HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2022 September 10.

Published in final edited form as:

Stat Med. 2021 September 10; 40(20): 4522–4539. doi:10.1002/sim.9077.

Estimation and Construction of Confidence Intervals for Biomarker Cutoff-Points Under the Shortest Euclidean Distance From the ROC Surface to the Perfection Corner

Brian R. Mosier[1,*], **Leonidas E. Bantis**[1]

[1]Department of Biostatistics & Data Science, The University of Kansas Medical Center, Kansas, United States

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is an aggressive type of cancer with a 5-year survival rate of less than 5%. As in many other diseases, its diagnosis might involve progressive stages. It is common that in biomarker studies referring to PDAC, recruitment involves three groups: healthy individuals, patients that suffer from chronic pancreatitis and PDAC patients. Early detection and accurate classification of the state of the disease are crucial for patients' successful treatment. ROC analysis is the most popular way to evaluate the performance of a biomarker and the Youden index is commonly employed for cutoff derivation. The so called generalized Youden index has a drawback in the three-class case of not accommodating the full data set when estimating the optimal cutoffs. In this paper, we explore the use of the Euclidean distance of the ROC to the perfection corner for the derivation of cutoffs in trichotomous settings. We construct an inferential framework that involves both parametric and nonparametric techniques. Our methods can accommodate the full information of a given data set and thus provide more accurate estimates in terms of the decision-making cutoffs compared to a Youden-based strategy. We evaluate our approaches through extensive simulations and illustrate them on a PDAC biomarker study.

Keywords

3-class; Box-Cox; Cutoffs; Euclidean Distance; Kernels; Perfection Corner; ROC; Youden Index

1 | INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is one of the leading causes of cancer related death in the world. Each year 350,000 people are diagnosed with PDAC. The annual resulting deaths are approximately 340,000. Its 5-year survival rate is less than 5%.[1] Due to the small window of time in which PDAC can be treated, early detection is a crucial part of a successful treatment.

[*]**Correspondence** Brian R. Mosier, 3901 Rainbow Blvd., Kansas City, KS. 66160, bmosier@kumc.edu.
Present Address
3901 Rainbow Blvd., Kansas City, KS 66160

A common biomarker that is used for the detection of PDAC is the protein CA19–9. Currently, there is a growing clinical interest for the derivation of additional promising biomarkers that could potentially outperform CA19–9.[2, 3, 4] PDAC studies include individuals that may belong to one of the following groups: healthy, chronic pancreatitis, and PDAC. It is common that in such studies the analysis is done in a pairwise fashion even though three groups are to be discriminated.[5] Following such a pairwise approach for deriving the cutoffs implies that we only accommodate two groups at a time, as opposed to all information provided. This may lead to having under-powered inferences. The resulting larger variances of the optimized cutoffs may also lead to a less informative set of confidence intervals for the cutoffs. Additionally, the higher variability of the point estimates of the cutoffs may lead to poor results when using external data to evaluate their performance.

In the two-class setting, the most popular tool of evaluating a continuous biomarker is the ROC curve. In such a setting, the ROC curve is the plot of *sensitivity* versus 1 − *specificity*. These are defined as $sens(c) = P(X_2 > c) = 1 - F_2(c)$ and $spec(c) = P(X_1 < c) = F_1(c)$, where c is an arbitrary cutoff value. We assume that the biomarker scores of the two groups, X_1 and X_2, come from two different distributions, F_1 and F_2. We are interested in the stochastic ordering $X_1 < X_2$. To construct the ROC curve we scan the possible values of biomarker scores with a continuous cutoff point, c. If the observed biomarker score is less than c, then we classify it as group 1. If the observed biomarker score is greater than c, then we classify it as group 2. Based on the proportion of each group we correctly classified for the selected value of c, we calculate the *sensitivity* and *specificity*. The Youden index is the maximum of the sum of *sensitivity* and *specificity* defined as[6]:

$$\begin{aligned} J_2 &= max_c\{sens(c) + spec(c) - 1\} \\ &= max_c\left\{F_1(c) - F_2(c)\right\}. \end{aligned} \tag{1}$$

ROC analysis has been extended to accommodate trichotomous settings.[7, 8] In the three-class case, we construct an ROC surface within the unit cube. In such a setting, we have three groups, X_1, X_2 and X_3, with stochastic ordering $X_1 < X_2 < X_3$, each distributed F_1, F_2, and F_3. Additionally, there are now two ordered cutoffs, $c_1 < c_2$. Using these cutoffs, the true-class rates are defined as follows $\mathrm{TCR}_1 = P(X_1 < c_1) = F_1(c_1)$, $\mathrm{TCR}_2 = P(c_1 < X_2 < c_2) = F_2(c_2) - F_2(c_1)$, and $\mathrm{TCR}_3 = P(X_3 > c_2) = 1 - F_3(c_2)$. The coordinates of the point on the surface are the triplet $(x, y, z) = (\mathrm{TCR}_1(c_1), \mathrm{TCR}_3(c_2), \mathrm{TCR}_2(c_1, c_2))$. The functional form of the ROC surface is $S = F_2\left(F_3^{-1}(c_2)\right) - F_2\left(F_1^{-1}(c_1)\right)$. The Youden index has been extended to the three-class setting by maximizing the sum of the true-classification rates as follows[9]:

$$\begin{aligned} J_3 &= max_{c_1, c_2; c_1 < c_2}\left\{\mathrm{TCR}_1(c_1) + \mathrm{TCR}_2(c_1, c_2) + \mathrm{TCR}_3(c_2) - 1\right\} \\ &= max_{c_1, c_2; c_1 < c_2}\left\{F_1(c_1) + F_2(c_2) - F_2(c_1) - F_3(c_2)\right\} \\ &= J_2(1, 2) + \mathrm{J}_2(2, 3). \end{aligned} \tag{2}$$

The result in (2) illustrates that $\boldsymbol{J}_3$ cannot simultaneously accommodate all data and is equivalent to proceeding with a pairwise Youden-based analysis.[9] Even though some alternatives to the Youden index for estimating cutoff values have been discussed in the literature, these mainly refer to the two-class case.[10, 11] Therein, the authors discuss the discrepancies of methods for selecting cutoff values and their abilities to estimate the true values. The literature regarding the three-class alternatives to the Youden index is more limited.[12, 13] In a recent paper the authors compare different measures for optimizing the pair of cutoffs using Monte Carlo simulations.[13] However, there is no inferential framework provided for the estimated cutoffs when dealing with non-Youden-based cutoffs. In addition, estimation was only performed using kernel-based estimates of the ROC surface. This leads to less efficient estimation when parametric methods are viable. In this paper, we focus on the shortest Euclidean distance of the ROC surface from the perfection corner, which has been previously referred to the northwestern corner (NWC) method in the two-class setting, and the closest to perfection (CP) method in the three-class setting. We fill in this literature gap by providing both parametric and nonparametric estimation and inferential frameworks for optimal cutoff values.[12, 13]

This paper is organized as follows: In Section 2 we discuss ROC analysis in the three-class setting under the normality assumption and construct confidence intervals for the cutoffs that correspond to the minimized Euclidean distance of the ROC surface from the perfection corner. In Section 3 we explore power transformations to allow for more flexibility when the biomarker scores deviate from normality. This involves taking into account the extra variability introduced by the estimated transformation parameter, which is often ignored in the literature.[14, 15, 16] In Section 4 we investigate nonparametric kernel-based approaches for estimating the ROC surface and the construction of the underlying confidence intervals of the cutoffs. In Section 5 we introduce generalization to the k-class setting. In Section 6 we present an extensive simulation study to evaluate the performance of our methods in comparison to the Youden index in terms of both confidence interval widths for the cutoffs, as well as total classification (sum of true classification rates). In Section 7 we implement our methods on a PDAC data set from a biomarker study conducted at the MD Anderson Cancer Center. Therein, we further highlight some advantages of our approach against the Youden index that could, in some settings, imply a collapse of the middle group to group 1 and group 3, failing to honor the progressive nature of the disease under study. We end with a discussion.

2 | THE PARAMETRIC APPROACH UNDER THE NORMAL ASSUMPTION

It might be the case that the biomarker scores are normally distributed. Let X_1, X_2, and X_3 be the biomarker scores for those who are healthy, those with mild/moderate disease progression, and those with severe disease progression, respectively. Under the normality assumption in the tri-normal setting we assume $X_1 \sim N\left(\mu_1, \sigma_1^2\right)$, $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$ and $X_3 \sim N\left(\mu_3, \sigma_3^2\right)$, where the underlying stochastic ordering is of the form $X_1 \prec X_2 \prec X_3$, with samples of size n_1, n_2, and n_3 respectively. The Youden index is therefore,

$$J_{3;(1,2,3)} = max_{c_1}\left\{\Phi\left(\frac{c_1-\mu_1}{\sigma_1}\right) - \Phi\left(\frac{c_1-\mu_2}{\sigma_2}\right)\right\} + max_{c_2}\left\{\Phi\left(\frac{c_2-\mu_2}{\sigma_2}\right) - \Phi\left(\frac{c_2-\mu_3}{\sigma_3}\right)\right\}. \quad (3)$$

The c_1 and c_2 that maximize $\boldsymbol{J}_3$ can be found from the following equations[17]:

$$c_1^* = \frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2 - \sigma_1\sigma_2\sqrt{(\mu_1-\mu_2)^2 + (\sigma_1^2-\sigma_2^2)\log\left(\frac{\sigma_1^2}{\sigma_2^2}\right)}}{\sigma_1^2-\sigma_2^2},$$
$$c_2^* = \frac{\mu_3\sigma_2^2 - \mu_2\sigma_3^2 - \sigma_2\sigma_3\sqrt{(\mu_2-\mu_3)^2 + (\sigma_2^2-\sigma_3^2)\log\left(\frac{\sigma_2^2}{\sigma_3^2}\right)}}{\sigma_2^2-\sigma_3^2}, \quad (4)$$

if $\sigma_1 \neq \sigma_2$ and $\sigma_2 \neq \sigma_3$. If $\sigma_1 = \sigma_2$, then $c_1^* = \frac{\mu_1+\mu_2}{2}$, and if $\sigma_2 = \sigma_3$, then, $c_2^* = \frac{\mu_2+\mu_3}{2}$. We see that when $\sigma_1 \neq \sigma_2$, c_1 is a function of only μ_1, σ_1, μ_2, and σ_2. It completely ignores any available information from the third group. Similarly, the expression for c_2 ignores any information from the first group. Thus, the Youden index does not take advantage of the full sample size. As an alternative, we investigate the use of the Euclidean distance, which is the closest point on the ROC surface to the perfection corner for choosing optimal cutoff values in the three-class case. This method takes into account all available data when estimating each of the optimal cutoff values. Under the tri-normal setting we can find the values of c_1 and c_2 that minimize this distance numerically by finding the values of c_1 and c_2 that minimize[12]:

$$\begin{aligned} D^* &= min_{c_1,c_2;c_1<c_2}\left\{(1-TCR_1)^2 + (1-TCR_2)^2 + (1-TCR_3)^2\right\}^{1/2} \\ &= min_{c_1,c_2;c_1<c_2}\left\{\left\{1 - \Phi\left(\frac{c_1-\mu_1}{\sigma_1}\right)\right\}^2 \right. \\ &\left. + \left(1 - \left\{\Phi\left(\frac{c_2-\mu_2}{\sigma_2}\right) - \Phi\left(\frac{c_1-\mu_2}{\sigma_2}\right)\right\}\right)^2 + \left\{\Phi\left(\frac{c_2-\mu_3}{\sigma_3}\right)\right\}^2\right\}. \end{aligned} \quad (5)$$

The ROC surface is a function of all six parameters: μ_1, σ_1, μ_2, σ_2, μ_3 and σ_3. A change in the value of any of these six parameters will modify the shape of the ROC surface; therefore, both c_1 and c_2 are a function of μ_1, σ_1, μ_2, σ_2, μ_3, and σ_3.

