Journal of Community Hospital Internal Medicine Perspectives

co-action PUBLISHING

BIOSTATISTICAL CONCEPTS AND TOPICS IN RESEARCH

Diagnostic testing: a key component of high-value care

Lucien J. Cardinal, MD, FACP*

Internal Medicine Residency Program, Stony Brook Medicine at Mather Hospital, John T. Mather Memorial Hospital, Port Jefferson, NY, USA

This is the fourth article of a series on fundamental concepts in biostatistics and research. In this article, the author reviews the fundamental concepts in diagnostic testing, sensitivity, and specificity and how they relate to the concept of high-value care. The topics are discussed in common language, with a minimum of jargon and mathematics, and with clinical examples. Emphasis is given to conceptual understanding. A companion article will follow focusing on predictive value and prior probability.

Keywords: *sensitivity*; *specificity*; *statistics*; *testing*; *diagnostic testing*; *harms*; *high-value care*; *false positive*; *false negative*

*Correspondence to: Lucien J. Cardinal, Internal Medicine Residency Program, Stony Brook Medicine at Mather Hospital, John T. Mather Memorial Hospital, 75 North Country Road, Port Jefferson, NY 11777, USA, Email: lcardinal@matherhospital.org

Received: 18 March 2016; Revised: 16 April 2016; Accepted: 2 May 2016; Published: 6 July 2016

Diagnostic testing entails the attempt to differentiate whether disease is present or absent. While tests exist that have more than two outcomes, classic diagnostic algorithms are dichotomous in their outcome, meaning that there are only two possible results. These are categorized as 'positive' and 'negative'. Each test varies in its ability to predict the presence or absence of disease. Results will sometimes be false. The ability to predict correctly often differs based upon whether testing is performed in those who have or those who do not have disease. Descriptive measures of the performance of the test in each group (disease present or absent) have been developed that inform the clinician as to strengths and weaknesses of the test.

High-value care

The term high-value care encompasses providing good quality care and reducing unnecessary costs. This includes the ability to weigh the benefits and harms associated with testing and the limitation of testing to those patients who are likely to experience a net benefit. Patients have a right, among other quality parameters, to care that is effective, efficient, and safe (1). The six general physician competencies were established by the Accreditation Council of Graduate Medical Education (ACGME), when their board of directors approved them in February 1999 (2). Medical Knowledge, Interpersonal & Communication Skills, and Systems-Based Practice are three of the six competencies that touch on high-value care and require the clinician to be able to demonstrate an understanding of diagnostic testing, shared decision making, and cost awareness.

The United States Preventive Services Task Force (USPSTF), established in 1984, makes recommendations regarding preventive services for healthy populations (3). Part of their effort is devoted to quantifying the impact of false results of screening tests (e.g., mammography and prostate-specific antigen) on patients. Recommendations for screening are made based on the balance of benefits and harms likely to be experienced by those tested. Although false-positive (FP) results are more frequently cited as a source of harm, false-negative (FN) results lead to harm as well. FP results are particularly concerning as they generate harms to healthy individuals, violating the clinician's traditional dictum, 'first do no harm' (4).

> False-positive result = relating to a test result that is erroneously classified in a positive category when the sought after condition is not present.
>
> False-negative result = relating to a test result that is erroneously classified in a negative category when the sought after condition is present.

When patients are subjected to testing, a proportion of the results will falsely indicate the presence of disease. These healthy individuals with FP results may be exposed to further tests and unnecessary treatment. This may result in significant harm in terms of injury, expense, anxiety, and loss of time (5, 6). FP results commonly occur when diagnostic testing is performed on a person at low risk of disease. In such cases, a test should be

Citation: Journal of Community Hospital Internal Medicine Perspectives 2016, **6**: 31664 - http://dx.doi.org/10.3402/jchimp.v6.31664

recommended only when evidence clearly demonstrates that benefits outweigh harms. Still, unnecessary testing of a person at low risk, and therefore high likelihood of a FP result, is common, for example, routine mammography of women less than 40 years of age (7). In some cases, diagnostic labeling resulting from FP results may result in emotional trauma or damage to personal relationships (8, 9). Contrarily, a FN test result indicates the absence of disease in an individual in whom disease is present. This may result in false reassurance, a cessation of further necessary testing and failure to initiate therapy.

Other organizations have attempted to identify commonly used nonproductive testing strategies in persons with suspected disease. High quality recommendations take into account the likelihood of false test results and the associated increase in cost and harm. One example, the 'Choosing Wisely' campaign, an initiative of the American Board of Internal Medicine Foundation (10), brings together recommendations from over 70 medical organizations (11). These recommendations are freely available via the Internet.

In order to utilize diagnostic and screening tests in an efficient and appropriate manner and to interpret those tests correctly, the physician must have a fundamental understanding of the principles underlying those tests.

