Acta Pædiatrica ISSN 0803–5253

REVIEW ARTICLE

Understanding diagnostic tests 1: sensitivity, specificity and predictive values

Anthony K Akobeng (tony.akobeng@cmmc.nhs.uk)

Department of Paediatric Gastroenterology, Booth Hall Children's Hospital, Central Manchester and Manchester Children's University Hospitals, Manchester, UK

Keywords
Disease prevalence, Negative predictive value, Positive predictive value, Sensitivity, Specificity

Correspondence
Dr. A. K. Akobeng, Department of Paediatric Gastroenterology, Central Manchester and Manchester Children's University Hospitals, Booth Hall Children's Hospital, Charlestown Road, Blackley, Manchester M9 7AA, United Kingdom.
Tel: 0161 220 5458 |
Fax: 0161 220 5072 |
Email: tony.akobeng@cmmc.nhs.uk

Received
8 June 2006; revised 4 December 2006; accepted 8 December 2006.

DOI:10.1111/j.1651-2227.2006.00180.x

Abstract

The usefulness of diagnostic tests, that is their ability to detect a person with disease or exclude a person without disease, is usually described by terms such as sensitivity, specificity, positive predictive value and negative predictive value. In this article, the first of the series, a simple, practical explanation of these concepts is provided and their use and misuse discussed. It is explained that while sensitivity and specificity are important measures of the diagnostic accuracy of a test, they are of no practical use when it comes to helping the clinician estimate the probability of disease in individual patients. Predictive values may be used to estimate probability of disease but both positive predictive value and negative predictive value vary according to disease prevalence. It would therefore be wrong for predictive values determined for one population to be applied to another population with a different prevalence of disease.

Conclusion: Sensitivity and specificity are important measures of the diagnostic accuracy of a test but cannot be used to estimate the probability of disease in an individual patient. Positive and negative predictive values provide estimates of probability of disease but both parameters vary according to disease prevalence.

INTRODUCTION

The usefulness of diagnostic tests, that is their ability to detect a person with disease or exclude a person without disease, is usually described by terms such as sensitivity, specificity, positive predictive value and negative predictive value (NPV). Many clinicians are frequently unclear about the practical application of these terms (1).

The traditional method for teaching these concepts is based on the 2 × 2 table (Table 1). A 2 × 2 table shows results after both a diagnostic test and a definitive test (gold standard) have been performed on a pre-determined population consisting of people with the disease and those without the disease. The definitions of sensitivity, specificity, positive predictive value and NPV as expressed by letters are provided in Table 1.

While 2 × 2 tables allow the calculations of sensitivity, specificity and predictive values, many clinicians find it too abstract and it is difficult to apply what it tries to teach into clinical practice as patients do not present as 'having disease' and 'not having disease'. The use of the 2 × 2 table to teach these concepts also frequently creates the erroneous impression that the positive and NPVs calculated from such tables could be generalized to other populations without regard being paid to different disease prevalence. New ways of teaching these concepts have therefore been suggested (2).

In this article, the first of the series, simple diagrams (not the 2 × 2 table) will be used to provide a practical explanation of what these concepts mean in clinical practice, and how they can be used to aid the diagnostic process.

HYPOTHETICAL POPULATION

To help understand the concepts of sensitivity, specificity and predictive values, imagine a hypothetical population of 100 people. Ten percent of the population (10 people) have a chronic disease, Disease A. We will assume that all the 100 people in the population have undergone bronchoscopy, the definitive method (gold standard) for diagnosing Disease A, so we are certain that the true prevalence of disease in this population is 10%. Figure S1 shows this population where white dots represent people without the disease and black dots represent people with the disease.

A new non-invasive test for diagnosing Disease A, Test A, has been developed which, hopefully, will help avoid the

Table 1 Defining sensitivity, specificity and predictive values from a 2 × 2 table

	Patients with disease	Patients without disease	
Test is positive	a	b	Total positive tests (a+b)
Test is negative	c	d	Total negative tests (c+d)
	Total number of patients with disease (a+c)	Total number of patients without disease (b+d)	Total number of patients (a+b+c+d)

Sensitivity: proportion of people with disease who will have a *positive* result {a/(a+c)}.
Specificity: the proportion of people without the disease who will have a *negative* result {d/(b+d)}.
Positive predictive value: the proportion of people with a positive test result who actually have the disease {a/(a+b)}.
Negative predictive value: the proportion of people with a negative test result who do not have disease {d/(c+d)}.

need for bronchoscopy (an invasive procedure) in some people being investigated for the disease. We will apply Test A to this population and use the hypothetical results to explain the concepts of sensitivity, specificity, NPV and positive predictive value.

