

An Introduction to Statistics: Diagnostic Tests

Priya Ranganathan 

ABSTRACT

Diagnostic tests are used to differentiate between those with and without disease. In this article, we examine some of the properties of diagnostic tests, such as sensitivity, specificity, predictive values, and receiver operating characteristic curves.

Keywords: Diagnostic test, Sensitivity, Specificity.

Indian Journal of Critical Care Medicine (2021): 10.5005/jp-journals-10071-24070

INTRODUCTION

Various tests are used in the field of medicine to diagnose clinical conditions. For example, in a previous issue of this journal, Abdelshafey studied the role of presepsin levels in the early identification of sepsis.¹ In this article, we look at some of the statistical properties of diagnostic tests.

Diagnostic tests are used to distinguish between individuals who have and do not have disease. Many conditions have a gold standard test—for example, liver biopsy and histopathology to diagnose liver cirrhosis. However, the test may have certain limitations: they may be invasive, risky, time-consuming, or expensive. Therefore, an alternate test that overcomes these limitations may be used in lieu of the gold standard—for example, ultrasound elastography of the liver. The new test needs to be compared to the gold standard to ensure adequate and acceptable performance.

SENSITIVITY OF A TEST

The sensitivity of a test refers to the ability of a test to detect disease when it is present. In other words, it is the probability that the test will be positive in an individual with disease.

$$\text{Sensitivity} = \frac{\text{Number of individuals with test positive}}{\text{Number of individuals with disease}}$$

A highly sensitive test is likely to be positive in almost all individuals with disease, and therefore if it is negative, it is useful to rule out disease. If a test has a sensitivity of 95%, it means that out of 100 individuals with disease, the test will be positive in 95 and only 5 of them will have a “false-negative” test. For such a test, a negative result is highly likely to be associated with the absence of disease.

In the study by Abdelshafey, presepsin with a cutoff value >640 pg/mL had a diagnostic accuracy of identifying septic cases with sensitivity of 73%.¹ This means that out of 100 patients *with* early sepsis, 73 would have a presepsin value of 640 pg/mL or more whereas 27 would have a presepsin level lower than 640 pg/mL and be incorrectly labeled as not in early sepsis.

SPECIFICITY OF A TEST

The specificity of a test refers to the ability of the test to rule out disease when it is absent. It is the probability that the test will be negative in an individual without disease.

Department of Anaesthesiology, Critical Care and Pain, Tata Memorial Centre, Homi Bhabha National Institute, Mumbai, Maharashtra, India

Corresponding Author: Priya Ranganathan, Department of Anaesthesiology, Critical Care and Pain, Tata Memorial Centre, Homi Bhabha National Institute, Maharashtra, India, Phone: +91-22-24177000, e-mail: drpriyaranganathan@gmail.com

How to cite this article: Ranganathan P. An Introduction to Statistics: Diagnostic Tests. *Indian J Crit Care Med* 2021; 25(Suppl 3): S283–S284.

Source of support: Nil

Conflict of interest: None

$$\text{Specificity} = \frac{\text{Number of individuals with test negative}}{\text{Number of individuals without disease}}$$

A highly specific test is likely to be negative in almost all individuals without disease and therefore, if it is positive, it is useful to rule in disease. A test with a specificity of 95% means that out of 100 individuals without disease, the test will be negative in 95 and only 5 individuals will have a “false-positive test.” For example, in the previously quoted study, presepsin level of 640 pg/mL or more had a specificity of 92% for detecting early sepsis.¹ This means that out of 100 patients *without* early sepsis, 92 would have a presepsin level less than 640 pg/mL, whereas only 8 of them would have a presepsin level higher than 640 pg/mL and be incorrectly labeled as at risk of sepsis.

The choice of test would depend on whether sensitivity or specificity is the priority. For screening studies, where identification of the majority with disease is crucial, a highly sensitive test is preferred. However, for confirmation of diagnosis, especially if expensive or toxic treatment is planned, a highly specific test should be used. D-dimer is a sensitive test for ruling out pulmonary embolism. However, it is not specific and a confirmatory test such as a CT pulmonary angiography is required before starting thrombolytic therapy.

Sensitivity and specificity are properties of the test and are useful parameters when we wish to determine the performance of the test against a gold standard—i.e., when we know whether the individual truly has disease or not. However, in clinical practice, we only know the results of the test, based on which we need to

predict whether the individual has disease or not. For this, we use predictive values.

