Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics Turkiye Klinikleri J Biostat. 2022;14(3):137-46

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH DOI: 10.5336/biostatic.2022-89951

A New Perspective for Area Under the Curve in Decision Making with Extra Safety Threshold Value: A Simulation Study

Ekstra Güvenlik Eşik Değeri ile Karar Vermede Eğri Altında Kalan Alan İçin Yeni Bir Bakış Açısı: Bir Simülasyon Çalışması

Selma GÜRLER[a], Özlem EGE ORUÇ[a], Dinçer GÖKSÜLÜK[b], Aslı SUNER[c]

[a]Department of Statistics, Dokuz Eylül University Faculty of Science, İzmir, Türkiye
[b]Department of Biostatistics, Erciyes University Faculty of Medicine, Kayseri, Türkiye
[c]Department of Bioistatistics and Medical Informatics, Ege University Faculty of Medicine, İzmir, Türkiye

ABSTRACT Objective: In studies conducted to detect a disease, making a false negative decision in cases such as detecting a deadly disease (Case I), or making a false positive decision in cases where diseases with high treatment costs (Case II) can lead to dangerous results. In this study, a new definition of the area under the curve (AUC) is proposed using a safety threshold value *t* for the diagnostic test to provide flexible decisions in critical cases. The alternative cut-off point for test diagnosis is evaluated by a simulation study in terms of sensitivity/specificity and relative efficiency. **Materials and Methods:** A simulation study was performed using different AUC values to obtain the cut-off point *c* shifted towards *c-t* for Case I and *c+t* for Case II. The normal distribution is used for the diseased (X) and non-diseased (Y) data. When obtaining the shift amount *t*, the gamma probability, which is the desired percentage of increase or decrease in the sensitivity/specificity value, is taken into account. **Results:** The results of our study showed that the relative efficiency is not significantly affected by working with the safety threshold *t* value when the test is less accurate and has a low AUC value. **Conclusion:** In this study, alternative cut-off points are obtained using the shift amount *t* determined by a predefined gamma probability. It is suggested that in critical situations, using the extra safety threshold *t*, determining the actual disease margin and safety standards for subjects can provide a more tolerant decision, especially in tests with low discrimination power.

Keywords: Area under curve; cut-off-value; sensitivity; Youden Index; maximum efficiency

ÖZET Amaç: Bir hastalığı saptamak için yapılan çalışmalarda, ölümcül bir hastalığın saptanması gibi durumlarda yanlış negatif karar verilmesi (Durum I), veya tedavi maliyetinin yüksek olduğu hastalıklarda yanlış pozitif karar verilmesi (Durum II) tehlikeli sonuçlara yol açabilir. Bu çalışmada, kritik durumlarda esnek karar sağlamak için tanı testi için bir güvenlik eşik değeri *t* kullanılarak eğri altındaki alanın (AUC) yeni bir tanımı önerilmiştir. Test tanısı için alternatif kesme noktası, duyarlılık/özgüllük ve göreli etkinlik açısından bir simülasyon çalışması ile değerlendirilmiştir. **Gereçler ve Yöntemler:** Durum I için *c-t*'ye ve Durum II için *c+t*'ye kaydırılan kesme noktası *c*'yi elde etmek için farklı AUC değerleri kullanılarak bir simülasyon çalışması yapılmıştır. Hasta (X) ve hastalıklı olmayan (Y) için normal dağılım kullanılmıştır. Kaydırma miktarı *t* elde edilirken, duyarlılık/özgüllük değerinde yüzde olarak istenen artış veya azalış miktarı olan gama olasılığı dikkate alınır. **Bulgular:** Çalışmamızın sonuçları, testin daha az doğru ve düşük AUC değerine sahip olduğu durumlarda, güvenlik eşiği *t* değeriyle çalışmanın göreli etkinliğinin önemli ölçüde etkilenmediğini göstermiştir. **Sonuç:** Bu çalışmada, önceden tanımlanmış bir gama olasılığı ile belirlenen kaydırma miktarı *t* kullanılarak alternatif kesme noktaları elde edilmiştir. Kritik durumlarda, ekstra güvenlik eşiği *t* kullanılarak, denekler için gerçek hastalık sınırı ve güvenlik standartlarının belirlenmesi, özellikle ayırt etme gücü düşük testlerde daha toleranslı bir karar sağlayabileceği önerilmektedir.

