



## Methodological letter

# Diagnostic tests. How to describe its validity?☆



## Pruebas diagnósticas. ¿Cómo describir su validez?

Juan-José Sancho-Insenser, Ana María González-Castillo\*

Servicio de Cirugía General y Digestiva, Hospital del Mar, Instituto Hospital del Mar de Investigaciones Médicas (IMIM), Universitat Autònoma de Barcelona, Barcelona, Spain

Surgeons routinely face the challenge of formulating a diagnosis, and to do so we use all available information, from anamnesis and physical examination to – increasingly exclusively – all kinds of diagnostic tests. In general terms, we understand diagnostic tests to be any source of information that can be used to confirm or rule out the presence of a disease. In addition to confirming the presence or absence of a disease, diagnostic tests also provide information on the intensity of a disease and/or its prognosis.<sup>1</sup> It is essential to know the validity of diagnostic tests, both when ordering and interpreting them for a particular patient, and when including them in a diagnostic and/or therapeutic algorithm within a clinical protocol or guideline.

### The need for research on diagnostic tests

Given the enormous potential benefit to be derived from the use of diagnostic tests, should we not use them massively until an accurate diagnosis is reached? Unfortunately, most diagnostic tests are not infallible and accurate at the same time and therefore rarely confirm or rule out a diagnosis in the entire population to which they are applied. In other words, they have limited validity. In addition to validity, any new diagnostic test should be investigated in terms of its resource requirements (both financial and professional, logistical and administrative time), the physical, emotional and economic impact on the patient in cases of false positives or negatives, as well as the safety derived from

the side effects inherent to the diagnostic test itself, which can range from a small ecchymosis in the elbow crease to a fatal outcome due to anaphylactic shock or perforation of a hollow viscera. Finally, when investigating a new diagnostic test, its clinical significance should be carefully assessed. Its usefulness will be proportional only to the change in attitude in patient management that follows from its result.<sup>2</sup>

Research on the validity and safety of diagnostic tests is essential for the surgeon, who very often bases his or her surgical indication and technique on the results of such tests. The validity of a diagnostic test is its ability to identify those who have the disease versus those who do not. The most commonly used type of study to assess the validity of a diagnostic test is the design of 2 groups of individuals, one group with the disease and one group without the disease. To determine the validity of the diagnostic test, the results obtained with the test to be evaluated in these individuals are compared with a reference criterion or gold standard.<sup>3,4</sup> Depending on the type of variable generated by the diagnostic test, we distinguish between two situations (Tables 1 and 2):

### Validity of the diagnostic tests with dichotomous outcomes

A dichotomous test classifies each patient as healthy or diseased depending on whether the test result is positive or

☆ Please cite this article as: Sancho-Insenser J-J, González-Castillo AM. Pruebas diagnósticas. ¿Cómo describir su validez? Cir Esp. 2022. <https://doi.org/10.1016/j.ciresp.2022.02.006>

\* Corresponding author.

E-mail address: [amgonzalezcastillo@psmar.cat](mailto:amgonzalezcastillo@psmar.cat) (A.M. González-Castillo).

**Table 1 – Relationship between the result of a diagnostic test and the true diagnosis of a disease.**

		True diagnosis	
		Diseased	Healthy
Test result	Positive	Sick patients with + test or TP	Healthy patients with + test or FP
	Negative	Sick patients with – test Or FN	Healthy patients with – test or TN
Sensitivity:		TP/TP + FN	TP/sick patients
Specificity:		TN/TN + FP	TN/healthy patients
Positive predictive value:		TP/TP + FP	TP/positives
Negative predictive value:		TN/TN + FN	TN/negatives
Positive probability ratio:		Sensitivity/100-Specificity	
Negative probability ratio:		100-Sensitivity/Specificity	
Youden index:		Sensitivity + Specificity-1	
Overall predictive value:		TN + TP/total	Accurate/all

FN: false negatives; TP: true positives.

