SVOA Medical Research
e-ISSN 2977-3644

ScienceVolks Open Access

Research Article

Understanding Test Sensitivity and Misunderstanding Test Specificity

Justin R.C. Abe, B.S[1*], Tyler J. Liu, B.S[1], Loren G. Yamamoto, MD, MPH, MBA[2]

[1]Medical Student at the University of Hawai`i John A. Burns School of Medicine, Honolulu, HI, USA.

[2]Professor of Pediatrics, University of Hawai`i John A. Burns School of Medicine, Honolulu, HI, USA.

***Corresponding Author:** Justin Abe, Medical Student at the University of Hawai`i John A. Burns School of Medicine, Honolulu, HI, USA.

DOI: https://doi.org/10.58624/SVOAMR.2026.04.005

Received: January 29, 2026

Published: February 23, 2026

Citation: Abe JRC, Liu TJ, Yamamoto LG. Understanding Test Sensitivity and Misunderstanding Test Specificity. *SVOA Medical Research* 2026, 4:1, 25-31. doi: 10.58624/SVOAMR.2026.04.005

Abstract

Background: Clinicians often assess test results by referencing sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). However, these concepts are frequently misinterpreted.

Objective: This study aimed to assess understanding of test sensitivity and specificity among physicians and medical trainees.

Methods: Questions were developed with the intention of testing participants on sensitivity and specificity in various ways: 1) a strict definition, 2) examples using numbers, and 3) within a clinical context. A survey containing these questions was distributed to medical students, residents, and physicians in 2023.

Results: 171 participants responded to the survey, which consisted of 83 trainees and 88 non-trainees. Since subjects who were contacted for the study were encouraged to share the study with other eligible participants that they knew, a response rate could not be obtained. Overall, participants performed better on questions that asked about sensitivity. 49% of participants correctly defined sensitivity, while 36% correctly defined specificity. The largest discrepancy was observed when participants were asked to define sensitivity and specificity within a clinical context: 63% answered correctly regarding sensitivity, as compared to 32% for specificity. When asked about specificity, the incorrect answer choices with the most participant responses referred to the positive predictive value.

Conclusion: Within this group of participants, understanding of sensitivity is moderate and understanding of specificity is poor. While this may be due to poor training or lack of clinical experience, we suspect it could also be due to inconsistencies in the English and epidemiological definitions of specificity.

Keywords: *Screening Test; Sensitivity; Specificity; Positive Predictive Value; Negative Predictive Value*

Introduction

Medical screening and diagnostic tests are often employed after forming a clinical differential diagnosis that requires further investigation. The results of these tests can range from slightly helpful to very helpful. Some tests form a gold standard which clearly establishes the diagnosis. For example, in a patient who presents with abdominal pain, an initial history and physical exam are helpful in assessing the risk for a serious condition such as acute appendicitis. While a white blood cell count, C-reactive protein, urinalysis, and ultrasound imaging all have varying degrees of diagnostic value, if the clinical risk is very high, an appendectomy is often performed and the histology of the appendix specimen forms the gold standard for the diagnosis of acute appendicitis.[1]

Clinicians assess the test results based on their diagnostic value which can be objectively measured by calculating their sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). [2]

Textbooks define these terms based on a 2 by 2 table of test results and the presence or absence of disease (based on a gold standard) as shown in Table 1. This table defines the true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). Sensitivity = TP / (TP + FN). Specificity = TN / (FP+TN). PPV = TP / (TP+FP). NPV = TN / (FN+TN).[3]

While this teaching is universal among clinicians, it is frequently misinterpreted [4]. The purpose of this study is to examine the level of understanding of the terms sensitivity and specificity among medical students and physicians.

Table 1. 2 by 2 table cross tabulating the test result (positive or negative) and the disease condition (positive or negative).

	Has disease (positive)	No disease (negative)
Test positive (abnormal)	True positive (TP)	False positive (FP)
Test negative (normal)	False negative (FN)	True negative (TN)

Methods

Subjects were invited via word of mouth and email to complete a survey testing knowledge of statistical principles in 2023. Subjects who were recruited for the study were also encouraged to share the study with other eligible participants that they knew. The survey was sent to medical students, residents, and attending physicians who were easily accessible by the study authors. Participants accessed the survey via online internet browser through an automated Google Form. Participation in the survey was voluntary and no compensation was given for completing the survey.

The first four questions in the survey tested the subject's conceptual understanding of sensitivity and specificity. Questions 1 and 2 assessed the subject's understanding of sensitivity in two different ways and Questions 3 and 4 assessed the subject's understanding of specificity in two different ways. The last two questions, Questions 5 and 6, tested the understanding of sensitivity and specificity in a clinical context (Supplemental Material 1).