Anahtar kelimeler: Eğri altında kalan alan; kesme noktası; duyarlılık; Youden Indeks; maximum etkinlik

Correspondence: Özlem EGE ORUÇ
Department of Statistics, Dokuz Eylül University Faculty of Science, İzmir, Türkiye
E-mail: ozlem.ege@deu.edu.tr

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 24 Mar 2022 ***Received in revised form:*** 25 Aug 2022 ***Accepted:*** 31 Aug 2022 ***Available online:*** 13 Sep 2022

A receiver operating characteristic (ROC) curve is a technique generally used as a standard method to select an optimal cut-off point for a diagnostic test (*c*), and to compare 2 or more diagnostic tests' accuracy.[1-3] ROC curves are first used for evaluating radar detection in the 1950s, and were very popular in signal detection theory to depict the tradeoff between hit rates and false alarm rates of classifiers, problems with radar, and analyzing the behavior of diagnostic systems.[4,5] However, this method had been started to use in medical decision-making in various areas of medicine as experimental psychology, psychophysics, radiology and clinical chemistry.[3,6,7]

Several indexes have been used to summarize the information contained in a ROC curve for instance area under the curve (AUC), partial area index, a true positive rate (TPR) for a given false positive rate (FPR).[8-10] The most common index is the AUC, an overall summary of diagnostic accuracy. It can be viewed as a measure based on pairwise comparisons between classifications of the 2 classes (disease and non-disease). Let X and Y denote the diagnostic test measurements for diseased and non-diseased subjects, respectively. AUC can be defined as the probability that the diagnostic test measurement is higher for the diseased subject in a randomly selected pair of non-diseased and diseased individuals, i.e., $Pr(X>Y)$.[11] Various parametric and non-parametric approaches exist for estimating and comparing the AUC in the literature. If the test values for both non-diseased and diseased populations follow the normal distribution, the standard parametric methods are used to estimate the AUC. When the normality assumption is not satisfied for at least one subset (i.e., diseased and healthy), the non-parametric approaches is used for estimating the AUC. The most common non-parametric approaches are the Mann-Whitney statistic and kernel smoothing. For the example studies on the estimation methods of AUC, see.[11-13]

A diagnostic test yields a measurement (criterion value) that is used to diagnose some condition of interest such as a disease. A person is assessed as diseased or non-diseased (healthy) depending on the situation of the corresponding test value, which is greater or less than a given cut-off point.[14] A positive or negative diagnosis is made by comparing the measurement to a cut-off point. If the measurement is less (or greater) than the cut-off point, the test is negative which is classified as non-diseased. Thus, the cut-off point helps determine the rates of false positives and negatives. In some critical situations (such as infectious or high cost of treatment diseases), the decision-maker may need additional safety threshold value for his/her decision. Because the cut-off point is not universal, and should be determined for each situation and each disease condition. The optimal cut-off point selection and classification accuracy evaluation were discussed for parametric and nonparametric approaches in the study of.[15] In addition, an R package named *OptimalCutpoints* is available to find the optimal cut-off point in diagnostic tests considering the diagnostic decision costs and the prevalence of the disease.[16] This package not only provides numerical output with optimal cut-off point and accuracy measures with confidence intervals, but also graphical output with ROC and predictive ROC curves. Habibzadeh et al. proposed a method for an index of diagnostic test effectiveness and showed that the costs incurred by misdiagnosis should be taken into account for determination of the cut-off point.[17] Modified AUC metrics such as pAUC, mAUC, sAUC or a weighted AUC were became prevalent in various studies.[18-20]

In this study, we proposed a new AUC definition denoted by *AUC-t* using the safety threshold t for 2 different cases to provide flexibility for interpretation of AUC. We investigate performances of the shifted c using threshold value t, i.e. $[c-t, c]$, or $[c, c+t]$, such as the amount of increase or decrease in sensitivity or specificity. In Section 2, we give some fundamental concepts and introduce the new definition of AUC, denoted by *AUC-t*, for 2 cases defining the critical situations. In Section 3, we provide numerical results on the value of cut-off points of a diagnostic test using estimation methods of the Youden Index and maximum efficiency. Finally, conclusions are given in Section 4.

MATERIAL AND METHODS

For a diagnostic test, we can define a null hypothesis that the subject is non-diseased. Generally, 4 possible decisions and 2 types of errors are made when comparing a test result with an actual state. As given in Table

1, if both actual state and test are positive, it is called a true positive. If the actual state is positive and the test is negative it is called a false negative (FN) then Type II error (β) has occurred. If the actual state is negative and the test is positive it is called a false positive (FP) then Type I error (α) has occurred.

TABLE 1: Relations between the disease status of a test result.

