Performance of Diagnostic Tests Based on Continuous Bivariate Markers

[1]Hani Samawi, [2]Ding-Geng Chen, [1]Jingjing Yin, [1]Marwan Alsharman

[1]Department of Biostatistics, Epidemiology

and Environmental Health Sciences

Jiann-Ping Hsu College of Public Health

Georgia Southern University, Statesboro, GA 30460, USA

[2]College of Health Solutions, Arizona State University, Phoenix, AZ 85040, USA

Department of Statistics, University of Pretoria, Pretoria, South Africa.

hsamawi@georgiasouthern.edu; dinchen@asu.edu; jyin@georgiasouthern.edu;

ma15387@georgiasouthern.edu

Abstract

In medical diagnostic research, it is customary to collect multiple continuous biomarker measures to improve the accuracy of diagnostic tests. A prevalent practice is to combine the measurements of these biomarkers into one single composite score. However, incorporating those biomarker measurements into a single score depends on the combination of methods and may lose vital information needed to make an effective and accurate decision. Furthermore, a diagnostic cut-off is required for such a combined score, and it is difficult to interpret in actual clinical practice. The paper extends the classical biomarkers' accuracy and predictive values from univariate to bivariate markers. Also, we will develop a novel pseudo-measures system to

maximize the vital information from multiple biomarkers. We specified these pseudo-and-or classifiers for the true positive rate, true negative rate, false-positive rate, and false-negative rate. We used them to redefine classical measures such as the Youden index, diagnostics odds ratio, likelihood ratios, and predictive values. We provide optimal cut-off point selection based on the modified Youden index with numerical illustrations and real data analysis for this paper's newly developed pseudo measures.

Keywords: Predictive values; bivariate analysis; likelihood ratios; Youden index; odds ratio

1. Introduction

For a clinician to provide reliable information about a patient's health condition and suggest a treatment plan for a patient, diagnostic tests (including those for screening) play a vital role in health care (Sox, Jr. et al., 1989; McNeil & Adelsten, 1976; Zhou et al., 2009). A good diagnostic test is essential to discriminate between diseased and non-diseased subjects. A biomarker will be dichotomized with a specific cut-off point to distinguish between diseased and non-diseased subjects. In this case, the diagnostic test and biomarker can be used interchangeably.

In binary two-stages disease (non-diseased and diseased), the most frequently used by clinicians, measures of diagnostic accuracy are true positive rate (TPR or sensitivity), true negative rate (TNR or specificity), false-positive rate (FPR), and false-negative rate (FNR), see for example Pepe (2003) and Zhou et al. (2009).

The Receiver Operating Characteristic (ROC) curve provides a graphical interpretation of TPR and FPR pairs over all possible cut-off points. On the other hand, other measures in the literature link some or all TPR, TNR, FPR, and FNR as a single index to summarize accuracy.

Among those are the area under the ROC curve (AUC), the Youden index, diagnostic odds ratio (OR), the overlap measure, and the Kullback–Leibler divergence measure (KL) (see, Samawi et al., 2017; Samawi et al., 2020; Yin and Tian 2014a).

Moreover, post-test accuracy measures such as positive predictive value (PPV, which is the likelihood of having the disease of interest in a subject given a positive test result), negative predictive value (NPV, which is the probability that a subject receives a negative result yet does not have the disease of interest), and diagnostic likelihood ratios (LRs) are the measures that provide information about the probability and the odds of disease given the diagnosis. They offer significant clinical implications for a diagnostic test (Altman & Bland, 1994; Samawi et al., 2021). The predictive values highly depend on the disease prevalence, which cannot be generalized among different populations with different disease prevalence. However, although the LRs can also provide information about the probability that a subject can be correctly diagnosed, they do not depend on the prevalence of the same disease; for example, see Boyko (1994) and Deeks & Altman (2004). Furthermore, LRs can be used as indicators for ruling-in/outpatients in a clinical setting and KL divergence (Gilbert et al., 2001; Boyko, 1994; Deeks & Altman, 2004; Lee, 1999; Samawi et al., 2020).

Recently, the medical community accepted the intuitive fact that diagnosis based on one single biomarker might not provide sufficient accuracy in decision-making (Sidransky, 2002; Kumar et al., 2006). Consequently, multiple biomarker tests are expected to be performed on everyone. The corresponding measurements are combined into a summary score to help clinicians make better diagnostic judgments (Yin and Tian 2014b). However, incorporating those biomarkers' measures into one score depends on the methods used to combine them and may lose vital information needed to make an effective and accurate decision.

Moreover, Bansal and Pepe (2013) pointed out that when a standard marker is not sufficiently accurate in classifying diseased and non-diseased subjects on its own, there is a need for new markers to achieve a combination of the markers to reach better performance. They found out that the vital principle for choosing new biomarkers is that they have decent performance on their own, and they preferred to be uncorrelated with the standard biomarker. In the literature, it is recommended to use linear combinations. Bansal and Pepe (2013) investigated the increase in performance when combining a new continuous marker with a moderately performing standard continuous marker under different biologically driven models for their joint distribution. Bansal and Pepe (2013) found that when a continuous marker is uncorrelated, a marker with moderate performance on its own most likely produces very minimal improvement in diagnostics performance. Moreover, they found other combinations of settings that may lead to better improvements in medical diagnostics (see Bansal and Pepe, 2013).

Furthermore, there is little literature on using bivariate biomarkers to improve the performance of medical diagnostics accuracy. Wang and Li (2012) extend the ROC function from univariate marker to bivariate marker case. They introduced a weighted ROC (WROC) function and proposed the area for bivariate biomarkers under the WROC (AUC). They indicated that since some biomarkers are identified from various biological sources, combining them using a linear combination may not be appropriate. Therefore, instead of combining markers, they propose to evaluate two markers in a bivariate setting and explore their respective roles and interaction. For analyzing biomarkers in bivariate settings, Baker (2000) and Etzioni et al. (2003) proposed new definitions for the ROC curve subject to the “and-or” classifiers. They divided the multidimensional marker space into small intervals and defined the true and false positive rates

based on the subsets of the divided region. The ROC curve was then defined as the objective function of finding the optimal true positive rate for each given false positive rate.

Instead of combining markers, this paper proposes to evaluate two markers in a bivariate setting and explore their respective roles, interaction, and advantages in medical diagnostics accuracy measures. Similarly, we divided the Bidimensional marker space into small intervals and defined the true and false positive and negative rates based on the subsets of the divided region. We determined the pseudo classifiers for the true positive rate (TPR), true negative rate (TNR), false-positive rate (FPR), and false-negative rate (FNR) based on those small intervals. We then use these newly developed pseudo-classification measures to redefine existing measures such as the Youden index, odds ratio, likelihood ratios, and predictive values. We will provide optimal cut-off point selection based on the modified pseudo-Youden index.

The paper unfolds as follows. Section 2 provides some preliminaries and motivations. Section 3 proposes the novel pseudo classifiers and the pseudo diagnostics accuracy measure, and the optimal cut-off points selection for the bivariate diagnostics tests or biomarkers. Derivations of the empirical estimates of the proposed measures with their variances are provided in Section 4. Numerical examples are presented in Section 5, with real data illustrations in Section 6. Final remarks and discussion are provided in Section 7.

2. Preliminaries: common diagnostic performance measures

In general, for clinical decision-making with continuous biomarkers, we are required to obtain a diagnostic cut-off point, *c*, to classify a subject either as diseased or non-diseased. Let X_0 and X_1 denote the marker values for non-diseased and diseased subjects, with cumulative distribution function (c.d.f.), $F_0(.)$ and $F_1(.)$ respectively. Higher marker values indicate greater

disease severity in most circumstances without the loss of generality. This assumption of directionality is essential for the Receiver Operating Characteristic curve (ROC) analysis to guarantee the validity of the ROC indices. The ROC curve is a graph of true positive rate (*TPR* or sensitivity), which is ($Se(c) = P(X_1 > c) = 1 - F_1(c)$), versus false-positive rate or 1- specificity as a measure of diagnostics accuracy ($FPR = 1 - Sp(c) = 1 - F_0(c)$), for a given cut point.

The classification matrix $\mathbf{P}$ in the binary disease setting, given $F_0(.)$ and $F_1(.)$ with threshold *c*, can be expressed as

$$\mathbf{P} = \begin{matrix} T=0 & T=1 \\ \end{matrix} \begin{matrix} \\ \end{matrix} \begin{matrix} T=0 & T=1 \\ \end{matrix} \\ \begin{bmatrix} Sp(c) & 1-Sp(c) \\ 1-Se(c) & Se(c) \end{bmatrix} = \begin{bmatrix} TNR & FPR \\ FNR & TPR \end{bmatrix} \begin{matrix} D=0 \\ D=1 \end{matrix}, \quad (1)$$

where $T = 0$ and $T = 1$, respectively, are the negative and positive test results, $D = 0$ and $D = 1$ are the disease stage, implying non-disease and disease, respectively.

