parametric approach with downward direction.

the largest Youden index. The performance of the derivation-based kernel-smoothing method is also reasonable.

In summary, among all proposed methods, under multivariate normality, the derivation-based kernel-smoothing method is the most preferred linear combination method; otherwise, stepwise-down approach is the optimal method.

5.2. *Comparison of combination methods based on Youden index and area under the receiving operating characteristic curve*

This paper proposes the idea of maximizing the Youden index for the purpose of obtaining a combined marker with the best possible diagnostic accuracy. Because *AUC* has been most widely used as the objective function for combining markers, exploring the gain in the total correct classification rate of the combined marker based on Youden index will provide us with some evidence about the importance of using Youden index instead of *AUC*. Hence, it is of interest to compare the proposed combination methods based on Youden index to combination methods based on *AUC*, in terms of diagnostic accuracy, that is, sensitivity, specificity, and total correct classification rate, at the optimal threshold.

It is well known that the optimal combination on *AUC* under multivariate normality has closed form by Su and Liu [10]. Therefore, it is straightforward to derive the cut-off point, Youden index, as well as sensitivity and specificity at the optimal cut-off point for the optimally combined score based on *AUC*. The appendix presents the details.

From simulation results in Sections 5.1.1 and 5.1.2, we observe that the derivation-based kernel-smoothing method is the optimal combination approach under multivariate normality. Therefore, this method is used for linear combination based on Youden index in this part of simulation. Denote the coefficient vector obtained by maximizing Youden index in (17) as $\boldsymbol{w}_J$ and that by maximizing *AUC* in (18) as $\boldsymbol{w}_A$. For each simulation, we empirically calculate the sensitivity, specificity, and total correct classification rate (i.e., sum of sensitivity and specificity) at the optimal threshold value for the combined scores Z_g $(g = 1, 2)$ based on $\boldsymbol{w}_J$ and $\boldsymbol{w}_A$, respectively.

We simulate multivariate normal distributions for different means and different variances. The mean vector for healthy group is set as $\boldsymbol{\eta}_2 = (0, 0, 0, 0)^T$. For diseased group, there are two settings of means: (i) $\boldsymbol{\eta}_1 = (0.2, 0.2, 0.5, 0.5)^T$ and (ii) $\boldsymbol{\eta}_1 = (0.1, 0.3, 0.4, 0.6)^T$. The variance matrices of diseased and healthy populations are set as $\boldsymbol{\Sigma}_1 = \gamma \boldsymbol{I}_{4\times4} + (1-\gamma)\boldsymbol{J}_{4\times4}$ and $\boldsymbol{\Sigma}_2 = (1-\gamma)\boldsymbol{I}_{4\times4} + \gamma \boldsymbol{J}_{4\times4}$ with $\gamma = 0.1, 0.5, 0.9$.

Table V presents means of total correct classification rate, sensitivity, and specificity for the combined score based on Youden index (denoted as C_J) and *AUC* (denoted as C_A), respectively, across 10,000 simulations.

Table V shows that the combination score based on Youden index has greater total classification rate at the optimal threshold than that based on *AUC* for all settings. When $\gamma = 0.1$, Youden index combination method performs better than that of *AUC* in terms of total correct classification rate, and it especially outperforms *AUC* for sensitivity. For example, as sample sizes are $(50, 50)$ and $\boldsymbol{\eta}_1$ takes setting $(0.2, 0.2, 0.5, 0.5)^T$, sensitivity is 0.94 and 0.79 for the combination on Youden index and *AUC*, respectively. As $\gamma = 0.9$, the performance of Youden index combination method is better than that of *AUC* in terms of total correct classification rate and specificity. When $\gamma = 0.5$, Youden index combination is slightly better than *AUC* combination in all aspects including sensitivity, specificity, and total correct classification rate. Therefore, the linear combination method using Youden index as an objective function yields a combined marker with better diagnostic accuracy compared with that using *AUC*.

6. Data example

Duchene muscular dystrophy (DMD) is a progressive recessive disorder passed from a mother to her children. Percy *et al.* [24] presented data of four different DMD biomarkers, namely serum creatine kinase (CK), hemopexin (HPX), pyruvate kinase (PK) and lactate dehydrogenase (LD). Complete data are available for 66 samples from female carriers and 127 samples from healthy female controls. As the variances of the four marker measurements are very different, standardization of the marker measurements to avoid different measurement units is performed before applying min–max combination approach.

