Outbreak, Surveillance, Investigation & Response (OSIR) Journal

Field Epidemiology Training Program, Division of Epidemiology
Department of Disease Control, Ministry of Public Health, Thailand
Tel: +6625903894, Fax: +6625903845, Email: osireditor@osirjournal.net, http://www.osirjournal.net

The Grammar of Science: Are You Positive that Your Test is Positive?

Jaranit Kaewkungwal*

Mahidol University, Thailand

*Corresponding author email: jaranitk@biophics.org

In the era of coronavirus disease 2019 (COVID-19) pandemic, we all have heard or ever been tested with "antigen test kit" (ATK). The test tells whether a person tests positive or negative for COVID-19. If you test negative for COVID-19 using ATK, you are likely not to be infected, provided you do not have any symptoms. On the other hand, if you test positive, you are likely to be infected, provided you have any symptoms. So, do we use such test for diagnostic or screening? Are you sure that the negative and positive test results valid? Is it possible that you have false positive/negative result?

Diagnostic vs. Screening Test

Diagnostic and screening tests differ regarding their intended usage and whether a person shows symptomatic signs or not. On the pathway of natural history of disease, after a person gets infected with the disease, there typically are time gaps at different stages of disease status as shown in Figure 1. The tests could be used for detecting the diseases at different staging. Screening tests are usually intended for asymptomatic persons whereas diagnostic tests are intended for those showing symptoms in need of a diagnosis.[1,2]

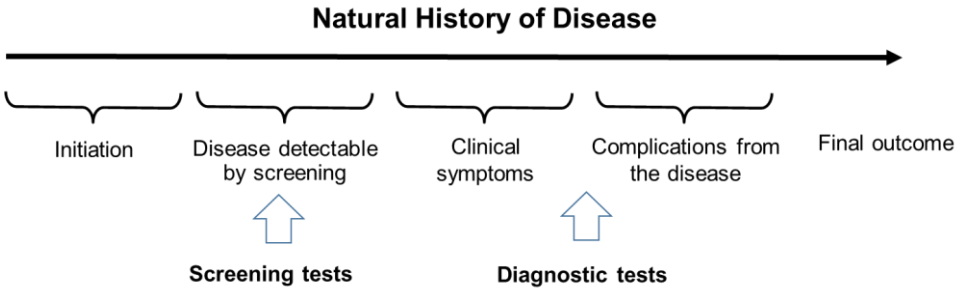

Figure 1. Screening and diagnostic tests at different stages of natural history of disease

The goal of screening is to detect disease as early as possible which is particularly useful for surveillance or reducing the risk of disease. Screening tests tend to be less invasive than diagnostic ones—and they are usually simpler to perform. ATK has become a quick tool for COVID-19 screening while other tests like RT-PCR tests are commonly used as diagnostic tests or confirmation tests. A positive result in a screening test usually requires a more accurate diagnostic test to confirm diagnosis.[3]

Validity and Efficacy of the Test

The screening or diagnostic test should have the ability to distinguish non-diseased and diseased persons. To do so, we need to compare the results of the test against some "gold standard" that establishes the true disease status. The gold standard may be a known test that provides a very accurate status but it may be more expensive and invasive or may take longer time to get the result. If there is no such gold standard, the assumed true disease status may be from the confirmed clinical signs and symptoms, or the acceptable consensus among experts. In some cases, the true disease status may be determined by following a certain group(s) of persons for a period of time to determine which patients ultimately develop the disease (Figure 2).[4–7]

The test is considered valid with two exquisitely qualifications. That is, the test is "sensitive" (true positive) when it gives high probability of detecting disease among diseased persons, and "specific" (true negative) when it gives high probability that those without the disease will have negative test result as shown in Figure 2.

The test is considered efficacious if it has favourable predictive qualification. The predictive value is the probability of having or not having disease according to test results. A positive predictive value (PPV) is the probability that the person has the disease when the test shows positive result while a negative predictive value (NPV) is the probability that the person does not have the disease when the test shows negative result.