POSITIVE PREDICTIVE VALUE

The positive predictive value (PPV) of a test is the probability that an individual with a positive test actually has disease.

PPV = Number of individuals with disease/Number of individuals with test positive

NEGATIVE PREDICTIVE VALUE

The negative predictive value (NPV) of a test is the probability that an individual with a negative test does not have disease.

NPV = Number of individuals without disease/Number of individuals with test negative

Agarwal looked at the Sepsis in Obstetrics Score (SOS) value of 6 for predicting severe sepsis in pregnant women.² This cutoff had a PPV of 83% and a NPV of 65%. This means that of 100 women with a SOS of 6 or more, 83 would have severe sepsis (and 17 would not develop severe sepsis despite a high SOS). Similarly, out of 100 women with a SOS less than 6, 65 would not develop severe sepsis (and 35 would have severe sepsis despite a low SOS).

Unlike sensitivity and specificity, which are properties of the test and are not affected by disease prevalence, NPV and PPV are dependent on disease prevalence.

RECEIVER OPERATING CHARACTERISTIC (ROC) CURVES

For tests where the results can give several values, the results of individuals with and without disease may have overlapping values. It becomes important to choose an appropriate cutoff to differentiate disease from no disease. In the presepsin example, it is essential to decide what cutoff value of presepsin will be chosen to define early sepsis. Choosing a very low cutoff will improve the sensitivity of the test (identify all with disease); however, some individuals without disease will also be classified as diseased (false positives). Choosing a high cutoff value will improve the specificity of the test (exclude those without disease) but may also result in some individuals with disease being missed (false negatives). As another example, procalcitonin levels are used to diagnose bacterial sepsis. A level of 2.0 ng/mL or more is very strongly predictive of

sepsis (high specificity); however, this may not pick up all patients with sepsis (low sensitivity). Choosing a cutoff of 1.0 ng/mL will improve the sensitivity but may also include some patients who do not have sepsis (lower specificity). Lowering the cutoff even further to 0.5 ng/mL will result in very high sensitivity with poor specificity. Thus, the choice of cutoff is a play between the sensitivity and the specificity and is determined using a method known as the receiver operating characteristic (ROC) curve. The ROC curve plots sensitivity (on the Y-axis) vs 1—specificity (on the X-axis) and identifies the cutoff point that offers the best balance between sensitivity and specificity.

The ability of a diagnostic test to differentiate between those with and without disease is given by the area under the ROC curve (AUROCC) and tests may be compared based on their AUROCC. A perfect test has an AUROCC of 1.0 and a worthless test has an AUROCC of 0.5. Values between 0.5 and 1.0 are considered moderate to excellent, with higher values representing a better test performance. In the study by Abdelshafey, presepsin with a cutoff value >640 pg/mL had an AUROCC of 0.85, which was higher than the SIRS (AUROCC 0.67) or qSOFA (AUROCC 0.65) criteria.¹

For further reading on this topic, readers are referred to more detailed articles.^{3,4}

ORCID

Priya Ranganathan  <https://orcid.org/0000-0003-1004-5264>

REFERENCES

1. Abdelshafey EE, Nasa P, Elgohary AE, Khalil MF, Rashwan MA, Ghezala HB, et al. Role of presepsin for the diagnosis of sepsis and ICU mortality: a prospective controlled study. *Indian J Crit Care Med* 2021;25(2):153–157. DOI: 10.5005/jp-journals-10071-23715.
2. Agarwal R, Yadav RK, Mohta M, Sikka M, Radhakrishnan G. Sepsis in Obstetrics Score (SOS) utility and validation for triaging patients with obstetric sepsis in the emergency department: evidence from a low income health care setting. *Obstet Med* 2019;12(2):90–96. DOI: 10.1177/1753495X18761522.
3. Ranganathan P, Aggarwal R. Common pitfalls in statistical analysis: understanding the properties of diagnostic tests – Part 1. *Perspect Clin Res* 2018;9(1):40–43. DOI: 10.4103/picr.PICR_170_17.
4. Aggarwal R, Ranganathan P. Understanding diagnostic tests – Part 3: Receiver operating characteristic curves. *Perspect Clin Res* 2018;9(3):145–148. DOI: 10.4103/picr.PICR_87_18.