Lessons in biostatistics

On determining the most appropriate test cut-off value: the case of tests with continuous results

Farrokh Habibzadeh*[1,2], Parham Habibzadeh[3], Mahboobeh Yadollahie[2]

[1]Adjunct Professor, Shiraz University of Medical Sciences, Shiraz, Iran,
[2]R&D Headquarters, Petroleum Industry Health Organization, Shiraz, Iran,
[3]Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran

*Corresponding author: Farrokh.Habibzadeh@theijoem.com

Abstract

There are several criteria for determination of the most appropriate cut-off value in a diagnostic test with continuous results. Mostly based on receiver operating characteristic (ROC) analysis, there are various methods to determine the test cut-off value. The most common criteria are the point on ROC curve where the sensitivity and specificity of the test are equal; the point on the curve with minimum distance from the left-upper corner of the unit square; and the point where the Youden's index is maximum. There are also methods mainly based on Bayesian decision analysis. Herein, we show that a proposed method that maximizes the weighted number needed to misdiagnose, an index of diagnostic test effectiveness we previously proposed, is the most appropriate technique compared to the aforementioned ones. For determination of the cut-off value, we need to know the pretest probability of the disease of interest as well as the costs incurred by misdiagnosis. This means that even for a certain diagnostic test, the cut-off value is not universal and should be determined for each region and for each disease condition.

Key words: ROC curve; diagnostic tests; sensitivity; specificity; prevalence

Introduction

Diagnostic tests are important clinical tools. If that is possible, we have to use gold-standard tests for the diagnosis of diseases. However, a gold-standard test either does not exist or is very difficult or expensive to perform for certain disease conditions (1). Therefore, we have to use alternative diagnostic tests as surrogates for gold-standard tests.

While interpretation of a test with binary results is straight forward, interpretation of a test with continuous results is not that simple. For instance, assume that the test is for discrimination of only two states, "diseased" (D^+) and "non-diseased" (D^-), and that the higher test values are more likely among D^+ persons. For discrimination of D^+ and D^- people, we need to set a cut-off value; test results equal to or greater than this value are considered positive (T^+), otherwise they are negative (T^-). The choice of the cut-off value determines the rates of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) test results (2). The sensitivity (Se) of a test is defined as the probability of a positive test (T^+) in a diseased person (D^+), that is (3):

$$Se = \frac{TP}{TP + FN}$$

The test specificity (Sp) is defined as the likelihood of a negative test (T^-) in a person without the disease (D^-), that is (3):

$$Sp = \frac{TN}{TN + FP}$$

Therefore, a sensitive test has a low FN rate – a negative result (T⁻) is very likely TN. Therefore, a sensitive test can be used to rule out a disease condition. Similarly, having a low FP rate, a specific test can be used to rule in a disease.

In a test with continuous (or multiple) results, every possible test value can be considered a cut-off point. This cut-off value determines the test Se and Sp. However, for a given test, we cannot increase the Se and Sp concomitantly; Se will be enhanced at the expense of Sp and *vice versa*. Decreasing the cut-off value to increase the test Se causes the Sp to decrease. If you want to have a more specific test (by increasing the cut-off value), you will have a less sensitive test.

Receiver Operating Characteristic (ROC) curve analysis

One of the most commonly used methods to analyze the effectiveness of a diagnostic test is receiver operating characteristic (ROC) curve analysis (4-6). Use of this method dates back to World War II when the ability of radar operators (receivers) was tested to determine whether a blip on the radar screen represented an object (signal, a TP result) or noise (a FP result), hence, the name (7). Several years later, the method was found useful in many other scientific disciplines including diagnostic medicine where a physician should discriminate a TP from a FP test result. The ROC curve offers a graphical illustration of the above-mentioned trade-off between a test Se and Sp and depicts TP rate (Se) against FP rate (1 - Sp) for each cut-off value (7).

The general structure of a ROC curve is simple. The curve is confined in a unit square (Figure 1). The left-lower corner (Se = 0, Sp = 1) corresponds to the highest possible test cut-off value. As the cut-off value decreases, the test Se increases and Sp decreases, moving on the curve from the left-lower corner up and to the right to ultimately reach the right-upper corner of the square where Se = 1 and Sp = 0, corresponding to the lowest possible

test cut-off value. In theory, we can think of a continuous curve with infinite number of points. However, in real world, a ROC curve is constructed based on a few discrete points. Although we can connect these points using various methods (line segments, spline, curve fitting, *etc*.), the curve is not differentiable and thus, in practice it is not possible to determine the exact slope at any point.