In practice, it is common to summarize the ROC curve into a single global value or index, such as the area under the curve (AUC). The AUC evaluates the discriminatory ability of a marker, where $ROC(q) = 1 - F_1[F_0^{-1}(1-q)]$ and $q = 1 - Sp(c)$. The AUC ranges [0.5, 1] and can be interpreted as the probability of the measurement from a random disease subject being greater than that from a random non-diseased subject. The AUC is calculated by the integral as the following equation:

$$AUC = P(X_1 > X_0) = \int_{-\infty}^{\infty} f_0(x)[1 - F_1(x)]dx. \quad (2)$$

Moreover, Li and Fine (2010), suggested the weighted area under the receiver operating characteristic curve and use it for gene selection.

Another frequently used measure that summarizes the sensitivity and specificity is the Youden index (J), which ranges [0, 1]. Also, J is commonly used as a criterion for optimal diagnostic cut-off point selection. For a given cut-off point c, the Youden index is given by

$$J = \underset{c}{Max}(TPR(c) + TNR(c) - 1) = \underset{c}{Max}(Se(c) + Sp(c) - 1) = \underset{c}{Max}(F_0(c) - F_1(c)). \quad (3)$$

Also, as a criterion for optimal diagnostic cut-off points (c) selection, using J, c is obtained as follows:

$$c = \arg \underset{c}{Max}(F_0(c) - F_1(c)). \quad (4)$$

However, recently a few popular methods were reviewed for cut-off selection (see Sande et al., 2021).

For a given cut-off point c, the diagnostic odds ratio (DOR) (Huang et al. 2018) for a given c is given by

$$DOR = \frac{Se(c)Sp(c)}{[1 - Se(c)][1 - Sp(c)]} = \frac{F_0(c)[1 - F_1(c)]}{F_1(c)[1 - F_0(c)]}, \quad (5)$$

where the range is from 0 to infinity. Higher values of DOR indicate higher discriminative power in diagnostic tests. A test is unsuitable when the value DOR is less than 1. It means more negative tests among the diseased population. A value of 1 indicates that a test cannot discriminate between patients with the disease and those without the disease (Glas et al., 2003).

For a given cut-off point c, the after-test performance measures of diagnostic tests, PPV and NPV are defined by:

$$PPV = \frac{Se(c)p}{Se(c)p + (1 - p)(1 - Sp(c))}, \quad (6)$$

and

$$NPV = \frac{Sp(c)(1-p)}{Sp(c)(1-p)+p(1-Se(c))}, \tag{7}$$

where $p = P(D=1)$, is the prevalence of the disease. Also, the inference procedures for PPV and NPV are presented by Li et al. (2007), and the use of NPV in personalized medicine as well to select patients with positive (or negative) is noted by Li et al. (2012).

Finally, for a given cut-off point c, the LRs (the positive likelihood ratio (LR_+) and the negative likelihood ratio (LR_-) are other after-test performance measures given by

$$LR_+ = \frac{sensitivity}{1- specificity} = \frac{TP}{FP} = \frac{Se(c)}{1-Sp(c)} \tag{8}$$

and

$$LR_- = \frac{1- sensitivity}{specificity} = \frac{FN}{TN} = \frac{1-Se(c)}{Sp(c)}. \tag{9}$$

3. Pseudo accuracy measures and the optimal cut-off criterion for bivariate diagnostics tests

With the preliminaries in Section 2, this section is aimed to consider situations when a correlated pair of continuous marker variables (X, Y) are available to classify the disease state with any logical classifier. Then based on both markers, we will extend the diagnostic accuracy and prediction measures to bivariate marker settings with our newly developed pseudo measures. Consequently, we may use any logical classifier based on some biological facts using the and-or system similar to that described by Etzioni et al. (2003). For example, Etzioni et al. (2003) used the and-or-logical system. They applied it to prostate cancer screening to define the true positive and the false-positive rates for constructing the ROC and calculated the AUC. They used some logical and biological facts about prostate cancer screening with Prostate-Specific Antigen (PSA). They indicated that high PSA levels are more likely to be associated with prostate cancer; however,

other benign conditions may also cause PSA elevation. Therefore, using (PSA > 4.0 ng ml−1) criterion for a positive result may yield a non-trivial number of false positives, which may then require unnecessary biopsies. However, Stenman et al. (1991) explained that PSA consists of two different subtypes, free and complex PSA, and while their sum tends to rise in the presence of a malignancy, the proportion of free PSA tends to decline. Therefore, it might be helpful to combine the free-to-total PSA (RPSA) ratio with the total PSA level (TPSA), improving prostate cancer screening.

As discussed by Gann et al. (2002) and Etzioni et al. (2003), a substantial number of the literature indicated a potential gain from using RPSA in combination with TPSA. Etzioni et al. (2003) used the combination of RPSA and TPSA to form a valid diagnostic test for prostate cancer where the diagnosis is positive if (TPSA > c) or ($d <$ TPSA $\leq c$ and RPSA < t), and c, d, t are determined cut-off points based on some optimization criteria. Therefore, by applying two correlated biomarkers through an and-or system, we can reduce false positives and false negatives and thus improve diagnostic accuracy.

To derive our proposed accuracy measures, let (X_0, Y_0) and (X_1, Y_1) denote the bivariate marker values for non-diseased and diseased subjects, with bivariate density functions, $f_0(x, y)$ and $f_1(x, y)$ and distribution functions $F_0(x, y)$ and $F_1(x, y)$ respectively. We assume that higher marker values indicate greater disease severity in most circumstances without the loss of generality.

Using the below basic bivariate calculus for given cut-off points, we have

$$\begin{aligned}1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_0(x,y)dxdy &= \int_{c_2}^{\infty}\int_{c_1}^{\infty} f_0(x,y)dxdy+\int_{-\infty}^{c_2}\int_{c_1}^{c_4} f_0(x,y)dxdy\\ &+\int_{-\infty}^{c_2}\int_{c_4}^{\infty} f_0(x,y)dxdy+\int_{-\infty}^{c_1}\int_{c_2}^{c_6} f_0(x,y)dydx\\ &+\int_{-\infty}^{c_1}\int_{c_6}^{\infty} f_0(x,y)dydx+\int_{-\infty}^{c_2}\int_{-\infty}^{c_1} f_0(x,y)dxdy,\end{aligned} \tag{10}$$

and

$$\begin{aligned}1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(x,y)dxdy &= \int_{c_2}^{\infty}\int_{c_1}^{\infty} f_1(x,y)dxdy+\int_{c_1}^{\infty}\int_{c_5}^{c_2} f_1(x,y)dydx\\ &+\int_{c_1}^{\infty}\int_{-\infty}^{c_5} f_1(x,y)dydxy+\int_{c_2}^{\infty}\int_{c_3}^{c_1} f_1(x,y)dxdy\\ &+\int_{c_2}^{\infty}\int_{-\infty}^{c_3} f_1(x,y)dxdy+\int_{-\infty}^{c_2}\int_{-\infty}^{c_1} f_1(x,y)dxdy\end{aligned} \tag{11}$$

where $(c_1, c_2, c_3, c_4, c_5, c_6 :\ c_3 < c_1 < c_4; c_5 < c_2 < c_6)$.

From (10) and (11), we can redefine the accuracy measures (TPR, TNR, FPR, FNR) by what we call pseudo accuracy measurers as follow:

$$\begin{aligned}TPR^* &= P(X_1 > c_1, Y_1 > c_2) + P(X_1 > c_1, c_5 \le Y_1 \le c_2) + P(c_3 \le X_1 \le c_1, Y_1 > c_2),\\ TNR^* &= P(X_0 \le c_1, Y_0 \le c_2) + P(c_1 \le X_0 \le c_4, Y_0 \le c_2) + P(X_0 \le c_1, c_2 \le Y_0 \le c_6),\\ FPR^* &= P(X_0 > c_1, Y_0 > c_2) + P(X_0 > c_4, Y_0 \le c_2) + P(X_0 \le c_1, Y_0 > c_6),\\ FNR^* &= P(X_1 \le c_1, Y_1 \le c_2) + P(X_1 > c_1, Y_1 \le c_5) + P(X_1 \le c_3, Y_1 > c_2),\\ FPR^* &= 1 - TNR^*,\\ FNR^* &= 1 - TPR^*.\end{aligned} \tag{12}$$

Therefore, as we explained above, using the PSA test, for example, TPR^* is the true positive rate denoted as $\{(X_1 > c_1)$ and $(Y_1 > c_2)\}$ or $\{(X_1 > c_1)$ and $(c_5 \le Y_1 \le c_2)\}$ or when $\{(Y_1 > c_2)$, while $(c_3 \le X_1 \le c_1)\}$. Similarly, we can interpret the other rates.

