Interpreting pathology tests
Screening versus diagnostic testing

PAT PHILLIPS MB BS, MA(Oxon), FRACP, MRACMA, GradDipHealthEcon(UNE)

The concepts of sensitivity, specificity and predictive value and how they can be applied to clinical testing are discussed in this third article in a series of four outlining a framework to interpret the results of pathology laboratory tests.

Tests may be performed because they provide a diagnosis in a patient with specific symptoms (diagnostic tests) or because they give information about the likelihood of an asymptomatic disease being present or likely to develop in a person without specific symptoms (screening tests). Many people believe that a positive laboratory test result is always positive for the disease and a negative laboratory test result excludes it. But sometimes a positive result is a false positive and a negative result is a false negative. The predictive value of a positive or negative result depends on the characteristics of the test itself (its sensitivity and specificity) and the likelihood of disease before the test is performed. In many cases of screening for a problem that has a low probability of being present, the sensitivity and specificity of the screening test are such that a positive result is more likely to be a false positive than an indicator of the presence of the disease.

This third article in a series of four outlining a framework to interpret laboratory results discusses the concepts of sensitivity, specificity and predictive value and how they can be applied to clinical testing. A case study using a hypothetical test for diabetes ('DiabDiag') is presented to illustrate the principles behind screening and diagnostic testing.

MedicineToday 2015; 16(9): 67-71
Amended October 2015

Dr Phillips is a Consultant Endocrinologist at the QE Specialist Centre, Adelaide, SA.

Sensitivity and specificity

For many laboratory tests, the distribution of results from people with and without the disease in question overlaps (Figure 1). Using the hypothetical DiabDiag test as an example, it is difficult to decide the glycosylated haemoglobin (HbA_{1c}) level above which diabetes is indicated and below which diabetes is considered absent – the decision point. If the HbA_{1c} level is set low enough to pick up all the people who have undiagnosed diabetes, many people who do not have diabetes but who have positive tests will also be identified as having diabetes (level A in Figure 2). On the other hand, if the level is set high enough to exclude all those who do not have diabetes, some of those who do have diabetes will be missed (level B in Figure 2). Clearly, there must be a compromise between a level that is so high that it misses many people with diabetes, or so low that it diagnoses many people who do not have diabetes.

Assuming a compromise level is chosen (level C in Figure 2), the test will be positive in most people with diabetes (e.g. 70%) and negative in most people without diabetes (e.g. 70%). The test is then said to have:

- a sensitivity of 70% (testing positive in 70% of those with diabetes)
- a specificity of 70% (testing negative in 70% of those without diabetes).

These characteristics allow the estimation of the likelihood that a positive test is a true positive (indicating diabetes) or a true negative (excluding diabetes), as indicated in Box 1.

Setting levels for tests

The level of a test considered as indicating the presence of a particular disease is decided by considering the following:

- the sensitivity and specificity of the test
- the likely presence of the disease in the tested population
- the importance of correctly diagnosing the disease if it is present

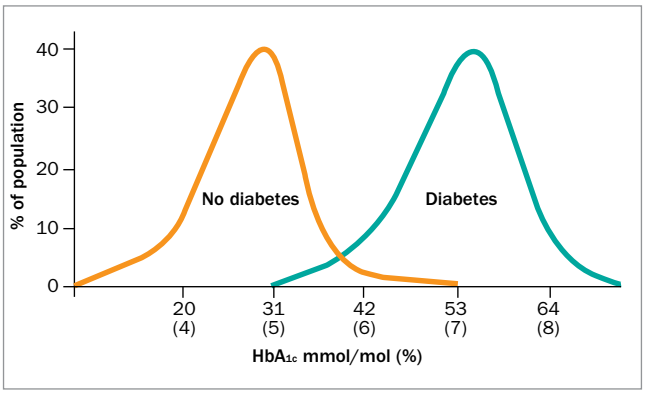

Figure 1. Distribution of HbA_{1c} levels in people with and without diabetes.

