On determining the sensitivity and specificity of a new diagnostic test through comparing its results against a non-gold-standard test

Farrokh Habibzadeh*

Global Virus Network, Middle East Region, Shiraz, Iran

*Corresponding author: Farrokh.Habibzadeh@gmail.com

Abstract

Diagnostic tests are important clinical tools. To assess the sensitivity and specificity of a new test, its results should be compared against a gold standard. However, the gold-standard test is not always available. Herein, I show that we can compare the new test against a well-established diagnostic test (not a gold-standard test, but with known sensitivity and specificity) and compute the sensitivity and specificity of the new test if we would have compared it against the gold-standard test. The technique presented is useful for situations where the gold standard is not readily available.

Keywords: biostatistics; diagnostic tests; prevalence; sensitivity and specificity

Submitted: September 22, 2022 *Accepted: October 31, 2022*

Introduction

Diagnostic tests are among the important means commonly used in clinical medicine. Before a new test can be used in clinical practice, it should be evaluated for clinical validity. Studies assessing the clinical validity of a test (also termed diagnostic accuracy studies) involve determining the test performance indices including the test sensitivity (Se) and specificity (Sp) (1). Other common performance indices are positive and negative predictive values, and likelihood ratios, which can be calculated based on the Se and Sp and the prevalence (pr) of the disease of interest (2,3). To determine a test performance, its results should be evaluated against another test, the so-called reference standard (4). The reference standard can be a gold-standard test, *i.e.*, a test with a Se and Sp of 1.0 (or 100%). The gold-standard test can thus correctly discriminate those with and without the disease or condition of interest. For a test with binary results, the outcome is clear – positive or negative. For tests with continuous results, however, we need to set a cut-off value to categorize the results into positive or negative (2). Compared to the gold standard, the obtained results can be categorized into true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) results (Table 1a). The tests Se and Sp are defined as follows (5):

$$Se = \frac{TP}{TP + FN}$$
$$Sp = \frac{TN}{TN + FP} \quad \text{Equation (Eq.) 1}$$

Both the Se and Sp follow the binomial distribution. Then, the squared standard errors (SE^2) for Se and Sp are:

$$SE_{Se}^2 = \frac{Se(1 - Se)}{TP + FN}$$
$$SE_{Sp}^2 = \frac{Sp(1 - Sp)}{TN + FP} \quad \text{Eq. 2}$$

©Copyright by Croatian Society of Medical Biochemistry and Laboratory Medicine. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/) which permits users to read, download, copy, distribute, print, search, or link to the full texts of these articles in any medium or format and to remix, transform and build upon the material, provided the original work is properly cited and any changes properly indicated.

The prevalence of the disease (π), is then:

$$\pi = \frac{TP + FN}{TP + FP + FN + TN} \quad \text{Eq. 3}$$

Combining Eq. 1 and Eq. 3, we have:

$$\begin{aligned} P(TP) &= \frac{TP}{TP + FP + FN + TN} = \pi\, Se \\ P(FN) &= \frac{FN}{TP + FP + FN + TN} = \pi(1 - Se) \\ P(TN) &= \frac{TN}{TP + FP + FN + TN} = (1 - \pi)\, Sp \\ P(FP) &= \frac{FP}{TP + FP + FN + TN} = (1 - \pi)(1 - Sp) \end{aligned} \quad \text{Eq. 4}$$

where P(x) designates the probability of x. To evaluate the Se and Sp of a new test, it is common to compare its test results against those obtained from a gold-standard test. Nonetheless, the gold-standard test may not always be available. It either does not exist or is very difficult or expensive to perform for certain disease conditions (6). The question arise is that whether it is possible to calculate the Se and Sp of the new test based on the results obtained from its comparison with a non-perfect reference standard – a well-established (but not a gold-standard) test? This is not a new question, and several solutions has so far been proposed (1). Herein, I wish to propose an analytical method to address the question raised.

Stating the question

Suppose that we have a well-established test, say T_1, with known Se and Sp (measured against a gold-standard test) of Se_1 and Sp_1 (Table 1a). Now, suppose that we have a new test, say T_2, the results of which were compared against T_1 (not against a gold standard), and that it had a Se and Sp (against T_1) of $Se_{2,1}$ and $Sp_{2,1}$ (Table 1b). We wish to derive the Se and Sp of T_2 (Se_2 and Sp_2), if it would have been tested against the gold standard (*e.g.*, Table 1c).

