Clinical review

Understanding sensitivity and specificity with the right side of the brain

Tze-Wey Loong

Can you explain why a test with 95% sensitivity might identify only 1% of affected people in the general population? The visual approach in this article should make the reason clearer

Department of Community, Occupational, and Family Medicine, National University of Singapore, Singapore
Tze-Wey Loong
clinical teacher (part time)

Correspondence to:
T-W Loong, King George's Medical Centre, Block 803 King George's Avenue, #01-144, Singapore 200803, Singapore
tzewey@singnet.com.sg

I first encountered sensitivity and specificity in medical school. That is, I remember my eyes glazing over on being told that "sensitivity = TP/TP+FN, where TP is the number of true positives and FN is the number of false negatives." As a doctor I continued to encounter sensitivity and specificity, and my bewilderment turned to frustration—these seemed such basic concepts; why were they so hard to grasp? Perhaps the left (logical) side of my brain was not up to the task of comprehending these ideas and needed some help from the right (visual) side. What follows are diagrams that were useful to me in attempting to better visualise sensitivity, specificity, and their cousins positive predictive value and negative predictive value.

Sensitivity and specificity

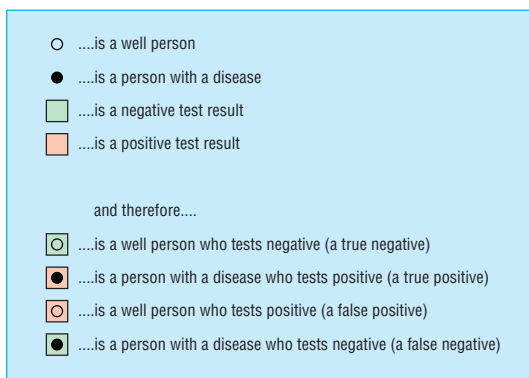I will be using four symbols in these diagrams (fig 1).

Fig 1 Key to symbols

Let us start by looking at a hypothetical population (fig 2). The size of the population is 100 and the number of people with the disease is 30. The prevalence of the disease is therefore 30/100 = 30%.

Now let us imagine applying a diagnostic test for the disease to this population and obtaining the results shown in figure 3. The test has correctly identified most, but not all of the people with the disease. It has also correctly labelled as disease free most, but not all, of the well people. Calculating sensitivity and specificity will allow us to quantify these statements.
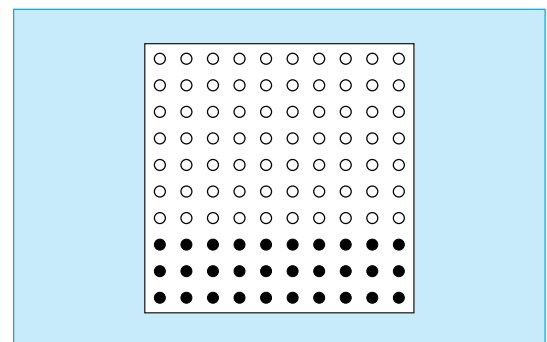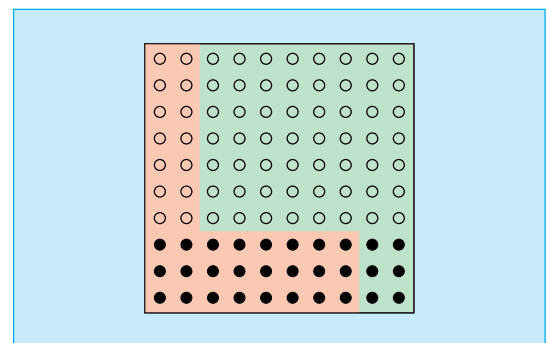

Fig 2 Hypothetical population

Fig 3 Results of diagnostic test on hypothetical population

Sensitivity refers to how good a test is at correctly identifying people who have the disease. When calculating sensitivity we are therefore interested in only this group of people (fig 4). The test has correctly identified 24 out of the 30 people who have the
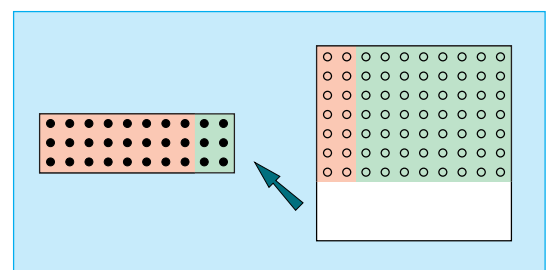

Fig 4 Sensitivity of test

disease. Therefore the sensitivity of this test is $24/30 = 80\%$.