	Diagnostic test result		Total
Actual state	Positive	Negative	
Positive	True positive (TP)	False negative (FN)	TP+FN
Negative	False positive (FP)	True negative (TN)	FP+TN
Total	TP+FP	FN+TN	TP+FP+FN+TN

The accuracy of a diagnostic test is evaluated by the sensitivity and the specificity. Sensitivity is described as the ability of a test to accurately identify those with the disease (true positive rate).[3] The sensitivity of a test (q) is the proportion of a positive test result among those with disease which can be defined as the True Positive Rate [TPR=TP/(TP+FN)]. This value provides the proportion of cases picked out by the test, relative to all cases who actually have the disease. Similarly, specificity is defined as detecting absence of the disease correctly when the disease is truly not present.[3] The specificity (1-p) is the proportion of a negative test result among those without disease. In other words, it is the ability of the test to pick out patients without the disease. Specificity is synonymous with the True Negative Rate [TNR=TN/(FP+TN)]. For the detailed studies one can see.[21-25]

The most important purposes of using ROC curve are defined as follows: (a) measuring the accuracy with sensitivity and specificity that are independent of disease prevalence, (b) comparing the diagnostic tests according to their FPRs, and (c) evaluating the accuracy by using both sensitivity and specificity in a single measure.[26] The ROC curve shows the characteristics of a diagnostic test by plotting the FPRs (i.e., 1-Specificity) on the horizontal axis (*X*) and the true positive rates (i.e., sensitivity) on the vertical axis (*Y*) for various cut-off points.[3,26] The AUC of the ROC curve is based on all possible cut-off points, and can be interpreted as non-informative when AUC is equal to 0.5, less accurate (0.5<AUC<0.7), moderately accurate (0.7<AUC<0.9), highly accurate (0.9<AUC<1) and perfect test (AUC=1).[27]

THE NEW DEFINITION OF AUC

Suppose that the measurement is less than the cut-off point, the test is negative, otherwise, the test is positive. Let $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ be the random samples from the diseased and non-diseased populations having cumulative distribution functions F and G, respectively. Then the sensitivity and the 1-specificity are defined as $q=1-F(c)$ and $p=1-G(c)$ for all possible cut-off points *c*, respectively, see.[13] Instead of using a single critical value for the diseased and non-diseased individuals, the diagnosis can be decided by considering a critical value such as $(c-t)$ or $(c+t)$. In this section, we propose new flexible AUC denoted by *AUC-t*. In the following, the determination of the extra safety threshold *t* value in terms of 2 different critical cases (Case I and Case II) is discussed in order to determine alternative cut-off points.

DETERMINATION OF THE EXTRA SAFETY THRESHOLD *t* Value

In cases defined as Case I and Case II, the safety threshold value, *t*, is used to shift the cut-off point *c*. For obtaining the shift amount *t*, the gamma probability (γ), which is the desired percentage of increase or decrease in the sensitivity/specificity value, is taken into account. Gamma expresses a probability that can take a value between 0 and 1 and it is determined by the researcher. The increase or decrease in the specified value of the gamma probability will determine the increase or decrease in the sensitivity/specificity value as much as determining the magnitude of the *t* value. Different definitions of the *AUC-t* arise from two cases are as follows.

Case I: Let D^+ and D^- represent positive and negative diagnoses, respectively. In the same way let T^+ and T^- denote the test results. In situations such as preventing epidemic diseases or detecting a fatal disease, making a FN decision has more hazardous results than making a FP decision. It is very dangerous to make non-diseased diagnosis to a person who is actually diseased in such cases. Here the test is required to be as sensitive as possible. In other words, Type II error (β) should be minimized. Here, the *AUC-t* can be defined as (2)

$$AUC\text{-}t=P(X+t>Y)=\int_y P(X > y-t \mid Y=y)dG(y)=\int_y (1-F(y-t))dG(y) \qquad (2)$$

The recommended critical value is the value changed by *t* value corresponding to γ100%, where γ is a certain probability. Here, it is suggested to make a one-sided biased classification in healthy individuals. The probability of a person who is assessed as non-diseased, but actually the person is diseased (denoted by $P_{(D^+|T^-)}$) can be defined as follows while c being the critical value.

$$\begin{aligned} P_{(D^+|T^-)} &= P(Y > c-t \mid Y < c) = \frac{P(Y > c-t, Y < c)}{P(Y < c)} \\ &= \frac{P(c-t < Y < c)}{P(Y < c)} = \frac{G(c)-G(c-t)}{G(c)} \\ &= 1-\frac{G(c-t)}{G(c)} = 1-\frac{1-p_t}{1-p} = \frac{p_t-p}{1-p} \end{aligned} \qquad (3)$$

where, $p_t = 1-G(c-t)$.

The *t*-value is called the safety threshold, which is obtained by equalizing the probability of $P_{(D^+|T^-)}$ to a certain probability gamma, γ. Thus, in order not to ignore the persons who are actually diseased, we defined a group of non-diseased persons as diseased. By increasing the Type I error (α), the safety threshold value is assigned to the cut-off point for the decision of diagnosis. This modification improves the reliability of identifying patients. However, precision and specificity are reduced.