The Euclidean distance can be visualized in Figure 1, along with the Youden index. The Youden index is visualized as the maximum vertical distance from the useless biomarker plane, which is also visualized in Figure 1. This plane is extended below 0 on the TCR_2 axis in order to visualize the Youden index more easily. As we can see, the optimal point

corresponding to the Euclidean method is not necessarily equal to the same optimal point corresponding to the Youden index.

Once we obtain $\widehat{D}^*$, we may construct confidence intervals for them. To construct a confidence interval for D^*, we can use the delta method to estimate the variance of $\widehat{D}^*$. The delta-based estimate of the variance is given by

$$Var\left(\widehat{D}^*\right) \approx \left\{\frac{\partial \widehat{D}^*}{\partial \mu_1}, \frac{\partial \widehat{D}^*}{\partial \sigma_1}, \frac{\partial \widehat{D}^*}{\partial \mu_2}, \frac{\partial \widehat{D}^*}{\partial \sigma_2}, \frac{\partial \widehat{D}^*}{\partial \mu_3}, \frac{\partial \widehat{D}^*}{\partial \sigma_3}\right\} \widehat{\Sigma} \left\{\frac{\partial \widehat{D}^*}{\partial \mu_1}, \frac{\partial \widehat{D}^*}{\partial \sigma_1}, \frac{\partial \widehat{D}^*}{\partial \mu_2}, \frac{\partial \widehat{D}^*}{\partial \sigma_2}, \frac{\partial \widehat{D}^*}{\partial \mu_3}, \frac{\partial \widehat{D}^*}{\partial \sigma_3}\right\}. \tag{6}$$

Under the normality assumption, $\sum = \text{diag}\left\{\frac{\sigma_1^2}{n_1}, \frac{\sigma_1^2}{2(n_1-1)}, \frac{\sigma_2^2}{n_2}, \frac{\sigma_2^2}{2(n_2-1)}, \frac{\sigma_3^2}{n_3}, \frac{\sigma_3^2}{2(n_3-1)}\right\}$. [19] We obtain $\widehat{\Sigma}$ by replacing σ_i with the respective maximum likelihood estimate, $\hat{\sigma}_i$. The corresponding 95% confidence interval is given by: $\widehat{D}^* \pm z_{1-\alpha/2}\sqrt{Var\left(\widehat{D}^*\right)}$.

After obtaining the estimated optimal cutoff values, $\hat{c}_1^*$ and $\hat{c}_2^*$, we can use the delta method to construct confidence intervals for c_1 and c_2. We can estimate the variance of $\hat{c}_1^*$ and $\hat{c}_2^*$ by using the delta method:

$$Var\left(\hat{c}_i^*\right) \approx \left\{\frac{\partial \hat{c}_i^*}{\partial \mu_1}, \frac{\partial \hat{c}_i^*}{\partial \sigma_1}, \frac{\partial \hat{c}_i^*}{\partial \mu_2}, \frac{\partial \hat{c}_i^*}{\partial \sigma_2}, \frac{\partial \hat{c}_i^*}{\partial \mu_3}, \frac{\partial \hat{c}_i^*}{\partial \sigma_3}\right\} \widehat{\Sigma} \left\{\frac{\partial \hat{c}_i^*}{\partial \mu_1}, \frac{\partial \hat{c}_i^*}{\partial \sigma_1}, \frac{\partial \hat{c}_i^*}{\partial \mu_2}, \frac{\partial \hat{c}_i^*}{\partial \sigma_2}, \frac{\partial \hat{c}_i^*}{\partial \mu_3}, \frac{\partial \hat{c}_i^*}{\partial \sigma_3}\right\}', \; i = 1, 2. \tag{7}$$

The corresponding confidence intervals are therefore $\hat{c}_i^* \pm z_{1-\alpha/2}\sqrt{Var\left(\hat{c}_i^*\right)}$, $i = 1, 2$. The assumption of normality can be restrictive, thus we explore the use of the Box-Cox transformation to transform the data to normality.

3 | THE BOX-COX APPROACH

In practice, it is often the case that the normality assumption is not justifiable by the data at hand. Note that the ROC surface is unchanged under transformations. As such, in these instances, a power transformation can potentially transform the data so that approximate normality is attained. Here, we explore the use of the Box-Cox transformation. This transformation has been used in the ROC literature for different methods of choosing optimal cutoffs.[21, 18] It has been shown to have robust performance in the ROC setting.[22] The Box-Cox transformation is defined by[23]:

$$X_\lambda = \begin{cases} \dfrac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(X) & \lambda = 0 \end{cases}. \tag{8}$$

Again, we will consider the biomarker scores for the different stages of disease, X_1, X_2, and X_3, where the underlying stochastic ordering is of the form $X_1 \prec X_2 \prec X_3$, with sample sizes of n_1, n_2, and n_3, for each of the groups respectively. The Box-Cox transformation uses a common λ to achieve approximate normality for each group. We denote the transformed biomarker scores with $X_1^{(\lambda)}$, $X_2^{(\lambda)}$, and $X_3^{(\lambda)}$ and assume that these are normally distributed.

The transformation parameter, λ, is common for all three groups, X_1, X_2, and X_3. To estimate the full parameter vector, $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3, \lambda)$, we can maximize the full likelihood which is of the following form[25]:

$$\begin{aligned} l(\theta) = & -\frac{n_1}{2}\log(2\pi) - n_1\log(\sigma_1) - \frac{1}{2\sigma_1^2}\sum_i \left(x_{1i}^{(\lambda)} - \mu_1\right)^2 \\ & -\frac{n_2}{2}\log(2\pi) - n_2\log(\sigma_2) - \frac{1}{2\sigma_2^2}\sum_j \left(x_{2j}^{(\lambda)} - \mu_2\right)^2 \\ & -\frac{n_3}{2}\log(2\pi) - n_3\log(\sigma_3) - \frac{1}{2\sigma_3^2}\sum_k \left(x_{3k}^{(\lambda)} - \mu_3\right)^2 \\ & +(\lambda - 1)\left(\sum_i \log(x_{1i}) + \sum_j \log(x_{2j}) + \sum_k \log(x_{3k})\right). \end{aligned} \tag{9}$$

Noting that the above likelihood yields closed form expressions for μ_1, σ_1, μ_2, σ_2, μ_3, and σ_3, we can plug them back into the likelihood function and derive the following profile likelihood that depends only on the transformation parameter λ.[24] The profile likelihood is given by:

$$l(\lambda) = -\frac{n_1}{2}\log\left(\frac{\sum_{i=1}^{n_1}\left(X_{1i}^{(\lambda)} - \frac{\sum_{i=1}^{n_1} X_{1i}^{(\lambda)}}{n_1}\right)^2}{n_1}\right) - \frac{n_2}{2}\log\left(\frac{\sum_{j=1}^{n_2}\left(X_{2j}^{(\lambda)} - \frac{\sum_{j=1}^{n_2} X_{2j}^{(\lambda)}}{n_2}\right)^2}{n_2}\right) - \frac{n_3}{2}\log\left(\frac{\sum_{k=1}^{n_3}\left(X_{3k}^{(\lambda)} - \frac{\sum_{k=1}^{n_3} X_{3k}^{(\lambda)}}{n_3}\right)^2}{n_3}\right) + (\lambda - 1)\left(\sum_{i=1}^{n_1}\log X_{1i} + \sum_{j=1}^{n_2}\log X_{2j} + \sum_{k=1}^{n_3}\log X_{3k}\right) + k\,. \tag{10}$$

To construct the confidence interval for D^*, we can use the delta method for the transformed biomarker scores, while taking into account the additional variability from estimating the transformation parameter, λ. The full 7×7 information matrix (see Web Appendix A) can be derived in closed form. By inverting the information matrix, we obtain the covariance matrix. Denote its upper left 6×6 portion with $\widehat{\Sigma}^{(\lambda)}$. Note that by using $\widehat{\Sigma}^{(\lambda)}$, we take into account the additional variability due to estimating the transformation parameter λ.[18] We approximate the variance of $\widehat{D}^*$ through the delta method as follows.

$$Var\left(\widehat{D}^*\right) \approx \left\{\frac{\partial \widehat{D}^*}{\partial \mu_{1^{(\lambda)}}}, \frac{\partial \widehat{D}^*}{\partial \sigma_{1^{(\lambda)}}}, \frac{\partial \widehat{D}^*}{\partial \mu_{2^{(\lambda)}}}, \frac{\partial \widehat{D}^*}{\partial \sigma_{2^{(\lambda)}}}, \frac{\partial \widehat{D}^*}{\partial \mu_{3^{(\lambda)}}}, \frac{\partial \widehat{D}^*}{\partial \sigma_{3^{(\lambda)}}}\right\} \widehat{\Sigma}^{(\lambda)} \left\{\frac{\partial \widehat{D}^*}{\partial \mu_{1^{(\lambda)}}}, \frac{\partial \widehat{D}^*}{\partial \sigma_{1^{(\lambda)}}}, \frac{\partial \widehat{D}^*}{\partial \mu_{2^{(\lambda)}}}, \frac{\partial \widehat{D}^*}{\partial \sigma_{2^{(\lambda)}}}, \frac{\partial \widehat{D}^*}{\partial \mu_{3^{(\lambda)}}}, \frac{\partial \widehat{D}^*}{\partial \sigma_{3^{(\lambda)}}}\right\}. \tag{11}$$

The corresponding confidence interval for D^* is therefore $\widehat{D}^* \pm z_{1-\alpha/2}\sqrt{Var\left(\widehat{D}^*\right)}$.