The proposed solution

When we compare T_2 against T_1, the calculated prevalence, pr, is not really the true prevalence, π, as T_1 is not a gold standard and thus would have FP and FN results. However, we can calculate the true prevalence, π, as follows (7):

$$\pi = \frac{pr + Sp_1 - 1}{Se_1 + Sp_1 - 1} \quad \text{Eq. 5}$$

Based on Eq. 4 and basic probability rules, we have (Table 1) (8,9):

$$\begin{aligned} P(TP_{2,1}) &= \frac{TP_{2,1}}{TP_{2,1} + FP_{2,1} + FN_{2,1} + TN_{2,1}} \\ &= pr\, Se_{2,1} \\ &= P(D^+)P(T_1^+ \mid D^+)P(T_2^+ \mid D^+) + P(D^-)P(T_1^+ \mid D^-)P(T_2^+ \mid D^-) \\ &= \pi\, Se_1\, Se_2 + (1 - \pi)(1 - Sp_1)(1 - Sp_2) \end{aligned} \quad \text{Eq. 6}$$

and

$$\begin{aligned} P(TN_{2,1}) &= \frac{TN_{2,1}}{TP_{2,1} + FP_{2,1} + FN_{2,1} + TN_{2,1}} \\ &= (1 - pr)\, Sp_{2,1} \\ &= P(D^-)P(T_1^- \mid D^-)P(T_2^- \mid D^-) + P(D^+)P(T_1^- \mid D^+)P(T_2^- \mid D^+) \\ &= (1 - \pi)\, Sp_1\, Sp_2 + \pi(1 - Se_1)(1 - Se_2) \end{aligned} \quad \text{Eq. 7}$$

where T^+ and T^- represent positive and negative test results; and D^+ and D^-, presence and absence of the disease, respectively. P(A|B) denotes the conditional probability of event A given event B.

Based on Eq. 6, we have:

$$pr\, Se_{2,1} = \pi\, Se_1\, Se_2 + (1 - \pi)(1 - Sp_1)(1 - Sp_2) \quad \text{Eq. 8}$$

Solving for Se_2, gives:

$$Se_2 = \frac{pr\, Se_{2,1} - (1 - \pi)(1 - Sp_1)(1 - Sp_2)}{\pi\, Se_1} \quad \text{Eq. 9}$$

Based on Eq. 7, we have:

$$(1-pr)Sp_{2,1} = (1-\pi)Sp_1 Sp_2 + \pi(1-Se_1)(1-Se_2)$$

Eq. 10

Then:

$$Sp_2 = \frac{(1-pr)Sp_{2,1} - \pi(1-Se_1)(1-Se_2)}{(1-\pi)Sp_1}$$

Eq. 11

Equations 9 and 11 are a system of two simultaneous equations. Substituting π from Eq. 5 and solving for Se_2 and Sp_2, yield:

$$Se_2 = \frac{pr\,Se_{2,1}\,Sp_1 - (1-pr)(1-Sp_{2,1})(1-Sp_1)}{pr + Sp_1 - 1}$$

$$Sp_2 = \frac{pr(1-Se_{2,1}) + Se_1\left[pr(Se_{2,1}-1) - (1-pr)Sp_{2,1}\right]}{pr - Se_1}$$

Eq. 12

If f is a function of k independent random variables, then the squared SE of f can be calculated as (10,11):

$$SE_f^2 = \sum_{i=1}^{k}\left(\frac{\partial f(x_1, \ldots, x_k)}{\partial x_i}\right)^2 SE_{x_i}^2$$

Eq. 13

Assuming that Se_2 is a function of independent random variables pr, $Se_{2,1}$, $Sp_{2,1}$, and Sp_1 (Eq. 12), using Eq. 13 and employing basic calculus, we have:

$$SE_{Se_2}^2 = \left(\frac{\partial Se_2}{\partial pr}\right)^2 SE_{pr}^2 + \left(\frac{\partial Se_2}{\partial Se_{2,1}}\right)^2 SE_{Se_{2,1}}^2 + \left(\frac{\partial Se_2}{\partial Sp_{2,1}}\right)^2 SE_{Sp_{2,1}}^2 + \left(\frac{\partial Se_2}{\partial Sp_1}\right)^2 SE_{Sp_1}^2$$
$$= \frac{\left[Sp_1(Se_{2,1}+Sp_{2,1}-1)(Sp_1-1)\right]^2}{(Sp_1+pr-1)^4} SE_{pr}^2 + \left(\frac{pr\,Sp_1}{Sp_1+pr-1}\right)^2 SE_{Se_{2,1}}^2 +$$
$$\frac{(1-pr)^2(1-Sp_1)^2}{(Sp_1+pr-1)^2} SE_{Sp_{2,1}}^2 + \frac{\left[pr(Se_{2,1}+Sp_{2,1}-1)(pr-1)\right]^2}{(Sp_1+pr-1)^4} SE_{Sp_1}^2$$