In a perfect test, both Se and Sp are equal to 1. The ROC curve corresponding to a perfect (i.e. the gold-standard) test is a line segment connecting the left-lower corner to the left-upper corner and to the right-upper corner (a curve coinciding with the left and top sides) of the unit square (8). On the other hand, the ROC curve corresponding to a test with no diagnostic value is the line segment connecting the lower-left corner to the right-upper corner – the 45° diagonal line (Figure 1). In practice, the curve lies somewhere between these two extremes. The area under the ROC curve (AUC) varies between 0.5 (for the 45° diagonal line representing an uninformative test) and 1.0 for a perfect test.

The AUC can be considered an index of discriminating ability of a test (1,8). Mathematically, the area is equivalent to the probability that the test result measured in a randomly selected D⁺ person is higher than that measured in a D⁻ person (7). A test with an AUC of 0.5 is equivalent to tossing a coin – an uninformative test. AUC is particularly useful when two or more diagnostic tests are compared. Having a higher AUC, a test with a ROC curve that lies completely above another curve, is clearly a better one (Figure 1). The methods for the calculation of the AUC are mainly based on a non-parametric statistical test, the Wilcoxon rank-sum test, proposed by DeLong *et al.* and Hanley *et al.* (8-10). The proposed methods can be used to test if the AUC of a curve is significantly higher than 0.5 (the AUC of an uninformative test), or to compare AUCs of two or more tests.

Criteria for selecting the most appropriate cut-off value

Choosing an appropriate cut-off value is of paramount importance in using a test effectively. Sev-
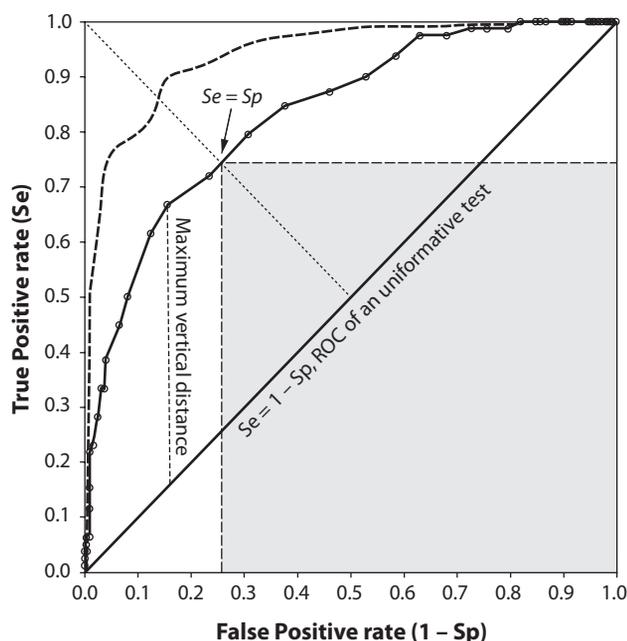
FIGURE 1. The general structure of a ROC curve. The curve (dashed line) which lies completely above another curve (solid line), is clearly a better test because it has a higher area under the curve. Having the left-upper corner moving on ROC curve (solid line), the area of the shaded rectangular region is maximum when its sides (Se and Sp) are equal. Se – sensitivity. Sp – specificity.

eral criteria, mostly based on ROC analysis, have so far been proposed for choosing the most appropriate cut-off value (2,5,11-13). Each point on a ROC curve corresponds to a cut-off value and is associated with a test Se and Sp. Locating the cut-off point thus requires a compromise between Se and Sp. In some cases, Se is more important than Sp, for example when a disease is highly infectious or associated with serious complications. On the other hand, in certain circumstances, Sp may be preferred over Se, say when the subsequent diagnostic testing is risky or costly (2). If there is no preference between Se and Sp, nonetheless, a reasonable approach would be to maximize both indices.

The lowest cut-off value corresponds to a Se = 1 and Sp = 0. As the cut-off value increases, the test Se decreases and the test Sp increases until a cut-off value corresponding to a test Se = 0 and Sp = 1. Over this interval, there is a cut-off value where the test Se is equivalent to the test Sp. One of the

frequently used criterion for determination of the test cut-off value is the one corresponding to this particular point, where Se = Sp. This point is mathematically the intersection of the line connecting the left-upper corner and the right-lower corner of the unit square (the line Se = Sp), and the ROC curve (Figure 1). This point of the curve is where the product of these two indices (Se x Sp) is maximum – the area of the shaded rectangle in Figure 1 is maximum when its sides (Se and Sp) are equal, a square.

Another approach to maximize both Se and Sp would be to maximize their summation (Se + Sp). At this point, the Youden's index (Se + Sp – 1) is also maximum (11,14-16). This is a commonly used technique to determine the most appropriate cut-off value and corresponds to a point on the ROC curve with the highest vertical distance from the 45° diagonal line (the ROC of an uninformative test). At this point, the difference between the test TP rate (Se) and FP rate (1 – Sp) is maximum too (15).