Regarding the construction of confidence intervals for the two cutoffs, one needs to account for the fact that the original scale of the data now matters, as the cutoffs lie on the real line and not in the ROC space. A natural approach would be to Box-Cox transform the data followed by the delta method to obtain confidence intervals for the cutoffs, and then back-transforming the intervals with the inverse Box-Cox transformation. However, during our simulations we found that this method led to poor coverage of the confidence intervals for the cutoffs. Here, we consider a resampling bootstrap-based algorithm to derive improved

confidence intervals for the cutoffs, rather than an approach based on the delta method. The recommended bootstrap-based approach is described in the following 7 steps:

1. Using the original data, obtain $\hat{\lambda}$ from equation (10). Apply the Box-Cox transformation and obtain the estimates for the means and variances of the transformed data. Then estimate $\hat{c}_1^{(\lambda)^*}$ and $\hat{c}_2^{(\lambda)^*}$ with equation (5). Lastly, back-transform the data to their original scale using:

$$T = \begin{cases} \left(T^{(\lambda)} \times \lambda + 1\right)^{\frac{1}{\lambda}} & \lambda \neq 0 \\ \exp\left(T^{(\lambda)}\right) & \lambda = 0 \end{cases}, \tag{12}$$

   to obtain $\hat{c}_1^*$ and $\hat{c}_2^*$.

2. Sample with replacement independently from X_{1i}, X_{2j}, and X_{3k}, $i = 1,\ldots,n_1$, $j = 1,\ldots,n_2$, and $k = 1,\ldots,n_3$, respectively. Denote these samples as $X_{1(b)}$, $X_{2(b)}$, and $X_{3(b)}$. These are the bootstrap samples for the b_{th} bootstrap iteration.

3. Substitute $X_{1(b)}$, $X_{2(b)}$, and $X_{3(b)}$ into equation (10) and obtain $\hat{\lambda}^{(b)}$, which is the estimate of λ for the b_{th} bootstrap iteration.

4. Apply the Box-Cox transformation using $\hat{\lambda}_{(b)}$ to $X_{1(b)}$, $X_{2(b)}$, and $X_{3(b)}$. Compute the corresponding means and variances and substitute them into (5) to obtain $\hat{c}_{1(b)}^{(\lambda)^*}$ and $\hat{c}_{2(b)}^{(\lambda)^*}$, which are the estimates for the optimal cutoff points for the current bootstrap sample.

5. Back-transform $\hat{c}_{1(b)}^{(\lambda)^*}$ and $\hat{c}_{2(b)}^{(\lambda)^*}$ using (12) to obtain the current bootstrap estimates of the cutoffs on the original scale, denoted by $\hat{c}_{1(b)}^{(\lambda)^*}$ and $\hat{c}_{2(b)}^{(\lambda)^*}$.

6. Repeat steps 2–5 B times and obtain the bootstrap-based estimate of the variance of $\hat{c}_1^*$ and $\hat{c}_2^*$, by computing $Var\left(\hat{c}_j^*\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{c}_{j(b)}^* - \bar{\hat{c}}_j^*\right)^2$, where $\bar{\hat{c}}_j^* = \frac{\sum_{i=1}^{B}\hat{c}_{j(b)}^*}{B}$, $j = 1, 2$.

7. The confidence intervals are then given by

$$\hat{c}_j^* \pm z_{1-\frac{\alpha}{2}}\sqrt{Var\left(\hat{c}_{j(b)}^*\right)}. \tag{13}$$

   As seen in our simulation section, this method provides a robust parametric approach that can oftentimes accommodate biomarker scores that are not in the power normal family.

4 | KERNEL-BASED APPROACH

In cases when normality is not approximated even after the Box-Cox transformation, a nonparametric strategy is required. Here, we explore the use of kernels and develop an algorithm for estimating D* and the associated cutoffs, along with their corresponding confidence intervals. Consider the biomarker scores, X_1, X_2, and X_3 where the underlying stochastic ordering is of the form $X_1 < X_2 < X_3$, and with sample sizes n_1, n_2, and n_3, respectively. The kernel-based cumulative distribution estimate for group $i = 1,2,3$ is given by:

$$\hat{F}_i^{(k)}(x) = \frac{1}{n_i}\sum_{j=1}^{n_i} \Phi\left(\frac{x - x_{ij}}{h_i}\right), \tag{14}$$

where n_i is the sample size of group i, x_{ij} is the score of the j_{th} individual in group i, and h_i is the bandwidth for group i. We employ a Gaussian kernel, and for the bandwidth we use[18]:

$$h_i = 0.9\min(sd(x_i),\ iqr(x_i)/1.34)n_i^{-0.2},$$

Traditional bootstrap methods did not perform well in terms of coverage for the cutoff values. The proposed corrected bootstrap method is denoted kernel w/ϵ (read kernel with ϵ) to differentiate it from a naive bootstrap approach. The coverage correction is made with the inclusion of step 3 in the following algorithm. Our proposed algorithm for the confidence intervals of the cutoffs is as follows:

1. Calculate h_1, h_2, and h_3 based on X_1, X_2, and X_3. Derive $\hat{c}_1^*$ and $\hat{c}_2^*$.
2. Sample with replacement independently from X_1, X_2, and X_3 to get the bootstrap samples $X_{1(b)}$, $X_{2(b)}$, and $X_{3(b)}$, which are of size n_1, n_2, and n_3, respectively.
3. Set $X_{i(b)} = X_{i(b)} + \epsilon_i$, $i = 1,2,3$, where each $\epsilon_i \sim N\left(0, h_i^2\right)$.
4. Obtain $\hat{c}_{1(b)}^*$ and $\hat{c}_{2(b)}^*$ by using the new $X_i's$.
5. Repeat steps 2–4 $\boldsymbol{B}$ times and obtain the bootstrap-based estimate of the variance of $\hat{c}_1^*$ and $\hat{c}_2^*$ by computing $Var\left(\hat{c}_j^*\right) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{c}_{j(b)}^* - \bar{\hat{c}}_j^*\right)^2$, where $\bar{\hat{c}}_j^* = \frac{\sum_{i=1}^{B}\hat{c}_{j(b)}^*}{B}$, $j = 1, 2$.
6. The confidence intervals are then given by

$$\hat{c}_j^* \pm z_{1-\frac{\alpha}{2}}\sqrt{Var\left(\hat{c}_{(b)}^*\right)}. \tag{15}$$

We can implement an analogous algorithm for the confidence interval of D^*, but without the inclusion of step 3. This is because the added variation from the ϵ terms adds bias to D^* and does not improve coverage. This can be seen in Web Appendix B. The kernel-based

approach for constructing confidence intervals provides a nonparametric robust framework that can handle data that the Box-Cox cannot.

5 | K-CLASS SETTING

Our methods can be extended beyond the three-class setting into the general k-class setting, so that there are k sets of biomarker scores, $X_1,\ldots,X_k$. Suppose $X_1 \sim F_1,\ldots, X_k \sim F_k$ with stochastic ordering $X_1 < \ldots < X_k$. By scanning for all possible values of the $k-1$ cutoffs, the k-dimensional hypersurface is derived through the k true-class rates in the k-dimensional space.

With the k groups we need to estimate $k-1$ optimal cutoff values. The Youden index has been extended to the k-class setting.[9] Similar to the three-class setting, using the Youden index to estimate the $k-1$ optimal cutoff values is equivalent to making $k-1$ pairwise comparisons. When using this to estimate the i_{th} optimal cutoff values, we utilize only the data from groups i and $i+1$, thus ignoring the data from the other $k-2$ groups. As the k increases, we ignore a larger portion of the available data when making these estimates.

Euclidean distance is defined for any number of dimensions, and thus while we cannot visualize it, we can find the point on the ROC hypersurface that is closest to the perfection corner. This distance is defined as,

$$D^* = \min_{c_j; c_j < c_{j+1}} \left\{ \sum_{i=1}^{k} \left(1 - TCR_i\right)^2 \right\}^{1/2}, j = 1, \ldots, k-1. \tag{16}$$

By minimizing this expression, we obtain estimates of the $k-1$ optimal cutoff values. To estimate the variance of these optimal cutoff values, we can implement the methods of the previous sections while accounting for all k groups. As k increases, the advantage of using the Euclidean method grows over the Youden index.

6 | SIMULATION STUDY

In this section we present an extensive simulation study to evaluate the performance of our approaches. All results for Section 6.1 are given in Web Appendix B. We explore scenarios that involve data generated from normal, log-normal, gamma, and normal mixture distributions. We consider sample sizes of (50,50,50), (100,100,100), (200,200,200), and (150,100,50). The parameters of the distributions are chosen in such a way that the proposed Euclidean-based cutoffs and the Youden-based cutoffs, when using the true model, are equal. These parameters are given in Table 1. Such a simulation setting allows for a "head to head" comparison regarding the coverage and the widths of the corresponding confidence intervals. In addition, the aforementioned parameters are chosen so that the volume under the surface (VUS) values explored are plausible and approximately equal to 0.4 and 0.6. The true values of the pairwise area under the curve (AUC) are also given in Table 1 and they span from 0.6106 to 0.8155, covering a wide range of potential scenarios that could be met in practice. The gamma distributions have scale parameter, α, and shape parameter, β. Our discussion in this section regarding the proposed kernel-based approach will focus on the approach that is

given in Section 4 that utilizes the ϵ's, which is the added noise in step 3 of the proposed algorithm (denoted as kernels w/ϵ or as proposed kernel-based method). Figure 2 includes the coverage from each of the proposed approaches. These include a naive kernel-based approach without the use of the ϵ's in order to contrast their performance in terms of coverage. This demonstrates the necessity of step 3 in the proposed algorithm presented in Section 4. Additional simulations in which the cutoffs of the two measures are not set to be equal are given and discussed in Section 6.2 and relevant tables are available in the Web Appendix C.