This study was determined to be exempt from institutional board review using the guidelines set by the Office of Human Research Protection (45 CFR 46.104(d)(2)).

Results

The demographic characteristics of the 171 survey participants are as follows: 75 medical students, 8 residents, 6 physicians that recently completed fellowship, 80 practicing physicians, and 2 retired physicians. There was a total of 83 trainees (students, residents) and 88 non-trainees. Since subjects who were recruited for the study were also encouraged to share the study with other eligible participants that they knew, it was not possible to calculate the response rate.

Survey results are presented in Table 2. Correct answer percentages were higher for all questions about sensitivity compared to the questions about specificity. Participants who answered incorrectly on questions about test sensitivity and specificity most often picked the answers that described the PPV or NPV. Non-trainees had higher correct percentages compared to trainees for all questions.

Table 2. Survey results. Bold italic rows are the correct answers.

Question	**Answer choices** (See Supplementary Materials for full stem)	**Total answers** (%)	**# Non-trainee answers** (%)	**# Trainee answers** (%)
Q1 – Define sensitivity?	A (PPV)	42 (25%)	18 (20%)	24 (29%)
	B (NPV)	27 (16%)	14 (16%)	13 (16%)
	C (Correct)	***83 (49%)***	***50 (57%)***	***33 (40%)***
	D (Specificity)	18 (10%)	6 (7%)	12 (14%)
	I don't know	1 (0%)	0	1 (1%)
	All	171	88	83
Q2 – A highly sensitive test is?	***A (Correct)***	***86 (50%)***	***49 (56%)***	***37 (45%)***
	B (Not applicable)	4 (2%)	1(1%)	3 (4%)
	C (PPV)	27 (16%)	11 (12%)	16 (19%)
	D (NPV)	36 (21%)	19 (22%)	17 (20%)
	E (Specificity)	18 (11%)	8 (9%)	10 (12%)
	I don't know	0 (0%)	0 (0%)	0 (0%)
	All	171	88	83
Q3 – Define specificity?	A (PPV)	44 (26%)	24 (27%)	20 (24%)
	B (NPV)	42 (25%)	19 (22%)	23 (28%)
	C (Sensitivity)	22 (13%)	12 (14%)	10 (12%)
	D (Correct)	***62 (36%)***	***33 (36%)***	***29 (35%)***
	I don't know	1 (0%)	0 (0%)	1 (1%)
	All	171	88	83
Q4 – A highly specific test is?	A (Sensitivity)	19 (11%)	8 (9%)	11 (13%)
	B (Not applicable)	4 (3%)	1 (1%)	3 (4%)
	C (PPV)	53 (31%)	29 (33%)	24 (29%)
	D (NPV)	33 (19%)	15 (17%)	18 (22%)
	E (Correct)	***62 (36%)***	***35 (40%)***	***27 (32%)***
	I don't know	0 (0%)	0 (0%)	0 (0%)
	All	171	88	83

Table 2 Continued...

Q5 – Clinical data. What does 85% sensitivity mean?	***A (Correct)***	***107 (63%)***	***61 (69%)***	***46 (55%)***
	B (Not applicable)	1 (1%)	0 (0%)	1 (1%)
	C (PPV)	25 (14%)	10 (11%)	15 (18%)
	D (NPV)	20 (12%)	10 (11%)	10 (12%)
	E (specificity)	16 (9%)	5 (6%)	11 (14%)
	I don't know	2 (1%)	2 (3%)	0 (0%)
	All	171	88	83
Q6 – Clinical data. What does 95% specificity mean?	A (Sensitivity)	20 (12%)	10 (12%)	10 (12%)
	B (Not applicable)	6 (3%)	1 (1%)	5 (6%)
	C (PPV)	46 (27%)	24 (27%)	22 (26%)
	D (NPV)	39 (23%)	20 (23%)	19 (23%)
	E (Correct)	***54 (32%)***	***31 (35%)***	***23 (28%)***
	I don't know	6 (3%)	2 (2%)	4 (5%)
	All	171	88	83

Discussion

Many scientific publications examining test performance characteristics frequently cite the sensitivity and specificity of a test. Clinicians frequently cite the sensitivity and specificity as well when justifying why they believe that a test is either useless, moderate, or good. Yet the diagnostic utility of the calculated specificity performance parameter is poorly understood as it provides little value to clinical practice compared to other parameters such as PPV.[5]

Both trainees and practicing physicians were more accurate at defining sensitivity than specificity. Our study confirms that understanding specificity is particularly difficult since roughly only one-third of the study subjects were able to correctly answer the specificity questions. We suspect that this is because the English definition of specificity is very different from its epidemiological calculation, which indicates that a highly specific test is negative in the patients who do not have the disease. Since the English definition of specificity has primed our expectations of what specificity should be, the epidemiological calculation is non-intuitive. For those who did not select the correct choice for specificity, the most common selection was the one that describes the PPV.