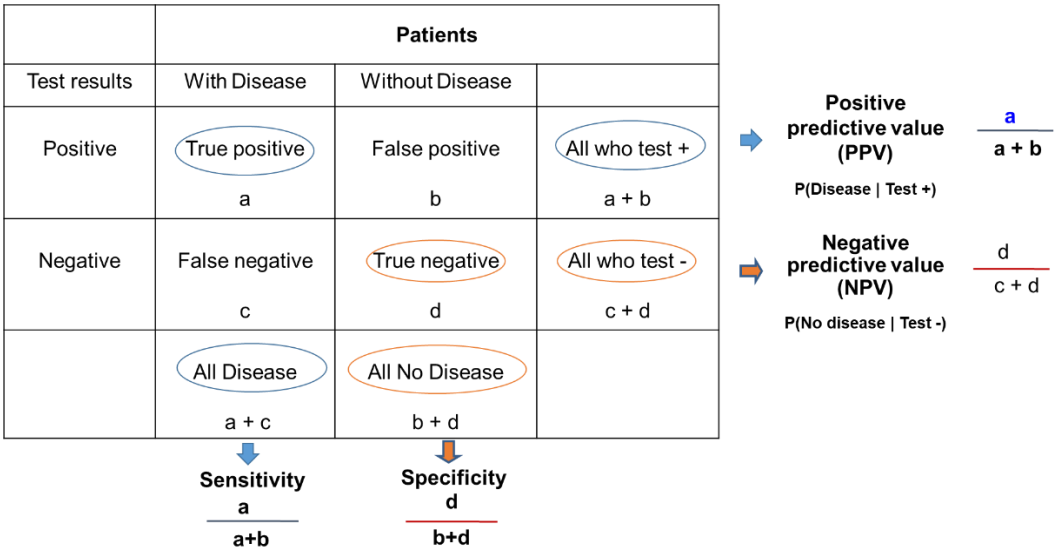

Figure 2. Sensitivity, specificity, positive predictive value and negative predictive value of a test

From the perspective of healthcare providers when we plan to use the test, we may prefer test with high sensitivity for screening purpose and the test with high specificity for confirmatory purpose. ***"If the person has the disease, how good that the test will show true positive result?" "If the person does not have the disease, how good that the test will confirm true negative result?"*** While sensitivity and specificity of a test are important considerations, predictive values are also important from the perspective of the patients. ***"If my test result is positive, what is the probability that I am truly positive?"***

While the values of sensitivity and specificity do not depend on the prevalence of the outcome in the population tested, the predictive values do. Given the same sensitivity and specificity, the PPV will increase and the NPV will decrease as the prevalence increases.[7] The relationships between these four terms are shown as follows.[8]

$$PPV = \frac{\text{Sensitivity} \times \text{Prevalence}}{(\text{Sensitivity} \times \text{Prevalence}) + (1-\text{Specificity}) \times (1-\text{Prevalence})}$$

$$NPV = \frac{\text{Specificity} \times (1-\text{Prevalence})}{[\text{Specificity} \times (1-\text{Prevalence})] + [(1-\text{Sensitivity}) \times \text{Prevalence}]}$$

Cutoff Point for Test Score

While some test results are qualitative—"positive" or "negative"—as discussed above, some are quantitative score. For example, the result on ATK for COVID-19 tells you whether you are negative (one stripe) or positive (two stripes). If you use RT-PCR test, it usually gives the result as cycle threshold (Ct) which is a semi-quantitative value that can broadly categorize the concentration of viral genetic materials in a patient sample as low, medium or high.[9] In diabetes diagnosis, we may use blood sugar level two hours after the last meal, which is continuous value for decision making.[10] The latter approach usually requires setting up a cut-off point from continuous data. The cut-off will affect the sensitivity and specificity of the test. Upon varying selected cutoff values, if the sensitivity increases, the specificity will decrease and vice versa (Figure 3).