SENSITIVITY

The sensitivity of a test is defined as the proportion of people ***with disease*** who will have a ***positive*** result. If we apply Test A to our hypothetical population, and 8 of the 10 people with Disease A test positive, then the sensitivity of the test is 8/10 or 80%. This is illustrated in Figure S2, where black dots in a red background represent people with disease who tested positive (true positives) and black dots which are not in a red background represent those with disease who tested negative (false negatives). Sensitivity is therefore calculated as the number of black dots in a red background (people with disease who tested positive) divided by the total number of black dots (all people with the disease).

Note that in defining sensitivity, we are only interested in the proportion of people ***with*** disease who test positive. Sensitivity can only be calculated from those people who have the disease (3). This means that the sensitivity of a test only tells us how good the test is for identifying people with disease when only looking at those ***with*** disease. Sensitivity tells us nothing about whether or not some people without the disease would also test positive and, if so, in what proportion.

SPECIFICITY

The specificity of a test is the proportion of people ***without*** the disease who will have a ***negative*** result. We can see from our hypothetical population (Fig. S1) that 90 people do not have Disease A. If we apply Test A to these 90 people and 85 of them test negative, then the specificity of the test is 85/90 = 94%. This is illustrated in Figure S3. All people without disease are represented by white dots. Those in a green background are without disease who tested negative. Specificity is calculated as number of white dots in a green background (people with disease who tested negative) divided by the total number of white dots (all people without disease).

Note that in defining specificity, we are only interested in the proportion of people ***without*** the disease who test negative. Specificity can only be calculated from those people who do not have the disease. Specificity tells us nothing about whether or not some people with the disease would also have a negative result and, if so, in what proportion.

USEFULNESS AND LIMITATIONS OF SENSITIVITY AND SPECIFICITY

Usefulness

A test with a high sensitivity is useful for 'ruling **out**' a disease if a person tests **negative** (4). Waisman and colleagues reported the sensitivity and specificity of the Uriscreen, a rapid diagnostic test for the early detection of urinary tract infection (UTI) to be, respectively, 100% and 68% in children presenting with symptoms suggestive of UTI (5). Imagine that we apply this test to a young girl who has developed symptoms suggestive of UTI and the test comes back positive. Does this mean she has UTI? The answer is maybe but maybe not. The 100% sensitivity means that the test will detect virtually every person who has UTI but its relatively low specificity means it will be falsely positive for a number of children who actually don't have UTI. The child's result might, therefore, be a false positive. What about if the child tests negative? Since 100% of children with UTI test positive, a person who tests negative is very *unlikely* to have the disease. A highly sensitive test is, therefore, most helpful to the clinician when the test result is negative. The mnemonic SnNout (high **Sen**sitivity, **N**egative test = rule **out**) is a useful way of remembering this principle (4).

A test with a high specificity is useful for 'ruling **in**' a disease if a person tests **positive** (4). Zaman et al. reported the sensitivity and specificity of the nitrite dipstick test in diagnosing UTI in hospitalized inpatients to be 27% and 94%, respectively (6). The sensitivity is pretty low but the specificity is high. If a person who presents with symptoms suggestive of UTI tests negative, does it mean he has not got UTI? We cannot tell. He actually might have UTI but the test's lack of sensitivity might have led to a false negative result. What about if he had tested positive? Since so many people without the disease test negative (94%), a person who returns a positive test is *likely* to have the disease. A highly specific test is, therefore, most helpful to the clinician when the test result is positive. The mnemonic for remembering this is SpPin (high **Sp**ecificity, **P**ositive test, rule **in**) (4).