Clearly, from (10) and (11), $TPR^* + FNR^* = 1$, and $TNR^* + FPR^* = 1$. Therefore, the maximum value that TPR^* or TNR^* could reach 1 where the minimum value is 0, and hence, the pseudo Youden index has a range of [0, 1], and it is defined as

$$J^* = \underset{c_1,c_2,c_3,c_4,c_5,c_6}{Max} (TPR^*(c_1,c_2,c_3,c_4,c_5,c_6) + TNR^*(c_1,c_2,c_3,c_4,c_5,c_6) - 1), \quad (13)$$

where

$$(c_1,c_2,c_3,c_4,c_5,c_6) = \arg \underset{c_1,c_2,c_3,c_4,c_5,c_6}{Max} (TPR^* + TNR^* - 1). \quad (14)$$

Consequently, the pseudo Youden index (J^*) can be treated as a measure that summarizes the pseudo correct classification rates in (12) and used it as a criterion for selecting the optimal diagnostic's cut-off points $(c_1,c_2,c_3,c_4,c_5,c_6)$, among other methods, given the constraints ($c_3 < c_1 < c_4; c_5 < c_2 < c_6$), using any available optimal rules with constraints. Note that when using the optimization rules, one needs to restrict the lower range of c_3 and c_5 and the upper range of c_4 and c_6 to the empirical range and the biological range of the biomarker if possible. Finding the diagnostic's cut-off points $(c_1,c_2,c_3,c_4,c_5,c_6)$ is challenging, therefore we propose using the following approach: A simple and logical approach to finding the optimal cut-points is first to find the optimal cut-point for each biomarker separately (c_1, c_2), (using any appropriate criterion such as the Youden index). Then fix c_1, c_2 and define (a,b,c_3,c_4,c_5,c_6) as follows: $c_3 = c_1 - a, c_4 = c_1 + a; c_5 = c_2 - b, c_6 = c_2 + b$, where a and b are increments parameters to control for the areas for false negative and false positive. Finally use optimal arguments $(a,b) = \arg \underset{a,b}{Max}(TPR^*(a,b) + TNR^*(a,b) - 1)$ to find the optimal cut-points (a, b) under the restrictions $c_3 < c_1 < c_4; c_5 < c_2 < c_6$.

For the pseudo-PPV and NPV, which are the after-test performance measures of diagnostic tests based on bivariate biomarkers, we propose the following:

$$PPV^* = \frac{p(TPR^*)}{p(TPR^*) + (1-p)(FPR^*)}, \tag{15}$$

and

$$NPV^* = \frac{(1-p)(TNR^*)}{(1-p)(TNR^*) + p(FNR^*)}, \tag{16}$$

where, $p = P(D = 1)$, is the prevalence of the disease.

Like the PPV^* and NPV^*, we propose the pseudo-likelihood ratios (LRs) as after-test performance measures as follows: The pseudo-likelihood ratio positive as

$$LR_+^* = \frac{TPR^*}{FPR^*} \tag{17}$$

and the pseudo-likelihood ratio negative as

$$LR_-^* = \frac{FNR^*}{TNR^*}. \tag{18}$$

Finally, the pseudo-odds-ratio that measures the diagnostic test accuracy at the threshold (c_1, c_2) can be defined as

$$DOR^* = \frac{LR_+^*}{LR_-^*} = \frac{(TPR^*)(TNR^*)}{(FPR^*)(FNR^*)}, \tag{19}$$

Similarly, the range DOR^* is from 0 to infinity, with higher values indicating higher discriminative power in diagnostic tests.

4. Estimation using the empirical distributions

The next logical step is to estimate these measures from observed data with the newly developed pseudo measures. This section presents a method of estimation based on the empirical distributions for the proposed pseudo-diagnostic measures in bivariate biomarkers.

Let $(X_{0,1},Y_{0,1}),(X_{0,2},Y_{0,2}),...,(X_{0,n},Y_{0,n_1})$ and $(X_{1,1},Y_{1,1}),(X_{1,2},Y_{1,2}),...,(X_{1,n_2},Y_{1,n_2})$ denote the two independent random samples from the bivariate markers' values for non-diseased and diseased subjects, with c.d.f. $F_0(x,y)$ and $F_1(x,y)$ respectively. Let the estimated cut-off points $(\hat{c}_1,\hat{c}_2,\hat{c}_3,\hat{c}_4,\hat{c}_5,\hat{c}_6:\ \hat{c}_3<\hat{c}_1\le\hat{c}_4;\hat{c}_5<\hat{c}_2\le\hat{c}_6)$ be determined by, for example, the pseudo-Youden index; we estimate first the classification rates as follows:

$$\begin{aligned}
\hat{TPR}^* &= \frac{1}{n_2}\sum_{i=1}^{n_2}[I(X_{1i}>\hat{c}_1,Y_{1i}>\hat{c}_2)+I(X_{1i}>\hat{c}_1,\hat{c}_5\le Y_{1i}\le\hat{c}_2)+I(\hat{c}_3\le X_{1i}\le\hat{c}_1,Y_{1i}>\hat{c}_2)],\\
\hat{TNR}^* &= \frac{1}{n_1}\sum_{i=1}^{n_1}[I(X_{0i}\le\hat{c}_1,Y_{0i}\le\hat{c}_2)+I(\hat{c}_1\le X_{0i}\le\hat{c}_4,Y_{0i}\le\hat{c}_2)+I(X_{0i}\le\hat{c}_1,\hat{c}_2\le Y_{0i}\le\hat{c}_6)],\\
\hat{FPR}^* &= \frac{1}{n_1}\sum_{i=1}^{n_1}[I(X_{0i}>\hat{c}_1,Y_{0i}>\hat{c}_2)+I(X_{0i}>\hat{c}_4,Y_{0i}\le\hat{c}_2)+I(X_{0i}\le\hat{c}_1,Y_{0i}>\hat{c}_6)],\\
\hat{FNR}^* &= \frac{1}{n_2}\sum_{i=1}^{n_2}[I(X_{1i}\le\hat{c}_1,Y_{1i}\le\hat{c}_2)+I(X_{1i}>\hat{c}_1,Y_{1i}\le\hat{c}_5)+I(X_{1i}\le\hat{c}_3,Y_{1i}>\hat{c}_2)],
\end{aligned} \quad , \quad (20)$$

where

$$I_A(x)=\begin{cases}1 & \text{if } x\in A\\ 0 & \text{otherwise}\end{cases}.$$

Note that ($\hat{TPR}^*,\hat{TNR}$, $\hat{FPR}^*$, and $\hat{FNR}^*$) are all unbiased estimators for (TPR^*,TNR^*, FPR^*, and FNR^*) in (12). Also, by using the WLLN, we have ($\hat{TPR}^*,\hat{TNR}$, $\hat{FPR}^*$, and $\hat{FNR}^*$) are consistent estimators for (TPR^*,TNR^*, FPR^*, and FNR^*), provided that $\hat{c}_j \xrightarrow{p} c_j;\ j=1,2,...,6$ using Slutsky's Theorem. Also, the asymptotic normality distributions of those (

$\hat{TPR}^*$, $\hat{TNR}$, $\hat{FPR}^*$, and $\hat{FNR}^*$) estimators are straightforward for large samples using the Central Limit Theorem. Furthermore, the empirical distribution converges uniformly as shown by Glivenko-Cantelli Theorem. Thus, all the above pseudo diagnostics measures of accuracy and prediction of accuracy can be estimated directly by substituting the results in (20) in their formulas.

Furthermore, to find the variances of the proposed estimators, we notice the following cross-tabulations as presented in Tables 1 and 2.

Table 1. Cross classification for the disease sample of a bivariate marker

X\Y	$Y_1 \le \hat{c}_2$	$\hat{c}_5 < Y_1 \le \hat{c}_2$	$\hat{c}_2 < Y_1 \le \hat{c}_6$	$Y_1 > \hat{c}_2$
$X_1 \le \hat{c}_1$	n_{111}	n_{112} (overlap with n_{111})	n_{113} (overlap with n_{114})	n_{114}
$\hat{c}_3 < X_1 \le \hat{c}_1$	n_{121}	n_{122} (overlap with n_{121})	n_{123} (overlap with n_{124})	n_{124}
$\hat{c}_1 < X_1 \le \hat{c}_4$	n_{131}	n_{132} (overlap with n_{131})	n_{133} (overlap with n_{134})	n_{134}
$X_1 > \hat{c}_1$	n_{141}	n_{142} (overlap with n_{141})	n_{143} (overlap with n_{144})	n_{144}

Table 2. Cross classification for a non-disease sample of the bivariate marker

X\Y	$Y_0 \le \hat{c}_2$	$\hat{c}_5 < Y_0 \le \hat{c}_2$	$\hat{c}_2 < Y_0 \le \hat{c}_6$	$Y_0 > \hat{c}_2$
$X_0 \le \hat{c}_1$	n_{011}	n_{012} (overlap with n_{011})	n_{013} (overlap with n_{014})	n_{014}
$\hat{c}_3 < X_0 \le \hat{c}_1$	n_{021}	n_{022} (overlap with n_{021})	n_{023} (overlap with n_{024})	n_{024}
$\hat{c}_1 < X_0 \le \hat{c}_4$	n_{031}	n_{032} (overlap with n_{031})	n_{033} (overlap with n_{034})	n_{034}
$X_0 > \hat{c}_1$	n_{041}	n_{042} (overlap with n_{041})	n_{043} (overlap with n_{044})	n_{044}