In medical education and communication among clinicians, we commonly use the terms "specific" and "specificity" in the following ways. 1) In discussing acute phase reactants, we often state that these tests are non-specific. 2) In discussing a leukocytosis, we often state that this is not specific for entities such as bacterial sepsis, appendicitis, pneumonia, viral etiologies, and other similar entities. 3) We ideally want a clinical test that is very sensitive and very specific. These are more consistent with the English definition as opposed to the epidemiological calculation of specificity.

The paradox between the English definition and the epidemiological calculation can be best illustrated with an example: The pregnancy test is 100% specific for testicular torsion in males. Table 3 shows some fictitious data. Since males cannot have a positive pregnancy test, only negative pregnancy tests occur in this theoretical cohort. Out of 200 fictitious male subjects, 4 males have testicular torsion. The epidemiological calculation for specificity is the TN / (FP+TN), which in this case is 196 / (0 + 196) = 100%.

This is absurd because we would never use a pregnancy test in the evaluation of a male for testicular torsion. It is absurd to state that the pregnancy test is highly specific for testicular torsion, yet according to the epidemiological calculation, it is. This is clearly incongruent with the English meaning of specific.

Table 3. Fictitious data in males using a pregnancy test to identify its specificity for testicular torsion.

	Testicular torsion present	Testicular torsion absent
Pregnancy test positive	0 (TP)	0 (FP)
Pregnancy test negative	4 (FN)	196 (TN)

As another example, we use the bradypnea test to identify cases of aortic dissection. A respiratory rate (RR) of less than or equal to 8 per minute is a positive test, while a RR of greater than 8 is a negative test. Table 4 describes fictitious data on a cohort of healthy college students with mild chest pain. The calculated specificity is again 100%. Although the test has a calculated specificity of 100%, it is clearly not a test that has any utility for clinicians who desire a highly specific test. This is because the English definition of the word “specific” is not equivalent to the specificity calculation.

Table 4. Fictitious data in college students using the bradypnea test to identify its specificity for aortic dissection.

	Aortic dissection present	Aortic dissection absent
Positive bradypnea	0 (TP)	0 (FP)
Negative bradypnea	1 (FN)	199 (TN)

The English definition of specificity is more consistent with the PPV. If a test is positive, and if this is associated with a high likelihood of having the disease condition, it becomes highly diagnostic for that condition. A high PPV indicates a high likelihood of having the disease that the patient is being evaluated for, making a test with a high PPV, a highly “specific” test for the disease condition most of the time. Often, PPV better informs clinicians about the implications of a diagnostic test, which may be why clinicians prefer to use the PPV even though they often erfer to this as specificity.

Previous studies suggest that sensitivity and specificity should not be emphasized as they don’t describe the impact on patient outcomes[5], and given the results of our study, this may help to resolve the confusion when distinguishing “specificity” and PPV. The calculated specificity largely relates to those without disease and a low number of false positives. It is sometimes called the true negative rate[6], which is probably why many subjects mistakenly selected the statement describing the NPV. In fact, the sum of the selections for PPV and NPV exceeded the correct responses for specificity.

The verbal use of the term “specific” in medicine is unclear because of its English definition and its non-similar epidemiological definition. It’s easy to confuse the statements “the probability a patient has a test result whether or not they have a disease” and “the probability of having a disease based on their test result”. Clinicians cite the calculated specificity from a journal article, but potentially use this value incorrectly in making clinical decisions and in teaching trainees to make clinical decisions.

Understanding sensitivity is more straightforward because its English definition is similar to its epidemiological calculation; however, only 49%, 50%, and 63% correctly selected the correct response for the three sensitivity questions. High sensitivity can be thought of as a large fish net. A very sensitive test catches all the fish that we are targeting, but it also potentially catches turtles and dolphins that are not the target species. Because of this, a test that is highly useful must have more than just high sensitivity. It must have a high PPV as well. Adding NPV and specificity (i.e., all four parameters) most objectively characterize the diagnostic value of a test.[7]

It is difficult to recommend a better way to communicate with other clinicians. Effective communication requires clarity, yet our use of the terms specific and specificity result in substantial non-clarity. It is not likely that we can ever change the names of the terms that we use, thus, we need to be cognizant of the definitions and their associated ambiguity.

Adding to this confusion is the receiver operator characteristic (ROC) curve which is used to assess the diagnostic accuracy of a test. ROC curves have a vertical axis and a horizontal axis. An ROC curve will frequently plot the sensitivity on the vertical axis and “1 minus the specificity” on the horizontal axis[7]. This adds to the generalization that clinicians want a test with high sensitivity and specificity.