The sensitivity and specificity of a test cannot be used to estimate the probability of disease in a patient (see below), but the two parameters could be combined into one measure called the likelihood ratio which may be used in

conjunction with disease prevalence to estimate an individual patient's probability of having disease. Likelihood ratios and how they can be used to estimate probability of disease will be discussed in the second article of the series.

Limitations

The major limitation of both sensitivity and specificity is that they are of no practical use when it comes to helping the clinician estimate the probability of disease in individual patients. When you see a patient in your clinic who returns a positive result for a particular test, the question that you and your patient would want an answer to is 'what is the chance (probability) of disease given the positive test?' Sensitivity and specificity cannot be used to answer such a question.

This is because both sensitivity and specificity are defined on the basis of people with or without a disease. However, because the patient would have presented to you with a set of symptoms rather than a diagnosis, you would not know at the time whether the patient has a disease or not and cannot, therefore, apply these parameters directly to them. What we need to know are predictive values which, in routine clinical practice, are more useful measures of diagnostic accuracy.

PREDICTIVE VALUES

The whole purpose of a diagnostic test is to use its results to make a diagnosis, so we need to know the probability that the test result will give the correct diagnosis (7). Positive and NPVs describe a patient's probability of having disease once the results of his or her tests are known.

Positive predictive value

The positive predictive value (PPV) of a test is defined as the proportion of people with a positive test result who actually have the disease. In the hypothetical population of 100 people, you will recall that 8 people with Disease A had a positive result for Test A, and 5 people without disease also tested positive. This means that a total of 13 people tested positive. Figure S4 shows these 13 people in a red background. You will realize that out of these 13 people, only 8 of them actually had the disease (black dots). From Figure S4, the PPV of Test A is calculated as the number of people with Disease A who tested positive (the number of black dots in red background) divided by the total number of people who tested positive (the total number of dots in red background) which is $8/13 = 0.62$ or 62%. This means that, in this hypothetical population, 62% of people who test positive will have Disease A, or put in another way, a person who has a positive test has a 62% chance of having Disease A. PPV is, sometimes, also referred to as the 'post-test probability of disease given a positive test.'

Negative predictive value

The NPV of a test is the proportion of people with a negative test result who do not have disease. In our hypothetical population of 100 people, 85 people who did not have Disease A tested negative, and 2 people who had Disease A also tested negative. Thus a total of 87 people tested negative. Figure S5 shows these 87 people in a green background. Out of these 87 people, 85 did not have the disease (white dots). From Figure S5, the NPV of test A is calculated as the number of white dots in a green background divided by the total number of dots in a green background which is $85/87 = 0.98$ or 98%. This means that 98% of people who test negative for Test A will not have Disease A, or put in another way, a person who has a negative test has a 98% chance of *not* having Disease A.

You can deduce from the above that NPV may also be defined as the probability of *not* having disease given a negative test. It is therefore important to note that the 'post-test probability of disease given a negative test' is not the same as the NPV but is the converse (1-NPV). In this example, the post-test probability of disease given a negative test will be $1- 0.98 = 0.02$ or 2%. This means that in this hypothetical population, a person who tests negative for Test A only has a 2% chance of having Disease A.

PREDICTIVE VALUES AND DISEASE PREVALENCE

The predictive value of a test is determined by the test's sensitivity and specificity and by the prevalence of the condition for which the test is used (8). Both PPV and NPV vary with changing prevalence of disease. It will therefore be wrong for clinicians to directly apply published predictive values of a test to their own populations, when the prevalence of disease in their population is different from the prevalence of disease in the population in which the published study was carried out.

To further understand the relationship between predictive values and disease prevalence, recall that I earlier calculated the PPV and NPV of Test A in our hypothetical population (with Disease A prevalence of 10%) to be 62% and 98%, respectively. Imagine that we now apply Test A to another population of 100 people but in whom the prevalence of Disease A is 20%. We already know that the sensitivity of Test A is 80%, which means that 80% of the 20 people with Disease A (16 people) in this population will test positive. The specificity of the test is 94%, which means that 94% of people without Disease A will test negative or that 6% of people without the disease will test positive. Thus 6% of the 80 people without Disease A (5 people) will test positive. Thus a total of 21 people will test positive, 16 with Disease A and 5 without.