The ROC of a perfect test passes through the left-upper corner of the unit square, the point where both Se and Sp are equal to 1 (a perfect test; the gold-standard). The closer a curve to this point, the better is a test. No surprise, another common criterion for choosing the most appropriate cut-off value is selecting the point on the ROC curve with the minimum distance from the left-upper corner of the unit square (8,15,16).

Although the aforementioned criteria are based on various assumptions and their usefulness is merely dependent on the validity of the presumptions made in the practical setting, some researchers prefer one method to another. For example, Perkins and Schisterman recommend the use of the Youden's index and warn about the use of the point with the minimum distance from the left-upper corner (16). Nonetheless, selection of the criterion to be used should be based on the situation the test to be applied and the importance of the test Se compared to Sp. For example in designing a screening test, we need a high enough Se, say 0.8 or more, to reduce the FN rate. Otherwise, many diseased persons will be missed.

All these methods are simple to use. However, in all of the above-mentioned methods, we inclusively assume that there is no difference between a FN and FP result. Neither do we consider the prior probability of the disease in question. Taking into account these variables, expectedly, makes the equations more complex (and hopefully more precise). This leads us to a related topic – the Bayesian decision analysis (17).

Bayesian approach in determining the cut-off value

Using a Bayesian approach, the odds of a disease before and after a diagnostic test can generally be related as follows:

$$\text{Post-test odds of } D^+ = \text{Pre-test odds of } D^+ \times \text{Bayes factor,}$$

where "Bayes factor" can be derived based on our assumptions. The Bayesian approach provides us with the information about how a test result would change the odds (and thus probability) of a disease (18).

The Bayes factor can be determined in various ways. For example, if we maximize the patient's expected utility for determination of the Bayes factor in the above equation, we come up to a condition suggesting that the most appropriate cut-off value corresponds to a point on the ROC curve where the slope of the tangent line to the curve satisfies the following equation (2,5):

$$\text{Slope of ROC curve} = \frac{H}{B} \times \frac{1-pr}{pr} \qquad \text{(Equation 1)}$$

where pr represents the pre-test (prior) probability of the disease, H is the net harms of treating people who do not have the disease (the harms of a FP result), and B the net benefit of treating those with the disease (in other words, the harms of a FN result).

The costs associated with harms of a FN and FP test result (B and H, respectively) and medical misdiagnosis have been the subject of growing number of articles (19). The Institute of Medicine (IOM), an American non-profit, non-governmental or-

ganization, reports that about 30% of health care costs spending in the US, around US$ 750 billion, is wasted on unnecessary services (20). In these types of analysis, a decision tree is constructed based on the available treatment options, and current evidence about risks and benefits associated with each option (2,21). Based on this structure, we can then estimate the cost-effectiveness and benefit-risk of making each decision and thus the probable outcome and harms associated with FN and FP results (21-23). Treatment protocols and screening programs are mainly shaped based on the results of such studies (24,25).

A limitation of Equation 1 is however that although it ascertains the slope at the most appropriate point, the point cannot always be easily located. In practice, as mentioned above, a ROC curve is constructed based only on a few discrete (non-differentiable) points (it is really not a continuous curve), and thereby finding the point with the given slope on the curve is generally difficult, if not impossible; we arrive to an approximation at best. Although theoretically correct, the method is not quite handy. It would therefore be feasible if we can figure out the coordination (instead of the slope) of the point on a ROC curve corresponding to the most appropriate cut-off value through an analytical method.

Analytical method for the calculation of the test cut-off value

Previously, we proposed a test index, the so-called "Number Needed to Misdiagnose" (NNM) (26), which is the number of patients who need to be tested in order for one to be misdiagnosed by the test, as follows:

$$NNM = \frac{1}{FN + FP}$$
$$= \frac{1}{pr(1 - Se) + (1 - pr)(1 - Sp)}$$

where pr represents the pre-test probability of the disease. For example, a NNM of 20 for a test means that one out of 20 people tested is misdiagnosed (either FP or FN). The higher the NNM of a test, the

closer is the test to the gold-standard, hence, a better test.

To determine the most appropriate cut-off value we can try to maximize the NNM. In the calculation of the NNM, however, the cost of FN and FP results are assumed equal. The cost of making a wrong diagnosis (either FP or FN) is nonetheless different in general. Note that here, the "cost" is referred to all costs incurred – the financial cost, time wasted on inappropriate treatments, missing the opportunity to cure a diseased person with consequences (complications, morbidities, disabilities, mortalities, *etc.*), and harms of treating people without disease with subsequent emotional harms to the patient, experiencing drug side effects, legal issues, *etc* (27,28). To consider this issue, we can assume that the cost of a FN result (misdiagnosing a D^+ person as D^-) is C times the cost of a FP one (diagnosing a D^- person as D^+) and define "weighted NNM" as follows:

$$\text{Weighted } NNM = \frac{1}{C \times FN + FP}$$

$$= \frac{1}{C \times pr(1 - Se) + (1 - pr)(1 - Sp)}$$

(Equation 2)

For example, if C = 5, then a FN result would cause five times more costs than a FP one; C = 1 means that costs for FN and FP results are equal. Then, to find the most appropriate cut-off value, we can maximize the weighted NNM – to take into account both closeness of the test results to the gold-standard results, and the costs of a misdiagnosis (either FP or FN).