Case II: Making FP decisions leads to more dangerous results than FN decision making when there are severe side effects of treatment and costly treatment diseases exist. Also applying the treatment, which has more severe side effects may bring heavy consequences. In this case the probability of a true negative (specificity) should be kept high. Especially, the diagnostic tests with moderate differentiation are more likely to diagnose the person as diseased who is actually healthy. Therefore, Type I error rate is higher for such tests. In this case, diagnosis is made by considering a critical value such as *c+t* in the diagnostic test value for the diseased persons. The critical value recommended for the diagnostic test value for diseased persons is modified by the value of *t* corresponding to *100%*γ, where γ is a certain probability. Here, it is recommended to perform a one-sided biased classification again in the diseased persons.

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be the random sample from the diseased and non-diseased population having cumulative distribution functions *F* and *G*, respectively. In this case, the sensitivity is defined as $P(X\text{-}t>c)=1\text{-}F(c+t)$ and the modified value of *q* is denoted by $q_t=1-F(c+t)$ for all possible cut-off points *c*. Thus, the *AUC-t* is defined as (4)

$$AUC\text{-}t=P(X\text{-}t>Y)=\int_y P(X > y+t \mid Y=y)dG(y)=\int_y (1-F(y+t))dG(y). \qquad (4)$$

The probability that a randomly selected person is actually assessed as non-diseased while diseased denoted by $P_{T^-|D^+}$, *c* being a critical value, can be defined as follow (5).

$$P_{T^-|D^+} = P(X < c+t | X > c) == \frac{F(c+t)-F(c)}{1-F(c)} = 1-\frac{1-F(c+t)}{1-F(c)} = 1-\frac{q_t}{q} = \frac{q-q_t}{q}, \quad (5)$$

where $q_t = 1-F(c+t)$.

The *t*-value is entitled as the safety threshold, which obtained by equating the probability of $P_{(T^-|D^+)}$ to a certain probability gamma, γ. In order not to ignore the persons who are actually non-diseased, we define some people as non-diseased from the group, which described as diseased. By increasing Type II error, a safe threshold value is assigned to the decision of the diagnosis. The increase in error improves the reliability of detecting a non-diseased person, but sensitivity is reduced in this situation. In case II, we want to determine actual bound for disease and safety standards for healthy subjects and it can allow decision-maker to know extra safety threshold *t*.

SIMULATION STUDY

We conducted a simulation study to evaluate the effect of safety threshold on the performance of the diagnostic tests using R programming language. For this purpose, we considered the combinations of 2 factors in the simulation set up such that,

- AUC levels – {0.65 and 0.85} as the *low* and *good* discrimination ability,
- Cut-off point determination method – {Youden and maximum efficiency}.

We repeated simulation study using above-mentioned scenarios for the Case I and Case II as described in the Material and Methods section. Since we assumed the values were normally distributed in diseased and healthy groups, sensitivity and specificity values for changing cut-off points were obtained from the cumulative normal distribution. ROC curve and the AUC of the diagnostic test were obtained using the parametric binormal ROC approach. Distributional parameters in the data generation were μ_1=240; σ_1=20; μ_2=229; σ_2=20 for diseased and healthy groups, respectively. The optimal cut-off point is parametrically estimated using one of the optimal cut-off point calculation methods. Assuming that a test result exceeding the cut-off value indicates that a person is diseased, one can shift the cut-off point *c* towards *c-t* to increase the sensitivity of a test such as in Case I. Similarly, the optimal cut-off point is shifted to *c+t* if the increase in specificity is of interest, as in Case II. The amount of shifting, i.e. the length of the interval [*c-t, c*], or [*c, c+t*] depends on pre-defined conditions such as amount of increase or decrease in sensitivity/specificity, relative change in the value of optimal criterion and the probability of the interval. The length of the interval (or the amount of shifting) is determined by considering a pre-defined probability gamma (γ). For each case, changing value γ is considered 0.05 as given in all tables. Within the interval [c, c + t] or [c – t, c], we evaluated the sensitivity and specificity of the test at 10 different cut-off points, and calculated the relative efficiency (RE) of the diagnostic test against the performances at the optimal cut-off point *c* without safety threshold.

RESULTS

In this study, we used Youden Index and maximum efficiency methods to find the optimal cut-off value of a diagnostic test. The Youden Index (*J*) is a function of sensitivity *(q)* and specificity *(1-p)* and generally used to measure overall diagnostic effectiveness.[28,29] *J* ranges between 0 and 1, higher value indicates that the diagnostic test's effectiveness is relatively large. *J* is calculated with (6) over all cut-off points of *c*, which ranges from -∞ to ∞,

$$J = \text{maximum}\{\text{sensitivity}(c) + \text{specificity}(c) - 1\}. \quad (6)$$

The other method, the maximum efficiency, determines a threshold value at the point where there was the greatest difference between the FPR and TPR across all possible threshold values.[30] Maximum efficiency

is calculated as in (7) where x_i is the percentage of true positives at threshold i, and y_i is the percentage of false positives at threshold i.