==Specificity, on the other hand, is concerned with how good the test is at correctly identifying people who are well== (fig 5). The test has correctly identified 56 out of 70 well people. The specificity of this test is therefore $56/70 = 80\%$.
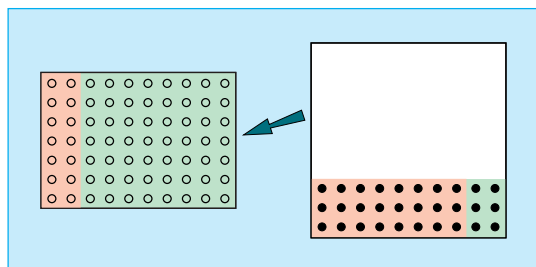

Fig 5 Specificity of test

Having a high sensitivity is not necessarily a good thing, as we can see from figure 6. This test has achieved a sensitivity of 100% by using the simple strategy of always producing a positive result. Its specificity, however, clearly could not be worse, and the test is useless. By contrast, Figure 7 shows the result a perfect test would give us.
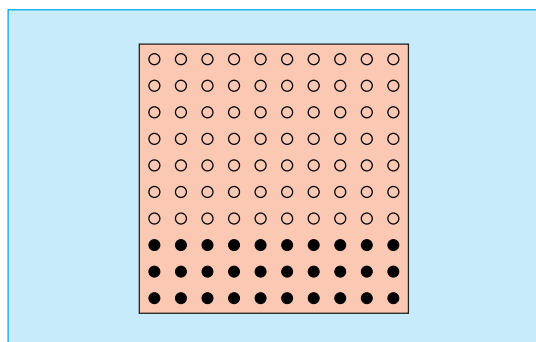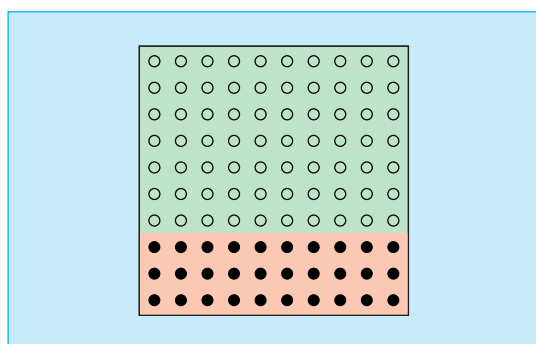

Fig 6 Test with 100% sensitivity

Fig 7 Perfect test

Predictive values

Now let us consider positive predictive value and negative predictive value. We will again use the population introduced in figure 3. Positive predictive value refers to the chance that a positive test result will be correct. That is, it looks at all the positive test results. Figure 8 shows that 24 out of 38 positive test results are correct. The positive predictive value of this test is therefore $24/38 = 63\%$.
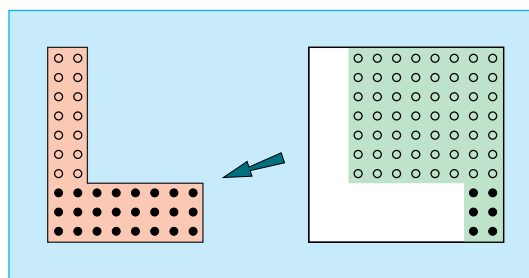

Fig 8 Positive predictive value

On the other hand, negative predictive value is concerned only with negative test results (fig 9). In our example, 56 out of 62 negative test results are correct, giving a negative predictive value of $56/62 = 90\%$.
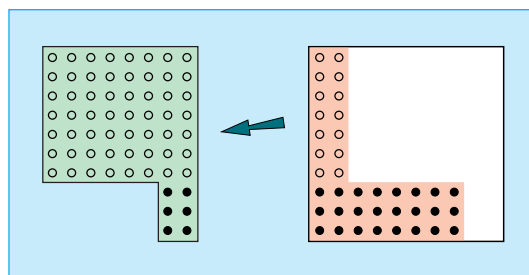
Fig 9 Negative predictive value