To find an analytical solution for the problem, let f(x) and g(x) designate the probability density function of a hypothetical diagnostic test with continuous results for D^+ and D^- population (Figure 2), respectively. Let the mean and standard deviation (SD) of the distribution be 0, and 1 for D^- people, and d and s for D^+ population, respectively. As it was mentioned earlier, Se and Sp are functions of the cut-off value. For a cut-off value of x, Se and Sp can be calculated as follows:

$$Se(x) = \int_{x}^{+\infty} f(t) \, dt$$
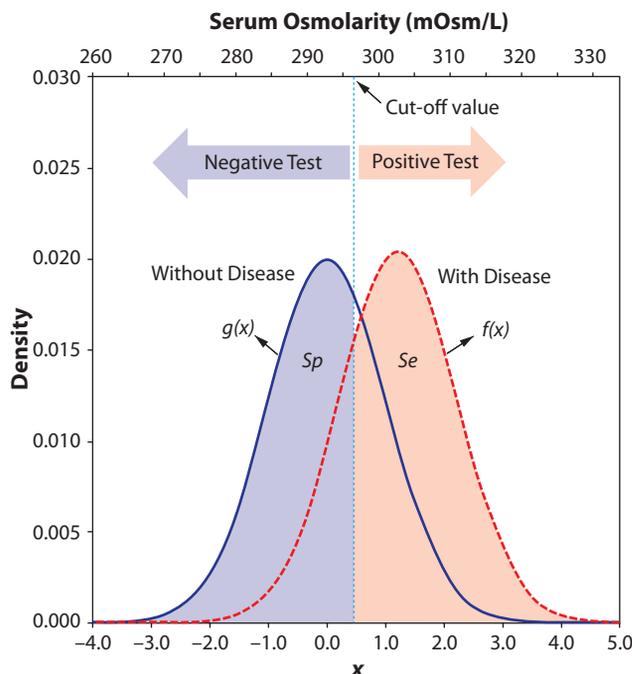
(Equation 3)

FIGURE 2. The probability density functions of a continuous diagnostic test for diseased (f(x), red dashed line) and non-diseased (g(x), blue solid line) persons. g(x) has a mean of 0 and a standard deviation of 1; f(x) has a mean of d and a standard deviation of s. The cut-off value is represented by the vertical dotted line. All test values equal or greater than this value are considered positive (T⁺), else they are considered negative (T⁻). Because f(x) and g(x) are probability density functions, the area under the curve for each of them is equal to one. The area under f(x) to the right of the cut-off value (the pink region) is Se, and the area under g(x) to the left of the cut-off value (the light blue region) is Sp. This figure is drawn based on the first data set (N = 400) presented in the text. There are two x axes: the upper axis indicates serum osmolarity of the studied people; the lower axis represents the corresponding standardized values.

and

$$Sp(x) = \int_{-\infty}^{x} g(t) \, dt$$

(Equation 4)

To maximize the weighted NNM (Equation 2), the denominator of the equation, should be minimized. Using basic calculus, to do so, the following equation should be solved:

$$\frac{\partial \left(\frac{1}{\text{Weighted } NNM} \right)}{\partial x} = -C \times pr \frac{\partial Se}{\partial x} - (1 - pr) \frac{\partial Sp}{\partial x} = 0$$

(Equation 5)

From Equations 3 and 4, we have:

$$\frac{\partial Se}{\partial x} = \lim_{\Delta x \to 0} \frac{\int\limits_{x+\Delta x}^{+\infty} f(t)\,dt - \int\limits_{x}^{+\infty} f(t)\,dt}{\Delta x} = -f(x)$$

$$\frac{\partial Sp}{\partial x} = \lim_{\Delta x \to 0} \frac{\int\limits_{-\infty}^{x+\Delta x} g(t)\,dt - \int\limits_{-\infty}^{x} g(t)\,dt}{\Delta x} = g(x)$$

(Equation 6)

