J Bras Pneumol. 2021;47(2):e20210139
https://dx.doi.org/10.36416/1806-3756/e20210139

Receiver operating characteristic analysis: an ally in the pandemic

Jezreel Pantaleón García[1,2], Juliana Carvalho Ferreira[1,3], Cecilia Maria Patino[1,4]

PRACTICAL SCENARIO

From a global public health perspective, a diagnostic test that accurately discriminates between positive and negative COVID-19 cases is critical to allocate human and material resources to manage the pandemic.(1) The ongoing COVID-19 pandemic has led to the expeditious development of multiple diagnostic tests to detect the SARS-CoV-2 infection. Thus, clinicians, researchers, and policy makers need to understand how to interpret the performance level of such diagnostic tests(1) to support the multilevel decision-making process. Here, we provide an overview of a commonly used tool to evaluate the accuracy of diagnostic or prognostic tests: the ROC curve.

ROC ANALYSIS

We use ROC analysis to graphically display, compare, and evaluate the accuracy of current and novel diagnostic tests. In order to do so, ROC curves integrate three related measures of accuracy: sensitivity (true positives), specificity (true negatives), and AUC.(2) These measures are calculated for any diagnostic test by comparing the test result (positive or negative) against a well-known gold standard that determines the true disease status in each case.

UNDERSTANDING ROC CURVES

ROC curves are created by plotting sensitivity (true positives) on the *y* axis against 1 − specificity (true negatives) on the *x* axis for every value found in a sample of subjects with and without the disease. It is expected that higher values would be more common among the subjects with the disease, and lower values would be more common among the subjects without the disease. In a perfect test, an obvious cutoff threshold can be identified that differentiates subjects with the disease from those without the disease, sensitivity and specificity being both 100%. Such a perfect differentiation is rarely the case for tests in real life, so ROC curves plot the trade-off between sensitivity and specificity for all possible cutoffs and the overall test accuracy. To express the diagnostic accuracy of a test numerically, we calculate the AUC, which estimates the probability of a random subject with the disease to have a higher value on the test than a subject without the disease. The probability ranges from 0% (AUC = 0) to 100% (AUC = 1).

USING ROC CURVES

Relative shapes of ROC curves within the plot are a quick approach to estimate and compare the accuracy between diagnostic tests (Figure 1). A perfect diagnostic test (AUC = 1.0) correctly identifies all positive and all negative results as diseased and non-diseased, respectively, and would reach the far top left. In contrast, a test that is inaccurate, or similar to flipping a coin, would result in a 45-degree line (AUC = 0.5). These two extremes (perfect test and uninformative test) are often used as references: ROC curves closer to a perfect diagnostic test have a higher AUC and are more accurate than are those closer to the random error line (AUC ~0.5).(2) Therefore, comparing multiple ROC curves may be an intuitive strategy to help us decide which the most accurate test for our clinical practice is. However, since there is always a trade-off between sensitivity and specificity, tests should not be evaluated by the AUC alone. In some cases, a test is more useful when it has high sensitivity (and, therefore, lower specificity), as when you cannot afford to miss the diagnosis. An example is when you are using a test to diagnose COVID-19. In that case, a test with lower

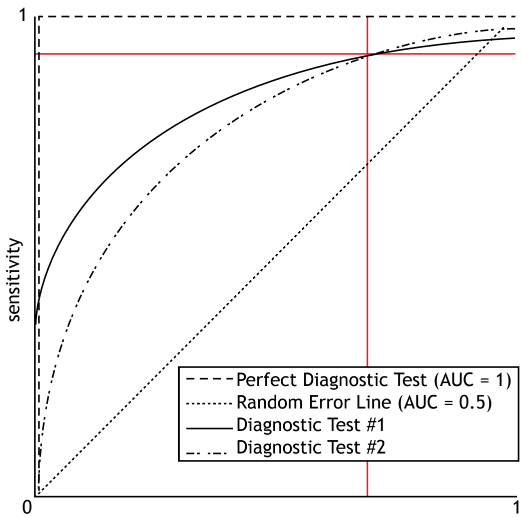

Figure 1. Comparative examples of ROC curves. ROC curve plots illustrating the accuracy performance of a perfect diagnostic test (AUC = 1), a random error line (AUC = 0.5) of an uninformative test, and two hypothetical diagnostic tests. Red lines depict a clinically relevant threshold of high sensitivity range in which the AUC of Diagnostic Test #2 outperforms Diagnostic Test #1.

1. Methods in Epidemiologic, Clinical, and Operations Research–MECOR–program, American Thoracic Society/Asociación Latinoamericana del Tórax, Montevideo, Uruguay.
2. Department of Pulmonary Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.
3. Divisão de Pneumologia, Instituto do Coração, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo (SP) Brasil.
4. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

AUC that has a high sensitivity may be more useful in certain clinical scenarios than a test with slightly higher AUC with lower sensitivity (and greater specificity).

REFERENCES

1. Butler-Laporte G, Lawandi A, Schiller I, Yao M, Dendukuri N, McDonald EG, et al. Comparison of Saliva and Nasopharyngeal Swab Nucleic Acid Amplification Testing for Detection of SARS-CoV-2: A Systematic Review and Meta-analysis [published correction appears in doi: 10.1001/jamainternmed.2021.0245]. JAMA Intern Med. 2021;181(3):353-360. https://doi.org/10.1001/jamainternmed.2020.8876
2. Ferreira JC, Patino CM. Understanding diagnostic tests. Part 3. J Bras Pneumol. 2018;44(1):4. https://doi.org/10.1590/s1806-37562018000000017