From Table 1, it is clear that n_{114}, n_{142} and n_{124} are mutually independent since they do not overlap, then provided that $\hat{c}_j \xrightarrow{p} c_j$; $j = 1,2,...,6$, we can show that the asymptotic variance is given by

$$\begin{aligned} Var(\hat{TPR}^*) = \frac{1}{n_2^2}\{ & n_{144}P(X_1 > c_1, Y_1 > c_2)[1 - P(X_1 > c_1, Y_1 > c_2)], \\ & + n_{142}P(X_1 > c_1, c5 \le Y_1 \le c_2)[1 - P(X_1 > c_1, c_5 \le Y_1 \le c_2)], \\ & + n_{124}P(c_3 \le X_1 \le c_1, Y_1 > c_2)[1 - P(c_3 \le X_1 \le c_1, Y_1 > c_2)]\}. \end{aligned} \quad (21)$$

Similarly, we can obtain the variance for the other pseudo classifiers as follows:

$$Var(T\hat{N}R^*) = \frac{1}{n_1^2}\begin{Bmatrix} n_{011}P(X_0 \le c_1, Y_0 \le c_2)[1-P(X_0 \le c_1, Y_0 \le c_2)] \\ +n_{031}P(c_1 \le X_0 \le c_4, Y_0 \le c_2)[1-P(c_1 \le X_0 \le c_4, Y_0 \le c_2)] \\ +n_{013}P(X_0 \le c_1, c_2 \le Y_0 \le c_6)[1-P(X_0 \le c_1, c_2 \le Y_0 \le c_6)] \end{Bmatrix}, \quad (22)$$

$$Var(F\hat{P}R^*) = \frac{1}{n_1^2}\begin{Bmatrix} n_{044}P(X_0 > c_1, Y_0 > c_2)[1-P(X_0 > c_1, Y_0 > c_2)] \\ +(n_{041}-n_{031})P(X_0 > c_4, Y_0 \le c_2)[1-P(X_0 > c_4, Y_0 \le c_2)] \\ +(n_{014}-n_{013})P(X_0 \le c_1, Y_0 > c_6)[1-P(X_0 \le c_1, Y_0 > c_6)] \end{Bmatrix}, \quad (23)$$

and

$$Var(F\hat{N}R^*) = \frac{1}{n_2^2}\begin{Bmatrix} n_{111}P(X_1 \le c_1, Y_1 \le c_2)[1-P(X_1 \le c_1, Y_1 \le c_2) \\ +(n_{141}-n_{141})P(X_{1i} > c_1, Y_1 \le c_5)[1-P(X_1 > c_1, Y_1 \le c_5)] \\ +(n_{114}-n_{124})P(X_1 \le c_3, Y_1 > c_2)[1-P(X_1 \le c_3, Y_1 > c_2)] \end{Bmatrix}. \quad (24)$$

We can obtain consistent estimators of the above variances by substituting the data values from Table 1 and Table 2 and the estimates of the pseudo classification rates.

Moreover, for fixed cut-off points (c_1, c_2, a, b), the estimate of pseudo Youden index is given by $\hat{J}^* = (T\hat{P}R^*(c_1, c_2) + T\hat{N}R^*(c_1, c_2) - 1)$, and its variance is given by:

$$Var(\hat{J}^*) = Var(T\hat{P}R^*) + Var(T\hat{N}R^*), \quad (25)$$

and it is easy to show that $\hat{J}^*$ has asymptotic normal distribution.

Also, the pseudo-LRs as after-test performance measures can be estimated by $L\hat{R}_+^* = \frac{T\hat{P}R^*}{F\hat{P}R^*}$, and the approximate variance by Delta-method of the log of $L\hat{R}_+^*$ (Asymptotically normally distributed) is given by

$$Var[\log(\hat{LR}_{+}^{*})] = \frac{Var(\hat{TPR}^{*})}{TPR^{*}} + \frac{Var(\hat{FPR}^{*})}{FPR^{*}} \tag{26}$$

and $\hat{LR}_{-}^{*} = \frac{\hat{FNR}^{*}}{\hat{TNR}^{*}}$ while an approximate variance by Delta-method for the log of $\hat{LR}_{-}^{*}$ (Asymptotically normally distributed) is given by

$$Var[Log(\hat{LR}_{-}^{*})] = \frac{Var(\hat{FNR}^{*})}{FNR*} + \frac{Var(\hat{TNR}^{*})}{TNR*} \tag{27}$$

Finally, the estimator of the pseudo-odds ratio that measures the diagnostic test accuracy at the threshold (c_1, c_2) is given by $\hat{DOR}^{*} = \frac{(\hat{TPR}^{*})(\hat{TNR}^{*})}{(\hat{FPR}^{*})(\hat{FNR}^{*})}$ (Asymptotically normally distributed), with an approximate variance by Delta-method, by using (26) and (27), is provided by

$$Var[Log(\hat{DOR}^{*})] = \frac{Var(\hat{TPR}^{*})}{TPR^{*}} + \frac{Var(\hat{TNR}^{*})}{TNR^{*}} + \frac{Var(\hat{FPR}^{*})}{FPR^{*}} + \frac{Var(\hat{FNR}^{*})}{FNR^{*}}. \tag{28}$$

5. Numerical examples and simulations

5.1 Numerical examples

We provide numerical examples for a disease with two stages (non-disease, disease). We used two underlying bivariate distributions (the bivariate normal and bivariate Gamma distributions). The bivariate Gamma distribution used in this numerical example is from Sumen et al. (2014) as follows:

$$f_{X,Y}(x,y) = \frac{\beta_1^{\alpha_1}\beta_2^{\alpha_2}x^{\alpha_1+\alpha_2-1}y^{\alpha_2-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}e^{-(\beta_1 x+\beta_2 xy)}, x>0, y>0, \beta_1, \beta_2 > 0, \alpha_1, \alpha_2 > 0. \tag{29}$$

Also, the correlation between X and Y is then given by,

$$\rho = \frac{\dfrac{-\alpha_2}{\beta_2(\alpha_1-1)}}{\sqrt{\left(\dfrac{\alpha_1}{\beta_1^2}\right)\left(\dfrac{\alpha_2\beta_1^2(\alpha_1+\alpha_2-1)}{\beta_2^2(\alpha_1-1)^2(\alpha_1-2)}\right)}} = \frac{\dfrac{-\alpha_2}{(\alpha_1-1)}}{\sqrt{\left(\dfrac{\alpha_1\alpha_2(\alpha_1+\alpha_2-1)}{(\alpha_1-1)^2(\alpha_1-2)}\right)}}. \tag{30}$$

As presented in Tables 5 and 6, we use the Youden index and the newly developed pseudo-Youden index to find the cut-off points $(c_1, c_2, c_3, c_4, c_5, c_6 :\ c_3 < c_1 \le c_4; c_5 < c_2 \le c_6)$.

Tables 3 and 4, provided the scenarios of different Parmenter's settings for bivariate normal and bivariate Gamma distribution respectively. Tables 5 and 6 show some numerical results from the bivariate normal distributions and bivariate Gamma distributions.

Table 3. Simulations scenarios for Bivariate normal distributions

μ_{X_0}	μ_{Y_0}	μ_{X_1}	μ_{Y_1}	$\sigma^2_{X_0}$	$\sigma^2_{Y_0}$	$\sigma^2_{X_1}$	$\sigma^2_{Y_1}$	Scenarios
0	0	1	1	1	1	1	1	S_1
0	0	1	2	1	1	1	1	S_2
0	0	2	1	1	1	2	1	S_3
0	0	2	1	2	1	1	1	S_4

Table 4. Simulations scenarios for Bivariate Gamma distributions

α_{X_0}	α_{Y_0}	α_{X_1}	α_{Y_1}	β_{X_0}	β_{Y_0}	β_{X_1}	β_{Y_1}	Scenarios
3	3	4	4	1	1	0.7	0.7	G_1
3	3	4	4	1	1	0.5	0.5	G_2
3	3	4	4	1	1	0.5	0.7	G_3
3	3	4	4	1	1	0.3	0.3	G_4
2	2	3	3	1	1	0.7	0.7	G_5
2	2	3	3	1	1	0.5	0.5	G_6
2	2	3	3	1	1	0.5	0.7	G_7
2	2	3	3	1	1	0.3	0.3	G_8

Table 5. A numerical example of all proposed utilizes measures for Bivariate distributions