The empirical estimates of Youden index for four biomarkers are 0.6086, 0.4129, 0.5025, and 0.5747, respectively. To ensure the optimal linear combination to have a unique solution, the coefficient of marker

Table V. Comparison of optimal linear combination methods based on Youden index and AUC combination criteria.

		$\eta_1 = (0.2, 0.2, 0.5, 0.5)^T$						$\eta_1 = (0.1, 0.3, 0.4, 0.6)^T$					
		Total rate		Specificity		Sensitivity		Total rate		Specificity		Sensitivity	
Parameter	Sample size	C_J	C_A	C_J	C_A	C_J	C_A	C_J	C_A	C_J	C_A	C_J	C_A
$\gamma = 0.1$	(10, 10)	1.6905	1.6006	0.7588	0.7139	0.9317	0.8867	1.7053	1.6192	0.7689	0.7223	0.9364	0.8969
	(10, 20)	1.6634	1.5701	0.7616	0.7275	0.9018	0.8426	1.6801	1.5920	0.7685	0.7434	0.9116	0.8486
	(30, 30)	1.5312	1.4332	0.5955	0.6187	0.9357	0.8144	1.5427	1.4507	0.6097	0.6220	0.9330	0.8287
	(50, 30)	1.5006	1.4009	0.5688	0.6130	0.9318	0.7879	1.5183	1.4294	0.5818	0.6261	0.9366	0.8033
	(50, 50)	1.4807	1.3822	0.5430	0.5944	0.9377	0.7879	1.5014	1.4047	0.5560	0.5933	0.9455	0.8114
$\gamma = 0.5$	(10, 10)	1.6516	1.5788	0.8182	0.7471	0.8334	0.8317	1.6605	1.5941	0.8161	0.7579	0.8444	0.8362
	(10, 20)	1.5876	1.5220	0.8294	0.7793	0.7582	0.7427	1.5954	1.5291	0.8321	0.7746	0.7633	0.7545
	(30, 30)	1.4601	1.4043	0.7187	0.678	0.7414	0.7263	1.4713	1.4162	0.7229	0.6765	0.7484	0.7397
	(50, 30)	1.4349	1.3819	0.6989	0.6777	0.7359	0.7042	1.4453	1.3940	0.7052	0.6739	0.7401	0.7201
	(50, 50)	1.3936	1.3503	0.6814	0.6522	0.7122	0.6981	1.4150	1.3730	0.6992	0.6655	0.7158	0.7075
$\gamma = 0.9$	(10, 10)	1.6894	1.6087	0.9102	0.8193	0.7792	0.7894	1.6985	1.6127	0.9162	0.8336	0.7823	0.7791
	(10, 20)	1.6185	1.5337	0.9278	0.8296	0.6907	0.7042	1.6245	1.5449	0.9342	0.8493	0.6904	0.6956
	(30, 30)	1.5269	1.4370	0.8999	0.7699	0.6270	0.6671	1.5411	1.4525	0.9051	0.7835	0.6360	0.6690
	(50, 30)	1.5127	1.4187	0.9103	0.7775	0.6024	0.6412	1.5342	1.4449	0.9203	0.8060	0.6140	0.6389
	(50, 50)	1.4811	1.3832	0.9081	0.7507	0.5730	0.6325	1.4976	1.4055	0.9228	0.7800	0.5749	0.6256

C_J, combination method based on Youden index (J); C_A, combination method based on AUC.
Total rate: the maximum total correct classification rate an linear combination can achieve, obtained by the sum of sensitivity and specificity at the optimal diagnostic threshold associated with Youden index.

Statistics in Medicine

Table VI. Summary of optimal linear combinations based on Youden index (*MVN*, *KS*, *MMX*, and *SWD* approaches) and the optimal linear combination based *AUC* (SULIU).

Methods	Optimal linear combination	Youden index	Sensitivity	Specificity
MVN	$CK + 1.2937 \times HPX + 1.8725 \times PK + 0.2970 \times LD$	0.7855	0.8485	0.9370
KS	$CK + 2.6980 \times HPX + 3.7766 \times PK + 0.4761 \times LD$	0.8164	0.8636	0.9528
MMX	$\max\{CK^*, HPX^*, PK^*, LD^*\} + 2.2222 \times \min\{CK^*, HPX^*, PK^*, LD^*\}$	0.7304	0.8485	0.8819
SWD	$CK + 0.3600 \times HPX + 2.7778 \times PK + 2.2778 \times LD$	0.8164	0.8636	0.9528
SULIU	$CK + 10.0834 \times HPX + 9.9794 \times PK + 44.3314 \times LD$	0.6849	0.8030	0.8819

MVN, derivation-based multivariate normal approach of Youden index criteria;
KS, derivation-based kernel-smoothing non-parametric method of Youden index criteria;
MMX, min–max non-parametric approach of Youden index criteria;
SWD, stepwise non-parametric approach with downward direction of Youden index criteria;
SULIU, Su and Liu's multivariate normal approach of *AUC* criteria.
*: CK^*, HPX^*, PK^* and LD^* values are standardized to be unit less for *MMX*.