6.1 | Simulation Results for the Scenarios with Equal Pairs of Cutoffs

When the data are generated from normal distributions, the proposed delta-based confidence intervals (CIs) are narrower than the Box-Cox-based intervals and the proposed kernel-based intervals, as expected. All results relating to scenarios generated from normal distributions with equal sets of cutoffs are available in Table B.1 of Web Appendix B. The Box-Cox-based CIs seem to be slightly inflated, indicating that there is only a small price to pay for estimating the additional transformation parameter, λ. This is not the case for the proposed kernel-based confidence intervals, which resulted in a two-fold increase of the CI width (see Figure 3). When the sample size is 200 and the $VUS = 0.4$, the confidence interval for c_1 using the kernel-based method results in a confidence interval 1.8 times that of the delta-based approach. Regarding the comparison of our proposed CIs versus the Youden-based ones under the same assumptions, we observe a substantial gain in terms of interval width. This is reflected in smaller widths in all but two scenarios, with large improvement in several scenarios. The two scenarios where the Youden index found narrower confidence intervals correspond to when the data was generated from normal distributions with a VUS of 0.6, and when the sample sizes were (150,100,50). For the delta method approach, the interval widths for c_2 were 0.4177 and 0.4295 for the Youden index and Euclidean methods, and for the Box-Cox approach, they were 0.4292 and 0.4504. Although, the average confidence interval widths for the cutoffs from these scenarios were smaller for the Euclidean method than the Youden index. The average interval widths for the Euclidean method and Youden index were 0.3356 versus 0.3826 for the delta method approach and 0.3618 versus 0.4066 for the Box-Cox approach. When data are generated from normal distributions with a VUS of 0.4 and pairwise AUCs equal to 0.6106 and 0.7103, the average CI widths of our approaches are 0.4366, 0.4544, and 0.7443 using the assumption of normality, the Box-Cox approach, and the kernels w/ϵ, respectively. The corresponding widths for the Youden, which exhibit a multifold increase are equal to 0.9337, 0.8674, and 1.6042, respectively. This resulted in intervals that are on average 2.14, 1.91, and 2.16 times wider for the Youden index than for the proposed delta, Box-Cox, and kernel-based methods. Note that the coverage of all proposed CIs are acceptable and close to the targeted level of 95% (see Figure 2). Regarding the CIs for the D^* itself, we observe that our CIs perform well in terms of coverage throughout all scenarios. The widths of which are similar when assuming normality and when applying the Box-Cox transformation. However, they are substantially inflated when using the proposed kernel approach, which reflects the cost to pay when working nonparametrically.

When the data are generated from log-normal distributions, the proposed Box-Cox-based confidence intervals are shorter than the proposed kernel-based ones. All results relating to scenarios generated from log-normal distributions with equal sets of cutoffs are available in Table B.2 of Web Appendix B. This is to be expected because the kernel-based method makes no parametric assumptions. This suggests that if the data can be transformed to normality, then the Box-Cox approach is preferred. The proposed CIs versus the Youden-based ones under the same assumptions show improved CI widths, with substantial improvement in several scenarios. For example, when the VUS is 0.4 and the pairwise AUCs are 0.6775 and 0.6670, the average CI widths of our approaches are 0.4248 and 0.6787 using the Box-Cox approach and the kernel w/ϵ approach, respectively. The corresponding widths for the Youden index exhibit a substantial increase and are equal to 0.7756 and 1.3902, respectively. This resulted in intervals that are on average 1.83 and 2.05 times wider for the Youden index than for the proposed Box-Cox and kernel-based methods. The coverage of the proposed CIs are acceptable and close to the targeted level of 95%. The CIs for D^* provide adequate coverage for all scenarios. The widths are substantially inflated when using the kernels rather than the Box-Cox approach. This again shows the cost to pay when using no parametric assumptions.

When the data are generated from gamma distributions, the Box-Cox-based confidence intervals are shorter than the kernel-based ones. All results relating to scenarios generated from gamma distributions with equal sets of cutoffs are available in Table B.3 of Web Appendix B. This demonstrates the benefits of using parametric assumptions. Even though the gamma distribution is not within the power normal family, we see that the coverage for the proposed CIs is still close to the targeted level of 95%. This demonstrates the robustness of the Box-Cox transformation to adequately transform the data to normality. The proposed CIs versus the Youden-based ones under the same assumptions show improved CI widths in all cases, with substantial improvement in several scenarios. For example, when the VUS is 0.4 and the pairwise AUCs are equal to 0.6049 and 0.6900, the average CI widths of our approaches are 0.6049 and 0.9760 for the Box-Cox and kernel-based approaches, while the corresponding widths for the Youden are 1.0152 and 1.8865. This resulted in intervals that are on average 1.68 and 1.93 times wider for the Youden than for the proposed Box-Cox and kernel-based approaches. We see that the coverage for the Euclidean method when using the Box-Cox approach is adequate in all scenarios and close to the targeted level of 95%. When using the proposed kernel-based method, when the VUS was 0.4 and the sample sizes were (200,200,200) and (150,100,50), the coverage was only 0.881 and 0.880, respectively. In all other scenarios the coverage for the proposed kernel-based method was at least 0.9.

When the data are generated from bimodal mixture distributions, the Box-Cox approach provided poor coverage for the cutoffs. All results relating to scenarios generated from mixture distributions with equal sets of cutoffs are available in Table B.4 of Web Appendix B. The average coverage over all of the mixture scenarios for our proposed Box-Cox method for c_1 and c_2 were 0.9997 and 0.0957, respectively. This demonstrates the necessity of having a more flexible nonparametric approach. With the proposed kernel-based approach we see coverage that is close to the targeted level of 95%. The proposed CIs versus the Youden-based ones under the same assumptions show greatly improved CI widths in all cases. The average CI width for the proposed method is 0.8918, while the average width for

the Youden approach is 1.5354. This corresponds to intervals that are 1.72 times wider on average for the Youden than the proposed method.

6.2 | Simulation Results for the Scenarios with Unequal Pairs of Cutoffs

We explored scenarios where the cutoffs were not equal by using the Box-Cox approach. All results relating to scenarios with unequal cutoffs are included in Table C.1 in Web Appendix C. This includes scores generated from normal, log-normal, and gamma distributions. The parameters of the distributions the data were simulated from are displayed in Table 2. Using the Box-Cox approach allowed us to compare the performance in interval widths for a wide variety of scenarios. Interestingly, there were 4 instances where the Youden index provided coverage below 0.90, while the Euclidean method had only 1 instance with coverage lower than 0.90.

For the normal scenarios, we used evenly spaced groups. We saw that on average the Youden index derived intervals that were 1.64 times wider than for the Euclidean method. For the log-normal scenarios the Youden index derived intervals that were on average 1.55 times wider than for the Euclidean method. For the gamma scenarios the Youden index derived intervals that were on average 1.40 times wider than the Euclidean method.

We observe that even though the two methods were not estimating the same set of cutoffs, the Euclidean method still saw narrower confidence intervals in almost all cases. There were a total of 4 instances where the confidence intervals for c_2 that were derived using the Youden index were narrower than for the Euclidean method. These corresponded to the gamma scenarios with a VUS of 0.6 for each of the sample sizes. For these scenarios, the intervals found by the Euclidean method were on average 1.12 times the width of those found by the Youden index for c_2. While the confidence intervals for c_2 were wider for the Euclidean method, the corresponding c_1 confidence intervals were shorter. When averaging the widths of both cutoffs for these scenarios, the Youden index had interval widths that were 1.17 times wider than for the Euclidean method. On average, the Euclidean method provide shorter confidence intervals, even when the Youden index had narrower confidence intervals for one of the cutoffs. In all other instances the Euclidean method found shorter intervals than the Youden index.

6.3 | Simulation Results to Compare Performance of Estimated Cutoff Values in Terms of True Classification Rates

When interest lies in maximizing the total classification of a biomarker (sum of TCRs), the Youden index is the obvious choice to estimate cutoff values as it seeks to maximize the sum of the true classification rates. By construction, in training data, the Youden index will achieve a higher total classification than the Euclidean method. Rather than only focusing on results from training data, it is important to examine how the cutoffs estimated by both methods perform with regard to total classification in the population. In simulation this can be accomplished by using the estimated cutoff values for both methods in conjunction with the true distributions of the biomarker scores. As has been shown in Section 6.1 and Section 6.2, the Youden index often provides substantially wider confidence intervals than the Euclidean method. Wider confidence intervals are a consequence of the under-utilization of

the data under a Youden-based approach in three-class settings. Such a consequence may not be apparent in training data, however, using simulations, we know the true distributions of the data. Therefore, by using the true distributions of the biomarker scores, we can evaluate the true performance of estimated cutoff values. Such simulations allow us to observe how biomarker scores from a study with limited sample size might perform in the population. In such a setting, we see that it is possible for the cutoffs derived by the Euclidean method to have higher total classification than those estimated by the Youden index.

Table 3 includes results from eight scenarios corresponding to two sets of normally distributed biomarker scores, each with sample sizes varying from 20 for each of the three groups to 100 for each group. The values in Table 3 are based on using the estimated cutoff values from both methods in conjunction with the true distributions of the biomarker scores. For the first four scenarios of Table 3, on average, the Youden index provides higher total classification than the Euclidean method. For a sample size of 20 per group, on average the Youden index yields a total classification of 0.0013 higher than the Euclidean method. For this scenario, the probability of having higher classification for the Youden index is 60.87%. This provides strong justification for using the Youden index in this case. One drawback for the Youden index in this case is that the worst performing iteration/occurrence in our simulation has total classification (Youden-based) of 1.2997, while it is 1.4115 for the Euclidean method. This is not surprising since there is higher variability when using the Youden index due to the under-utilization of data. As the sample size increases, we see that the Youden index has a greater difference in average total classification, as well as a higher probability of achieving a larger total classification. In addition, while the Euclidean method provides minimum total classification that is larger than that of the Youden index, this difference decreases as sample size increases. In the last four scenarios in Table 3, we see that on average, the Euclidean method provides higher total classification for sample sizes of 20, 30, and 50 per group. When the sample size for each group is 20, the total classification for the Euclidean method is 0.0081 higher than for the Youden index. Even with sample sizes of 50 for each group, the Euclidean method still seems to have a higher total classification than the Youden index. Furthermore, when the sample size for each group is 20, the probability of having higher total classification than the Youden index is 63.73%. For larger sample sizes, equal to 30 and 50 per group, the probabilities of having higher total classification were 59.49% and 51.48% respectively. It is only when the sample sizes reach 100 for each group that the Youden index obtained higher total classification than the Euclidean method, and even then, the Youden index only provided higher total classification by 0.0007.