	Optimal $c_1, c_2\,(a, b)$	$J^*(J_X, J_Y)$	PPV^*	NPV^*	LR_+^*	LR_-^*	DOR^*
(P=0.1, $\rho = 0.1$)							
S_1	0.500, 0.500 (4.837,4.837)	0.785 (0.383, 0.383)	0.479	0.987	8.280	0.121	68.56
S_2	0.500, 1.000 (5.383, 4.558)	0.885 (0.383, 0.683)	0.645	0.993	16.336	0.061	266.86
S_3	1.000, 1.181 (4.062, 7.271)	0.882 (0.683, 0.345)	0.808	0.989	37.832	0.096	392.24
S_4	0.762, 0.500 (6.703, 5.057)	0.912 (0.541, 0.383)	0.517	0.995	9.632	0.042	229.10
(P=0.1, $\rho = 0.5$)							
S_1	0.500, 0.500 (4.983, 4.983)	0.673 (0.383, 0.383)	0.363	0.979	5.123	0.195	26.24
S_2	0.500, 1.000 (5.025, 4.311)	0.805 (0.383, 0.683)	0.507	0.988	9.259	0.108	85.73
S_3	1.000, 1.181 (4.560, 8.610)	0.829 (0.683, 0.345)	0.659	0.986	17.386	0.127	137.09
S_4	0.762, 0.500 (6.631, 4.815)	0.856 (0.541, 0.383)	0.430	0.992	6.800	0.076	89.00
(P=0.1, $\rho = 0.9$)							
S_1	0.500, 0.500 (2.244, 2.244)	0.509(0.383, 0.383)	0.255	0.965	3.076	0.325	9.463
S_2	0.500, 1.000 (2.709, 1.832)	0.701(0.383, 0.683)	0.387	0.981	5.681	0.176	32.28
S_3	1.000, 1.181 (3.663, 9.623)	0.759 (0.683, 0.345)	0.497	0.983	8.880	0.160	55.53
S_4	0.762, 0.500 (5.506, 7.162)	0.804 (0.541, 0.383)	0.356	0.987	4.963	0.116	42.807
(P=0.4, $\rho = 0.1$)							
S_1	0.500, 0.500 (4.837,4.837)	0.785 (0.383, 0.383)	0.847	0.926	8.280	0.121	68.56
S_2	0.500, 1.000 (5.383, 4.558)	0.885 (0.383, 0.683)	0.916	0.961	16.336	0.061	266.86
S_3	1.000, 1.181 (4.062, 7.271)	0.882 (0.683, 0.345)	0.962	0.940	37.832	0.096	392.24
S_4	0.762, 0.500 (6.703, 5.057)	0.912 (0.541, 0.383)	0.865	0.973	9.632	0.042	229.10
(P=0.4, $\rho = 0.5$)							
S_1	0.500, 0.500 (4.983, 4.983)	0.673 (0.383, 0.383)	0.774	0.885	5.123	0.195	26.24
S_2	0.500, 1.000 (5.025, 4.311)	0.805 (0.383, 0.683)	0.861	0.933	9.259	0.108	85.73
S_3	1.000, 1.181 (4.560, 8.610)	0.829 (0.683, 0.345)	0.921	0.922	17.386	0.127	137.09
S_4	0.762, 0.500 (6.631, 4.815)	0.856 (0.541, 0.383)	0.819	0.952	6.800	0.076	89.00
(P=0.4, $\rho = 0.9$)							
S_1	0.500, 0.500 (2.244, 2.244)	0.509(0.383, 0.383)	0.672	0.822	3.076	0.325	9.463
S_2	0.500, 1.000 (2.709, 1.832)	0.701(0.383, 0.683)	0.791	0.895	5.681	0.176	32.28
S_3	1.000, 1.181 (3.663, 9.623)	0.759 (0.683, 0.345)	0.856	0.904	8.880	0.160	55.53
S_4	0.762, 0.500 (5.506, 7.162)	0.804 (0.541, 0.383)	0.768	0.928	4.963	0.116	42.807

Table 6. A numerical example of all proposed utilizes measures for Gamma distributions

Scenarios	Optimal $c_1, c_2\,(a, b)$	$J^*(J_X, J_Y)$	PPV^*	NPV^*	LR_+^*	LR_-^*	DOR^*
(P=0.1, $\rho = -0.447$)							
G_1	3.890, 3.890(3.800, 3.800)	0.716 (0.454, 0.454)	0.742	0.972	25.809	0.263	98.11
G_2	4.663, 4.663 (4.600,4.600)	0.801 (0.637, 0.637)	0.874	0.979	62.674	0.189	332.28
G_3	4.663, 3.890 (4.600, 3.800)	0.785 (0.637, 0.454)	0.834	0.978	45.170	0.201	224.53
G_4	5.911, 5.911 (5.900, 5.900)	0.901 (0.829, 0.829)	0.954	0.990	187.949	0.095	1987.48
(P=0.1, $\rho = 0$)							
G_5	2.641, 2.641 (2.600, 2.600)	0.728 (0.458, 0.458)	0.475	0.979	8.136	0.189	43.00
G_6	3.212, 3.212 (3.100, 3.100)	0.787 (0.612, 0.612)	0.597	0.983	13.305	0.159	83.803
G_7	3.212, 2.641 (3.100, 2.600)	0.760 (0.612, 0.458)	0.534	0.981	10.293	0.172	60.018
G_8	4.126, 4.126 (4.100, 4.100)	0.881(0.788, 0.788)	0.745	0.990	26.255	0.087	300.69
(P=0.4, $\rho = -0.447$)							
G_1	3.890, 3.890(3.800, 3.800)	0.716 (0.454, 0.454)	0.945	0.851	25.809	0.263	98.11
G_2	4.663, 4.663 (4.600,4.600)	0.801 (0.637, 0.637)	0.977	0.888	62.674	0.189	332.28
G_3	4.663, 3.890 (4.600, 3.800)	0.785 (0.637, 0.454)	0.968	0.882	45.170	0.201	224.53
G_4	5.911, 5.911 (5.900, 5.900)	0.901 (0.829, 0.829)	0.992	0.941	187.949	0.095	1987.48
(P=0.4, $\rho = 0$)							
G_5	2.641, 2.641 (2.600, 2.600)	0.728 (0.458, 0.458)	0.844	0.888	8.136	0.189	43.00
G_6	3.212, 3.212 (3.100, 3.100)	0.787 (0.612, 0.612)	0.899	0.904	13.305	0.159	83.803
G_7	3.212, 2.641 (3.100, 2.600)	0.760 (0.612, 0.458)	0.873	0.897	10.293	0.172	60.018
G_8	4.126, 4.126 (4.100, 4.100)	0.881(0.788, 0.788)	0.946	0.945	26.255	0.087	300.69

All provided measures were affected when increasing the location shift and any change in the variations in the diseased group and the increase in the absolute value of the correlation between markers. Table 5 provided the optimal cut-off points (c_1, c_2, a, b) as well as the proposed measures.

On the other hand, Table 6 provides the skewed bivariate Gamma distribution results. Clearly, from (30), the association between the markers is always negative. In general, as the rates of the distributions decrease (means increase) for fixed shape parameters, the provided measures of test accuracy values are increased. Also, when the absolute value of the association increases,

all provided accuracy pseudo measures increase. Tables 5 and 6 showed that using our proposed measures provided a higher pseudo-Youden Index than the individual Youden Index which indicates higher correct classification rates. Finally, as expected, with higher disease prevalence, only pseudo-PPV and NPV will be affected.

5.2 Simulations

We conducted a simulation study to investigate the behavior of the asymptotic variances of our proposed pseudo-measures of medical diagnostics accuracy. We selected the parameters setting in Table 3 for the bivariate normal distribution for different values of the correlation coefficient between the two biomarkers. We simulate 5000 bivariate samples from the bivariate normal distributions of small, moderate, and large sample sizes $(n_1 = n_2 = 50, 100, 500)$. The prevalence of the assumed underlying disease is selected to be p=0.1. The results of our simulations are provided in Table 7. In Table 7 we calculated the empirical variances presented in section 4 and compared them with the simulated counterpart variances. Clearly from Table 7, the empirical variances of all presented measures decrease as the sample sizes increase for the same set of parameters. This indicates that as expected that empirical estimators of the asymptotic variances presented in section 4 are consistent estimators. Similarly, we observed the same behavior of the simulated variances for the proposed measures.

Moreover, Table 7 indicates that the empirical variances decrease slightly as the correlation coefficient between the two biomarkers increases. Also, variances of Log(LR_+^*), Log(LR_-^*) and Log(DOR^*) increase as the values of these parameters increase by comparing scenarios S_1 and S_2 when fixing all other parameters. Finally, the empirical and the simulated variances are compatible even for small sample sizes.