From varying cut-off points, a receiver operating characteristic curve (ROC curve) can be created representing a graphical plot that illustrates the diagnostic ability of a binary classifier outcome. The term "ROC" was originally developed for operators of military radar receivers starting in 1941. The ROC curve is created by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. The apex of the ROC curve, toward the upper left corner, represents a greater discriminatory ability of the test with high true-positive rate and low false-positive rate (Figure 3). It is important to note that ROC performance may change upon different conditions, such as patient populations and staging or severity levels of the disease. It has been suggested in a number of literature that we should explore and (if possible) pool the results of several studies that examined the same test in different situations, then generate averaged specificity and sensitivity and construct the ROC.[6,11,12]

The area under the ROC curve (AUC) is another indicator of the test's discriminatory power. When the test provides perfect performance (i.e., 100% sensitive and 100% specific), the AUC is 1.0. When the test has no discriminative value (i.e., 50% sensitive and 50% specific), the AUC is 0.5, representing by the area under a straight, diagonal line. In general, it is noted that an AUC ≤0.75 indicates that the test is not clinically useful, while an AUC >0.96 indicates excellent discriminatory ability.[13–15]

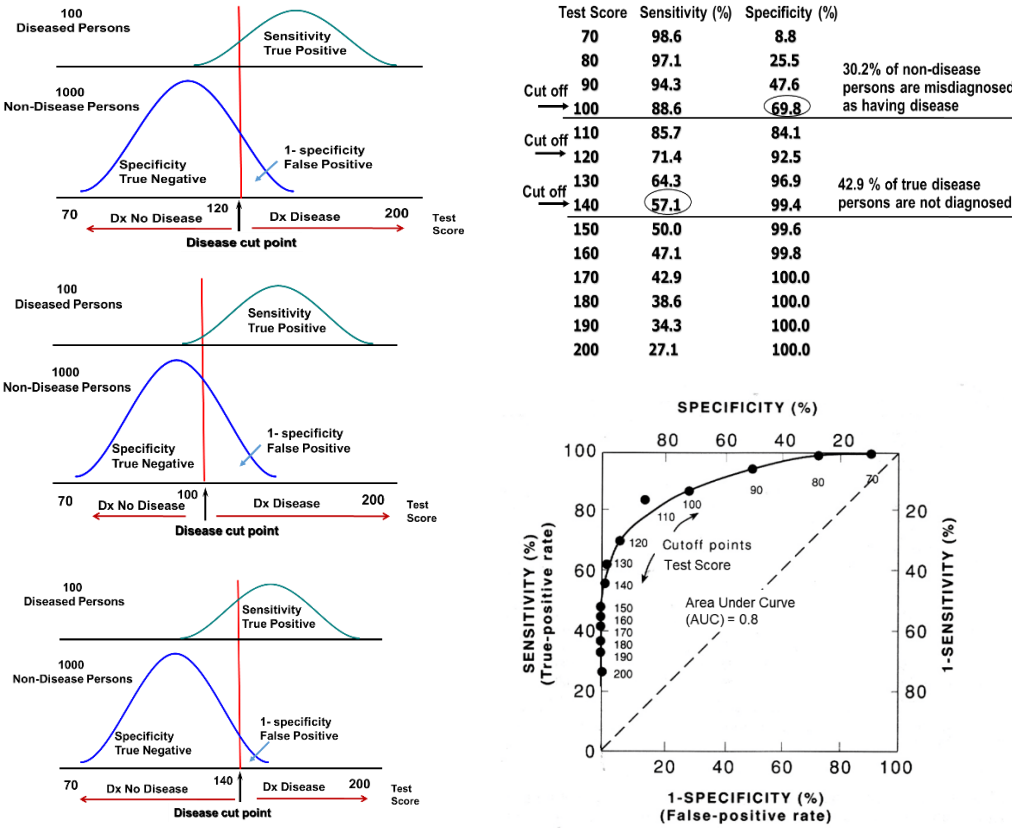

Figure 3. Sensitivity, specificity, receiver operating characteristic curve (ROC), and area under curve (AUC)

So, are You Positive that You Test Positive?