The interesting thing about positive and negative predictive values is that they change if the prevalence of the disease changes. Let's assume that the prevalence of disease in our population has fallen to 10%. If we were to use the same test as before, we would obtain the results in figure 10. The sensitivity and

Negative predictive value

Positive predictive value

Fig 10 Results of testing population with disease prevalence of 10%

specificity have not changed (sensitivity $= 8/10 = 80\%$ and specificity $= 72/90 = 80\%$), but the positive predictive value is now $8/26 = 31\%$ (compared with 63% previously) and the negative predictive value is $72/74 = 97\%$ (compared with 90% previously).

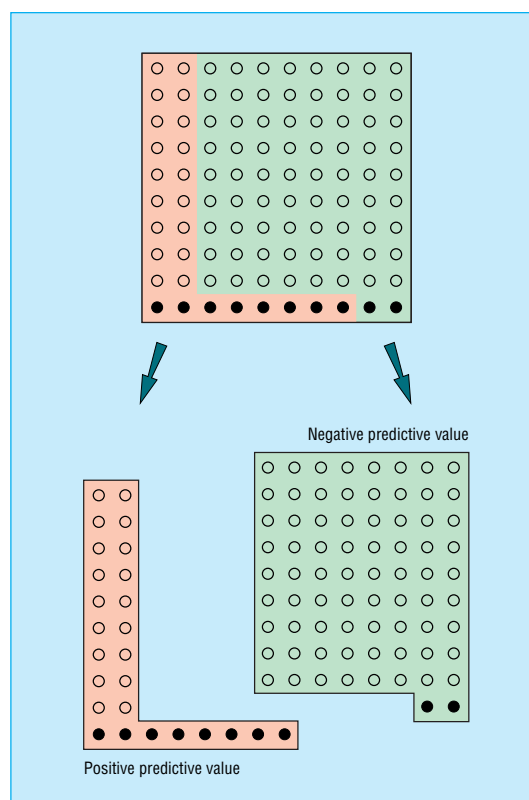In fact, for any diagnostic test, the positive predictive value will fall as the prevalence of the disease falls while the negative predictive value will rise. This is not really so mystifying if we consider the prevalence to be the probability that a person has the disease before we do the test. A low prevalence simply means that the person we are testing is unlikely to have the disease and therefore, based on this fact alone, a negative test result is likely to be correct. The following real example should make this clearer.

A real example

So far we have been discussing hypothetical cases. Let us now take a look at the use of the antinuclear antibody test in the diagnosis of systemic lupus erythematosus. I have massaged the numbers slightly to make them easier to illustrate, but they are close to reported figures in both the United Kingdom and Singapore.[1][2] The prevalence of systemic lupus erythematosus is 33 in 100 000, and the antinuclear antibody test has a sensitivity of 94% and a specificity of 97%. To visualise this we need to imagine 1000 of the 10 by 10 squares used in the earlier figures (fig 11). Only one of these squares contains some patients with the disease.
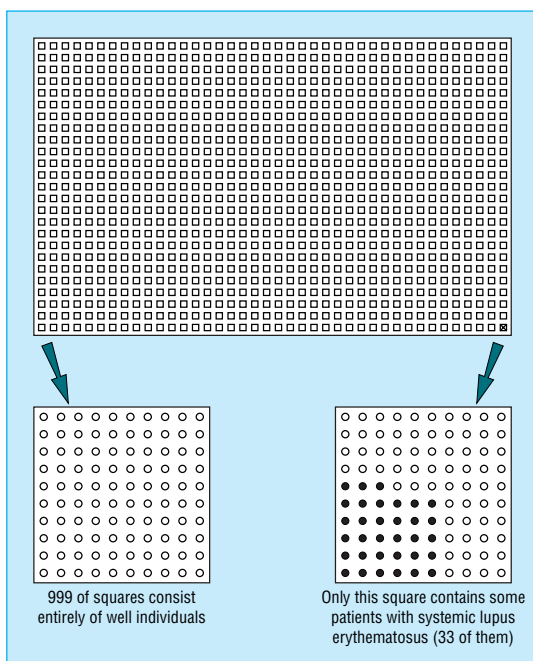