$$\text{Maximum efficiency} = \text{Max}[x_i - y_i] \quad (7)$$

SIMULATION RESULTS FOR CASE I

This case includes critical situations where making a FN decision is inconvenient or dangerous, such as preventing epidemics or detecting a deadly disease. Hence, the cut-off point c is shifted to c-t to increase the sensitivity of the test in this case assuming a test result exceeding the cut-off value indicates that a person is not healthy. Thus, in order not to ignore the persons who are actually diseased, we defined a group of non-diseased persons as diseased. In this simulation study, we investigate the performance of the shifted c using threshold value t, i.e. [c-t, c], for Case I. Recall that the t is called the safety threshold, which is obtained by equalizing the probability of $P_{(D^+|T^-)}$ given in Equation (3) to a certain probability, γ=0.05. Table 2 gives the simulation results for cut-off points, sensitivity, specificity, criterion and RE using Youden Index and maximum efficiency methods for γ=0.05. For the maximum efficiency method, we chose the prevalence as 0.30. The values in bold in the cut-off column in Table 2 are the optimal c value obtained by methods. The new cut-off value c-t is obtained as 221.4362 and 233.7068 for the cases where the AUC value is 0.86 and 0.65, respectively.

TABLE 2: Shifted cut-off points of a diagnostic test for both Youden Index and maximum efficiency methods with different AUC values (γ: 0.05).

AUC	Youden Index					Maximum efficiency (prevalence: 0.30)				
	Cut-off	Sens.	Spec.	Criterion	RE	Cut-off	Sens.	Spec.	Criterion	RE
0.86	221.4362	0.8233	0.7163	0.5396	0.9870	233.7068	0.6235	0.8821	0.8045	0.9981
	221.8115	0.8184	0.7226	0.5410	0.9896	233.9792	0.6183	0.8847	0.8048	0.9985
	222.1869	0.8134	0.7289	0.5423	0.9919	234.2516	0.6131	0.8874	0.8051	0.9988
	222.5622	0.8084	0.7350	0.5434	0.9939	234.5240	0.6079	0.8899	0.8053	0.9991
	222.9375	0.8032	0.7411	0.5443	0.9956	234.7964	0.6026	0.8925	0.8055	0.9994
	223.3128	0.7980	0.7472	0.5451	0.9971	235.0688	0.5974	0.8950	0.8057	0.9996
	223.6882	0.7926	0.7531	0.5458	0.9982	235.3412	0.5921	0.8974	0.8058	0.9997
	224.0635	0.7872	0.7590	0.5463	0.9991	235.6136	0.5868	0.8998	0.8059	0.9999
	224.4388	0.7817	0.7648	0.5466	0.9997	235.8860	0.5815	0.9022	0.8060	1.0000
	224.8142	0.7762	0.7706	0.5467	1.0000	236.1584	0.5762	0.9046	0.8060	1.0000
	225.0018	0.7733	0.7734	0.5467	1.0000	**236.2946**	0.5735	0.9057	0.8060	1.0000
0.65	231.8389	0.6584	0.5564	0.2148	0.9914	260.4907	0.1528	0.9423	0.7055	0.9985
	232.1188	0.6532	0.5620	0.2152	0.9931	260.9987	0.1469	0.9452	0.7057	0.9988
	232.3987	0.6481	0.5675	0.2155	0.9946	261.5066	0.1411	0.9480	0.7059	0.9991
	232.6785	0.6428	0.5730	0.2158	0.9960	262.0145	0.1355	0.9506	0.7061	0.9993
	232.9584	0.6376	0.5784	0.2161	0.9971	262.5225	0.1301	0.9531	0.7062	0.9995
	233.2383	0.6324	0.5839	0.2163	0.9981	263.0304	0.1248	0.9556	0.7063	0.9997
	233.5182	0.6271	0.5894	0.2164	0.9988	263.5384	0.1196	0.9579	0.7064	0.9998
	233.7981	0.6218	0.5948	0.2166	0.9994	264.0463	0.1146	0.9601	0.7065	0.9999
	234.0780	0.6164	0.6002	0.2166	0.9998	264.5542	0.1098	0.9623	0.7065	1.0000
	234.3579	0.6111	0.6056	0.2167	1.0000	265.0622	0.1051	0.9643	0.7065	1.0000
	234.4978	0.6084	0.6083	0.2167	1.0000	**265.3161**	0.1028	0.9653	0.7065	1.0000

AUC: Area under the curve; Sens: Sensitivity; Spec: Specificity; RE: Relative efficiency.