Figure 4 is a histogram of the percent difference in the total classification for the Euclidean method versus the Youden index when evaluated using the true biomarker distributions. Positive values along the *x*-axis correspond to better performance for the Euclidean method. Scenarios (a)-(d) in Figure 4 correspond to the first four scenarios in Table 3, and illustrate a case where the Youden index provides better performance on average than the Euclidean method for all explored sample sizes. As the sample size increases, the benefit of using the Youden index increases. In scenarios (e)-(h) in Figure 4, which correspond to scenarios 5–8 in Table 3, we see better performance for the Euclidean method with smaller sample sizes. This benefit decreases as the sample size increases due to the lower variability in the

estimates of the Youden-based cutoffs. An interesting note for each of the plots is that we see a skewed distribution for the percent difference in total classification. We see that due to the higher variability in the estimates of the Youden-based optimal cutoff values, for the scenarios we explored, the potential gains in total classification when using the Euclidean method are greater than the potential gains when using the Youden index. For instance, in Figure 4 (a), the iteration that corresponds to the maximum benefit of the Euclidean method yields a total classification that is 8% higher than the Youden index. For the same simulation, the iteration that corresponds to the maximum benefit of the Youden index yields a total classification that is 4% higher than the Euclidean method.

In addition to the scenarios generated from normal distributions, we also explore scenarios corresponding to lognormal and gamma distributions. Results are available in Web Appendix D. Therein, we provide results that demonstrate that the Euclidean method may outperform the Youden index in terms of total classification (based on the true distributions of the data).

As suggested by an anonymous referee, we explored the use of the Euclidean method and the Youden index in the two-class setting. Similar to our approach for the three-class setting, we examined the performance of the estimated cutoff values in terms of total classification based on the true distributions of the biomarker scores. Results are displayed in tables D.3, D.4, and D.5 in Web Appendix D. In the two-class setting, we see that the Euclidean method tends to provide a more balanced set of TCRs than the Youden index. This is in line with the what we observed in the three-class setting and in line with the discussion presented in Hua & Tian[13]. Additionally, in some of the explored scenarios, the Euclidean method provides a higher total classification when the performance of the cutoffs are gauged with the true distributions of the biomarker scores. This implies that in such scenarios, on average, the estimated cutoff outperforms the Youden index in terms of testing based on total classification. Our testing strategy of using the true distributions of the biomarker scores is equivalent to using infinite sample sizes for both groups of the testing set.

When the optimal TCR_1 (specificity) is close in value to the optimal TCR_2 (sensitivity) for both methods, they produce similar sets of TCRs. In such instances, the Euclidean method tends to outperform the Youden index in terms of total classification (sum of TCRs) in the population. This is observed in Table D.3 of Web Appendix D, where $X_1 \sim N(0,1^2)$ and $X_2 \sim N(1,1.1^2)$. For the Euclidean method, the true optimal values of TCR_1 and TCR_2 are 0.6940 and 0.6729, respectively. For the Youden index, they are 0.7190 and 0.6487. When the sample size is 20 for both groups, the Euclidean method produces a total classification of 1.4635, while for the Youden index it is 1.4607. In this scenario, both methods produce a relatively balanced set of TCRs, and the Euclidean method outperforms the Youden index in terms of total classification. Similar results are observed in Table D.4, where $X_1 \sim Lognormal(1,0.4^2)$ and $X_2 \sim Lognormal(1.5,0.45^2)$, and in Table D.5, where $X_1 \sim Gamma(2,1)$ and $X_2 \sim Gamma(4,1)$.

When the optimal set of TCRs is unbalanced for both methods, the Euclidean method and Youden index produce much different sets of TCRs. In such instances, the Youden index tends to outperform the Euclidean method in terms of total classification in the population.

This is observed in Table D.3 of Web Appendix D, where $X_1 \sim N(0,1^2)$ and $X_2 \sim N(1,1.7^2)$. For the Euclidean method, the true optimal values of TCR_1 and TCR_2 are 0.7236 and 0.5945, respectively. For the Youden index, they are 0.8486 and 0.4929. When the sample size is 20 for both groups, the Euclidean method produces a total classification of 1.3116, while for the Youden index it is 1.3314. The Youden index provides a much larger value of TCR_1 than TCR_2. The Euclidean method also has a larger value of TCR_1 than TCR_2, but it attempts to balance the two, which sacrifices total classification. There is a large difference in the optimal TCRs for both methods, and in in such instances, the Youden index outperforms the Euclidean method. Similar results are observed in Table D.4, where $X_1 \sim Lognormal(1,0.4^2)$ and $X_2 \sim Lognormal(1.5,0.8^2)$, and in Table D.5, where $X_1 \sim Gamma(2,1)$ and $X_2 \sim Gamma(1.5,3)$.

This part of the simulation provides strong justification for using the Euclidean method in the three-class setting when sample size is limited. While the Youden index seeks to provide the highest total classification, its higher variability may lead to inferior performance in the population than the Euclidean method when the study sample size is limited. The simulations regarding the two-class setting provide justification for using the Euclidean method when the optimal TCR_1 is close in value to the optimal TCR_2. In such instances the Euclidean method tends to provide higher total classification in the population.

7 | PANCREATIC DUCTAL ADENOCARCINOMA SCREENING

CA19–9 is a commonly used protein for the detection of PDAC. In a recent study investigators evaluated the performance of additional proteins on PDAC patients, of which, we focus on three: TIMP-1, LRG1, and IGFBP-2.[5] The design of the study refers to a trichotomous setting, as the data set includes 60 healthy controls (or group 1), 60 patients suffering from benign pancreatic disease (chronic pancreatitis or group 2), and 73 PDAC cases (or group 3). Initially, we test for normality before and after the Box-Cox transformation to determine which approach is appropriate for each biomarker. Both LRG1 and IGFBP-2 had approximately normally distributed scores and thus there is no need for applying the Box-Cox transformation. The p-values for the healthy, chronic pancreatitis, and PDAC groups for the LRG1 biomarker were found to be 0.8945, 0.3147, and 0.9875 using the Kolmogorov-Smirnov test (KS test). For the IGFBP-2 biomarker, the corresponding p-values were 0.0577, 0.0628, and 0.3197. Regarding the scores of TIMP-1 and CA19–9, we observe that they do not conform with the assumption of normality. The KS-based p-values for TIMP-1 for the healthy, chronic pancreatitis, and PDAC groups are 0.0770, 0.9759, and 0.0463, respectively. For the CA19–9 biomarker the corresponding p-values were 0.0530, 0.0758, and 0.0264. This remains even after the Box-Cox transformation, as the corresponding KS-based p-values for the transformed TIMP-1 scores are equal to 0.0476, 0.8592, and 0.0185 respectively. The corresponding p-values for CA19–9 were 0.0298, 0.0304, and 0.0808. Figure 5 includes the distributions of the biomarker scores and their ROC surfaces.

In Table 4 we observe that both the Youden index and the D^* measure rank the CA19–9 biomarker as the best performing one (YI = 0.6248, D^* = 0.8639).

Table 5 includes the estimated cutoffs with corresponding confidence intervals and interval widths for both the Euclidean and Youden methods. The Euclidean method provides shorter confidence intervals for both c_1 and c_2 for every one of the biomarkers. For the CA19–9 biomarker, the interval width for c_1 given by the Euclidean method is 3.58 times shorter than for the Youden index. For the CA19–9, TIMP-1, and IGFBP-2 markers, the Euclidean-based confidence intervals for c_1 are less than half the width of the Youden-based intervals. For the LRG1 marker, the Euclidean-based confidence interval for c_2 is less than half the width of the Youden-based interval. In terms of interval width, the Euclidean method outperforms the Youden index by a substantial margin.

An interesting observation is that for the CA19–9 protein, our proposed cutoffs are equal to −0.4634 and 1.6731 with 95% CIs of (−1.0459, 0.1191) and (1.3285, 2.0177), respectively. However, the Youden index-based approach yields estimated cutoffs of 1.1546 for both c_1 and c_2. This implies a true-class triplet equal to (TCR_1, TCR_2, TCR_3) = (0.8800,0.0000,0.7432). We provide all true classification rates in Table 6. Thus, the Youden index results in a pair of cutoffs that completely collapses the middle group of patients that suffer from benign pancreatic disease to the remaining two groups (healthy and PDAC). Such cutoff estimates would drive a decision-making process that does not allow for monitoring this group, as they would either be regarded falsely as healthy or falsely as having PDAC. Both of these misclassifications may have severe consequences in this setting. The proposed cutoffs for CA19–9 yield estimates that correspond to a true-class triplet equal to (TCR_1, TCR_2, TCR_3) = (0.5220,0.3703,0.6518). That is, the Euclidean-based cutoffs seem to perform better with respect to the aforementioned aspect, acknowledging the existence of a "middle" group. Unfortunately, this comes at a cost of reduced TCR_1 and TCR_3, as expected. An analogous situation is observed with the TIMP-1 protein. Again, the middle group is completely collapsed by the Youden index. For the IGFBP-2 marker, the Youden index provides $TCR_2 = 0.0033$, which has also effectively collapsed the middle group. Only one out of four of the biomarkers retains a middle group when using the Youden index. This issue is not observed with any of the four biomarkers when using the Euclidean method. Our observation that the Euclidean method provides a more balanced set of TCRs is in line with other literature that has observed a more balanced set of TCRs from the Euclidean method than from the Youden index.[13]

We further explored an additional parametric approach to fit the TIMP-1 biomarker. The normality of the first two groups could not be rejected based on the KS-test, while this is not the case for the third group (p-value = 0.0463). Following the suggestion of a referee, we noted that for the third group, one could successfully fit a skew normal density (KS-test p-value = 0.0517 and thus the skew normal assumption is not rejected). Relevant goodness of fit plots are available in Web Appendix E. The estimated cutoffs, c_1 and c_2, along with their corresponding 95% confidence intervals for the Euclidean method are 0.0575 (−0.2573, 0.3722) and 1.2298 (1.0058, 1.4537), respectively. The corresponding Youden-based cutoffs and 95% confidence intervals are 0.0402 (−0.5979, 0.6782) and 1.1639 (0.6137, 1.7142). The confidence intervals for c_1 when using the Youden index is 2.0275 times wider than when using the Euclidean method. The confidence interval for c_2 when using the Youden index is 2.4570 times wider than when using the Euclidean method. We see that the Euclidean method provides substantially shorter confidence intervals compared to the

Youden index under this setting. For c_1, the kernel-based confidence interval widths are 0.5039 and 1.1542 for the Euclidean method and Youden index, respectively. When using the skew normal distribution, the corresponding confidence interval widths are 0.6295 and 1.2761. For c_2, the kernel-based confidence interval widths are 0.6437 and 0.7286 for the Euclidean method and the Youden index, respectively. When using the skew normal distribution, the corresponding confidence interval widths are 0.4479 and 1.1005.