Table 7. Results of the simulation of the proposed accuracy measures variances for Bivariate distributions (Simulated variance)

	Var(TPR^*)	Var(TNR^*)	Var(J^*)	Var(PPV^*)	Var(NPV^*)	Var(Ln LR_+^*)	Var(Ln LR_-^*)	Var(Ln DOR^*)
(P=0.1, $\rho = 0.1$, $n_1 = n_2 = 50$)								
S_1	0.0035 (0.0073)	0.0037 (0.0021)	0.0072 (0.0056)	(0.0281)	(0.00008)	0.0807 (0.3614)	0.0385 (0.0053)	0.1211 (0.4900)
S_3	0.0037 (0.0045)	0.0036 (0.0012)	0.0073 (0.0037)	(0.0269)	(0.00005)	0.1047 (0.3376)	0.0525 (0.0035)	0.1624 (0.5518)
(P=0.1, $\rho = 0.5$, $n_1 = n_2 = 50$)								
S_1	0.0033 (0.0085)	0.0035 (0.0033)	0.0068 (0.0061)	(0.0229)	(0.00010)	0.0529 (0.3255)	0.0265 (0.0057)	0.0800 (0.4157)
S_3	0.0036 (0.0048)	0.0034 (0.0020)	0.0070 (0.0040)	(0.0270)	(0.00006)	0.0713 (0.3651)	0.0409 (0.0037)	0.1137 (0.5057)
(P=0.1, $\rho = 0.9$, $n_1 = n_2 = 50$)								
S_1	0.0030 (0.0112)	0.0031 (0.0055)	0.0061 (0.0080)	(0.0159)	(0.00013)	0.0304 (0.2508)	0.0172 (0.0067)	0.0477 (0.3415)
S_3	0.0034 (0.0056)	0.00030 (0.0034)	0.0064 (0.0047)	(0.0231)	(0.00006)	0.0437 (0.3370)	0.0307 (0.0042)	0.0747 (0.4354)
(P=0.1, $\rho = 0.1$, $n_1 = n_2 = 100$)								
S_1	0.0018 (0.0036)	0.0019 (0.0017)	0.0037 (0.0023)	(0.0154)	(0.00004)	0.0345 (0.2846)	0.0187 (0.0027)	0.0533 0.2725
S_3	0.0019 (0.0020)	0.0019 (0.0008)	0.0038 (0.0016)	(0.0155)	(0.00002)	0.0575 (0.3741)	0.0241 (0.0016)	0.0812 (0.3928)
(P=0.1, $\rho = 0.5$, $n_1 = n_2 = 100$)								
S_1	0.00168 (0.00442)	0.00175 (0.00246)	0.00343 (0.00284)	(0.0101)	(0.00005)	0.0200 (0.1759)	0.0128 (0.0031)	0.0327 (0.1923)
S_3	0.0018 (0.0025)	0.0017 (0.0014)	0.0035 (0.0019)	(0.0136)	(0.00003)	0.0315 (0.2685)	0.0183 (0.0020)	0.0499 (0.2803)
(P=0.1, $\rho = 0.9$, $n_1 = n_2 = 100$)								
S_1	0.00153 (0.00641)	0.00156 (0.00416)	0.00308 (0.00710)	(0.0056)	(0.00007)	0.0119 (0.1119)	0.0085 (0.0039)	0.0203 (0.1524)
S_3	0.0017 (0.0030)	0.0016 (0.0024)	0.0033 (0.0025)	(0.0101)	(0.00004)	0.0180 (0.1844)	0.0140 (0.0023)	0.0319 (0.2183)
(P=0.1, $\rho = 0.1$, $n_1 = n_2 = 500$)								
S_1	0.00037 (0.00110)	0.00038 (0.00083)	0.00074 (0.00045)	(0.00403)	(0.00001)	0.00437 (0.06670)	(0.00376 0.00087)	0.00813 (0.0447)
S_3	0.00038 (0.00060)	0.00039 (0.00030)	0.00077 (0.00034)	(0.00462)	(0.000007)	0.00845 (0.09790)	0.00428 (0.00049)	0.01273 (0.0684)
(P=0.1, $\rho = 0.5$, $n_1 = n_2 = 500$)								
S_1	0.00034 (0.00142)	0.00035 (0.00113)	0.00069	(0.00216)	(0.00002)	0.00286 (0.03838)	0.00256 (0.00102)	0.00542 (0.03137)
S_3	0.00036 (0.00072)	0.00037 (0.00047)	0.00073 (0.00043)	(0.00339)	(0.000009)	0.00494 (0.05679)	0.00329 (0.00087)	0.00823 (0.04810)
(P=0.1, $\rho = 0.9$, $n_1 = n_2 = 500$)								
S_1	0.00031 (0.00207)	0.00031 (0.00173)	0.00062 (0.00091)	(0.00091)	(0.00002)	0.00186 (0.02229)	0.00171 (0.00125)	0.00358 (0.02837)
S_3	0.00034 (0.00088)	0.00034 (0.00083)	0.00068 (0.00071)	(0.00244)	(0.00001)	0.0031 (0.0415)	0.0025 (0.0007)	0.0057 (0.0479)

6. Illustration using WBCD data

This section applies the proposed measures to the Diagnostic Wisconsin Breast Cancer Database (WBCD) created by the University of Wisconsin (http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29). For diagnosing breast cancer, summary features of digitized images of a fine needle aspirate (FNA) of a breast mass are considered biomarkers for diagnosing purposes. This section aims to use WBCD data to investigate how using bivariate markers would improve those biomarkers' diagnostic abilities and select the proper biomarkers for breast cancer diagnosis. The WBCD data set contains 569 observations and 30 features as candidate univariate biomarkers for selection. The reference/gold standard is indicated as the "Diagnosis" variable in the data, and it has two values: B = benign ($n_1 = 357$) or M = malignant ($n_2 = 212$).

Hyuna et al. (2021) provided an update on the global cancer burden using the GLOBOCAN 2020 estimates of cancer incidence and mortality produced by the International Agency for Research on Cancer. They reported that female breast cancer had surpassed lung cancer as the most diagnosed cancer, with an estimated 2.3 million new cases (11.7%), followed by lung (11.4%). So, we will use 0.117 as the prevalence of breast cancer in this application.

Tables 8 and 9 present the results of analyzing WBCD data for all 30 biomarkers using the Youden Index (J) to select the optimal cut-off points. Table 8 is sorted in ascending order concerning the objective function J. Also, Figure1 shows selected biomarkers from Table 8 plots of their distributions. From Figure 1 some of the plots show the symmetry and the skewness of the underlying distribution of those biomarkers.

Table 8. Results of univariate data analysis WBCD using Youden Index for selecting the cut-off point for each of the markers listed.

Markers	Cut-off points	J	TPR	TNR	PPV	NPV	*DOR*
perimeter_worst (M1)	106.000	0.839	0.920	0.919	0.60	0.99	130.48
concave.points_mean (M2)	0.049	0.828	0.915	0.913	0.58	0.99	112.97
radius_worst (M3)	16.820	0.814	0.844	0.969	0.78	0.98	169.11
concave.points_worst (M4)	0.136	0.812	0.868	0.944	0.67	0.98	110.85
area_worst (M5)	739.300	0.811	0.948	0.863	0.48	0.99	114.84
concavity_mean (M6)	0.090	0.761	0.868	0.894	0.52	0.98	55.46
perimeter_mean (M7)	90.430	0.758	0.887	0.871	0.48	0.98	53.00
area_mean (M8)	698.800	0.736	0.764	0.972	0.78	0.97	112.38
area_se (M9)	31.330	0.732	0.830	0.902	0.53	0.98	44.94
radius_mean (M10)	15.050	0.729	0.759	0.969	0.76	0.97	98.44
concavity_worst (M11)	0.261	0.728	0.896	0.832	0.41	0.98	42.67
perimeter_se (M12)	2.765	0.616	0.745	0.871	0.43	0.96	19.73
radius_se (M13)	0.386	0.614	0.759	0.854	0.41	0.96	18.42
compactness_mean (M14)	0.102	0.607	0.825	0.782	0.33	0.97	16.91
compactness_worst (M15)	0.267	0.566	0.717	0.849	0.39	0.96	14.25
concavity_se (M16)	0.021	0.502	0.892	0.611	0.23	0.98	12.97
texture_mean (M17)	19.320	0.472	0.755	0.717	0.26	0.96	7.81
concave.points_se (M18)	0.011	0.470	0.807	0.664	0.24	0.96	8.26
texture_worst (M19)	24.890	0.463	0.830	0.633	0.23	0.97	8.42
smoothness_worst (M20)	0.136	0.420	0.689	0.731	0.25	0.95	6.02
compactness_se (M21)	0.022	0.391	0.708	0.683	0.23	0.95	5.22
symmetry_worst (M22)	0.299	0.360	0.604	0.756	0.25	0.94	4.73
smoothness_mean (M23)	0.090	0.344	0.868	0.476	0.18	0.96	5.97
symmetry_mean (M24)	0.172	0.310	0.797	0.513	0.18	0.95	4.14
fractal_dimension_worst (M25)	0.082	0.305	0.627	0.678	0.21	0.93	3.54
fractal_dimension_se (M26)	0.003	0.225	0.651	0.574	0.17	0.93	2.51
texture_se (M27)	0.821	0.099	0.830	0.269	0.13	0.92	1.80
fractal_dimension_mean (M28)	0.067	0.080	0.288	0.793	0.16	0.89	1.55
smoothness_se (M29)	0.004	0.054	0.906	0.148	0.12	0.92	1.67
symmetry_se (M30)	0.045	0.054	0.057	0.997	0.72	0.89	20.09