Eq. 14

In the same way, assuming that Sp_2 is a function of independent random variables pr, $Se_{2,1}$, $Sp_{2,1}$, and Se_1 (Eq. 12), we have:

$$SE_{Sp_2}^2 = \left(\frac{\partial Sp_2}{\partial pr}\right)^2 SE_{pr}^2 + \left(\frac{\partial Sp_2}{\partial Se_{2,1}}\right)^2 SE_{Se_{2,1}}^2 + \left(\frac{\partial Sp_2}{\partial Sp_{2,1}}\right)^2 SE_{Sp_{2,1}}^2 + \left(\frac{\partial Sp_2}{\partial Se_1}\right)^2 SE_{Se_1}^2$$
$$= \frac{\left[Se_1(1-Se_1)(Se_{2,1}+Sp_{2,1}-1)\right]^2}{(Se_1-pr)^4} SE_{pr}^2 + \frac{pr^2(1-Se_1)^2}{(Se_1-pr)^2} SE_{Se_{2,1}}^2 +$$
$$\frac{Se_1^2(1-pr)^2}{(Se_1-pr)^2} SE_{Sp_{2,1}}^2 + \frac{\left[pr(Se_{2,1}+Sp_{2,1}-1)(pr-1)\right]^2}{(Se_1-pr)^4} SE_{Se_1}^2$$

Eq. 15

The SE for the Se and Sp of the tests can be calculated using Eq. 2.

Discussion

It was shown that the test Se and Sp can be determined with acceptable accuracy even if the gold-standard test is not available. The Se and Sp of the new test (T_2) derived by transforming the values obtained from its comparison with a non-gold-standard test ($Se_{2,1}$ and $Sp_{2,1}$) are acceptably close to the values if the test would have been compared with the gold-standard (Se_2 and Sp_2). The variances of the calculated Se_2 and Sp_2 (Eqs. 14 and 15) are higher than those you might obtain if you would have compared T_2 directly against the gold standard, instead of

Table 1. Results of a hypothetical test validity study

a		**Gold-standard test**		
		Positive	Negative	**Total**
T_1	Positive	TP: 85 πSe_1	FP: 40 $(1-\pi)(1-Sp_1)$	125
	Negative	FN: 15 $\pi(1-Se_1)$	TN: 360 $(1-\pi)Sp_1$	375
	Total	100	400	500
b		**T_1**		
		Positive	Negative	**Total**
T_2	Positive	107 $pr\,Se_{2,1}$	104 $(1-pr)(1-Sp_{2,1})$	211
	Negative	43 $pr(1-Se_{2,1})$	346 $(1-pr)Sp_{2,1}$	389
	Total	150	450	600
c		**Gold-standard test**		
		Positive	Negative	**Total**
T_2	Positive	76 πSe_2	64 $(1-\pi)(1-Sp_2)$	140
	Negative	4 $\pi(1-Se_2)$	256 $(1-\pi)Sp_2$	260
	Total	80	320	400

a) a well-established test, T_1, against the gold-standard test; **b**) a new test, T_2, against T_1; note that here, the true prevalence, π, is replaced by the apparent prevalence, pr (7) as T_1 is not a gold standard; and **c**) another hypothetical study if T_2 would have been tested against the gold standard. TP – True positive. FP – False positive. FN – False negative. TN – True negative. π – True prevalence. pr – Apparent prevalence. Sp – specificity. Se – sensitivity.

T_1. This is attributed to the uncertainty exist in the variables used for the calculation (Eq. 12). To examine the application of the technique proposed let us apply it to an example.