CK is fixed to be 1. Table VI lists the optimal linear combination as well as the values of Youden index, sensitivity, and specificity at the optimal diagnostic threshold for each combination method. The empirical combination method based on kernel smoothing and stepwise-down approach dominate all other methods and give exactly same values of Youden index, sensitivity, and specificity at the optimal threshold. Furthermore, the four combination methods using Youden index as the objective function all outperform Su and Liu's [10] method based on *AUC* in every aspect, that is, Youden index, sensitivity, and specificity, at the optimal diagnostic threshold. To conclude, both the kernel smoothing method and the stepwise-down approach provide optimal linear combinations of the largest sensitivity ($= 0.8636$), the largest specificity ($= 0.9528$) at the optimal threshold, as well as the largest Youden index ($= 0.8164$), which indicates the largest total correct classification rate at the optimal threshold equals 1.8164. The two equally optimal linear combinations of the four DMD markers are $CK + 2.6980 \times HPX + 3.7766 \times PK + 0.4761 \times LD$ and $CK + 0.3600 \times HPX + 2.7778 \times PK + 2.2778 \times LD$ for kernel smoothing method and stepwise-down approach, respectively.

7. Summary and discussion

In this paper, we propose a new optimization criteria, that is, Youden index, for obtaining optimal linear combination of multiple biomarkers. We present several methods including two empirical searching methods: a min–max non-parametric method and a stepwise-down non-parametric approach, and two derivation-based methods: a parametric approach assuming multivariate normality and a non-parametric method implemented by kernel smoothing technique. Simulation results show that the stepwise-down non-parametric approach is superior than other three methods for samples that are distributed much differently from multivariate normal distributions and the derivation-based kernel-smoothing non-parametric method performs the best for multivariate normal data. Therefore, for multivariate normal distributions, we can utilize the derivation-based kernel-smoothing non-parametric approach. And when multivariate normality assumption is not met, we recommend stepwise-down non-parametric approach. The reason for the parametric approach assuming multivariate normality being inferior to the non-parametric kernel smoothing approach for normal data might be due to the re-substitution error, and for future work, the proposed approaches can be investigated further by cross-validation [25]. Another possible reason is that different methods are used for the two numerical searching methods to determine the optimal cut-off point and for which, the kernel smoothing approach is more flexible: For parametric approach, the cut-off point is a function of combination coefficients (Equation (14)), while for the kernel smoothing method, the optimal cut-off point is numerically searched along with the combination coefficients. Furthermore, the reason of proposed kernel smoothing approach to be inferior than the stepwise-down approach for non-normal data is that only the Gaussian kernel function (which is the most popular choice for kernel smoothing) is applied in this paper. For future work, different types of kernel functions could be explored to see if the performance can be improved. Moreover, the derivation-based kernel-smoothing method has the advantage of being much more efficient to implement than the stepwise-down approach when the number of biomarkers to be combined is not small (e.g., $p > 5$).

In the min–max approach proposed by Liu *et al.* [12], the minimum and the maximum values of the p biomarkers are linearly combined to a single score, which offers the maximum combination criterion, for

example, *AUC*, or in our paper, the Youden index. In our simulation, the min–max approach generally has lower chance to achieve optimality as compared with other methods that utilize information of all biomarkers. This indicates that only combining the minimum and the maximum biomarkers might not be enough. Such min–max approach can be easily extended to a min–median–max method, by adding a third summary statistic, the median, into the linear combination. However, the min–median–max method is more computationally challenging, as it requires to search a point in the plane, that is, (w_1, w_2), instead of a single coefficient as for the min–max procedure.

Furthermore, Tang *et al.* [26] proposed to maximize a discriminatory measure based on AUC to accommodate situations when a random true diseased subject does not necessarily have higher marker value than a random healthy subject. The discriminatory measure is defined as $i = \max(AUC, 1-AUC)$. We may propose some similar measures based on Youden index (J) such as $\max(J, -J)$ or $|J|$. Furthermore, their paper proposed to combine markers using an objective function rewarding large discriminatory score and penalizing the number of markers. So far, our paper only focuses on methods to combine all biomarkers; however, when the number of candidate markers is extremely large, an objective function based on Youden index, which also penalizes the number of markers, will be useful.