Introduction to diagnostic testing and terminology

When considering a patient for diagnostic testing, each individual is considered to be representative of a group that is composed of a mix of patients with similar characteristics, both with and without disease. This must be taken into consideration because test predictive ability differs based on the presence or absence of disease. Difficulties with test interpretation and indications for testing are often related to a failure to consider the patient's likelihood of disease prior to testing. Adding to difficulties is the fact that the likelihood of disease may change based on circumstance and setting. The following discussion is organized under two headings, with and without disease. This will reinforce to the reader that tests perform differently in those with and without disease and that the decision whether or not to order a test must consider the likelihood of disease in the given patient.

The term prevalence may lead to confusion. Prevalence is an epidemiologic term expressing the likelihood of disease in a population in a given time period. This term is typically applied to large populations, for example, the percent of women with breast cancer in the United Kingdom in 2015. While high-order prevalence may be useful to define the background likelihood of disease, its use is usually inappropriate to define a given patient's likelihood of disease. In the clinical setting, patient level characteristics, for example, personal and family medical history, social factors and suggestive symptoms, are used to more accurately approximate likelihood.

The meaning of the terms sensitivity and specificity are commonly confused or forgotten. The words, in and of themselves, do not suggest their meaning. It is recommended that the clinician become familiar with the following descriptive terms: true negative (TN), FP, true positive (TP), and FN. These terms are relatively transparent in referring back to the patient groups to which they relate (disease present or absent).

Abbreviation Key

Persons With Disease	Persons Without Disease
TP=True Positive	TN=True Negative
FN=False Negative	FP=False Positive
TPR=True Positive Rate	TNR=True Negative Rate
TPR=Sensitivity	TNR=Specificity
FNR=False Negative Rate	FPR=False Positive Rate

Testing in the disease-free group

Results are classified as true negative or false positive

When a group of individuals without the disease in question is subjected to testing, there are two possible results: negative and positive. A correct negative result in an individual without disease is referred to as a 'TN'. An incorrect positive result is referred to as a 'FP'. The number of disease-free persons tested is equal to the sum of the TN and FP.

$$\text{Disease free} = \text{TN} + \text{FP}$$

Quantifying predictive ability

Diagnostic test performance in the disease-free group may be quantified as the 'specificity'. The specificity is the ability of a test to correctly classify persons without the disease as being disease free.[1] A terminology that is more transparent is 'true negative rate (TNR)'. The terms are equivalent. The TNR can be calculated by dividing the number of TN results by the number of disease-free persons tested. The false positive rate (FPR) can be calculated by dividing the number of FP results by the number of disease-free persons tested.

$$\text{Specificity} = \text{TNR} = \text{TN/disease free} \quad \text{FPR} = \text{FP/disease free}$$

[1]For our purposes, 'disease free' means that there is an absence of the disease that the test is designed to detect. 'Disease present' in this context means that there is presence of the disease or condition that the test was designed to detect.

Citation: Journal of Community Hospital Internal Medicine Perspectives 2016, 6: 31664 - http://dx.doi.org/10.3402/jchimp.v6.31664

The TNR and FPR always sum to 100% of those that are disease free.

$$TNR + FPR = 100\% \qquad TN + FP = \text{Disease free}$$

Because of these equivalencies, the TNR can be used to calculate the FPR and vice versa. Likewise, the absolute number of TN can be used to deduce the FP, and vice versa given that the number of disease free is known. For example, if a test in a healthy population is (true) negative in 160 of 200 cases, then the remaining 40 cases must be (false) positive.

$$\text{Disease free} = TN + FP$$

$$200 = 160\ (TN) + FP \rightarrow FP = 200 - 160 = 40$$

One can further deduce from this information the TNR (specificity) of 80% and FPR of 20%.

$$TNR = 160/200 = 80\% \qquad FPR = 40/200 = 20\%$$

$$TNR + FPR = 80\% + 20\% = 100\%$$

Because the TNR and FPR sum to 100%, one can easily compute one from the other.

$$TNR + FPR = 100\% \quad TNR = 100\% - FPR \quad FPR = 100\% - TNR$$

$$80\% + 20\% = 100\% \quad TNR = 100\% - 20\% \quad FPR = 100\% - 80\%$$

Testing in the disease-present group

Results are classified as true positive or false negative

When a group of individuals with the disease in question is subjected to testing, there are two possible results: positive and negative. A correct positive result in an individual with disease is referred to as a 'TP'. An incorrect negative result found in an individual with disease is referred to as a 'FN'.

$$\text{Disease present} = TP + FN$$

Quantifying predictive ability

Analogous to the case in the disease-free population, the test accuracy in the disease population is characterized in terms of the positive rate and the negative rate. The true positive rate (TPR) is also called 'sensitivity'. Analogous to quantification of test results in the disease-free population, the sum of the TPR and false-negative rates (FNR) in the disease-present group is 100%. Knowing either value allows the deduction of the other. If the TPR is 25%, then the FNR is 75%.

Stability of sensitivity and specificity

It is commonly stated that sensitivity and specificity are stable regardless of the likelihood of disease in the population tested (12–15). By definition, TPR (sensitivity) and FNR apply to test performance in the disease-present subgroup (likelihood of disease of 100%). Analogously, FPR and TNR (specificity) apply to test performance only in the disease-absent subgroup (likelihood of disease of 0%). Therefore, regardless of how common disease is in the population tested, the calculation of the sensitivity or specificity is only based on performance of the test in the relevant subgroup (see Example 1).

