Sensitivity, specificity and a used car salesman: which one should you trust?

SALLY MURRAY MB BS, MPHTM, MHSc(PHP), FAFPHM

Almost daily, GPs are confronted with decisions about initiating tests or interpreting test results. Making decisions in this setting requires understanding the characteristics of the tests as well as the population in which the tests are being used.

Tests may be very good at detecting disease or, alternatively, they may be good at excluding disease. The value of a test depends also on how common the disease is in a population. The best screening test in the world for prostate cancer is not of much use in most women!

Additionally, being able to share information about the use and interpretation of tests with your patients is essential for providing them with the ability to give informed consent for undergoing any screening or diagnostic testing. This should include helping patients to understand the ramifications of positive and negative results.

Terminology

There are several important terms and concepts relevant to describing tests. These include sensitivity, specificity, and positive and negative predictive values (see the box on this page). Likelihood ratios, and pre- and post-test probabilities are also important in assessing what a result means for the patient in front of you.

Sensitivity

Sensitivity is the likelihood that when someone has a disease, the test returns a positive result. If a test has a high sensitivity, it means that it will detect nearly all cases of a disease: that is, it has a high rate of true positive results. Or put another way, it has a low false negative rate. Tests with high sensitivity are used when the goal of the test is to rule out disease. A sensitive test (S) that is negative (N) rules disease out (OUT), SNOUT. For example, the ELISA test for HIV is very sensitive. When it is negative, it rules out the possibility that a patient has HIV.

Dr Murray is Senior Research Fellow, Department of Paediatrics, University of Melbourne, Vic.

Specificity

Specificity is the likelihood that when someone does not have a disease, a test correctly identifies that the disease is absent. A test with high specificity has a low false positive rate, and a high rate of true negatives. Specific tests are used when it is important to be quite definite that disease is present. A specific test (S) that is positive (P) rules disease in (IN), SPIN. For example, because of the many ramifications of a patient testing positive for HIV, it is important to be sure that a positive test means he or she has the disease. Hence, to confirm the findings from a positive ELISA test, a highly specific Western Blot test is used. A positive Western Blot test rules HIV in.

Positive and negative predictive values

While sensitivity and specificity are fixed characteristics of a test, positive and negative predictive values depend on how common a disease is in a given population. The predictive values provide information on how likely having a disease is in an individual depending on both whether the result is positive as well as the population from which the individual comes. Characteristics of the individual, such as gender, help define the population in which we are testing.

Positive predictive value (PPV) is the likelihood that when the test is positive, a person really does have the disease.

Sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV)

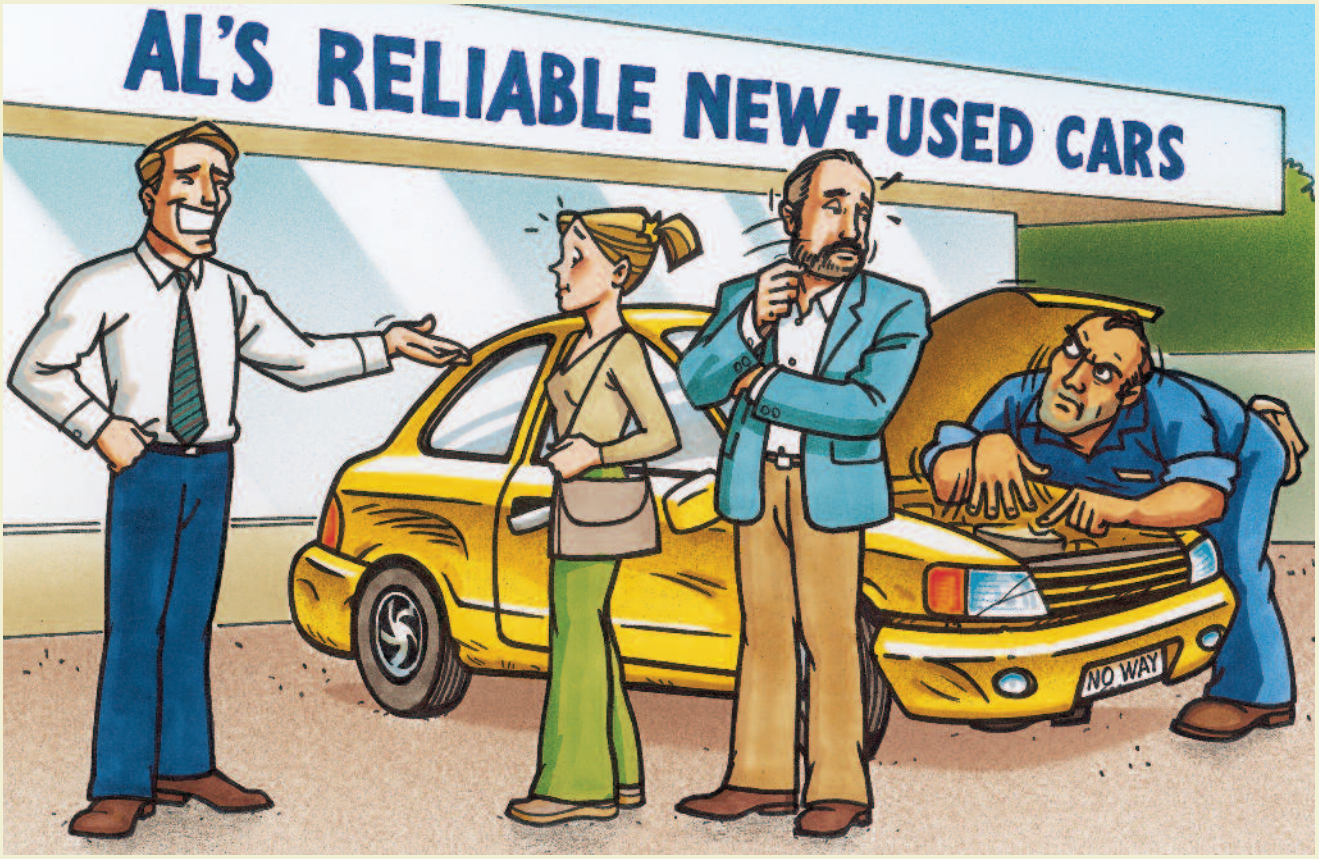
	Disease present	Disease absent
Positive test result	True positive (A)	False positive (B)
Negative test result	False negative (C)	True negative (D)