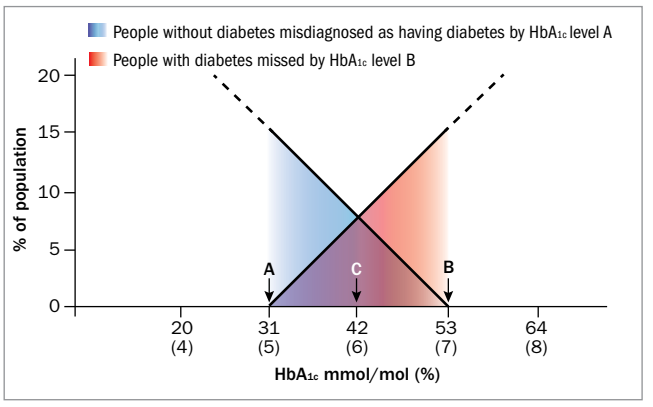

Figure 2. Testing strategies to screen for undiagnosed diabetes. An HbA_{1c} level set low enough to pick up all the people who have undiagnosed diabetes will identify as having diabetes those who do not have diabetes but who have positive tests (level A). A level set high enough to exclude people who do not have diabetes will not identify some of the people who do have diabetes (level B). A compromise diagnostic level (level C) is set so that the test will be positive in most people with diabetes and negative in most people without diabetes. The perfect test is always positive in those with the condition and negative in those without (sensitivity and specificity both 100%); tests in the real world rarely have sensitivity or specificity greater than 80%.

- the importance of not diagnosing the disease in those without it.

For example, if the likely prevalence of a disease is quite high and it is important not to miss the disease, a low test level might be considered positive (i.e. the sensitivity will be increased even though the specificity will be decreased). A higher test level with a higher specificity and lower sensitivity will be set for a disease with a low prevalence and where it is important not to falsely suggest disease in those without it.

The difference between screening and diagnostic tests is that a screening test is performed for diseases with low prevalence in the population being tested and yields many false positives, whereas a diagnostic test is used for diseases with a higher prevalence in the tested population and will yield fewer false positives.

Pre-test probability of disease and predictive value

Sometimes a test is performed because it provides a diagnosis in a patient with specific symptoms; examples are tissue biopsy in cancer and measurement of human chorionic gonadotropin (hCG) levels in early pregnancy. Sometimes testing is done in people without specific symptoms and which does not make a diagnosis itself but makes the presence of a disease more or less likely. This can be referred to a screening test. Often these tests are used to help determine those individuals needing more follow-up testing and those for whom nothing further is required.

1. SENSITIVITY AND SPECIFICITY

Within a population some people have a condition (like diabetes) and others do not. The sensitivity of a test is the frequency the test is positive in those with the condition. The specificity is the frequency the test is negative in those without the condition.

Other terms used are:

- true positive (TP) and false positive (FP) rates – the frequency of positive tests in those with and without the condition, respectively
- true negative (TN) and false negative (FN) rates – the frequency of negative tests in those without and with the condition, respectively.

A condition may be present or absent and the test for it may be positive or negative.

True positive rates, i.e. sensitivity
= Number of TP ÷ Number with disease (i.e. TP + FN)

False positive rates
= Number of FP ÷ Number without disease (i.e. FP + TN)

True negative rates, i.e. specificity
= Number of TN ÷ Number without disease (i.e. FP + TN)

False negative rates
= Number of FN ÷ Number with disease (i.e. TP + FN)

The perfect test is always positive in those with the condition and negative in those without (sensitivity and specificity both 100%). Tests in the real world rarely have sensitivity or specificity over 80%.

2. PROBABILITIES: PERCENTAGES AND ODDS RATIOS

Probabilities can be expressed in at least two ways:

- percentage – indicating the proportion of people under consideration who have the disease being tested for
- odds ratio – indicating the odds that a person might have the disease being tested for.

For example, if the prevalence of undiagnosed diabetes in the population being tested is 0.4%, four in 1000 people will have diabetes and 996 will not, which is equivalent to an odds ratio of 4:996.

3. CASE SCENARIO ILLUSTRATING PREDICTIVE VALUES

Case scenario

Don and his daughter Donna have come to see you because they want diabetes checks. At a health-screening day at Donna's university they were both told that they could have diabetes because Don's father, mother, elder brother and one of his two sisters have type 2 diabetes.

Don is 52 years old, 181 cm tall and weighs 100 kg (BMI, 30.5 kg/m^2).