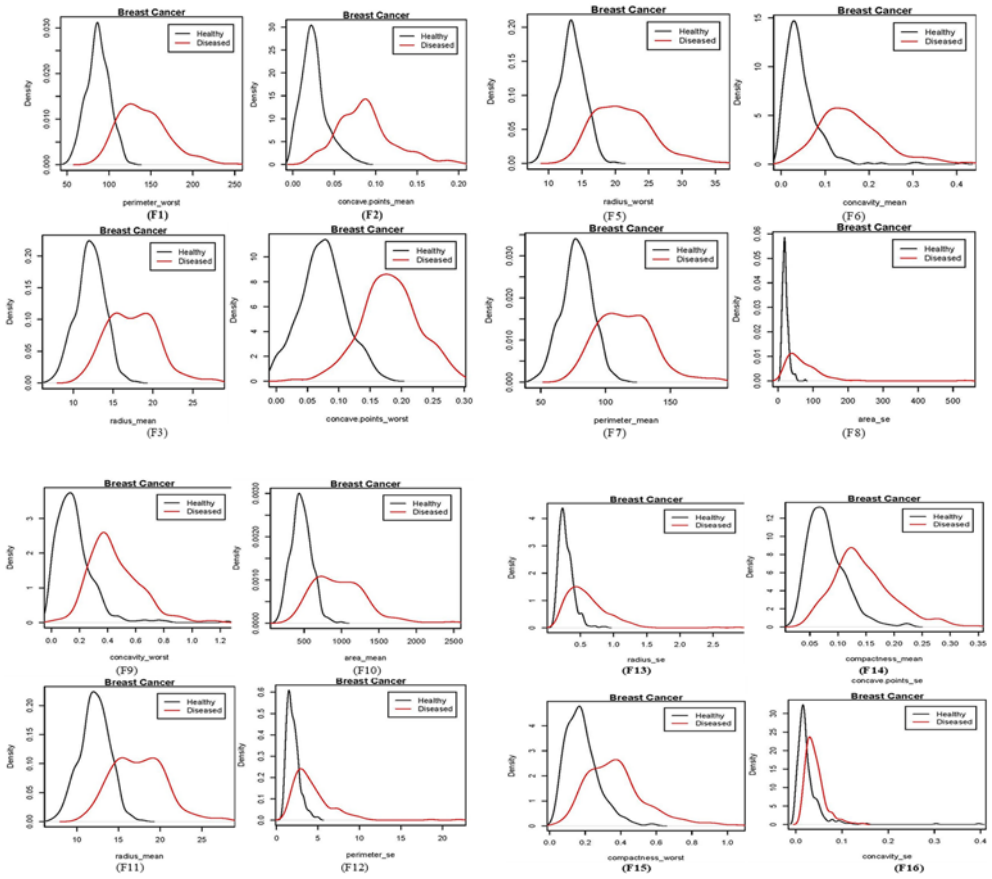

Figure 1. Probability density function plot of selective biomarkers from Table 7.

Table 8 provides the cut-off points of each marker independently along with some tests' accuracy and prediction ability. In contrast, Table 8 provides the bivariate analysis and the estimate of pseudo accuracy measures. To reduce the number of tables, we chose perimeter_worst (M1), which has the highest value of *J*, as a good marker and associated it with all other markers. Also, to further demonstrate the improvement of overall diagnostic accuracy through the proposed measures, we chose two markers with relatively less satisfying diagnostic performance in terms of the *J* value, namely, smoothness_mean (M23) and symmetry_se (M30), with *J* values of 0.344 and 0.054, respectively. The results from combining M23 and M30 are listed at the end of Table 9.

From Table 9, we can see that the bivariate analysis of (M1 with M2), (M1 with M28), (M1 with M22), (M1 with M4), (M1 with M19), and (M1 with M23) have the highest pseudo Youden

index values. However, (M30 with M23) has the highest DOR^*. Also, this combination improved the TPR of M23 0.868 and of M30 0.057 to 0.869 for (M23 with M30). Similarly, the TRN of M23 0.476 and M30 0.997 improved to 0.997 for (M23 with M30). This combination brought the best of both markers. The other pseudo-measure values depend on the pseudo-TPR and TNR and the balance between them.

We can also see that the DOR* values rise steeply when combining those markers. This behavior of DOR is commonly observed in the univariate case (Yin and Vogel 2017). We can see such behavior as more evident for the bivariate analysis. So, in this sense, we recommend using bounded pseudo accuracy measures such as J*, PPV*, or NPV* as they are within [0,1] over unbounded measures such as DOR* or LR* for bivariate biomarker evaluations.

From the end of Table 8, combining two weak markers; namely, smoothness_mean (M23) and symmetry_se (M30), with J values are 0.344 and 0.054, has the advantage over using these markers separately. Consequently, other accuracy measures will also be changed and will balance between rule-out and rule-in of patients. Using M23 alone indicates that this marker has more rule-out ability and a high negative predictive value of the disease. In comparison, M30 alone has more rule-in ability and closely balances predictive positive and negative values. However, using the proposed bivariate approach has much higher accuracy based on the J value than using each marker separately. Also, using the bivariate approach provided a more balanced rule-out and rule-in ability, higher negative predicted value, and above-average positive predicted value.

Tables 8 and 9 indicate that using two markers in bivariate analysis through the and-or system improved the diagnostic accuracy of individual makers alone. It seems that the proposed measures

have the potential of helping clinicians to decide more accurately, especially when using a single biomarker that could put a healthy person in the gray area of false positives.

Table 9. Results of bivariate data analysis WBCD using Pseudo Youden Index for cut-off points (a, b).

Markers	r	Optimal Cut-off points	J^* (Var)	TPR^*	TNR^*	PPV^*	NPV^*	DOR^* (95% C.I)
M1 & M2	0.856	106.00, 0.049 (22, 0.04)	0.954 (0.00085)	0.976	0.978	0.852	0.997	1806.08 (1233.85, 2643.70)
M1& M3	0.994	106.00, 16.82 (9, 2)	0.884 (0.00075)	0.915	0.961	0.756	0.988	264.06 (212.57, 328.01)
M1 & M4	0.816	106.00, 0.136 (11, 0.09)	0.932 (0.00092)	0.962	0.966	0.791	0.995	733.13 (533.13, 1006.70)
M1 & M5	0.978	106.00, 739.30 (9, 102)	0.880 (0.00068)	0.953	0.927	0.634	0.993	257.16 (206.03, 320.98)
M1 & M6	0.730	106.00, 0.09 (10, 0.07)	0.918 (0.00098)	0.958	0.961	0.764	0.994	552.61 (405.69, 752.74)
M1 & M7	0.970	106.00, 90.43 (9, 6)	0.843 (0.00079)	0.925	0.919	0.601	0.989	138.55 (112.78, 170.21)
M1 & M8	0.959	106.00, 698.80 (9, 180)	0881 (0.00094)	0.906	0.975	0.826	0.987	371.20 (287.95, 478.53)
M1 & M9	0.761	106.00, 31.33 (10, 14)	0.890 0.00097	0.929	0.961	0.758	0.990	321.77 (246.66, 419.75)
M1& M10	0.965	106.00, 15.05 (9, 3.3)	0.890 (0.00097)	0.915	0.975	0.828	0.989	416.74 (320.05, 542.64)
M1& M11	0.618	106.00, 0.261 (10, 0.12)	0.909 (0.0010)	0.962	0.922	0.619	0.995	299.63 (223.84, 401.07)
M1& M12	0.721	106.00,2.765 (12, 1.6)	0.903 (0.0012)	0.934	0.969	0.801	0.991	444.86 (323.70, 611.36)
M1& M13	0.720	106.00, 0.386 (12, 0.26)	0.908 (0.0012)	0.939	0.969	0.802	0.992	481.50 (347.31, 667.53)
M1& M14	0.590	106.00, 0.102 (10, 0.04)	0.861 (0.0012)	0.948	0.913	0.591	0.993	192.16 (144.96, 254.72)
M1& M15	0.529	106.00, 0.267 (10, 0.13)	0.881 (0.0013)	0.948	0.933	0.651	0.993	253.53 (186.10, 345.41)
M1& M16	0.227	106.000, 0.021 (10, 0.02)	0.758 (0.0011)	0.962	0.796	0.384	0.994	99.21 (45.38, 130.56)
M1& M17	0.358	106.00, 19.32 (12, 12)	0.934 (0.0015)	0.962	0.972	0.820	0.995	884.85 (583.90, 1340.91)
M1& M18	0.395	(106.00, 0.011) (12, 0.01)	0.877 (0.0014)	0.967	0.910	0.588	0.995	297.43 (208.20, 424.92)
M1&M19	0.365	106.00, 24.89 (10, 14)	0.928 (0.0014)	0.967	0.961	0.766	0.996	717.50 (483.29, 1065.21)
M1& M20	0.237	106.00, 0.136 (11, 0.1)	0.447 (0.0011)	0.811	0.636	0.228	0.962	7.51 (6.44, 8.76)
M1& M21	0.261	106.00, 0.022 (10, 0.01)	0.840	0.958	0.882	0.519	0.994	169.17 (121.70, 235.15)

Table 9 (Continue). Results of bivariate data analysis WBCD using Pseudo Youden Index for cut-off points (a, b).