The PPV of test A for this population is therefore calculated to be 16/21 or 76%. In a similar way, we can work out that a total of 79 people would test negative (75 without the disease and 4 with the disease). Thus the NPV (the proportion of people with a negative test who do not have disease) is 75/79 or 95%. When we repeat these calculations on other populations with different Disease A prevalence, we will see clearly that the PPV of the test increases with increasing prevalence of disease and the NPV decreases with increasing prevalence (Table 2).

Clinical implications

You can gather from Table 2 that the higher the disease prevalence, the higher the PPV, that is the more likely a

Table 2 Relationship between predictive values and disease prevalence (a test with a sensitivity of 80% and a specificity of 94%)

Prevalence (%)	Positive predictive value (%)	Negative predictive value (%)
5	40	99
10	62	98
20	76	95
40	89	87
50	93	82
60	96	76

positive result is able to predict the presence of disease. When the prevalence of disease is low, the PPV will also be low, even when using a test with high sensitivity and specificity. In such a situation, a significant proportion of people who have a positive test may not necessarily have disease.

What this means in clinical practice is that the usefulness of a test result for an individual patient depends on the prevalence of the disease in the population being tested. The diagnostic value of a test will be much improved if, based on our history and clinical assessment, we limit the use of the test to those patients who are likely to have the disease in question. A positive or a negative result is then more likely to be meaningful, than when the test is indiscriminately applied to patients. A diagnostic test should be used to supplement rather than as a substitute for clinical judgement.

Defining the population

You should be aware that the term 'population' as used in the above context does not necessarily refer to people in a specified geographical area. It could also refer to a constellation of people with similar symptoms and/or signs. For example, the prevalence of bacteraemia in the population of 10-month-old infants with high temperature in Manchester will be higher than the prevalence of bacteraemia in 10-month-old infants in the same city who are well with no symptoms.

Thus the PPV of, say, the white blood cell count in diagnosing bacteraemia will be higher in the former group of babies meaning that a 10-month-old baby in Manchester with a high temperature who has an elevated white blood cell count will be more likely to have bacteraemia than a well 10-month-old baby in the same city who also has an elevated white blood cell count on routine testing.

It must also be pointed out that the same test result (positive or negative) may yield different predictive values in primary care, secondary care or tertiary care settings in the same geographical region according to the prevalence of disease in these settings.

CONCLUSION

The sensitivity and specificity of a test have limited clinical usefulness as they cannot be used to estimate the probability of disease in an individual patient. Predictive values may be used to estimate this but both PPV and NPV vary according to disease prevalence, and published predictive values should not be applied to populations whose prevalence of disease is different from the population in the published study. There are simple, practical ways of estimating probability of disease (predictive values) for individual patients in routine clinical practice. These will be discussed in the second article of the series.

ACKNOWLEDGEMENT

I thank Akosua Manu Akobeng for helping to design the figures in this article.

References

1. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002; 324: 824–6.
2. Loong TW. Understanding sensitivity and specificity with the right side of the brain. *BMJ* 2003; 327: 716–9.
3. Mayer D. *Essential evidence based medicine*. Cambridge: Cambridge University Press, 2004.
4. Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine: how to practice and teach EBM*. London: Churchill-Livingstone, 2000.
5. Waisman Y, Zerem E, Amir L, Mimouni M. The validity of the uriscreen test for early detection of urinary tract infection in children. *Pediatrics* 1999; 104: e41.
6. Zaman Z, Borremans A, Verhaegen J, Verbist L, Blanckaert N. Disappointing dipstick screening for urinary tract infection in hospital inpatients. *J Clin Pathol* 1998; 51: 471–2.
7. Altman DG, Bland JM. Diagnostic tests 2: predictive values. *BMJ* 1994; 309: 102.
8. Last JM. *A dictionary of epidemiology*. New York: Oxford University Press, 2001.

Supplementary material

The following supplementary material is available for this article:

Figure S1 Hypothetical population of 100 people with Disease A prevalence of 10%
Figure S2 Sensitivity of Test A
Figure S3 Specificity of Test A
Figure S4 Positive predictive value
Figure S5 Negative predictive value

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1651-2227.2006.00180.x