**Table 2 – Main characteristics and limitations of the tests used to describe the validity of a diagnostic test with dichotomous outcomes.**

	Characteristic	Limitation
Sensitivity and Specificity	Describe well the validity of a diagnostic test.	They require knowledge of the actual diagnosis before they can be calculated.
Predictive value (positive and negative)	Ideal for the epidemiologist.	They are very sensitive to extreme values of prevalence.
Probability ratios (positive and negative)	They do not need to know the actual diagnosis before being calculated.	They are not useful for comparing tests in different populations.
Overall predictive value	Ideal for the clinician.	They must be interpreted correctly.

negative. A positive result indicates the presence of disease and a negative result indicates the absence of disease.

Diagnostic validity is quantified with the classical ratios: sensitivity and specificity. They are usually expressed as percentages, and they objectify the ability of the test to correctly or erroneously classify a person according to the presence or absence of a disease.

The data obtained distribute the individuals into 4 groups, usually represented in a  $2 \times 2$  table in which the result of the diagnostic test (in the rows) is cross-referenced with the actual status of the individuals (in the columns). The test result can be correct (true positive and true negative) or incorrect (false positive and false negative).

On many occasions the true and absolute diagnosis of all individuals is unknown, and the result of the gold standard test used for diagnosis is taken as a comparator until the new diagnostic test to be evaluated is available.

**Sensitivity** is the ability of the test to detect the disease. Mathematically, it is the proportion of individuals with the disease who obtain a positive result, i.e., the probability of correctly classifying a diseased individual.

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True positives} + \text{false negatives}}$$

$$= \frac{\text{True positives}}{\text{All the sick patients}}$$

It can therefore be referred to as the Proportion of true positives.

Screening-oriented diagnostic tests must have a high sensitivity in order to be able to detect all patients, even at the cost of some false positives. A highly sensitive test is particularly appropriate for situations where failure to diagnose the disease is dangerous for the patient, such as in the case of dangerous but treatable diseases, or in diseases where a false positive does not cause serious patient distress. Before treatment is initiated, a confirmatory test, ideally with high specificity, is usually required.

**Specificity** is the ability of a test to detect healthy individuals. Mathematically, it is the proportion of individuals without the disease who have a negative test result and indicates the usefulness of the test in identifying individuals who do not have the disease.

$$\begin{aligned} \text{Specificity} &= \frac{\text{True Negatives}}{\text{True negatives} + \text{false positives}} \\ &= \frac{\text{True negatives}}{\text{All the healthy}} \end{aligned}$$

It can also be referred to as the Proportion of true negatives.

Tests confirmatory to a diagnosis must have a high specificity in order to avoid false positives. High specificity tests are essential when there is a strong interest in the absence of disease, in serious but untreatable diseases that are curable, or when diagnosing a patient with a disease that he or she does not actually have may have serious consequences, whether physical, psychological or financial.<sup>5</sup>

Sensitivity and specificity are inherent properties of the diagnostic test and are therefore independent of the prevalence of the disease in the study population.<sup>2</sup> Sensitivity and specificity assess the validity of a diagnostic test, but are concepts that are not directly applicable in daily clinical practice. These parameters may be used to recommend a test in a clinical protocol or guideline, but they are not useful for assessing the outcome of a test in a specific patient. The reason is that both sensitivity and specificity assess the probability of obtaining a particular result (positive or negative) based on the true diagnosis. However, when a patient undergoes a diagnostic test, we do not know the true diagnosis (that is why we ask for the test!). The clinician therefore asks the opposite question: if the test is positive, what is the probability that the patient is really ill? We must calculate predictive values to obtain this information.<sup>3</sup>

### Predictive positive and negative values

The **predictive positive value** is the probability that an individual with a positive result has the disease. Mathematically it is the proportion of patients with a positive result from the test who are actually ill.<sup>4</sup>

$$\begin{aligned} \text{Predictive positive value} &= \frac{\text{True Positives}}{\text{True positives} + \text{false positives}} \\ &= \frac{\text{True positives}}{\text{All positives}} \end{aligned}$$

The **negative predictive value** is the probability that an individual with a negative result does not actually have the disease. Mathematically it is the proportion of patients with a negative result who are actually healthy.<sup>4</sup>

$$\begin{aligned} \text{Predictive negative value} &= \frac{\text{True Negatives}}{\text{True negatives} + \text{false negatives}} \\ &= \frac{\text{True negatives}}{\text{All the negatives}} \end{aligned}$$