SIMULATION RESULTS FOR CASE II

When the treatment has serious side effects and/or high treatment costs, fatal results can occur. In particular, diagnostic tests with moderate accuracy are more likely to diagnose a healthy person as having the disease. Therefore, the Type I error rate is higher in such tests. In this case, the probability of a true negative (specificity) should be kept high. In this case, the diagnosis is made by considering a critical value such as c+t in the diagnostic test value for people with the disease. Thus, reducing the Type I error rate results in a lower FP decision. Similar to Case I, we can find the *t*-value which is obtained by equating the probability of $P_{(T^-|D^+)}$ in Equation (5) to a certain probability γ. Table 3 and Table 4 give simulation results for cut-off points of a diagnostic test with different AUC values with the specificity increased by 5% for Youden Index method and maximum efficiency method respectively. In Table 3, the values 225.0018 and 228.5680 in the cut-off column is the optimal and shifted cut-off values respectively obtained by the Youden Index method. In Table 4, the values in bold in the cut-off column are the optimal *c* values obtained by the maximum efficiency when the sensitivity is increased with a gamma probability of 0.05. For the prevalence is 0.40, the shifted cut-off values *c+t* is obtained as 235.2038 and 253.9945 for the cases where the AUC value is 0.86 and 0.65, respectively. When the prevalence is increased, the cut-off values are decreased.

TABLE 3: Shifted cut-off points of a diagnostic test (AUC: 0.86, Method: Youden, γ: 0.05).

Cut-off	Sens.	Spec.	Criterion	RE
228.5680	0.7162	0.8234	0.5396	0.9869
228.1926	0.7225	0.8185	0.5410	0.9895
227.8172	0.7288	0.8135	0.5423	0.9918
227.4418	0.7350	0.8084	0.5434	0.9939
227.0665	0.7411	0.8033	0.5443	0.9956
226.6911	0.7471	0.7980	0.5451	0.9971
226.3157	0.7531	0.7927	0.5458	0.9982
225.9403	0.7590	0.7873	0.5462	0.9991
225.5649	0.7648	0.7818	0.5466	0.9997
225.1895	0.7705	0.7762	0.5467	1.0000
225.0018	0.7733	0.7734	0.5467	1.0000

AUC: Area under the curve; Sens: Sensitivity; Spec: Specificity; RE: Relative efficiency.

DISCUSSION

For the Case I, the results in Table 2 show that working with the *c-t* value does not significantly affect the RE especially when the test is less accurate and has low AUC value. Therefore, in tests with low discrimination power, changing the value of *c* as much as γ probability can lead to a more tolerant decision. For the same AUC values, the cut-off point, specificity and criterion values, except sensitivity, are lower in the Youden Index method compared to the maximum efficiency method.

For Case II, the shifted cut-off value (*c+t*) is obtained as 228.5680 by the Youden index method. It was observed from Table 3 that the specificity increased to 0.8234, while the RE decreased. For the maximum efficiency method with 0.40 prevalence, the shifted cut-off values are obtained as 235.2038 and 253.9945 for 0.86 and 0.65 AUC values, respectively. When the prevalence is increased, the cut-off values are decreased.

Maximum efficiency method determined an efficient cut-off value to be highly sensitive in the case of increased prevalence.

TABLE 4: Shifted cut-off points of a diagnostic test for maximum efficiency method with different AUC values and prevalence (γ: 0.05).

AUC	Maximum efficiency									
	Prevalence: 0.40					Prevalence: 0.60				
	Cut-off	Sens.	Spec.	Criterion	RE	Cut-off	Sens.	Spec.	Criterion	RE
0.86	235.2038	0.5948	0.8962	0.7756	0.9926	222.5174	0.8090	0.7343	0.7791	0.9970
	234.6987	0.6045	0.8916	0.7768	0.9940	222.2098	0.8131	0.7292	0.7796	0.9976
	234.1937	0.6142	0.8868	0.7778	0.9953	221.9022	0.8172	0.7241	0.7800	0.9981
	233.6886	0.6238	0.8819	0.7787	0.9964	221.5946	0.8213	0.7190	0.7804	0.9986
	233.1835	0.6334	0.8768	0.7794	0.9974	221.2870	0.8253	0.7137	0.7807	0.9990
	232.6784	0.6428	0.8716	0.7801	0.9983	220.9794	0.8292	0.7085	0.7809	0.9993
	232.1733	0.6522	0.8662	0.7806	0.9990	220.6718	0.8331	0.7032	0.7811	0.9996
	231.6682	0.6615	0.8607	0.7810	0.9995	220.3642	0.8369	0.6978	0.7813	0.9998
	231.1631	0.6707	0.8550	0.7813	0.9998	220.0566	0.8407	0.6925	0.7814	0.9999
	230.6580	0.6798	0.8492	0.7814	1.0000	219.7489	0.8444	0.6870	0.7814	1.0000
	230.4054	0.6843	0.8462	0.7814	1.0000	**219.5951**	0.8462	0.6843	0.7814	1.0000
0.65	253.9945	0.2420	0.8943	0.6334	0.9969	222.4688	0.8096	0.3720	0.6346	0.9988
	253.4946	0.2499	0.8897	0.6338	0.9975	222.1834	0.8135	0.3666	0.6347	0.9990
	252.9948	0.2579	0.8849	0.6341	0.9980	221.8980	0.8173	0.3613	0.6349	0.9993
	252.4949	0.2661	0.8800	0.6344	0.9985	221.6127	0.8210	0.3559	0.6350	0.9994
	251.9950	0.2743	0.8749	0.6347	0.9989	221.3273	0.8248	0.3506	0.6351	0.9996
	251.4952	0.2827	0.8697	0.6349	0.9993	221.0419	0.8284	0.3454	0.6352	0.9997
	250.9953	0.2912	0.8643	0.6351	0.9995	220.7565	0.8320	0.3401	0.6353	0.9998
	250.4954	0.2999	0.8588	0.6352	0.9998	220.4712	0.8356	0.3349	0.6353	0.9999
	249.9956	0.3086	0.8531	0.6353	0.9999	220.1858	0.8391	0.3297	0.6353	1.0000
	249.4957	0.3175	0.8473	0.6353	1.0000	219.9004	0.8425	0.3246	0.6354	1.0000
	249.2458	0.3219	0.8443	0.6354	1.0000	**219.7577**	0.8443	0.3220	0.6354	1.0000