The questions used in this survey were developed with the intention of testing participants on sensitivity and specificity in various ways: 1) a strict definition, 2) examples using numbers, and 3) examples of a commonly employed test using data from a journal publication. The poor performance of our participants relative to other studies may be due to the wording of our questions and answer choices. This makes it challenging to compare our findings to other studies that may have used different methods to test understanding of these concepts. In a study by Bergus, 88% of 4th year medical students and 1st-year interns were able to correctly identify sensitivity and specificity[8]. While these numbers are much higher than our study, participants in the Bergus study were asked to critically appraise a study and to identify sensitivity and specificity. This may have skewed their answers positively compared to our study.

The answer choices used in multiple choice questions regarding sensitivity and specificity may also influence the degree of difficulty of questions in other studies. A study by Steurer found that 76% of physicians were able to define sensitivity[9], while 49% of physicians in our study were able to define sensitivity. However, in Steurer’s questionnaire, the other incorrect answers were “the percentage of persons without the disease having a negative test result” (the definition of specificity), “the false positive rate,” and the “false negative rate”. This essentially leaves two pairs of similar answers, and since the false positive and false negative choices are easily understood as not being the sensitivity, it renders this into a selection between two choices. In our questionnaire, the other incorrect answers were the definition of specificity, negative predictive value, and positive predictive value. This may suggest that while practitioners are able to distinguish sensitivity from specificity and false positive/negative rate, they struggle to differentiate sensitivity and specificity from positive and negative predictive value.

This study is not without limitations. Our study could have been affected by sampling bias if subjects who chose to complete the study differed significantly from the subjects who chose not to complete the survey. It is possible that the participants who chose to complete the study were more confident in their ability to identify statistical terms and thus the true understanding of specificity and sensitivity in the practicing physician community is worse than our study indicates. Additionally, a large portion of the physicians who were contacted were faculty physicians, so if mostly academic physicians responded (noting that the degree of academic training and accomplishments vary greatly at any faculty), it could have falsely elevated the scores further. On the other hand, we did not assess research involvement amongst physicians. If most physicians who responded are not involved in research, our scores could be falsely deflated since these physicians might not calculate sensitivity and specificity as frequently.

Conclusion

In conclusion, our study suggests that amongst trainees and practicing physicians, the understanding of sensitivity is moderate, and the understanding of specificity is poor. Additionally, our study suggests that that trainees and practicing physicians struggle to differentiate sensitivity and specificity from positive and negative predictive values and supports the notion that commonly used metrics of test accuracy are poorly understood by health professionals.

It was particularly concerning that the number of respondents who selected “I don’t know” was typically zero or a miniscule fraction of the number of respondents who selected an incorrect response, which indicates that the trainees and physicians may not be aware of their misunderstandings. As sensitivity and specificity are frequently used in clinical practice to characterize screening tests, educational interventions should be implemented so physicians can accurately interpret and communicate the validity of tests to patients.

Funding

None

Conflict of Interest

The authors declare no conflict of interest.

Prior Presentation

Previously presented as a poster at the Health Professions Education Conference at the John A. Burns School of Medicine on February 1, 2025.

Abbreviations

Positive predictive value **(PPV)**, Negative predictive value **(NPV)**, True positive **(TP)**, False positive **(FP)**, True negative **(TN)**, False negative **(FN)**, Respiratory rate **(RR)**, Receiver operator characteristic **(ROC)**.

References

1. Khan U, Kitar M, Krichen I, Maazoun K, Ali Althobaiti R, Khalif M, Adwani M. 2018. To determine validity of ultrasound in predicting acute appendicitis among children keeping histopathology as gold standard. Ann Med Surg (Lond). 2018;38:22-27.

2. Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. Acta Paediatr. 2007;96(3):338-341.

3. Celentano D, Szklo M, Farag YMK. Assessing the validity and reliability of diagnostic and screening tests. In: Celentano D, Szklo M, Farag YMK (eds). Gordis Epidemiology, 2025, Philadelphia: Elsevier Inc. pp:96-124.

4. Whiting PF, Davenport C, Jameson C, Burke M, Sterne J, Hyde C, et al. How well do health professionals interpret diagnostic information? A systematic review. BMJ Open. 2015;5(7):e008155.

5. Moons KG, Harrell FE. Sensitivity and specificity should be de-emphasized in diagnostic accuracy studies. Acad Radiol. 2003;10(6):670-672.

6. Zou KH, O’Malley AJ, Mauri L. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. Circulation. 2007;115(5):654–657.

7. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Front Public Health. 2017;5:307.

8. Bergus G, Vogelgesang S, Tansey J, Franklin E, Ronald F. Appraising and applying evidence about a diagnostic test during a performance-based assessment. BMC Med. Educ. 2004;4:20.

9. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. BMJ. 2002;324:824-826.