Donna is 23 years old, 162 cm tall and 63.5 kg (BMI, 24.2 kg/m^2).

What are Don and Donna's risks of diabetes?

Risk factors for type 2 diabetes include overweight, inactivity, increasing age, family history (in a first-degree relative), ethnicity and social disadvantage, the major drivers being the F words – Forty, Family and Fat.[1]

Don has all three 'F words' for type 2 diabetes: he is over Forty, has a positive Family history in a first-degree relative and is Fat (BMI, 30.5 kg/m^2).

Donna has none of the 'F words' for diabetes: she is 23 years old, has no positive family history of diabetes and her BMI is in the healthy weight range.

Several algorithms exist to predict the risk of a person developing diabetes in the next five or 10 years. One of these is the QDScore developed by Hippisley-Cox and colleagues in 2009.[2] The paper relating to this score, published in the *BMJ*, contains graphical representations of age interactions for adults at risk of type 2 diabetes, displayed as hazard ratios. From these, the effects of the 'F words' on Don and Donna's risks of developing diabetes can be calculated, assuming a prevalence of undiagnosed diabetes in the general population of 0.4% (i.e. 4 in 1000 men and women aged 25 years who have no family history of diabetes and are a healthy weight will have undiagnosed diabetes):[2]

- Don's risk is increased 15-fold by his age of 52 years, a further 2.7-fold by his positive family history, and a further threefold by his fatness. This is a total increase in risk of 125-fold, and Don's odds ratio (OR) for having undiagnosed diabetes is increased from 4:996 to 500:996 (and his percentage probability is increased from 0.4% to 500 ÷ [500 + 996], or 33%).
- Donna's risk is not increased by any 'F words' and remains 0.4%.

Case continued

Performing the hypothetical DiabDiag test for diabetes gives HbA_{1c} results of 34 mmol/mol (5.3%) for both Don and Donna.

What are the likelihoods of Don and Donna having diabetes?

Let it be assumed that the DiabDiag test used for determining HbA_{1c} level has a sensitivity and specificity for diabetes of 70% at an HbA_{1c} of 34 mmol/mol (5.3%) – i.e. 70% of those with diabetes will have a positive test result and 70% of those without diabetes will have a negative result, the decision point for a positive result being equal to or greater than 34 mmol/mol.

It is often useful to consider 1000 or 10,000 people like the one individual being tested so there are whole numbers of people in the calculation and not fractions.

Likelihood of Don having diabetes – positive predictive value

If a population of 1000 people with the same pre-test probability as Don is tested (i.e. a 33% risk of having diabetes), 33% (330) will have diabetes and 67% (670) will not.

In the 330 with diabetes, from the sensitivity for diabetes of 70%, 70% will have a positive HbA_{1c} test result (231, the true positives [TPs]) and 99 will not (the false negatives [FNs]).

In the 670 without diabetes, from the specificity for diabetes of 70%, 30% will have a positive HbA_{1c} test result (201, the false positives [FPs]) and 70% will not (469, the true negatives [TNs]).

There will be a total of 432 positive tests (231 in those with diabetes and 201 in those without).

The likelihood that a positive test result indicates diabetes (i.e. the predictive value of a positive test result) is the number of TPs divided by the total number of positive tests (i.e. TPs plus FPs). For Don:

231 ÷ 432 = 53%

Likelihood of Donna having diabetes – positive predictive value

If a population of 10,000 people with the same pre-test probability as Donna is tested (i.e. a 0.4% risk of having diabetes), 0.4% (40) will have diabetes and 99.6% (9960) will not.

In the 40 with diabetes, 70% will have a positive HbA_{1c} test result (28, the TPs) and 30% will not (12, the FNs).

In the 9960 without diabetes, there will be 30% with a positive HbA_{1c} test result (2988, the FPs) and 70% without (6972, the TNs).

There will be a total of 3016 positive tests (28 in those with diabetes and 2988 in those without).

The likelihood that a positive test result will indicate diabetes in Donna is:

28 ÷ 3016 = 0.9%

Predictive values

The predictive value of a positive result in this hypothetical test in Don is 53% and in Donna is 0.9%, despite their having identical HbA_{1c} results. This is because of the higher pre-test probability of diabetes for Don.