You can see now that the test kit could be used for different purposes–screening, diagnosis or confirmation. Many of those statistics related to the test kit depend on prevalence of the disease in the tested areas, characteristics of the tested populations, staging of the diseases, etc. Even though the test may not be perfect as it could not give 100% sensitivity and 100% specificity, or it does not show 100% correct predictive values, it is still useful in helping the clinician to estimate the probability that a person has disease. If your ATK test for COVID-19 shows positive, it may be true or false positive depending on the quality of the ATK itself. However, on the safe side, you should still follow the national practice guideline for COVID-19 case management.

Suggested Citation

Kaewkungwal J. The grammar of science: are you positive that your test is positive? OSIR. 2022 Jun;15(2):64–7.

References

1. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence Based Medicine. New York: Churchill Livingstone, 1997.
2. Machin D, Campbell MJ, Walters SJ. Medical Statistics: A Textbook for the Health Sciences. Chichester (UK): John Wiley & Sons, 2007.
3. Ruf M, Morgan O, Mackenzie K. Differences between screening and diagnostic tests and case finding [Internet]. Buckinghamshire (UK): HealthKnowledge; 2017 [cited 2022 Jun 25]. <https://www.healthknowledge.org.uk/public-health-textbook/disease-causation-diagnostic/2c-diagnosis-screening/screening-diagnostic-case-finding>
4. LaMorte WW. Screening for Disease: Test Validity [Internet]. Boston: School of Public Health, Boston University;[updated 2016 Jun 15, cited 2022 Jun 25]. <https://sphweb.bumc.bu.edu/otlt/mph-modules/ep/ep713_screening/EP713_Screening3.html>

5. Cardoso JR, Pereira LM, Iversen MD, Ramos AL. What is gold standard and what is ground truth? Dental Press J Orthod. 2014;19(5):27–30. doi:10.1590/2176-9451.19.5.027-030.ebo

6. Bachmann LM, Juni P, Reichenbach S, Ziswiler HR, Kessels AG, Vogelin E. Consequences of different diagnostic 'gold standards' in test accuracy research: Carpal Tunnel Syndrome as an example. Int J Epidemiol. 2005 Aug;34(4):953–5. <https://doi.org/10.1093/ije/dyi105>

7. Ford MP, Springman SR. Preanesthetic Evaluation: False-Positive Tests. In: Atlee JL, editor. Complications in Anesthesia. 2nd ed. Philadelphia: Saunders; 2007. p. 141-143.

8. Kanchanaraksa S, Johns Hopkins University. Evaluation of Diagnostic and Screening Tests: Validity and Reliability [Internet]. Baltimore (MD): Johns Hopkins University; 2008 [cited 2022 Jun 25]. <http://ocw.jhsph.edu/courses/fundepi/PDFs/lecture11.pdf>

9. Public Health England. Understanding cycle threshold (Ct) in SARS-CoV-2 RT-PCR: A guide for health protection teams. London: Public Health England; 2020 Oct [cited 2022 Jun 25]. 12 p. <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/926410/Understanding_Cycle_Threshold__Ct__in_SARS-CoV-2_RT-PCR_.pdf>

10. International Diabetes Federation. A Guide to National Diabetes Programmes [Internet]. Belgium: International Diabetes Federation; 2010 [cited 2022 Jun 25]. 122 p. <https://www.worlddiabetesfoundation.org/sites/default/files/Guide-to-NDP_web.pdf>

11. Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. Clin Biochem Rev. 2008;29 Suppl 1(Suppl 1):S83-7.

12. Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. Canadian Journal of Emergency Medicine. 2006;8(1):19–20. <https://doi.org/10.1017/S1481803500013336>

13. Zhu W, Zeng N, Wang N. Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS® Implementations [Internet]. SAS Conference Proceedings; 2010 Nov 14–17: Baltimore. [place unknown]: SAS Institute Inc; [cited 2022 Jun 25]. 9 p. <https://www.lexjansen.com/nesug/nesug10/hl/hl07.pdf>

14. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982 Apr;143(1):29–36.

15. Whiting PF, Sterne JAC, Westwood ME, Bachmann LM, Harbord R, Egger M & Deeks JJ. Graphical presentation of diagnostic information. BMC Med Res Methodol. 2008 Apr 11;8:20.