Our results indicate that our approaches might reveal such interesting behaviors that are eye-opening to clinicians and public health practitioners. Our inferences, as shown in this application, may reveal an alternative decision making framework which may be more beneficial depending on the setting under study and the associated follow-up implied for patients.

8 | DISCUSSION

As opposed to typical case-control biomarker studies, there are many settings in which patients are to be classified in three, rather than two, categories. When interest lies in deriving decision-making cutoffs, the Youden index and its recent three-class generalization are commonly used in the literature.[9] In this paper, we explore an alternative measure for dealing with such trichotomous settings and the derivation of its associated cutoffs. This measure refers to the distance of the ROC surface from the perfection corner of the unit cube. Even though such a measure is known in the literature, an inferential framework regarding its associated cutoffs is not currently available.[12, 13]

In this paper, we fill this literature gap and propose parametric, flexible parametric, and kernel-based nonparametric approaches for the estimation of the pair of the cutoffs and the derivation of their confidence intervals. We evaluate our approaches through extensive simulations that indicate narrower confidence intervals for the Euclidean-based cutoffs compared to the Youden-based ones in the explored scenarios, and we demonstrate scenarios where the Euclidean method provides higher total classification than the Youden index when cutoffs are evaluated using the true distributions of the biomarker scores. We demonstrate settings in both simulation and real-world application where multifold differences of such widths are present, making the use of our method the recommended choice since point estimates for the Youden index are uncertain to be accurate. In a given data set, the Euclidean and the Youden-based cutoffs will not necessarily agree. Such discrepancies have been discussed in the literature before.[11, 13] This may imply substantial discrepancies regarding the associated true-class rates. The Euclidean-based cutoffs tend to yield more evenly valued true-class rates compared to the Youden index. As seen in our application, the three-class Youden-based cutoffs may completely ignore the existence of one group. This can be problematic when monitoring is crucial for a group of patients, since they could be misclassified either as healthy or with aggressive cancer. This, in turn, could imply unnecessary invasive follow ups. Thus, our approaches may also serve as an eye-opening alternative decision-making process that imply decisions which are clinically more appealing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported in part by the COBRE grant (NIH) P20GM130423 and by the NIH Clinical and Translational Science Award (UL1TR002366) to the University of Kansas. The authors would like to thank Dr. Samir Hanash for providing the data used in the application section.

References

1. Foucher ED, Ghigo C, Chouaib S, Galon J, Iovanna J, Olive D. Pancreatic ductal adenocarcinoma: a strong imbalance of good and bad immunological cops in the tumor microenvironment. Front Immunol. 2018; 9:1044. [PubMed: 29868007]
2. Wingren C, Sandstrom A, Segersvard R, Carlsson A, Andersson R, Lohr M, Borrebaeck CA. identification of serum biomarker signatures associated with pancreatic cancer. Cancer Res. 2012; 72(10): 2481–2490. [PubMed: 22589272]
3. Gold DV, Goggins M, Modrak DE, et al. Detection of early-stage pancreatic adenocarcinoma. Cancer Epidemiol, Biomarkers and Prevention. 2010; 19(11): 2786–2794.
4. Chan A, Dimitromanolakis A, Brand RE, Serra S, Diamandis EP, Blasutig IM. Validation of biomarkers that complement CA19.9 in detecting early pancreatic cancer. Clin Cancer Res. 2014; 20(22): 5787–5795. [PubMed: 25239611]
5. Capello M, Bantis LE, Ghislaine S, et al. Sequential validation of blood-based protein biomarker candidates for early-stage pancreatic cancer. J Natl Cancer Inst. 2017; 109(4): djw266.
6. Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3(1): 32–35. [PubMed: 15405679]
7. Mossman D Three-way ROCs. Med Decis Making. 1999; 19(1): 78–89. [PubMed: 9917023]
8. Nakas CT, Yiannoutsos CT. Ordered multiple-class ROC analysis with continuous measurements. Stat Med. 2004; 23(22): 3437–3449. [PubMed: 15505886]
9. Nakas CT, Dalrymple-Alford JC, Anderson T, Alonzo TA. Generalization of Youden index for multiple-class classification problems applied to the assessment of externally validated cognition in Parkinson disease screening. Stat Med. 2013; 32: 995–1003. [PubMed: 22949169]
10. Unal I. Defining an optimal cut-point value in ROC analysis: an alternative approach. Comput Math Methods Med. 2017; 2017: 3762651. [PubMed: 28642804]
11. Perkins NJ, Schisterman EF. The inconsistency of “optimal” cut-points using two ROC based criteria. American J Epidemiol 2006; 163(7): 670–675.
12. Attwood K, Tian L, Xiong C. Diagnostic thresholds with three ordinal groups. J Biopharm Stat. 2014; 24(3): 608–633. [PubMed: 24707966]
13. Hua J, Tian L. A comprehensive and comparative review of optimal cut-points selection methods for diseases with multiple ordinal stages. J Biopharm Stat. 2020; 30(1): 46–68. [PubMed: 31250693]
14. Molodianovitch K, Faraggi D, Reiser B. Comparing the areas under two correlated ROC curves: Parametric and nonparametric approaches. Biom J. 2006; 48(5): 745–757. [PubMed: 17094340]
15. Schisterman EF, Reiser B, Faraggi D. ROC analysis for markers with mass at zero. Stat Med. 2006; 25(4): 623–638. [PubMed: 16345033]
16. Molanes-Lopez EM, Leton E. Inference of the Youden index and associated threshold using empirical likelihood for quantiles. Stat Med. 2011; 30(19): 2467–2480. [PubMed: 21751233]
17. Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. Biom J. 2005; 47(4): 458–472. [PubMed: 16161804]
18. Bantis LE, Nakas CT, Reiser B, Myall D, Dalrymple-Alford JC. Construction of joint confidence regions for the optimal true class fractions of receiver operating characteristic (ROC) surfaces and manifolds. Stat Methods Med Res. 2015; 26(3): 1429–1442. [PubMed: 25911331]

19. Schisterman EF, Perkins N. Confidence Intervals for the Youden index and corresponding optimal cut-point. Commun Stat Simul Comput. 2007; 36(3): 549–563.

20. Silverman BW. Density estimation for statistics and data analysis. London: Chapman Hall/CRC; 1998.

21. Bantis LE, Nakas CT, Reiser B. Construction of confidence intervals for the maximum of the Youden index and the corresponding cutoff point of a continuous biomarker. Biom J 2018; 61(1): 138–156. [PubMed: 30408224]

22. Hanley JA. The robustness of the 'binormal' assumptions used in fitting ROC curves. Med Decis Making. 1988; 8(3): 197–203. [PubMed: 3398748]

23. Box GEP, Cox DR. An analysis of transformations. J R Stat Soc Series B Stat Methodol. 1964; 26(2); 211–243.

24. Zou KH, Hall WJ. (2000). Two transformation models for estimating an ROC curve from continuous data. J Appl Stat. 2000; 27(5); 621–631.

25. Hernandez F, Johnson RA. The large-sample behavior of transformations to normality. J Am Stat Assoc. 1980; 75(372): 855.

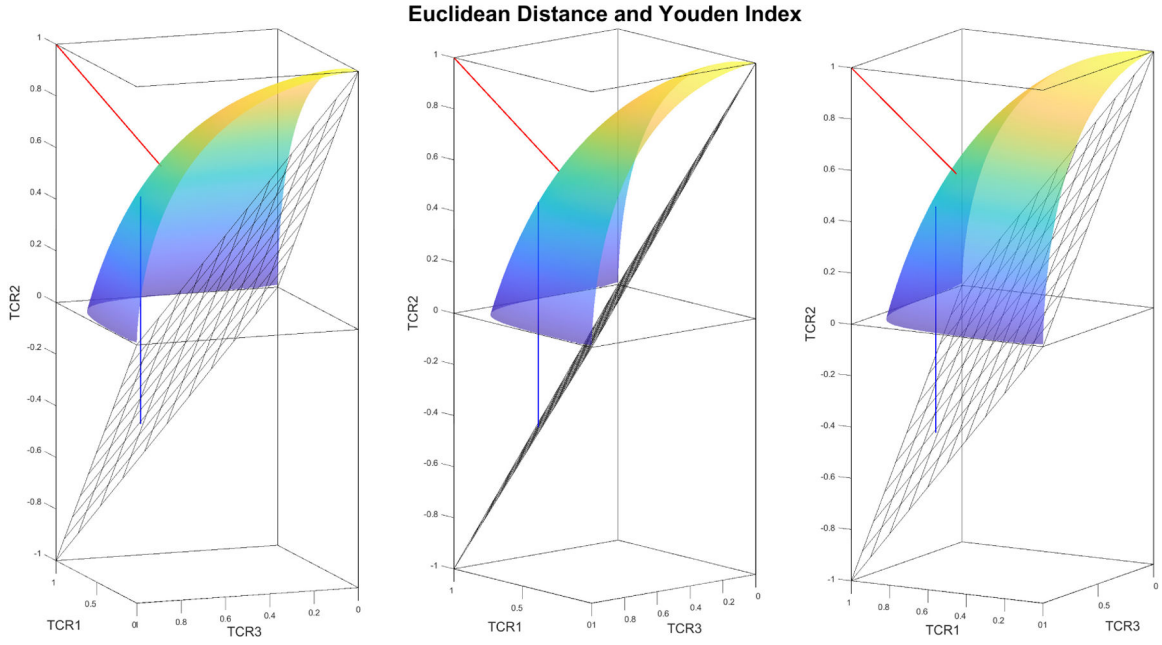

FIGURE 1.

Plots comparing the Euclidean distance and the Youden index of a biomarker score viewed from different perspectives for convenience. Euclidean distance is represented by the red line segment and Youden index is represented by the blue line segment. The useless biomarker plane is represented by the diagonal gridded plane.