Moreover, in this paper, we use Youden index as the optimization criteria to search for the optimal cut-off point for diagnosis as well as the combination coefficients. There exist a variety of selection methods besides Youden index for the optimal cut-off point, for example, by approaching the northwest corner, that is, the point $(0, 1)$ on *ROC* curve [27], by diagnostic odds ratio [28], and by PCdx statistic [29]. A recent paper by Zou *et al.* [30] provides more details of various methods to determine the optimal threshold in ROC analysis. Our proposed combinations are only 'optimal' in terms of total correct classification rate, and under different optimizing criteria, the linear combination coefficients should be obtained by maximizing or minimizing the corresponding metrics.

To conclude, the empirical comparison in terms of the optimal overall classification rates between the proposed combination based on Youden index and the traditional one based on *AUC* demonstrated a significant gain in diagnostic accuracy for the proposed approach. Therefore, for the purpose of making diagnosis, combing biomarkers based on Youden index is a more appropriate way.

An R program is available upon request from Dr. Lili Tian (ltian@buffalo.edu).

Appendix: Youden index for the combined score using *AUC* as the objective function under multivariate normality

Under multivariate normal assumption, that is, $\boldsymbol{Y}_g \sim MVN_p\left(\boldsymbol{\eta}_g, \boldsymbol{\Sigma}_g\right), g = 1, 2$ for diseased and healthy groups, respectively, Su and Liu [10] showed that the optimal linear combination coefficients, that is, $\boldsymbol{w}_A$, based on *AUC* criteria is

$$\boldsymbol{w}_A \propto \left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right)^{-1}\left(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\right). \tag{18}$$

Similar to Section 4.1, the combined scores of diseased and healthy groups, $Z_g = \boldsymbol{w}_A^T \boldsymbol{Y}_g$ $(g = 1, 2)$, are distributed as

$$Z_g \sim N\left(\mu_g, \sigma_g^2\right)$$

where

$$\begin{aligned} \mu_g &= \left(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\right)^T \left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right)^{-1} \boldsymbol{\eta}_g, \\ \sigma_g^2 &= \left(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\right)^T \left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right)^{-1} \boldsymbol{\Sigma}_g \left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right)^{-1} \left(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\right). \end{aligned} \tag{19}$$

For such combined score, the optimal cut-point c and corresponding Youden index can be obtained using the results of Schisterman and Perkins [18] presented in Section 4.1. When $\sigma_1^2 \neq \sigma_2^2$, that is, $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$,

$$\begin{aligned} a &= \mu_1 - \mu_2 \\ &= \left(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\right)^T \left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right)^{-1} \left(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\right), \end{aligned} \tag{20}$$

and

$$b = \sqrt{\sigma_1^2/\sigma_2^2} = \sqrt{\frac{(\eta_1-\eta_2)^T(\Sigma_1+\Sigma_2)^{-1}\Sigma_1(\Sigma_1+\Sigma_2)^{-1}(\eta_1-\eta_2)}{(\eta_1-\eta_2)^T(\Sigma_1+\Sigma_2)^{-1}\Sigma_2(\Sigma_1+\Sigma_2)^{-1}(\eta_1-\eta_2)}}. \quad (21)$$

Henceforth, by substituting a and b in (14) and (15), we obtain the Youden index J_{W_A} using combination coefficient w_A in terms of η_g's and Σ_g's for $g = 1, 2$. When variances for diseased and healthy groups are the same (i.e., $\Sigma_1 = \Sigma_2 = \Sigma$), Youden index is reduced to

$$J_{W_A} = 2\Phi\left[\frac{1}{2}\sqrt{(\eta_1-\eta_2)^T\Sigma^{-1}(\eta_1-\eta_2)}\right] - 1.$$

Finally, the maximum total correct classification rate that the combined score can achieve is obtained at the optimal threshold associated with Youden index, which can be readily calculated as $J_{W_A} + 1$.