However, unusual population characteristics may alter sensitivity or specificity by increasing FNs or FPs, respectively. For example, the test for syphilis, known as the rapid plasma regain test (RPR), has a high FPR when rheumatologic disease is present. An effort to screen for venereal disease in a busy rheumatologic practice using

Example 1

TPR = 25%; TNR = 50%

Likelihood of disease 40%

Number of patients = 10

With disease	4	*Without disease*	6
True positives	1	True negatives	3
False negatives	3	False positives	3

TPR = TP/with disease = 1/4 = 25%

TNR = TN/without disease = 3/6 = 50%

Likelihood of disease 80%

Number of patients = 10

With disease	8	*Without disease*	2
True positives	2	True negatives	1
False negatives	6	False positives	1

TPR = TP/with disease = 2/8 = 25%

TNR = TN/without disease = 1/2 = 50%

In the example above, the likelihood of disease varies from 40 to 80%. Note that although the number of persons that test positive or negative for disease change in each scenario, the TPR and the TNR remain the same.

the RPR would be expected to result in a higher FPR than that typically described for the general population (lowering the specificity) (16). Similarly, screening mammography has an increase in the FNR in women with increased breast density (lowering the sensitivity) (17, 18). Unrelated conditions may alter test accuracy. It is important to keep in mind that tests have limitations and are always subject to interpretation.

Finally, one can never use any of the parameters used to characterize the disease-present group (TPR, sensitivity, FNR, TP, or FN) to calculate any value used to characterize the disease-absent group (TNR, specificity, FPR, TN, or FP) and vice versa.

Acknowledgements

The author gratefully acknowledge the contributions of Yashodan Chivate and Sunna Zia, whose thoughtful reviews and constructive criticisms contributed to the completion of this document.

Conflict of interest and funding

The author has not received any funding or benefits from industry or elsewhere to conduct this study.

References

1. Institute of Medicine (US) (2001). Committee on quality of health care in America. Crossing the quality chasm: A new health system for the 21st century. Washington, DC: National Academy Press; 337 p.
2. Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. Health Aff (Millwood) 2002; 21(5): 103–11.
3. Lawrence RS, Mickalide AD, Kamerow DB, Woolf SH. Report of the US Preventive Services Task Force. JAMA 1990; 263(3): 436–7.
4. Jatoi I, Miller AB. Breast cancer screening in elderly women: Primum Non Nocere. JAMA Surg 2015; 150(12): 1107–8.
5. Klaas PB, Berge KH, Klaas KM, Klaas JP, Larson AN. When patients are harmed, but are not wronged: Ethics, law, and history. Mayo Clin Proc 2014; 89(9): 1279–86.
6. Sackett DL, Sackett DL. Clinical epidemiology: A basic science for clinical medicine. 2nd ed. Boston, MA: Little, Brown and company; 1991, xvii, 441 p.
7. Wilt TJ, Harris RP, Qaseem A. High value care task force of the American College of P. Screening for cancer: Advice for high-value care from the American College of Physicians. Ann Intern Med 2015; 162(10): 718–25.
8. Feig DS, Chen E, Naylor CD. Self-perceived health status of women three to five years after the diagnosis of gestational diabetes: A survey of cases and matched controls. Am J Obstet Gynecol 1998; 178(2): 386–93.
9. Meyer KB, Pauker SG. Screening for HIV: Can we afford the false positive rate? N Engl J Med 1987; 317(4): 238–41.
10. Cassel CK, Guest JA. Choosing wisely: Helping physicians and patients make smart decisions about their care. JAMA 2012; 307(17): 1801–2.
11. Foundation ABoIM (2015). Choosing wisely: About the campaign. Available from: http://www.choosingwisely.org/wp-content/uploads/2015/04/About-Choosing-Wisely.pdf [cited 23 May 2016].
12. Kramer MS. Clinical epidemiology and biostatistics: A primer for clinical investigators and decision-makers. Berlin: Springer-Verlag; 1988, xii, 286 p.
13. Fletcher RH, Fletcher SW, Fletcher GS. Clinical epidemiology: The essentials. 5th ed. Philadelphia, PA: Wolters Kluwer; 2014, 253 p.
14. Sox HC, Higgins MC, Owens DK. Medical decision making, 2nd ed. Chichester: Wiley; 2013.
15. Feinstein AR. Clinical epidemiology: The architecture of clinical research. Philadelphia, PA: W.B. Saunders Co; 1985, xii, 812 p.
16. Peter CR, Thompson MA, Wilson DL. False-positive reactions in the rapid plasma reagin-card, fluorescent treponemal antibody-absorbed, and hemagglutination treponemal syphilis serology tests. J Clin Microbiol 1979; 9(3): 369–72.
17. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA 1996; 276(1): 33–8.
18. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology 1992; 184(3): 613–17.