### The effect of prevalence on the predictive values

Predictive values vary according to the prevalence of the disease and, therefore, they assess both the validity of the diagnostic test and the prevalence of the disease in the population studied and, therefore, the clinician must know when assessing them the prevalence of the disease he/she intends to confirm or rule out in the population to which the individual studied belongs, and consequently correct his/her assessment in the light of the prevalence. Very low prevalences mean that a negative result can more reliably rule out the disease and will therefore have an artificially high negative predictive value. In the opposite scenario, very high prevalences mean that a positive test is very likely to be true and will therefore have a misleadingly high positive predictive value if applied to another population with a lower prevalence.

All these percentages have a confidence interval that can be easily calculated and should always accompany the specificity, sensitivity, positive and negative predictive values. All these percentages should always be evaluated together: sensitivity with specificity and positive predictive value with negative predictive value, as each pair is like the two sides of a coin. However, in order to obviate the dependence on prevalence, a series of parameters have been proposed that, combining the above, reflect the validity of a diagnostic test in a single figure and are independent of the prevalence of the disease to be diagnosed: the likelihood ratios.

### Likelihood ratios

As prevalence is a very influential factor in the predictive values of a test, it cannot be used as a measure to compare two different diagnostic methods. Therefore, it is necessary to use other assessment parameters that are both clinically useful and do not depend on the prevalence of the disease in the population to be studied: these are the likelihood ratio, likelihood ratio or positive and negative likelihood ratio, which compare the probability of obtaining a given result (positive or negative) in an individual with the disease with that of obtaining it in a subject in whom the presence of the disease has been ruled out.<sup>6</sup>

The **Positive Likelihood Ratio (PLR)** is calculated by dividing the proportion of cases that test positive (sensitivity) by the proportion of people who do not have the disease, but in whom the test has given a positive result (100-specificity, false positives).

$$\text{PLR} = \frac{\text{Sensitivity}}{100 - \text{Specificity}} = \frac{\text{Proportion of true positives}}{\text{Proportion of false positives}}$$

The **Negative Likelihood Ratio (NLR)** is calculated by dividing the proportion of cases that are negative in the

presence of disease (false negatives) by the probability of a negative result in the absence of disease.

$$NLR = \frac{100 - \text{Sensitivity}}{\text{Specificity}} = \frac{\text{Proportion of false positives}}{\text{Proportion of true positives}}$$

Likelihood ratios offer the advantage that they relate sensitivity and specificity of a test in a single index and do not depend on prevalence. This allows them to be used to compare different tests for the same diagnosis. Likelihood ratios can be calculated at various levels of a new measure and it is not necessary to express the information in a dichotomous form (positive or negative).<sup>7</sup>

### Other synthetic values to assess the validity of a diagnostic test

#### Overall predictive value

Another parameter that was used to describe the validity of a test in a single figure is the overall predictive value (Overall Accuracy). Despite its impressive name, it is the least objective parameter for describing the validity of a diagnostic test with a dichotomous result.

$$\text{Total predictive value} = \frac{\text{True positives} + \text{True negatives}}{\text{Total}}$$

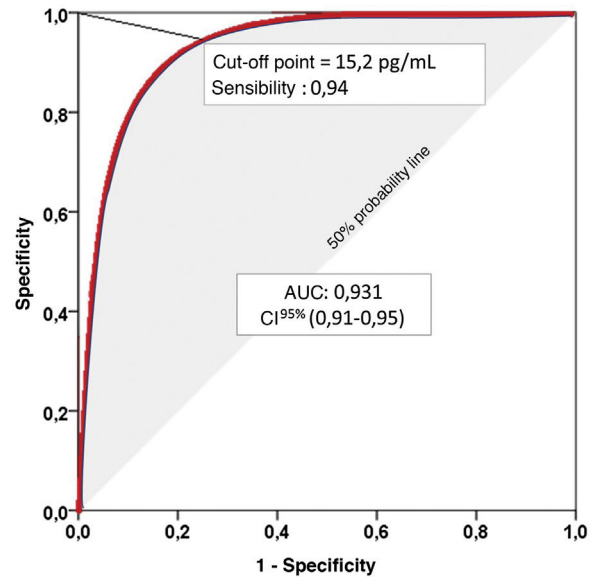
As synthetic, single-digit measures of the validity of a diagnostic test, they assume that sensitivity and specificity are of equal importance for the case at hand and are therefore of limited use in specific cases where only one test is available to suspect and confirm the diagnosis. They are clearly inferior to the likelihood ratios and should currently be disused.<sup>8</sup>