$$\textbf{Sensitivity} = \frac{\text{number of true positives}}{\text{number of all people with disease}} = \frac{A}{A+C}$$

$$\textbf{Specificity} = \frac{\text{number of true negatives}}{\text{number of people without the disease}} = \frac{D}{B+D}$$

$$\textbf{Positive predictive value} = \frac{\text{number of true positives}}{\text{number of positive results}} = \frac{A}{A+B}$$

$$\textbf{Negative predictive value} = \frac{\text{number of true negatives}}{\text{number of negative results}} = \frac{D}{C+D}$$



Whereas negative predictive value (NPV) is the likelihood that when the test is negative, the person really does not have the disease.

These values are highly dependent on how common the disease is in the population. If a test is positive but a disease is very rare, then it is quite likely that the test is falsely positive (i.e. has a low PPV). For example, an elevated prostate specific antigen (PSA) level in a woman is likely to be a false positive! On the other hand, if a disease is common, such as diabetes, and the test (e.g. elevated fasting blood sugar level) is positive then the result is likely to be a true positive (i.e. the test has a high PPV).

A good example of how to increase the predictive value of tests is the use of screening in high risk populations. Screening high risk populations, rather than the whole population, increases the prevalence of the disease in the screened population, thereby increasing the PPV of the test. This is the basis of many screening programs (e.g. breast screening for women aged 50 to 69 years).

Likelihood ratios and pre- and post-test probabilities

Another way of thinking about test results and what a positive or negative test means for the patient is to consider likelihood ratios and pre-test and post-test probabilities.

Likelihood ratios (LRs) reflect properties of tests themselves and provide a measure of whether a test can be used to rule in a disease, which occurs if the LR is high (e.g. above 10), or rule out a disease, when the LR is low (e.g. less than 0.1). Likelihood ratios can also be used to calculate post-test probability. Post-test probability, like positive predictive value, determines how likely your patient is to have a disease given a positive test result. The likelihood ratio for a positive test result (LR+) is equal to sensitivity/(1–specificity) while the likelihood ratio for a negative result (LR–) is (1–sensitivity)/specificity.

Prevalence (the amount of disease in your population) can be thought of as the pre-test probability that your patient has the disease for which you are testing. You will make this judgement based on his or her risk factors, history and any clinical findings. Take the example of detecting disease from a breast lump biopsy. The pre-test probability in a 60-year-old woman with a breast lump, palpable lymph nodes and a positive family history is much greater than in an otherwise well 19-year-old with a breast lump and no relevant history.

Having determined the pre-test probability and likelihood ratio, you can then calculate the post-test probability using the calculation sequence outlined in the box on page 94.

Deciding whether to use a test

Tests are usually either more sensitive or more specific; that is, there is a trade-off between the two. Choosing which test is appropriate for a patient depends on both of these properties as well as what additional tests or procedures will follow from

a positive result, and what the ramifications are of missing a case of the disease. It will also depend on how common the disease is in the relevant population at that time.

If you are testing with the aim of detecting every case of a disease, a highly sensitive test is used; however, this is likely to result in an increased number of false positive results. If the repercussions of a positive test are significant (e.g. very invasive follow up tests), then a test with high specificity can be chosen to avoid too many false positives.

Combining tests can improve the specificity or sensitivity of testing. One way to increase sensitivity of testing is by using parallel screening tests. This is when multiple tests are administered at the same time – for example, prostate assessment with both digital rectal examination and PSA testing. Specificity of testing can be increased by the use of serial screening. The earlier example of ELISA and Western Blot testing for HIV is an example of serial screening.