References

1. Shapiro DE. The interpretation of diagnostic tests. *Statistical Methods in Medical Research* 1999; **8**(2):113–134.
2. Zhou X, McClish DK, Obuchowski NA. *Statistical Methods in Diagnostic Medicine*, Vol. 569. Wiley-Interscience: New York City, NY, 2009.
3. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press: USA, 2004.
4. Zou KH, Liu A, Bandos A, Ohno-Machado L, Rockette H. *Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis*. CRC Press: Boca Raton, FL, 2010.
5. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**(3):837–845.
6. Wieand S, Gail MH, James BR, James KL. A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika* 1989; **76**(3):585–592.
7. Obuchowski NA, Lieber ML. Confidence intervals for the receiver operating characteristic area in studies with small samples. *Academic radiology* 1998; **5**(8):561–571.
8. Tian L. Confidence intervals for $p(y_1 > y_2)$ with normal outcomes in linear models. *Statistics in medicine* 2008; **27**(21):4221–4237.
9. Qin G, Hotilovac L. Comparison of non-parametric confidence intervals for the area under the ROC curve of a continuous-scale diagnostic test. *Statistical Methods in Medical Research* 2008; **17**(2):207–221.
10. Su JQ, Liu JS. Linear combinations of multiple diagnostic markers. *Journal of the American Statistical Association* 1993; **88**(424):1350–1355.
11. Pepe MS, Thompson ML. Combining diagnostic test results to increase accuracy. *Biostatistics* 2000; **1**(2):123–140.
12. Liu C, Liu A, Halabi S. A min–max combination of biomarkers to improve diagnostic accuracy. *Statistics in medicine* 2011; **30**(16):2005–2014.
13. Kang L, Xiong C, Crane P, Tian L. Linear combinations of biomarkers to improve diagnostic accuracy with three ordinal diagnostic categories. *Statistics in Medicine* 2012; **32**(4):631–43.
14. Gao F, Xiong C, Yan Y, Yu K, Zhang Z. Estimating optimum linear combination of multiple correlated diagnostic tests at a fixed specificity with receiver operating characteristic curves. *Journal of Data Science* 2008; **6**(1):105–123.
15. Liu A, Schisterman EF, Zhu Y. On linear combinations of biomarkers to improve diagnostic accuracy. *Statistics in Medicine* 2005; **24**(1):37–47.
16. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; **3**(1):32–35.
17. Perkins NJ, Schisterman EF. The inconsistency of optimal cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American Journal of Epidemiology* 2006; **163**(7):670–675.
18. Schisterman EF, Perkins N. Confidence intervals for the Youden index and corresponding optimal cut-point. *Communications in Statistics‐Simulation and Computation®* 2007; **36**(3):549–563.
19. Silverman BW. *Density Estimation for Statistics and Data Analysis*, Vol. 26. Chapman & Hall/CRC: Boca Raton, FL, 1986.
20. Zou KH, Tempany CM, Fielding JR, Silverman SG. Original smooth receiver operating characteristic curve estimation from continuous data: statistical methods for analyzing the predictive value of spiral CT of ureteral stones. *Academic Radiology* 1998; **5**(10):680–687.
21. Faraggi D, Reiser B. Estimation of the area under the ROC curve. *Statistics in Medicine* 2002; **21**(20):3093–3106.
22. Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. *Biometrical Journal* 2005; **47**(4):458–472.
23. Proschan F, Sullo P. Estimating the parameters of a multivariate exponential distribution. *Journal of the American Statistical Association* 1976; **71**(354):465–472.
24. Percy ME, Andrews DF, Thompson MW. Duchenne muscular dystrophy carrier detection using logistic discrimination: serum creatine kinase, hemopexin, pyruvate kinase, and lactate dehydrogenase in combination. *American Journal of Medical Genetics* 1982; **13**(1):27–38.

25. Kang L, Liu A, Tian L. Linear combination methods to improve diagnostic/prognostic accuracy on future observations. *Statistical Methods in Medical Research* 2013; **22**(4). DOI: 10.1177/096228021348.
26. Tang LL, Kang L, Liu C. An additive selection of markers to improve diagnostic accuracy based on a discriminatory measure. *Academic Radiology* 2013; **20**(7):854–862.
27. Karve S, Cleves MA, Helm M, Hudson TJ, West DS, Martin BC. Good and poor adherence: optimal cut-point for adherence measures using administrative claims data. *Current Medical Research and Opinion* 2009; **25**(9):2303–2310.
28. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology* 2003; **56**(11):1129–1135.
29. Davila M, Christenson LJ, Sontheimer RD. Epidemiology and outcomes of dermatology in-patient consultations in a midwestern US university hospital. *Dermatology Online Journal* 2010; **16**(2):12.
30. Zou KH, Yu C-R, Liu K, Carlsson MO, Cabrera J. Optimal thresholds by maximizing or minimizing various metrics via ROC-type analysis. *Academic Radiology* 2013; **20**(7):807–815.