M1 with M22	0.270	106.00, 0.299 (10, 0.14)	0.934 (0.0016)	0.962	0.972	0.820	0.995	884.85 (574.65, 1362.50)
M1 with M23	0.239	106.00, 0.090 (11, 0.05)	0.922 (0.0013)	0.967	0.955	0.741	0.995	624.15 (428.31, 909.55)
M1 with M24	0.219	106.00, 0.172 (10 0.171)	0.914 (0.0015)	0.967	0.947	0.707	0.995	520.98 (353.79, 767.18)
M1 with M25	0.139	106.00, 0.082 (9, 0.02)	0.915 (0.0016)	0.962	0.952	0.728	0.995	510.00 (341.98, 760.58)
M1 with M26	-0.001	106.00, 0.003 (10, 0.0029)	0.823 (0.0016)	0.958	0.980	0.486	0.994	145.2 (103.86, 203.00)
M1 with M27	-0.102	106.00, 0.821 (11, 0.82)	0.748 (0.0013)	0.972	0.776	0.365	0.995	118.88 (83.91, 168343)
M1 with M28	-0.205	106.00, 0.067 (10, 0.05)	0.936 (0.0013)	0.953	0.983	0.883	0.994	1181.70 (767.56, 1819.29)
M1 with M29	-0.217	106.00, 0.004 (11, 0.0039)	0.611 (0.0009)	0.967	0.644	0.265	0.993	53.04 (41.16, 68.34)
M1 with M30	-0.104	106.00, 0.045 (10, 0.044)	0.911 (0.0007)	0.920	0.992	0.936	0.989	1353.53 (960.05, 1908.29)
M23 with M30	0.200	0.090, 0.045 (0.089, 0.044)	0.865 (0.0013)	0.868	0.997	0.976	0.983	2339.43 (870.27, 6288.76)

7. Final remarks and discussion

It is customary, in medical research, to collect information on multiple continuous biomarkers to improve the effectiveness of diagnostics tests (markers) in clinical decisions and the diagnostic tests' accuracy. Recently, the trend in practice has been to combine the measurements of these biomarkers into one single score. However, incorporating those biomarkers' measures into one score depends on the methods used to combine them and may lose vital information needed to make an effective and accurate decision.

We provided extensions to some accuracy measures and predictive values from univariate to bivariate markers and named them pseudo measures in the paper. We defined pseudo-and-or classifiers for the true positive rate, true negative rate, false-positive rate, and false-negative rate. Then we use them to redefine some existing measures such as the Youden index, odds ratio, likelihood ratios, and predictive values. We provided optimal cut-off points selection based on the

modified Youden index. Also, we offered the derivation for the empirical estimators and their variances.

Our numerical examples and the real data analysis indicated that the new pseudo classification measures, including J^*, TPR^*, TNR^*, PPV^*, NPV^*, and DOR^*, using bivariate markers, preserve the quality information provided by each marker separately. Consequently, using bivariate markers will increase the effectiveness, accuracy, and predictive values for rule-in and rule-out patients. We, therefore, recommend using the developed pseudo measures in this paper for diagnostics whenever bivariate biomarkers are available.

References

Altman, D. G., & Bland, J. M. (1994). Statistics Notes: Diagnostic tests 2: predictive values. *B*MJ, 309(6947), 102.

Baker, S. J. (2000). Identifying combinations of cancer markers for further study as triggers of early intervention. *Biometrics* 56, 1082- 1087.

Bansal, A, and Pepe, M. S. (2013). When does combining markers improve classification performance and what are implications for practice? Stat Med. 32(11),1877-1892.

Boyko, E. J. (1994). Ruling out or ruling in disease with the most sensitive or specific diagnostic test: Short cut or wrong turn? *Medical Decision Making*, 14(2), 175-179.

Deeks, J. J., & Altman, D. G. (2004). Diagnostic tests 4: likelihood ratios. *BMJ,* 329(7458), 168-169.

Etzioni, R., Kooperberg, C., Pepe, M., Smith, R., and Gann, P.H. (2003). Combining biomarkers to detect disease with application to prostate cancer. *Biostatistics* 4, 523-538.

Gann, P. H., Ma, J., Catalona, W. J.and Stampfer, M. J. (2002). Strategies for Combining Total and Percent Free PSA for Detection of Prostate Cancer: a Prospective Evaluation. *Journal of Urology,* 167, 2427–2434.

Gilbert, R., Logan, S., Moyer, V. A., & Elliott, E. J. (2001). Assessing diagnostic and screening tests: Part 1. Concepts. *The Western journal of medicine*, 174(6), 405-409.

Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., and Bossuyt, P. M. M. (2003). The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology*, 56(11), 1129-1135.

Huang, Y., Yin, J., & Samawi, H. (2018). Methods improving the estimate of diagnostic odds ratio. *Communications in Statistics-Simulation and Computation*, 47(2), 353-366.

Hyuna, S., Jacques, F., Rebecca L. S., Mathieu, L., Isabelle, S., Ahmedin, J., and Freddie B.,

(2021). Global cancer statistics 2020: GLOBOCAN estimates worldwide incidence and mortality for 36 cancers in 185 countries. *CA Cancer J Clin*,0, 1-41.

Kumar, S., Mohan, A., and Culeria, R. (2006). Biomarkers in cancer screening, research, and detection: present and future: a review. *Biomarkers,* 11, 385-405.

Lee, W. C. (1999). Selecting diagnostic tests for ruling out or ruling in disease: the use of the Kullback-Leibler distance. *International journal of epidemiology*, 28(3), 521-525.

Li, J., Fine, J. P., and Safdar, N. (2007). Prevalence-dependent diagnostic accuracy measures. *Statistics in Medicine*, 26, 3258-3273.

Li, J., and Fine, J. P. (2010). Weighted area under the receiver operating characteristic curve and its application to gene selection. *Journal of Royal Statistical Society, Applied Statistics Series C,* 59, (4), 673-692.

Li, J., Kuk, A.Y.C., and Rush, A.J. (2012). A practical approach to the early identification of antidepressant medication non-responders. *Psychological Medicine* 42 (2), 309-316.

McNeil, B. J., & Adelstein, S. J. (1976). Determining the value of diagnostic and screening tests. *Journal of Nuclear Medicine*, 17(6), 439-448.

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford.

Samawi, H., Yin, J., Rochani, H., and Panchal, V. (2017). Notes on overlap measure as an alternative to Youden Index: How are they related? *Statistics in Medicine,* 36 (26), 4230–4240.

Samawi, H., Yin, J., Zhang, X., Yu, L., Rochani, H., Vogel, R., and Mo, C. (2020). Applications of Kullback-Leibler Information as a Measure of Diagnostic Accuracy. *Journal of Applied Bioinformatics & Computational Biology* 9:2, 1-10.

Samawi, H., Yin, J., Rochani, H., Mo, C., & Kersey, J. (2021). Post-test diagnostic accuracy measures of a continuous test with a disease of ordinal multi-stages. *Statistics in Biopharmaceutical Research*, 1-12. https://doi.org/10.1080/19466315.2021.1873841

Sande, Z. S., Loraine Seng, L., Li, J., and D'Agostino, R. (2021). Statistical Learning in Medical Research with Decision Threshold and Accuracy Evaluation. *Journal of Data Science* 19 (4), 634–657

Sidransky, D. (2002). Emerging molecular markers of cancer. *Nature Reviews Cancer,* 2, 210-219.

Sox J. r., H. C., Koran, L. M., Sox, C. H., Marton, K. I., Dugger, F., & Smith, T. (1989). A medical algorithm for detecting physical disease in psychiatric patients. *Psychiatric Services*, 40(12), 1270-1276.

Stenman, U. H., Leionen, J., Alfthen, H., Ranniko, S., Tuhkanen, K. and Alfthen, O. (1991). A complex between prostate-specific antigen and alpha-1-antichymotrypsin is the major form of prostate-specific antigen in serum of patients with prostatic cancer: Assay of the complex improves clinical sensitivity for cancer. *Cancer Research,* 51, 222–226.

Sumen, S., Rajan, L., and Norou, D. (2014). A Bivariate Distribution with Conditional Gamma and its Multivariate Form *Journal of Modern Applied Statistical Methods*. 13 (2), 169-184.

Yin, J., & Tian, L. (2014a). Joint confidence region estimation for area under ROC curve and Youden index. *Statistics in medicine*, 33(6), 985-1000.

Yin, J., & Tian, L. (2014b). Optimal linear combinations of multiple diagnostic biomarkers based on Youden index. *Statistics in medicine*, 33(8), 1426-1440.

Yin, J., & Vogel, R. L. (2017). Using the roc curve to measure association and evaluate prediction

accuracy for a binary outcome. *Biometrics and Biostatistics International Journal*, 5(3), 1.

Wang, M., and Li., S. (2012). Bivariate Marker Measurements and ROC Analysis. *Biometrics*, 68(4), 1207-1218.

Wong, H. B., and Lim, G. H. (2011). Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV. *Proceedings of Singapore Healthcare*, 20(4), 316-318.

Zhou, X. H., McClish, D. K., & Obuchowski, N. A. (2009). Statistical methods in diagnostic medicine (569). John Wiley & Sons.