The minus sign before f(x) is because Se is a decreasing function of x; Sp is increasing. Then, Equation 5 becomes:

$$C \times pr \times f(x) - (1 - pr)\,g(x) = 0 \quad \text{(Equation 7)}$$

For simplicity, let g(x) has a normal distribution. Considering its mean and SD are 0 and 1, respectively, we have (29):

$$g(x) = \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$$

Let f(x) also has a normal distribution and taking into account its mean, and SD are d, and s, we have (29):

$$f(x) = \frac{1}{s\sqrt{2\pi}}\,e^{-(x-d)^2/2s^2}$$

Solving Equation 7 for x:

$$\frac{C \times pr}{s\sqrt{2\pi}}\,e^{-(x-d)^2/2s^2} - \frac{1-pr}{\sqrt{2\pi}}\,e^{-x^2/2} = 0$$

gives:

$$x = \frac{s\sqrt{2(s^2-1)\,Ln\left(\dfrac{s(1-pr)}{C \times pr}\right) + d^2} - d}{s^2 - 1} \quad \text{(Equation 8)}$$

if s≠1. If s=1, then x becomes:

$$x = \frac{d}{2} + \frac{Ln\left(\dfrac{(1-pr)}{C \times pr}\right)}{d} \quad \text{(Equation 9)}$$

This value corresponds to the most appropriate test cut-off value.

Generality of the analytical method

Many of the aforementioned commonly used techniques in ROC analysis can be considered special cases of the proposed analytical method (Equations 8 and 9). As an example, if we assume the pre-test probability (pr) is 0.5, if FN and FP costs are equal (C = 1), and if the dispersions (SDs) of the test values for diseased and non-diseased people are equal (s = 1), then the cut-off value predicted by the proposed analytical method (Equation 9 which assumes s = 1), reduces to:

$$x = \frac{d}{2}$$

the value that is obtained from one of the most commonly used approaches to ROC analysis, i.e., a point where Se = Sp.

It can also be shown that the optimum cut-off point derived from the proposed analytical method (Equation 8) has exactly the slope calculated by Equation 1. Using Equation 6, and substituting values for f(x) and g(x), the slope of ROC curve is:

$$\text{Slope of ROC curve} = \frac{\partial Se}{\partial(1-Sp)} = -\frac{\partial Se}{\partial Sp} = \frac{f(x)}{g(x)}$$

$$= \frac{1}{s}\,e^{[-(x-d)^2 - s^2 x^2]/2s^2}$$

Substituting x from Equation 8 (the coordination of the derived cut-off point) in the above equation yields:

$$\text{Slope of ROC curve} = \frac{1}{C} \times \frac{1-pr}{pr}$$

But, 1/C is the cost of a FP result divided by the cost of a FN result, and equals H/B (Equation 1). Therefore, these two methods are technically equivalent. This means that maximizing either patient's expected utility or weighted NNM results in the same cut-off value.

The advantage of the proposed analytical method (Equation 8) over Equation 1, is however, its ease of use: although finding the point on a ROC curve is generally not possible and accurate solely based on the slope of the point (Figure 3), calculation of the cut-off value by the proposed analytical method (Equations 8 and 9) is straight forward – you just need to know the test result means in diseased and non-diseased, SDs, pre-test probability of the disease (an estimate of the disease prevalence, if no other information is available), and an estimate of the costs of FN and FP test results (Equation 1 also needs the last two variables).

Example

To compare the results obtained from different methods for the derivation of the most appropriate test cut-off value, we used the data set provided by Hooper *et al.*, who studied the diagnostic accuracy of calculated serum osmolarity to predict dehydration in people aged 65 years or more (30). They used the directly measured serum/plasma osmolality of 595 participants to determine if they had dehydration (serum/plasma osmolality > 300 mOsm/kg) or not (considered the gold-standard test). They then calculated serum osmolarity for each participant based on their serum sodium, potassium, glucose, and urea by an equation and used the calculated value as the test result. The calculated serum osmolarity was rounded off to the nearest integer value (31). Then, for each cut-off value, the test was compared against the gold-standard test result. The prevalence of dehydration among the studied population was considered 0.19 (30). Hooper *et al.* also estimated that the cost of a FN result (missing a dehydrated person and its health consequences) was five times the cost of a FP result (labelling a person as dehydrated, when he or she is actually not, resulting in a more blood test to directly measure serum osmolality or encouraging them to drink more) (30).

We randomly divided the data set into a 400-person and 195-person subsets. The groups sizes were arbitrary chosen. The first data set was used to calculate the cut-off values using the above-mentioned techniques. The second data set was
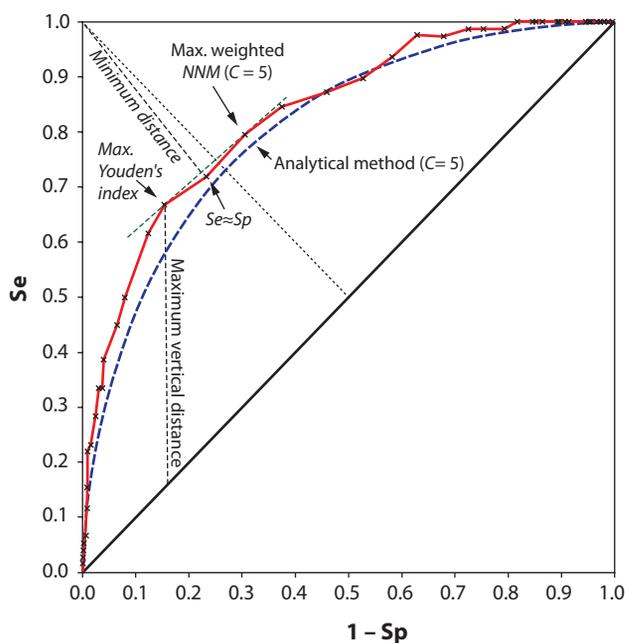