The general formulas for determining positive and negative predictive values of tests are:

Positive predictive value of test = TPs ÷ all positive tests

= (pre-test probability of condition x TP rate) ÷ [(pre-test probability of condition x TP rate) + (pre-test probability of no condition x FP rate)]

Negative predictive value = TNs ÷ all negative tests

= (pre-test probability of no condition x TN rate) ÷ [(pre-test probability of no condition x TN rate) + (pre-test probability of condition x FN rate)]

(Note: TN rate = specificity and FN rate = 1 – TP rate, where TP rate = sensitivity)

Examples of negative predictive value:

If the HbA_{1c} test result were negative for Don (i.e. below 34 mmol/mol [5.3%]), the negative predictive value would be:

TN ÷ all negative tests

From Don's pre-test probability of diabetes of 33% and the test's specificity of 70% and sensitivity of 70%,

TN = (pre-test probability of no condition x TN rate), where TN rate = specificity

TN = (67% x 70%) = (0.67 x 0.7)

and

All the negative tests = (pre-test probability of no condition x TN rate) + (pre-test probability of condition x FN rate), where TN rate = specificity and FN rate = 1 – TP rate

All the negative tests = (67% x 70%) + [33% x (1 - 70%)]

= (0.67 x 0.7) + (0.33 x 0.3)

Therefore the predictive value of a negative test for Don is:

TN tests (0.67 x 0.7) ÷ All the negative tests (0.67 x 0.7 + 0.33 x 0.3)

= 0.47 ÷ (0.47 + 0.099)

= 0.83 or 83%

For Donna, the negative predictive value would be (from her pre-test probability of diabetes of 0.4% and the test's specificity of 70% and sensitivity of 70%):

TN ÷ all negative tests

TN = (1 - 0.4%) x 70%) = 0.996 x 0.7

and

All the negative tests = [(1 - 0.4%) x 70%) + [0.4% x (1 - 70%)]

= (0.996 x 0.7) + (0.004 x 0.3)

Therefore the predictive value of a negative test for Donna is:

TN tests (0.996 x 0.7) ÷ All the negative tests (0.996 x 0.7 + 0.004 x 0.3)

= 0.6972 ÷ (0.6972 + 0.0012) = 0.998 or 99.8%

The negative test in Donna makes the likelihood of her not having diabetes much higher than Don's likelihood because her pre-test probability was lower.

4. PRACTICE POINTS: SCREENING VERSUS DIAGNOSTIC TESTING

- The sensitivity and specificity of tests and pre-test probability of the condition being tested for all need to be considered when interpreting 'positive' and 'negative' test results.
- The predictive value of a positive or negative test result can be calculated using the sensitivity, specificity and pre-test probability.
- Likelihood ratios are a simple way to work out the post-test probability of the condition from the sensitivity, specificity and pre-test probability.
- For diseases with a low likely prevalence, a positive test is likely to be false positive.
- For diseases with a high likely prevalence, a negative test is likely to be false negative.
- When comparing the clinical performance of different tests it is important to consider the decision points used, as sensitivity and specificity can be adjusted by using higher or lower decision points.

Before a test is performed, the healthcare professional will have some idea of how likely the condition is in the population to be tested (the pre-test probability of the condition). Probabilities can be expressed as either percentages or odds ratios, as discussed in Box 2. An individual's test result adds information to this pre-test probability – the pre-test probability of the disease in that individual.

The likelihood that a positive test result indicates a particular condition (i.e. the predictive value of a positive test result) and that a negative test result indicates the condition is not present (i.e. the predictive value of a negative test result) can be calculated from the pre-test probability of the individual and the test's sensitivity and specificity for that condition.

The case scenario of two people (Don and Donna) being tested for diabetes illustrates how the same test result may have different predictive values in different settings (Box 3).

Likelihood ratios

Some people prefer to calculate the predictive value of a positive or negative test, as discussed above. Others want something a bit easier. The likelihood ratio is the ratio of the probability that a result is correct to the probability that the result is incorrect.