Fig 11 Prevalence of systemic lupus erythematosus

Figure 12 shows the result of applying the antinuclear antibody test to this population. There are many more true negative results than false negative results and many more false positive than true positive results. The test therefore has a superb negative predictive value of 99.99% and a depressingly low positive predictive value of about 1%. In practice, since most diseases have a low prevalence, even when the tests we

No of true negatives = 96 900

No of false negatives = 2

Negative predictive value = $\dfrac{96\ 900}{96\ 900 + 2}$

$= 99.99\%$

No of true positives = 31

No of false positives = 3067

Positive predictive value = $\dfrac{31}{31 + 3067}$

$\cong 1\%$
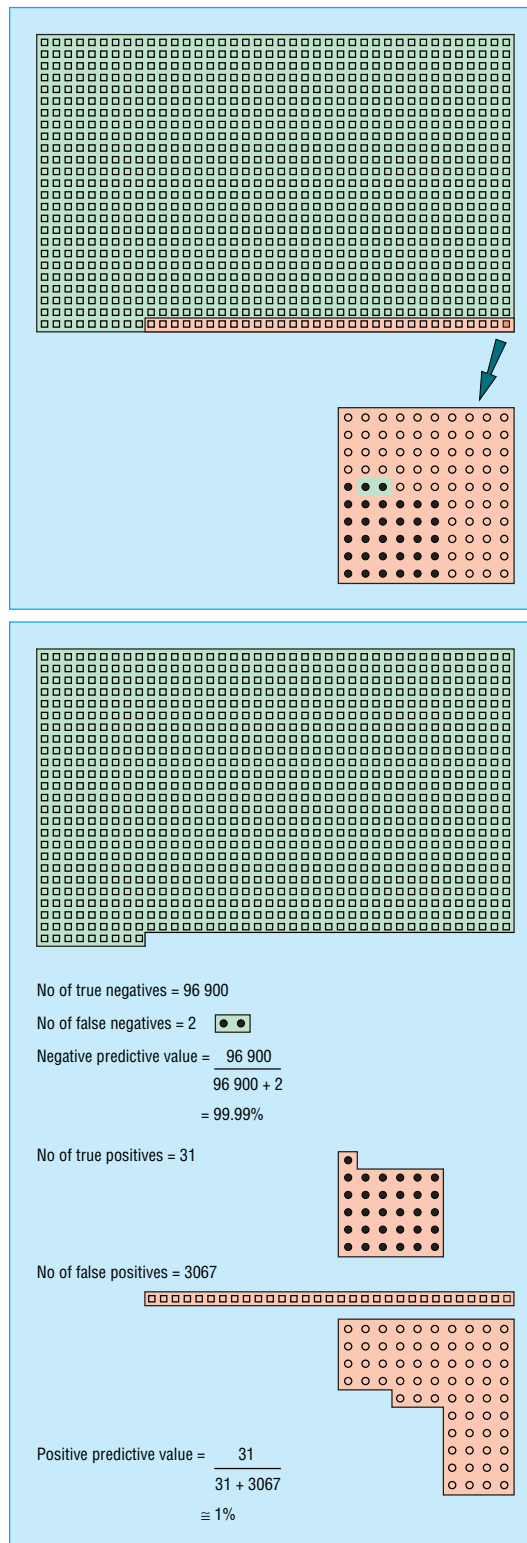
Fig 12 (top) Results of antibody nuclear test in systemic lupus erythematosus; (bottom) negative and positive predictive values

use have apparently good sensitivity and specificity we may end up with dismal positive predictive values.

Knowing that the positive predictive value of this test is 1%, we may then ask: does a positive test result in a female patient with arthritis, malar rash, and proteinuria really mean that she has only a 1% chance of actually having systemic lupus? The answer is no.