FIGURE 3. The ROC curve constructed based on the first data set (N = 400) presented in the text: the real data set are presented as red solid curve; the values predicted from the proposed mathematical model are presented as the blue dashed curve. The arrows indicate points corresponding to cut-off values derived by various methods (Table 1). The green dashed line is the tangent line with a slope of 0.853. Note that the tangent line intersects the ROC curve at two points. Se – sensitivity. Sp – specificity.

used to test the effectiveness of each method to classify the participants. SPSS® for Windows®, ver. 17 (SPSS Inc, Chicago, IL, USA), was used for dividing the data at random into the two subsets, and data analyses including ROC analysis.

Table 1 shows cut-off values derived by each of the previously described criteria. Theoretically, the intersection of the ROC curve (red solid line) and the line Se = Sp (Figure 3) corresponds to the point where Se = Sp. However, there is no point in our data set satisfying this equation and the closest point is where Se and Sp are 0.718 and 0.767, respectively, corresponding to a serum osmolarity cut-off value of 298 mOsm/L. This point also has the minimum distance from the left-upper corner of the unit square (Figure 3, Table 1). A cut-off value of 299 mOsm/L maximizes the Youden's index (Figure 3, Table 1).

TABLE 1. Test cut-off values calculated based on the first group data set using different criteria

Criterion	Cut-off value (mOsm/L)	Se	Sp	Cost of misdiagnosis (US$) in the second data set (N = 195)
Se = Sp	298	0.718	0.767	10,500
Maximum Youden's index	299	0.667	0.845	11,300
Minimum distance from the left-upper corner of the unit square	298	0.718	0.767	10,500
Slope of the ROC curve (slope = 0.853, C = 5)	?*	?*	?*	?*
Analytical method (C = 5)	297	0.795	0.693	9800
Maximum weighted NNM (C = 5)	297	0.795	0.693	9800

*Cannot be located accurately (see the tangent line in Figure 3).

Because there was no other information about the participants, the best estimate for the pre-test probability was the prevalence of dehydration, 0.19. Based on Equation 1, the slope of the tangent line to the ROC curve at the most appropriate cut-off point is 0.853 (Figure 3, green dashed line), presuming that H/B equals to 1/5, i.e., the costs of harms of a FN result is five times the harms of a FP result (30). However, because of the discrete (non-differentiable) data set, we could not find the corresponding point solely based on knowing its slope (without curve fitting). To figure out the point of interest according to an instruction described previously (4), we passed a line with the slope through the left-upper corner of the unit square and moved it toward the ROC curve (red solid line) until it first intersected the curve. However, the line intersected the curve at two points (Figure 3); practically, it was very hard to locate the point of interest visually with enough accuracy.

The mean serum osmolarity (the test) in the first group (N = 400) was 292.3 (SD 8.2) mOsm/L in 322 participants without dehydration (D⁻), and 302.2 (SD 8.0) mOsm/L in 78 patients with dehydration (D⁺). Using the analytical method we proposed (Equation 8), we have:

$$d = \frac{302.2 - 292.3}{8.2} = 1.21 \text{ and } s = \frac{8.0}{8.2} = 0.98$$

Assuming that the pre-test probability (an estimate of the prevalence of dehydration) is 0.19, if

the cost of a FN result is five times the cost of a FP result (C = 5), Equation 8 yields x = 0.463 corresponding to a cut-off value of 296.1 (292.3 + 0.463 x 8.2) mOsm/L for the serum osmolarity that corresponds to a Se of 0.777 and a Sp of 0.678 (Figure 3, dashed blue curve). Based on the calculation, the closest available cut-off value in our data set is 297 mOsm/L, corresponding to a test Se of 0.795 and a Sp of 0.693 (Table 1). This is where weighted NNM is also maximum (Figure 3, Table 1).

Let the cost of labelling a person as dehydrated, when he or she is actually not, (FP result) be approximately US$ 100 (more blood test to directly measure serum osmolality, encouraging them to drink more, waste of time), and the cost of missing a dehydrated person and its health consequences be about US$ 500. If we use the above-mentioned cut-off values to test the second data set (N = 195) and calculate the costs incurred by FN or FP test results as cost of FN plus cost of FP, the cut-off value obtained by the analytical method and maximizing the weighted NNM (which in this case are the same) is associated with lower costs compared to other methods (Table 1).