#### Validity of diagnostic tests with continuous value results

The results of many diagnostic tests are continuous numerical values (e.g., serum lipase). In these tests, we must decide at which numerical value or cut-off point the results will be considered positive or negative. A compromise between higher sensitivity and lower specificity or vice versa should be sought, depending on the specific intention of the diagnostic test. To gauge their validity, they can be dichotomised from that cut-off point, but it is much better to use the functional characteristic curves.<sup>5</sup>

Receiver-operating characteristic (ROC) curves graphically represent the pairs of sensitivity and specificity values resulting from the continuous variation of cut-off points over the range of observed results. The Area Under the Curve (AUC) ranges from 0 to 1, and is an excellent indicator of the diagnostic validity of the test, so that the higher the AUC, the greater the ability of the test to discriminate correctly between diseased and non-diseased.<sup>9</sup>

The graphical representation of several ROC curves allows a direct appreciation of the differences between several diagnostic tests or between several cut-off points of the same diagnostic test. A diagonal line representing an



**Fig. 1 – Representation of a Receiver-Operating Characteristic (ROC) curve. Area Under the Curve (AUC): Area under the curve represents the positive diagnostic validity of the test. The closer to unity, the better the ability to distinguish between sick and healthy, without prior knowledge of the true diagnosis and without influence of the prevalence of the disease to be diagnosed. 95% confidence interval (95% CI) for the AUC.**

AUC of 0.5, i.e., the probability of a positive result in the flip of a coin, is usually incorporated into the graphical representation. Therefore, to correctly interpret the AUC of a ROC curve we must subtract 0.5 from the AUC. Thus, an AUC of 0.750 may seem convincing, but it means that the diagnostic test we are evaluating improves by only 25% on the result we would obtain by flipping a coin (shaded area in Fig. 1).

In conclusion, it is essential to assess the validity and safety of diagnostic tests in order to select the most appropriate test in each clinical situation.

#### REFERENCES

1. Sackett D, Haynes R, Tugwell P. *Epidemiología clínica*. Madrid: Díaz de Santos; 1989.
2. Escrig-Sos J, Martínez-Ramos D, Miralles-Tena J. Pruebas diagnósticas: nociones básicas para su correcta interpretación y uso. *Cir Esp*. 2006;79:267-73. [http://dx.doi.org/10.1016/s0009-739x\(06\)70904-6](http://dx.doi.org/10.1016/s0009-739x(06)70904-6).
3. Altman D, Bland J. Diagnostic tests. 1: sensitivity and specificity. *BMJ*. 1994;308:1552. <http://dx.doi.org/10.1136/bmj.308.6943.1552>.
4. Altman D, Bland J. Diagnostic test — predictive values. *BMJ*. 1994;309:102. <http://dx.doi.org/10.1016/j.ebobjyn.2005.10.001>.
5. Pita-Fernández S, Pértegas-Díaz S. Pruebas diagnósticas. *Cad Aten Primaria*. 2003;10:120-4.

6. Moskowitz CS, Pepe MS. Comparing the predictive values of diagnostic tests: sample size and analysis for paired study designs. *Clin Trials*. 2006;3:272-9. <http://dx.doi.org/10.1191/1740774506cn147oa>.
7. María Eugenia Burgos D, Carlos Manterola D. Cómo interpretar un artículo sobre pruebas diagnósticas. *Rev Chil Cir*. 2010;62:301-8. <http://dx.doi.org/10.4067/s0718-40262010000300018>.
8. Alberg AJ, Park JW, Hager BW, Brock MV, Diener-West M. The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. *J Gen Intern Med*. 2004;19:460-5. <http://dx.doi.org/10.1111/j.1525-1497.2004.30091.x>.
9. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*. 1993;39:561-77. <http://dx.doi.org/10.1093/clinchem/39.4.561>.