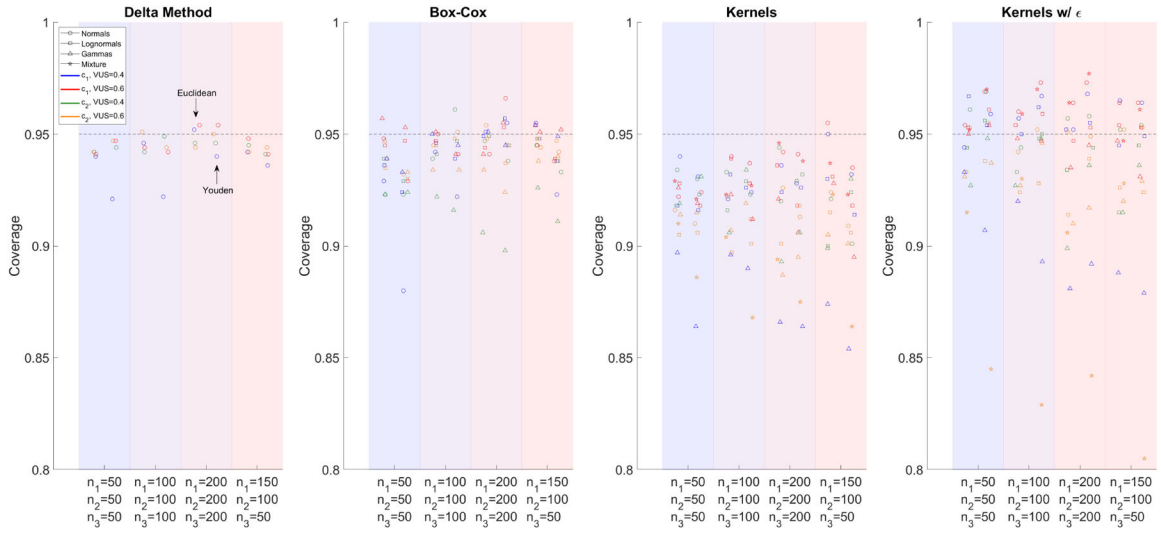

FIGURE 2.

Plot of coverage for c_1 and c_2 derived using the Euclidean method for all scenarios. The four plots, in order from left to right contain coverages for the following methods: **(1)** the delta method approach based on the assumption of normally distributed biomarker scores, **(2)** the Box-Cox approach, which assumes biomarker scores can be transformed to normality, **(3)** the naive kernel-based approach without the coverage correction of added noise for the bootstrap, and **(4)** the proposed kernel-based method with the coverage correction (denoted kernels w/ϵ). Each subplot contains four shaded regions, representing sample sizes of $(n_1,n_2,n_3) = (50,50,50)$, $(n_1,n_2,n_3) = (100,100,100)$, $(n_1,n_2,n_3) = (200,200,200)$, and $(n_1,n_2,n_3) = (150,100,50)$. Each shaded region contains two vertical bands of points, the left of which corresponds to the coverages of the Euclidean method, and the right of which corresponds to the Youden index, as illustrated in the leftmost subplot.

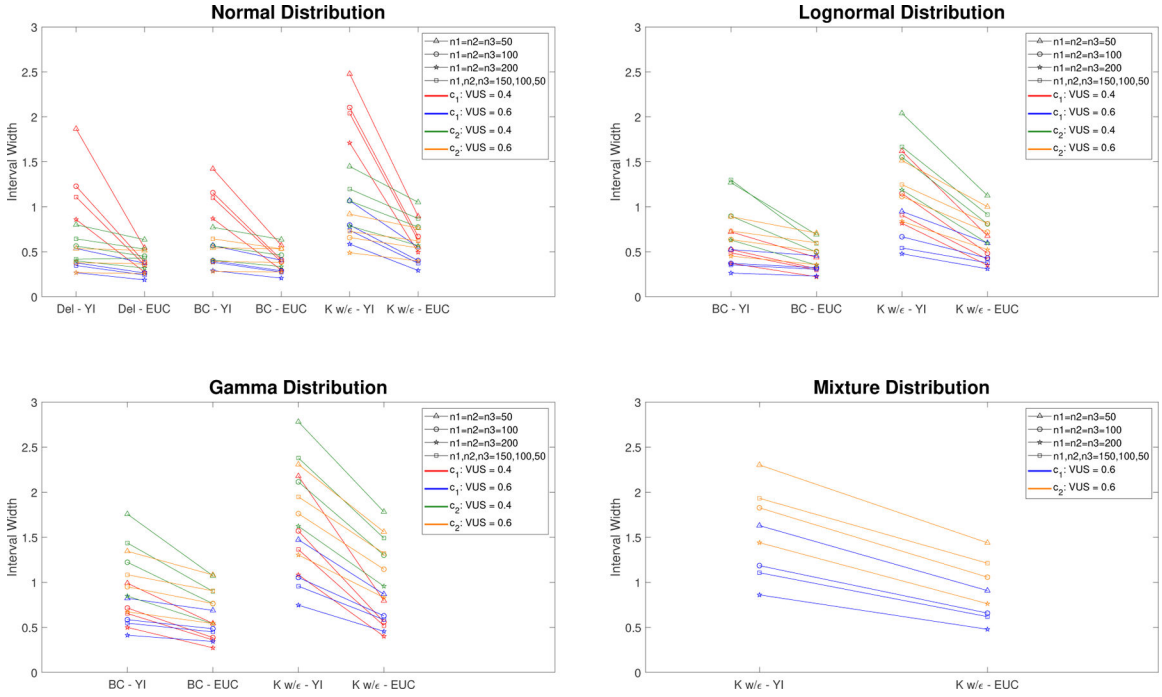

FIGURE 3.

Interval widths for biomarker scores for all simulated scenarios. Scenarios include data generated from normal, lognormal, gamma, and normal mixture distributions. Sample sizes of $(n_1,n_2,n_3) = (50,50,50)$, $(n_1,n_2,n_3) = (100,100,100)$, $(n_1,n_2,n_3) = (200,200,200)$, and $(n_1,n_2,n_3) = (150,100,50)$ are included, as well as scenarios corresponding to VUS = 0.4, and VUS = 0.6. We include interval widths for the delta method approach (Del), the Box-Cox approach (BC), and the proposed kernel-based approach with added noise that was described in Section 4 (K w/ ϵ).

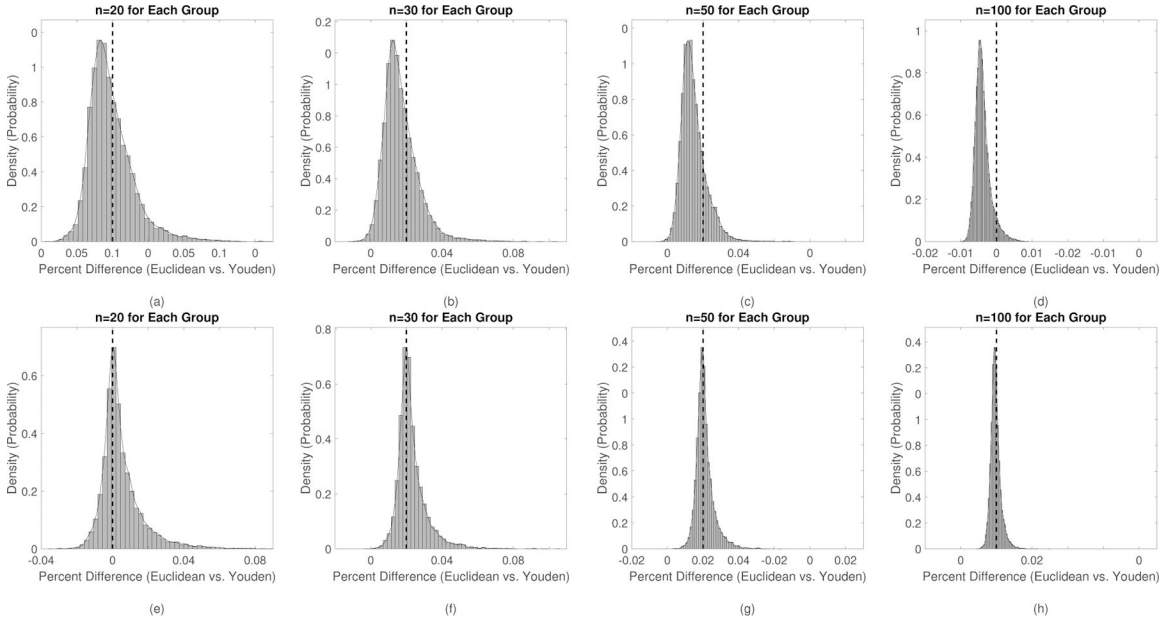

FIGURE 4.

The plots are histograms of the percent difference in total classification (sum of TCRs) for the Euclidean method and Youden index. Optimal cutoff values were estimated using the Euclidean and Youden methods based on training data. The estimated cutoff values were then used to derive true classification rates using the true distributions of the biomarker scores. The area to the right of 0 corresponds to a higher sum of true classification rates for the Euclidean method than for the Youden index. Plots (a)-(d) correspond to biomarkers simulated from normal distributions with equally spaced means of (μ_1,μ_2,μ_3) = (5,5.6592,6.3184), and standard deviations of $(\sigma_1,\sigma_2,\sigma_3)$ = (1,1,1). Plots (e)-(h) correspond to normal distributions with means of (μ_1,μ_2,μ_3) = (10,10.8,12), and standard deviations of $(\sigma_1,\sigma_2,\sigma_3)$ = (1,1,1.7)

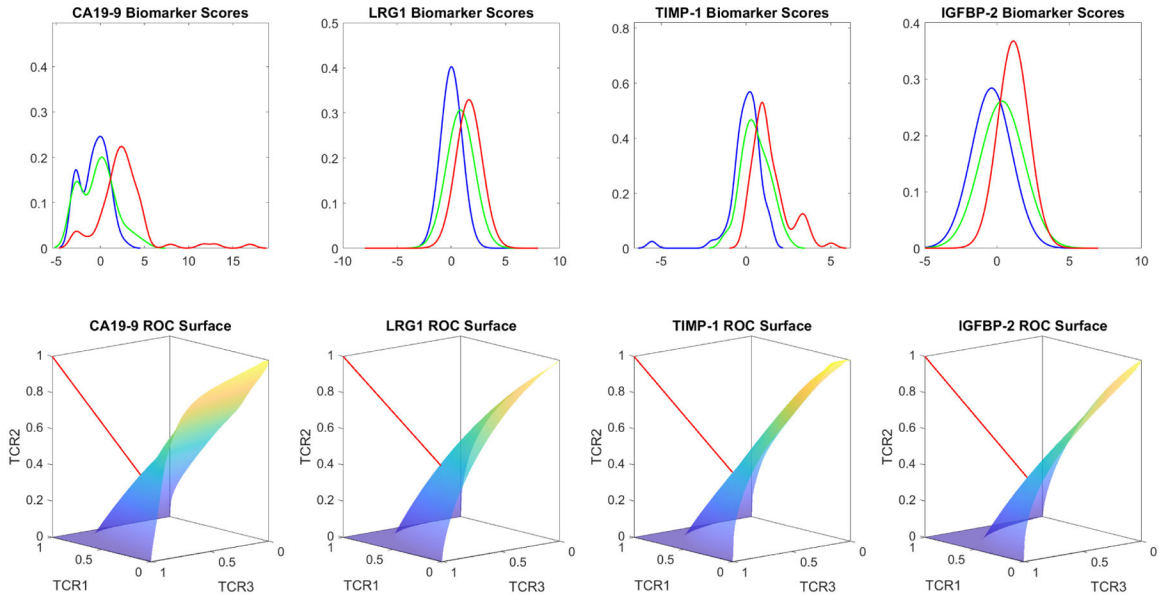

FIGURE 5.
Plots of biomarker scores and ROC surfaces for each of the biomarkers.