Conclusions

The proposed analytical method gives a cut-off value that depends on the pre-test probability of the disease of interest. In the absence of any previous information or test results in a person, the pre-

test probability can be estimated as the prevalence of the disease of interest. According to the proposed method, the cut-off value is higher in places where the disease is less prevalent.

Taking the pre-test probability (or prevalence) of the disease of interest into account would result in major clinical implications. The appropriate cut-off point depends on the place where the test is going to be used. For example, considering Equation 8, the cut-off value for serum osmolarity for the diagnosis of dehydration in a tropical region, where the prevalence of the disease is high, should be lower (a more sensitive test) than that in a cold region, where the prevalence of dehydration is lower – we need a more sensitive test to diagnose dehydration in an endemic area. Even in a given place, the appropriate cut-off value depends on the group of people who need to be tested. For example, the cut-off value for a group of athletes exercising (higher risk/prevalence of dehydration) should be lower than that in general population.

The cut-off value is also different for the diagnosis of diseases with different prevalence rates in a region. As an example, if in a region the prevalence of dehydration is different from the prevalence of diabetes insipidus, if we want to use serum osmolarity as a diagnostic test, we need to set two different cut-off values for the diagnosis of these two conditions. This finding supports the importance of the recommendations of the Clinical and Laboratory Standards Institute (CLSI) and the International Federation of Clinical Chemistry (IFCC) C28-A3 guideline published in 2008, stating that the reference intervals for laboratory analyses should be validated locally, using specimens taken from healthy local people (32,33). Reference intervals are different from clinical decision limits; while the former is based on the test results in the normal population, the latter is a cut-off value derived from one of the above-mentioned methods and is based on test results distribution in both the normal and diseased population (32). Equations 8 and 9 clearly describe this association.

Employing a Bayesian approach, the post-test (posterior) probability of a disease depends on the pre-test probability of the disease and the test re-sult. The post-test probability of a disease after the patient is tested can however be considered the pre-test probability of the next test to be done. Based on what has been presented, the cut-off value of the second test should be different for two patients suspicious for the same disease but having different results on their first test, hence different post-test probabilities.

In our analytical method to derive Equations 8 and 9, we assumed the test results followed a normal distribution for D^+ and D^- persons. This assumption, though supported by extensive data from psychophysical and medical studies (9), may not be true in general. Nevertheless, we have shown that the analytical method proposed, which is based on maximizing the weighted NNM, is mathematically equivalent to Equation 1, the derivation of which is based on maximizing the patient's expected utility (2,5). As maximizing either of patient's expected utility or the weighted NNM would result in the same result, it seems that maximizing the weighted NNM (Equation 2) is the best available method for determination of the most appropriate test cut-off value. This can easily be done by having an estimation of the pre-test probability of the disease, the relative cost of a FN to FP test result (C), and Se and Sp values for each cut-off point, which are readily available in most statistical software output. Using the weighted NNM mentioned above also abolishes the presumption of normal distribution of test values in diseased and non-diseased people.

Only by taking the pre-test probability (prevalence, in lack of other information) of the disease of interest in the study population into account, and considering the cost (not just financial) of FN and FP results, we can find the most appropriate cut-off value for a diagnostic test. All these make it imperative to study more on the prevalence (as an estimate of the pre-test probability in lack of any information) and the cost of FN and FP test results in various populations. Besides the specimen to be analyzed, future autoanalyzers need to be fed with an estimate of pre-test probability (based on the previous test results), the disease of interest, and the associated cost of misdiagnosis. They are

also equipped with a global positioning system so that they can retrieve important relevant data (e.g., prevalence of a disease) to determine if a test is positive or not for a certain disease.

Acknowledgements

Potential conflict of interest

None declared.

References

1. Newman TB, Kohn MA, eds. Evidence-Based Diagnosis. New York: Cambridge University Press, 2009. http://dx.doi.org/10.1017/CBO9780511759512.

2. Sox HC, Higgins MC, Owens DK, eds. Medical Decision Making. 2nd ed. Oxford, UK: Wiley-Blackwell, 2013. http://dx.doi.org/10.1002/9781118341544.

3. Fletcher RH, Fletcher SW. Clinical Epidemiology: The Essentials. 4th ed. Philadelphia: Lippincott Williams & Wilkins, 2005.

4. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993;39:561-77.

5. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978;8:283-98. http://dx.doi.org/10.1016/S0001-2998(78)80014-2.

6. Swets JA. Measuring the accuracy of diagnostic systems. Science 1988;240:1285-93. http://dx.doi.org/10.1126/science.3287615.

7. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ 1994;309:188. http://dx.doi.org/10.1136/bmj.309.6948.188.

8. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-45. http://dx.doi.org/10.2307/2531595.

9. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36. http://dx.doi.org/10.1148/radiology.143.1.7063747.

10. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839-43. http://dx.doi.org/10.1148/radiology.148.3.6878708.

11. Doi SAR. Using and interpreting diagnostic tests with quantitative results. In: Doi S, Williams G, eds. Methods of Clinical Epidemiology. London: Springer; 2013. p.67-78. http://dx.doi.org/10.1007/978-3-642-37131-8_6.

12. McNeil BJ, Keller E, Adelstein SJ. Primer on certain elements of medical decision making. N Engl J Med 1975;293:211-5. http://dx.doi.org/10.1056/NEJM197507312930501.

13. Metz CE, Goodenough DJ, Rossmann K. Evaluation of receiver operating characteristic curve data in terms of information theory, with applications in radiography. Radiology 1973;109:297-303. http://dx.doi.org/10.1148/109.2.297.

14. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32-5. http://dx.doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.

15. Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. Acta Paediatr 2007;96:644-7. http://dx.doi.org/10.1111/j.1651-2227.2006.00178.x.

16. Perkins NJ, Schisterman EF. The inconsistency of «optimal» cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 2006;163:670-5. http://dx.doi.org/10.1093/aje/kwj063.

17. Kornak J, Lu Y. Bayesian decision analysis for choosing between diagnostic/prognostic prediction procedures. Stat Interface 2011;4:27-36. http://dx.doi.org/10.4310/SII.2011.v4.n1.a4.

18. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. Ann Intern Med 1999;130:1005-13. http://dx.doi.org/10.7326/0003-4819-130-12-199906150-00019.

19. Balogh EP, Miller BT, Ball JR, eds. Improving Diagnosis in Health Care. Washington, DC: The National Academies Press, 2015. http://dx.doi.org/10.17226/21794.

20. Smith M, Saunders R, Stuckhardt L, McGinnis JM, eds. Best Care at Lower Cost: the Path to Continuously Learning Health Care in America. Washington, DC: The National Academies Press, 2013.

21. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med 1980;302:1109-17. http://dx.doi.org/10.1056/NEJM198005153022003.

22. Vickers AJ. Decision analysis for the evaluation of diagnostic tests, prediction models and molecular markers. Am Stat 2008;62:314-20. http://dx.doi.org/10.1198/000313008X370302.

23. Graber ML, Wachter RM, Cassel CK. Bringing diagnosis into the quality and safety equations. JAMA 2012;308:1211-2. http://dx.doi.org/10.1001/2012.jama.11913.

24. Canadian Task Force on Preventive Health Care. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. CMAJ 2011;183:1991-2001. http://dx.doi.org/10.1503/cmaj.110334.

25. Lin JS, Piper MA, Perdue LA, Rutter C, Webber EM, O'Connor E et al. Screening for Colorectal Cancer: A Systematic Review for the U.S. Preventive Services Task Force. Evidence Synthesis No. 135. AHRQ Publication No. 14-05203-EF-1. Rockville, MD: Agency for Healthcare Research and Quality, 2016.

26. Habibzadeh F, Yadollahie M. Number needed to misdiagnose: a measure of diagnostic test effectiveness. Epidemiology 2013;24:170. http://dx.doi.org/10.1097/EDE.0b013e31827825f2.

27. Beerworth E. Clinical investigations: legal implications. Aust Prescr 1996;19:87-8. http://dx.doi.org/10.18773/austprescr.1996.084.

28. Fields MM, Chevlen E. Screening for disease: making evidence-based choices. Clin J Oncol Nurs 2006;10:73-6. http://dx.doi.org/10.1188/06.CJON.73-76.

29. Topping J, editor. Errors of Observation and Their Treatment. 4th ed. London: Chapman and Hall, 1972. http://dx.doi.org/10.1007/978-94-011-6928-8.

30. Hooper L, Abdelhamid A, Ali A, Bunn DK, Jennings A, John WG et al. Diagnostic accuracy of calculated serum osmolarity to predict dehydration in older people: adding value to pathology laboratory reports. BMJ Open 2015;5:e008846. http://dx.doi.org/10.1136/bmjopen-2015-008846.

31. Habibzadeh F, Habibzadeh P. How much precision in reporting statistics is enough? Croat Med J 2015;56:490-2. http://dx.doi.org/10.3325/cmj.2015.56.490.

32. Ozarda Y. Reference intervals: current status, recent developments and future considerations. Biochem Med (Zagreb) 2016;26:5-16. http://dx.doi.org/10.11613/BM.2016.001.

33. CLSI and IFCC. Defining, establishing and verifying reference intervals in the clinical laboratory. Approved guideline - third edition. C28-A3 document. Wayne, PA, USA: CLSI, 2008.