Summary points

- For a given test, the lower the prevalence of the disease, the lower the positive predictive value
- Since most diseases have a low prevalence in the general population, even a test with an apparently good sensitivity and specificity (>90%) may have a very low positive predictive value
- However, if this test is applied to a person with symptoms or signs of the disease, the positive predictive value will be higher, as that person is from a population with a higher prevalence of the disease

Look at it this way—the patient is not a member of the general population. She is from the population of people with symptoms of systemic lupus erythematosus, and in this population the prevalence is much higher than 33 in 100 000. Hence the positive predictive value of the test in her case is going to be much higher than 1%.

Using both sides of the brain

I hope that having worked through sensitivity and specificity from scratch you will be wondering why it initially seemed so confusing. It may be because of our dependence on the left (linguistic) side of the brain. When told that a test has a sensitivity of 94% and a positive predictive value of 1%, our left brain has difficulty grasping how a test can be 94% sensitive and yet be correct only 1% of the time. It is partly misled by the huge difference between prevalence, on the one hand, and sensitivity and specificity on the other. The prevalence of systemic lupus erythematosus is 0.033% while the sensitivity and specificity of the test are about 95%; this difference is of several orders of magnitude. If, for example, we developed a test with sensitivity and specificity of 99.999% rather than 95%, we would be able to boast of a positive predictive value of 97%.

Competing interests: None declared.

1 Johnson AE, Gordon C, Palmer RG, Bacon PA. The prevalence and incidence of SLE in Birmingham, England. Relationship to ethnicity and country of birth. *Arthritis Rheum* 1995;38:551-8.
2 Boey ML. Systemic lupus erythematosus. *Singapore Med J* 1992;33:291-3.

Who invented that bleeping thing?

While preparing a talk on Thomas Fogarty, of balloon catheter fame, I stumbled on information about a different gentleman who was a joint winner with Fogarty of the much coveted MIT-Lemelson prize. This person is someone who affects nearly all doctors every day. Indeed, if he had not recently died, I am sure many of us would love to get our hands on him. However, as you read on and discover what a truly remarkable man he was, you may see him and his invention in a different light.

Al Gross was born in 1918 in Toronto but grew up in Cleveland, Ohio. He had a childhood interest in amateur radio and went on to study for a diploma in electronics. He was a bright student, and his area of interest lay in unexplored radio frequencies above 100 MHz. He wanted to invent a small, mobile, two way radio, and by1938, two years into his diploma, he had invented the first handheld radios ("walkie-talkies"), which could communicate for up to 30 miles. These caused quite a stir with the military, who deemed their invention as "top secret." They quickly commandeered the idea and furthered its use to introduce ground to air communication for fighter pilots and to detonate bombs at a distance, such as for blowing up bridges.

After time, these long range radios were made public knowledge, and, as a result, in 1946 citizen band (CB) radio was invented, the familiar mode of communication of taxi drivers and truckers.

A more sinister twist in Gross's career occurred in 1949, when he invented the telephone pager system. However, his first large scale attempt to sell pagers to doctors did not meet with the success he had anticipated. "In Philadelphia, there was a hospital convention, and we set up the pager there. We demonstrated the pager to all the hospital administrators, doctors, and nurses, and they absolutely refused to go along with the idea," said Gross. "They claimed it would disturb the patient, the nurses wouldn't want to carry it, and the doctors would be disturbed in their game of golf."

Although the idea initially failed to catch on, New York's Jewish Hospital did install his paging system in 1950, and the Federal Communications Commission officially approved it in 1958, marking the era of mass production. The name "pager" is derived from the Motorola Pageboy 1, one of the first commercially available models.

As we are daily reminded, pagers are here to stay. Recent estimates suggest that there are now over 60 million pagers in use worldwide. All I can say in their defence is that, when you are out on a Saturday night in the rain, and taxi control gets a cab to you in five minutes using CB radio communication, perhaps you will see your pager in a different light.

Fraser Smith research registrar, St James's Hospital, Dublin, Republic of Ireland