Positive likelihood ratio (LR) = sensitivity ÷ (1-specificity)
or True positive (TP) rate ÷ False positive (FP) rate

Negative LR = (1 – sensitivity) ÷ specificity
or False negative (FN) rate ÷ True negative (TN) rate

For tests with likelihood ratios above 1, the post-test probability of disease is higher than the pre-test probability. For tests with the likelihood ratio below 1, the post-test probability is decreased below the pre-test probability. The process is quite understandable, and more information is given in *Clinical Epidemiology – How To Do Clinical Practice Research*.[4]

Returning to Don and Donna in the case scenario (Box 2), the likelihood ratio for an HbA_{1c} of 34 mmol/mol (5.3%), following the previous assumption that this level of HbA_{1c} has a sensitivity and specificity for diabetes of 70%, is:

TP ÷ FP = 70 ÷ 30
= 2.3

- For Don, the likelihood ratio of 2.3 increases his odds ratio from 1:2 to 2.3:2 or, as a percentage, (2.3 ÷ 4.3) x 100 = 53%.
- For Donna, the likelihood ratio of 2.3 increases her odds ratio from 0.4:99.6 to 0.9:99.6 or, as a percentage, (0.9 ÷ 100.5) x 100 = 0.9%.

These are the same results as those derived in Box 3, and are much easier calculations.

Nomograms

The post-test probability can be determined without any calculation by using a nomogram (the Fagan nomogram, from Bayes' theorem).[5] The pre-test probability and the positive or negative likelihood ratio can be noted on the vertical axes on the left and in the middle of the nomogram; the post-test probability is indicated by where the line joining these points passes through the right-hand vertical axis.

Predictive values and likelihood ratios in practice

Both the predictive value and likelihood ratio approaches require that you know the sensitivity and specificity of the relevant test (which may not be easily available) and the pre-test probability estimate (your clinical judgement or a decision tool based on clinical features). However, the following key points are suggested by an understanding of the principles:

- do not perform a test if the implications of positive and negative results have not been thought through
- sometimes the test being thought of should not be performed and/or a different test is needed that will give more relevant information.

For example, if the pre-test probability is low, a positive result is likely to be a false positive unless the test is very specific, and if the pre-test probability is high, a negative result is quite likely to be a false negative unless the test is very sensitive (a high true positive rate). But can the result be relied on or should the test be repeated? This can be put in a slightly different way. If the test result is positive, are both the doctor and patient prepared to proceed to the next step – e.g. an angiogram if a stress test is positive, and then stenting or bypass? If the test result is negative, are both the doctor and patient prepared to exclude the condition – e.g. a negative faecal occult blood test for a person with a strong family history of bowel cancer and a recent history of change in bowel habit? In these two cases, it may be better to not perform

the stress test in the first and to perform a colonoscopy instead of a faecal occult blood test in the second.

SUMMARY

Screening tests give information about the likelihood of a disease being present in a person without specific symptoms whereas diagnostic tests provide a diagnosis in a patient with specific symptoms.

The sensitivity and specificity of tests and the pre-test probability of the condition being tested for all need to be considered when interpreting 'positive' and 'negative' test results. In many cases of screening for a problem that has a low probability of being present, the sensitivity and specificity of the screening test are such that a positive result is more likely to be a false positive than an indicator of the presence of the disease. Conversely when diagnosing a disease in a population that has a high probability of having the disease, the sensitivity and specificity of the diagnostic tests are such that a negative result is more likely to be a false negative rather than indicating the absence of the disease. Some practice points are listed in Box 4. MT

References

1. Phillips P. Diana's diabetes. Med Today 2011; 12(3): 57-64.
2. Hippisley-Cox J, Coupland C, Robson J, Sheikh A. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. BMJ 2009; 338: b880.
3. Australian Institute of Health and Welfare 2008. Diabetes: Australian facts 2008. Diabetes series No. 8. Cat. no. CVD40. Canberra: AIHW; 2008.
4. Haynes RB, Sackett DL, Guyatt GH, Tugwell P. Clinical epidemiology – how to do clinical practice research. Philadelphia: Lippincott Williams and Wilkins; 2006.
5. Fagan TJ. Nomogram for Bayes' theorem. N Engl J Med 1975; 293: 257-261.

COMPETING INTERESTS: Dr Phillips has received research and travel grants, acted on advisory boards and been involved with clinical trials and seminars sponsored by a range of pharmaceutical companies. He does not think these associations have influenced the content of this article.