Example

Suppose that in a validity study of 500 (arbitrary chosen) randomly selected people, a diagnostic test (let us call it T_1) was tested against the gold standard (Table 1a), and that the test could correctly identify 85 of 100 diseased people, hence a Se (Se_1) of 0.85, and 360 of 400 disease-free individuals, hence a Sp (Sp_1) of 0.90 (Table 1a). The calculated SE^2 for the Se_1 and Sp_1 are 1.3×10^{-3} and 2.3×10^{-4}, respectively (using Eq. 2). Also, suppose that in a validity study on 600 (arbitrary chosen) randomly selected people, the results of a new diagnostic test, T_2, was compared against T_1 (Table 1b). Based on the information provided, the apparent prevalence, pr, is 0.25 ($SE^2 = 3.1\times10^{-4}$). Using Eq. 5, the true prevalence (π) is:

$$\pi = \frac{pr + Sp_1 - 1}{Se_1 + Sp_1 - 1} = \frac{0.25 + 0.90 - 1}{0.85 + 0.90 - 1} = 0.20 \quad \text{(Eq. 11)}$$

which is correct when the disease prevalence is measured by a gold-standard test (Table 1a). The Se and Sp (along with their SE^2) of T_2 against T_1 (Table 1b), are then:

$$Se_{2,1} = \frac{TP_{2,1}}{TP_{2,1} + FN_{2,1}} = \frac{107}{107 + 43} = 0.713 \left(SE^2 = 1.4 \times 10^{-3}\right)$$

$$Sp_{2,1} = \frac{TN_{2,1}}{TN_{2,1} + FP_{2,1}} = \frac{346}{346 + 104} = 0.769 \left(SE^2 = 3.9 \times 10^{-4}\right)$$

(Eq. 12)

Plugging in the values in equations 12, 14 and 15, estimations of Se_2 and Sp_2 are 0.95 ($SE^2 = 8.0 \times 10^{-3}$; 95% confidence interval (CI): 0.77 to 1.00) and 0.80 ($SE^2 = 5.4 \times 10^{-4}$; 95% CI: 0.75 to 0.85), respectively, which are compatible with the results if T_2 would have been compared against the gold-standard test – 0.95 ($SE^2 = 5.9 \times 10^{-4}$; 95% CI: 0.90 to 1.00) and 0.80 ($SE^2 = 5.0 \times 10^{-4}$; 95% CI: 0.76 to 0.84), respectively (Table 1c). Note that the 95% CI of the calculated Se_2 and Sp_2 when they are derived through comparing the results with T_1 is wider than those if they are directly compared against a gold-standard test.

In conclusion, it seems that this technique is useful, particularly where the gold-standard test is not readily available or is expensive. Further studies are needed to elaborate on the conditions of the validity study where the Se_1 and Sp_1 are estimated, the minimum number of data points examined, the probable effect of the prevalence of the disease or condition of interest on the choice of the reference test, among other things.

Potential conflict of interest

None declared.

References

1. *Umemneku Chikere CM, Wilson K, Graziadio S, Vale L, Allen AJ. Diagnostic test evaluation methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - An update. PLoS One. 2019;14:e0223832. https://doi.org/10.1371/journal.pone.0223832*
2. *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;26:297-307. https://doi.org/10.11613/BM.2016.034*
3. *Habibzadeh F, Habibzadeh P. The likelihood ratio and its graphical representation. Biochem Med (Zagreb). 2019;29:020101. https://doi.org/10.11613/BM.2019.020101*
4. *Kohn MA, Carpenter CR, Newman TB. Understanding the direction of bias in studies of diagnostic test accuracy. Acad Emerg Med. 2013;20:1194-206. https://doi.org/10.1111/acem.12255*
5. *Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ. 1994;308:1552. https://doi.org/10.1136/bmj.308.6943.1552*
6. *Newman TB, Kohn MA eds. Evidence-Based Diagnosis Newman. 2nd ed.New York: Cambridge University Press; 2009.*
7. *Habibzadeh F, Habibzadeh P, Yadollahie M. The apparent prevalence, the true prevalence. Biochem Med (Zagreb). 2022;32:020101. https://doi.org/10.11613/BM.2022.020101*
8. *Enøe C, Georgiadis MP, Johnson WO. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. Prev Vet Med. 2000;45:61-81. https://doi.org/10.1016/S0167-5877(00)00117-3*
9. *Toft N, Jørgensen E, Højsgaard S. Diagnosing diagnostic tests: evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. Prev Vet Med. 2005;68:19-33. https://doi.org/10.1016/j.prevetmed.2005.01.006*
10. *Champac V, Gervacio JG. Appendix A: Variance of a Function of Random Variables Approximated with Taylor's Theorem. In: Champac V, Gervacio JG, eds. Timing Performance of Nanometer Digital Circuits Under Process Variations: Springer; 2018.*
11. *Topping J ed. Errors of Observation and their Treatment: Springer Science & Business Media; 2012.*