AUC: Area under the curve; Sens: Sensitivity; Spec: Specificity; RE: Relative efficiency.

There are some studies in the literature that focus on creating different models for AUC to be used for specific purposes and in critical situations. Karaismailoglu et al. have used the change in the area under the receiver operating characteristic curve (ΔAUC) to investigated the impact of correlation structure, prevalence and effect size on the risk prediction model.[31] Thompson and Zucchini and McClish have suggested the partial area under ROC curve (pAUC) as an alternative to AUC for evaluating a diagnosis.[32,33] pAUC is used when a restricted region of the ROC curve is clinically relevant for a critical disease, for example in cancer and the high specificity is necessary. However, pAUC presents a specific range of ROC curve and it ignores the information on the other portion of AUC. Weng and Poon defined a weighted AUC which consider the majority class in severe imbalanced datasets.[18] Alternatively, Yu et al. propose a modified AUC (mAUC) as a weighted average of 2 AUCs.[19] Our AUC model considers a threshold value (t) for subjects to provide extra security to the decision maker for 2 different critical situations. Perkins and Schisterman have demonstrated the intuitive similarity of the 2 criteria used to choose an "optimal" cut-off point.[34] They showed that the criteria agreed in some cases and disagreed in others. Most importantly, cut-off points chosen through criteria that are less than "optimal" or "optimal" in an arbitrary sense can lead to unnecessary misclassifications, resulting in needlessly missed opportunities for disease diagnosis and intervention, they noted. Therefore, we believe that the safe threshold value suggested in our study will provide an important

support to the decision makers for the diagnosis and intervention of the disease. As Habibzadeh et al. points out, costs of FP and FN results from misdiagnosis must be considered for the cut-off value.[17] They also indicate that the cut-off value for a particular diagnostic test is not universal and should be determined for each region and each disease condition. The results of our study showed that the RE is not significantly affected by working with the safety threshold, *t* value. Therefore, changing the value of *c* as much as γ probability can lead to a more tolerant decision depending on the value of the discrimination power.

CONCLUSION

Different parametric and nonparametric approaches for AUC have been compared in various studies. However, none of these studies evaluated the AUC with a safety threshold (*t*) for diagnostic testing for both diseased and non-diseased persons to provide extra safety to the decision-maker for 2 different critical situations. The proposed method considers a new definition of AUC giving a threshold value (*t*) for the diagnostic test for subjects to provide extra safety to the decision-maker for 2 different critical situations. We conducted a simulation study to find the optimal and shifted cut-off values using Youden Index and maximum efficiency methods for a diagnostic test in 2 critical cases. We calculated the shifted cut-off points under the assumption that a test result exceeds the cut-off value indicates that a person is diseased. It is seen that working with shifted cut off point causes a decrease in RE for both cases. However, considering the desired increases in sensitivity and specificity values depending on the cases, it provides an advantage in critical situations.