When deciding whether to use a new or unfamiliar test, there are a few other considerations. These include the expertise required to perform or interpret a test. Factors such as these may alter the usefulness of a test in your own setting. Similarly, if your patients are very different from those described in the original study that trialled a specific test (e.g. at a much later stage of disease), the test may not be as useful.

Calculation of pre-test and post-test probabilities

1. Calculate pre-test probability
   Pre-test probability = prevalence of disease
2. Convert to pre-test odds
   Pre-test odds = prevalence/(1–prevalence)
3. Calculate post-test odds
   Post-test odds = pre-test odds x likelihood ratio*
4. Convert to post-test probability
   Post-test probability = post-test odds/(post-test odds +1)

*Likelihood ratio (LR) = sensitivity/(1-specificity)

Table. How good is your mechanic? His last 100 assessments

	Car is a bomb (Disease present)	Car is OK (Disease absent)
Assessed as 'old bomb' (Positive test result)	20	30
Assessed as 'OK' (Negative test result)	5	45

Can you trust the used car salesman?

A familiar example may help lighten something that you swore you would never revise again. Suppose you have succumbed to your 17-year-old daughter or son's pleading to buy her or him a second hand car. It would be reasonable to seek a mechanic's evaluation of the car to increase your confidence that you were choosing a mechanically sound car (for your child's safety and your pocket). However, to really assess your chances of finding a reliable car, you would need to know about the mechanic (the test) and how common old bombs are in the second hand market (prevalence). Regarding your mechanic, you would want to know how good he is at detecting car trouble (sensitivity) and how reliable his assessment is that a car is 'OK' (specificity).

A sensitive mechanic

The Table on this page shows your mechanic's performance for his last 100 customers seeking an assessment of a second hand car. From the table you can see that the prevalence of bombs is 0.25, or 25%. Sensitivity, the likelihood that when there are problems with the car the mechanic detects them, is 20/25=0.80, or 80%. On the other hand, specificity, the likelihood that when there are no problems with the car the mechanic correctly says it's OK, is 45/75=0.60, or 60%.

Predicting a good deal

The positive predictive value in this example is the likelihood that the car has a problem when the mechanic says it does; PPV=20/50=0.40, or 40%. The negative predictive value is the likelihood that when the mechanic says there is no problem, the car is indeed fine; NPV 45/50=0.90, or 90%. From this, you can be fairly confident that when the mechanic says 'no worries' you are in fact going to get a good deal. On the other hand it is likely that you will miss a few bargains because the mechanic is only right 40% of the time when he says the car has a problem. These values are predictive in this community, that is, where the prevalence of old bombs is 0.25. If you were selecting from a pool of 1-year-old Mercedes, the positive predictive value would be less because mechanical problems would be unlikely. Conversely, negative predictive value would be greater.

Now, if you really cannot stand getting ripped off, you will need to increase the sensitivity of your test. This means you will increase the chance of correctly identifying the cars with problems. To do this maybe you can request a more senior mechanic or a mechanic who specialises in the relevant make of car. Alternatively you can have several mechanics look over

it at the same time (parallel testing) to increase the likelihood that if the car has a problem it will be detected (sensitivity).

Probably avoiding a dud

What about likelihood ratios and post-test probabilities? Using the calculations outlined in the box on page 94, it is possible to calculate the pre and post-test probabilities. The likelihood ratio (LR) comes from the known sensitivity and specificity of our local mechanic's testing procedures. (In this case, LR=0.80/[1–0.60]=2.0.) The steps are outlined as follows:

1. The pre-test probability is the prevalence of bombs on the market; this is 0.25
2. Pre-test odds = prevalence/(1–prevalence)
   = 0.25/0.75 = 0.33
3. Post-test odds = pre-test odds × the likelihood ratio
   = 0.33 × 2.0 = 0.66
4. Post-test probability = post-test odds/(post-test odds+1)
   = 0.66/1.66 = 0.40

What the post-test probability tells us is the likelihood of a car being a bomb after the mechanic states there are big problems with it. Here, if a car gets the thumbs down then there is only a 40% chance the car has a problem. In this scenario you are likely to miss some bargains. By the same token, you are not too likely to buy a dud.

Summary

When using a screening or diagnostic test in any population it is important to remember why we do so in the first place. Will the results change your management? Will your patient be better off as a result of the test? Screening and diagnostic tests are usually done when there is treatment available for a disease and with the belief that early diagnosis can improve a patient's outcome. The costs of unnecessary or imperfect screening or diagnostic processes borne by a patient must always be considered. Anxiety and discomfort can be associated with the test or future follow-on tests. Anxiety also results from unclear or false positive test results. Moreover, our health system can ill afford the financial costs of unnecessary tests. If the pre-test probability of disease is very low or the test results will not alter your management, why test?

Doctors face daily the need to describe tests and their results to patients. In a setting where no test is perfect, it is important we have the skills to ensure that our patients as well as our lawyers understand the limitations of investigations! MT

INDEX TO ADVERTISERS