Table 1

Parameter values of the distributions used in settings with equal cutoffs.

	VUS	AUC_{12}	AUC_{23}	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	c_1	c_2
Normal Scenarios:	0.4	0.6106	0.7103	10.0000	1.3029	10.5000	1.2130	11.9020	2.2200	10.0396	11.8342
	0.6	0.7658	0.8155	11.0000	1.0400	12.0000	0.90584	14.0000	2.0334	11.4066	13.2051
	VUS	AUC_{12}	AUC_{23}	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	c_1	c_2
Lognormal Scenarios:	0.4	0.6775	0.6670	0.9000	0.4500	1.1617	0.3465	1.4000	0.4299	2.4869	4.0126
	0.6	0.8035	0.7839	0.9000	0.46004	1.3645	0.2900	1.7400	0.3800	2.9059	4.9641
	VUS	AUC_{12}	AUC_{23}	α_1	β_1	α_2	β_2	α_3	β_3	c_1	c_2
Gamma Scenarios:	0.4	0.6639	0.6900	1.1000	1.5000	2.0000	1.2149	1.6500	2.7013	1.2288	3.4954
	0.6	0.7914	0.7837	1.7800	1.3442	5.4200	0.7811	4.1800	1.7604	2.5510	5.6888
	VUS	AUC_{12}	AUC_{23}	Group 1: Normal		Group 2: Gamma		Group 3: Mixture Normal		c_1	c_2
Mixture Scenarios:	0.6	0.7909	0.7372	$\mu =$ 1.8000	$\sigma =$ 1.6557	$\alpha = 6.0$	$\beta = 0.6$	$\mu_a =$ 4.1903	$\sigma_a = 1.5$	2.1928	4.4929
								$\mu_b =$ 9.1903	$\sigma_b = 1.5$		

Table 2

Parameter values of the distributions used in settings with unequal cutoffs.

	VUS	AUC_{12}	AUC_{23}	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	**Method**	c_1	c_2
Normal Scenarios:	0.4	0.6794	0.6794	5.0000	1.0000	5.6592	1.0000	6.3184	1.0000	Euclidean	5.0730	6.2454
										Youden	5.3296	5.9888
	0.6	0.7955	0.7955	5.0000	1.0000	6.1675	1.0000	7.3350	1.0000	Euclidean	5.3899	6.9451
										Youden	5.5838	6.7512
	VUS	AUC_{12}	AUC_{23}	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	Method	c_1	c_2
Log-Normal Scenarios:	0.4	0.6865	0.6633	0.9000	0.4700	1.2000	0.4000	1.4475	0.4299	Euclidean	2.5075	4.1685
										Youden	2.6274	3.9302
	0.6	0.7442	0.8445	0.9000	0.4700	1.3050	0.4000	1.9000	0.42991	Euclidean	2.7926	5.4215
										Youden	2.8464	5.0157
	VUS	AUC_{12}	AUC_{23}	α_1	β_1	α_2	β_2	α_3	β_3	Method	c_1	c_2
Gamma Scenarios:	0.4	0.6286	0.7146	1.2000	1.5060	2.0000	1.2000	1.7783	2.6000	Euclidean	1.2954	3.5600
										Youden	1.2328	3.4081
	0.6	0.6763	0.9163	1.2000	1.2860	2.0000	1.2000	3.0000	2.8000	Euclidean	1.3850	4.8639
										Youden	1.3193	4.1753

Table 3

Summary statistics for the true classification rates from the Euclidean method and Youden index based on the estimated cutoff values from both methods in conjunction with the true distributions of the biomarker scores. This allows us to determine the true performance of the cutoff values from both methods to see how they would perform in the population. Our testing strategy is equivalent to using infinite sample sizes for both groups of the testing set. Included are the mean TCRs for each group for both methods, the mean sum of TCRs (Sum(TCRs)), the probability of the method having the higher sum of TCRs ($P(> Sum)$), the minimum sum of TCRs (Min(Sum)), and the maximum sum of TCRs (Max(Sum)). By construction, in training data, the Youden index will achieve higher total classification (sum of TCRs) than the Euclidean method.

Parameters	n_1, n_2, n_3	Method	TCR_1	TCR_2	TCR_3	Sum($TCRs$)	$P(> Sum)$	$Min(Sum)$	$Max(Sum)$
$X_1 \sim N(5,1^2)$ $X_2 \sim N(5.6592,1^2)$ $X_3 \sim N(6.3184,1^2)$	20, 20, 20	Euclidean	0.5252	0.4421	0.5266	1.4939	0.3913	1.4115	1.5150
		Youden	0.5984	0.2965	0.6003	1.4952	0.6087	1.2996	1.5166
	30, 30, 30	Euclidean	0.5270	0.4416	0.5275	1.4962	0.2991	1.4494	1.5140
		Youden	0.6087	0.2822	0.6110	1.5010	0.7009	1.3583	1.5166
	50, 50, 50	Euclidean	0.5288	0.4404	0.5288	1.4980	0.1826	1.4653	1.5132
		Youden	0.6206	0.2654	0.6202	1.5061	0.8174	1.4143	1.5166
	100, 100, 100	Euclidean	0.5296	0.4410	0.5288	1.4993	0.0610	1.4825	1.5106
		Youden	0.6277	0.2569	0.6264	1.5110	0.9390	1.4707	1.5166
$X_1 \sim N(10,1^2)$ $X_2 \sim N(10.8,1^2)$ $X_3 \sim N(12,1.7^2)$	20, 20, 20	Euclidean	0.5993	0.5424	0.5456	1.6874	0.6369	1.5239	1.6992
		Youden	0.6480	0.4878	0.5435	1.6793	0.3631	1.4411	1.6992
	30, 30, 30	Euclidean	0.6001	0.5452	0.5450	1.6903	0.5966	1.5775	1.6992
		Youden	0.6507	0.4927	0.5421	1.6854	0.4034	1.5190	1.6992
	50, 50, 50	Euclidean	0.6006	0.5476	0.5443	1.6926	0.5117	1.6164	1.6992
		Youden	0.6533	0.4965	0.5410	1.6909	0.4883	1.5962	1.6992
	100, 100, 100	Euclidean	0.6010	0.5502	0.5432	1.6944	0.3659	1.6652	1.6992
		Youden	0.6548	0.5010	0.5392	1.6951	0.6341	1.6494	1.6992

Table 4

Estimated D^*, Youden index, and VUS for each of the biomarkers. Also included are the corresponding confidence intervals.

Biomarker	D*	Youden index	VUS
CA19–9	0.8639 (0.7629, 0.9648)	0.6248 (0.5054, 0.7442)	0.3852 (0.3036, 0.4669)
LRG1	0.8763 (0.7938, 0.9587)	0.5490 (0.4406, 0.6575)	0.3964 (0.2995, 0.4933)
TIMP-1	0.9006 (0.8092, 0.9921)	0.5447 (0.4299, 0.6595)	0.4018 (0.3309, 0.4752)
IGFBP-2	0.9375 (0.8564, 1.0185)	0.4642 (0.3519, 0.5764)	0.3397 (0.2534, 0.4260)

Table 5

Point estimates, confidence intervals, and widths for each of the biomarkers using proposed and Youden methods.

		c_1		c_2	
Biomarker	**Method**	**Confidence Inverval**	**Width**	**Confidence Inverval**	**Width**
CA19–9	Euclidean	−0.4634 (−1.0459, 0.1191)	1.1650	1.6731 (1.3285, 2.0177)	0.6891
	Youden	1.1546 (−0.8335, 3.1428)	3.9763	1.1546 (0.7431, 1.5662)	0.8231
LRG1	Euclidean	0.2073 (0.0088, 0.4057)	0.3969	1.5638 (1.3540, 1.7736)	0.4195
	Youden	0.7348 (0.3975, 1.0721)	0.6746	1.1033 (0.6237, 1.5828)	0.9591
TIMP-1	Euclidean	0.1129 (−0.1390, 0.3649)	0.5039	0.9936 (0.6718, 1.3155)	0.6436
	Youden	0.6375 (0.0604, 1.2146)	1.1542	0.6375 (0.2732, 1.0018)	0.7286
IFGBP-2	Euclidean	−0.4507 (−0.7250, −0.1765)	0.5485	0.9450 (0.7280, 1.1621)	0.4341
	Youden	0.2133 (−0.5030, 0.9296)	1.4327	0.2261 (−0.1480, 0.6002)	0.7482

Table 6

True-classification rates for each of the biomarkers using the proposed and Youden methods of choosing cutoffs.

Biomarker	Method	TCR_1	TCR_2	TCR_3
CA19–9	Euclidean	0.5220	0.3703	0.6518
	Youden	0.8800	0.0000	0.7432
LRG1	Euclidean	0.5828	0.3977	0.5194
	Youden	0.7707	0.1124	0.6659
TIMP-1	Euclidean	0.5023	0.3776	0.5804
	Youden	0.7847	0.0000	0.7600
IGFBP-2	Euclidean	0.4797	0.3513	0.5673
	Youden	0.6635	0.0033	0.7974