This study is mainly focused on the parametric ROC curves. Therefore, the proposed method does not reflect the findings from a nonparametric ROC curve such as Mann-Whitney methodology. We leave this topic as the limitation of this study and also a further research topic.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and/or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

Idea/Concept: *Selma Gürler, Özlem Ege Oruç;* ***Design:*** *Selma Gürler, Özlem Ege Oruç, Dinçer Göksülük;* ***Control/Supervision:*** *Özlem Ege Oruç, Aslı Suner;* ***Analysis and/or Interpretation:*** *Selma Gürler, Dinçer Göksülük;* ***Literature Review:*** *Özlem Ege Oruç, Aslı Suner;* ***Writing the Article:*** *Selma Gürler, Özlem Ege Oruç, Dinçer Göksülük, Aslı Suner;* ***Critical Review:*** *Selma Gürler, Özlem Ege Oruç.*

REFERENCES

1. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ. 1994;309(6948):188. [Crossref] [PubMed] [PMC]
2. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8(4):283-98. [Crossref] [PubMed]
3. Obuchowski NA, Lieber ML, Wians FH Jr. ROC curves in clinical chemistry: uses, misuses, and possible solutions. Clin Chem. 2004;50(7):1118-25. [Crossref] [PubMed]
4. Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988;240(4857):1285-93. [Crossref] [PubMed]
5. Swets JA, Dawes RM, Monahan J. Better decisions through science. Sci Am. 2000;283(4):82-7. [Crossref] [PubMed]
6. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review for radiologists. Korean J Radiol. 2004;5(1):11-8. [Crossref] [PubMed] [PMC]
7. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993;39(4):561-77. [Crossref] [PubMed]
8. Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. Invest Radiol. 1979;14(2):109-21. [Crossref] [PubMed]

9. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36. [Crossref] [PubMed]
10. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med. 2013;4(2):627-35. [PubMed] [PMC]
11. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. Journal of Mathematical Psychology. 1975;12(4):387-415. [Crossref]
12. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45. [Crossref] [PubMed]
13. Faraggi D, Reiser B. Estimation of the area under the ROC curve. Stat Med. 2002;21(20):3093-106. [Crossref] [PubMed]
14. Schisterman EF, Faraggi D, Reiser B, Hu J. Youden Index and the optimal threshold for markers with mass at zero. Stat Med. 2008;27(2):297-315. [Crossref] [PubMed] [PMC]
15. Liu X. Classification accuracy and cut point selection. Stat Med. 2012;31(23):2676-86. [Crossref] [PubMed]
16. López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C, Gude-Sampedro F. Optimal Cutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. J Stat Soft. 2014;61(8):1-36. [Crossref]
17. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. Biochem Med (Zagreb). 2016;26(3):297-307. [Crossref] [PubMed] [PMC]
18. Weng CG, Poon J. A new evaluation measure for imbalanced datasets. Proceedings of the 7th Australasian Data Mining Conference. Australia: 27-28 November 2008 V87. (AusDM'08). p.27-32. [Link]
19. Yu W, Chang YI, Park E. A modified area under the ROC curve and its application to marker selection and classification. J Korean Stat Soc. 2014;43(2):161-75. [Crossref]
20. Yu W, Park T. Two simple algorithms on linear combination of multiple biomarkers to maximize partial area under the ROC curve. Computational Statistics & Data Analysis. 2015;88:15-27. [Crossref]
21. Hlatky MA, Pryor DB, Harrell FE Jr, Califf RM, Mark DB, Rosati RA. Factors affecting sensitivity and specificity of exercise electrocardiography. Multivariable analysis. Am J Med. 1984;77(1):64-71. [Crossref] [PubMed]
22. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med. 1997;16(9):981-91. [Crossref] [PubMed]
23. Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. Epidemiology. 1997;8(1):12-7. [Crossref] [PubMed]
24. Ridker PM, Cook NR, Lee IM, Gordon D, Gaziano JM, Manson JE, et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. N Engl J Med. 2005;352(13):1293-304. [Crossref] [PubMed]
25. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007;115(7):928-35. [Crossref] [PubMed]
26. Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. Acta Paediatr. 2007;96(5):644-7. [Crossref] [PubMed]
27. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med. 2000;45(1-2):23-41. [Crossref] [PubMed]
28. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J. 2005;47(4):458-72. [Crossref] [PubMed]
29. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. Epidemiology. 2005;16(1):73-81. [Crossref] [PubMed]
30. Lippitt CD, Rogan J, Toledano J, Sangermano F, Eastman JR, Ronald J, et al. Incorporating anthropogenic variables into a species distribution model to map gypsy moth risk. Ecological Modelling. 2008;210:339-50. [Crossref]
31. Karaismailoglu E, Konar NM, Goksuluk D, Karaagaoglu AE. Factors effecting the model performance measures area under the ROC curve, net reclassification improvement and integrated discrimination improvement. Communications in Statistics-Simulation and Computation. 2019;48(9):2586-98. [Crossref]
32. Thompson ML, Zucchini W. On the statistical analysis of ROC curves. Stat Med. 1989;8(10):1277-90. [Crossref] [PubMed]
33. McClish DK. Analyzing a portion of the ROC curve. Med Decis Making. 1989;9(3):190-5. [Crossref] [PubMed]
34. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol. 2006;163(7):670-5. [Crossref] [PubMed] [PMC]