Interpreting pathology tests

Testing tests in clinical diabetes care

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

There is more to interpreting pathology test results than the listing on the pathology report indicating whether a test is positive or negative and which results are abnormal. The basic principles of interpreting pathology test results are summarised in this final article in a short series outlining a framework for such interpretation.

Reports from pathology laboratories tell us whether a test is positive or negative and which results are abnormal but there is more to interpreting results than this. For example, we may want to know if a test reported as positive is a true positive for the problem tested for or a false positive in someone without the problem, how abnormal an abnormal result is and whether a difference between consecutive results indicates a real change or background variability.

This article is the last in a short series about providing a framework for interpreting the results of pathology tests. A summary of the principles discussed in the previous articles is given in Box 1, and a case study is presented to illustrate the use of the framework in practice.[1-3]

Screening versus case finding

Case scenario

David is 58 years old and wants to be tested for diabetes because both of his regular golf buddies have recently been diagnosed with type 2 diabetes.

How likely is David to have diabetes?

Although there is not much information to go on, David has the most important 'F' word for diabetes risk – 'over Forty'.[4] Based on his age, his AUSDRISK score is 9 and he would be at intermediate risk (approximately 3%) of developing type 2 diabetes within five years.[5] In addition, his risk of having undiagnosed diabetes (his pre-test probability of having the undiagnosed disease) would be about one-fifth of his risk of developing diabetes (i.e. 0.6%). (The AUSDRISK tool is available online at www.health.gov.au/preventionoftype2diabetes.)

Case continued

David's history reveals he has both of the other 'F' words (i.e. 'Family history' and 'Fatness'): his mother developed type 2 diabetes at the age of 62 years and he is very overweight ('over waist' – waist circumference 102 cm or greater) with a BMI of 33 kg/m^2 and a waist circumference of 104 cm. Apart from his monthly round of golf, David takes little exercise. He is taking medication for hypertension and dyslipidaemia.

Now how likely is David to have diabetes?

With this additional information, David's AUSDRISK score is 20. This puts him in the top high-risk category for diabetes, in which a person has a one in three chance of developing type 2 diabetes within five years, and a risk of having undiagnosed diabetes of about 7% (one-fifth of 33%).

MedicineToday 2015; 16(10): 59-63

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

1. TESTING TESTS – AN INTERPRETIVE TOOL KIT

The laboratory reference interval[1]

The laboratory reference interval (RI) includes results from 95% of a healthy population and is usually the range lying between ± 2 standard deviations (SDs) of the healthy population mean (i.e. a total of 4 SDs). (The term 'reference interval' is now preferred to 'reference range'.) The SD of the RI is therefore the span (top minus bottom) of the RI divided by 4.

Checking how far away a test result is from the population mean in terms of SDs is a way of assessing the abnormality of the test result (Table 1). This method provides a useful approximation of the degree of abnormality of test results for many analytes (although it may not be appropriate for analytes with a skewed, i.e. non-Gaussion, distribution).

TABLE 1. LIKELIHOOD THAT TEST RESULTS WILL LIE OUTSIDE DIFFERENT RANGES

	Result range			
	Mean ± 1 SD	**Mean ± 2 SD (i.e. the RI)**	**Mean ± 3 SD**	**Mean ± 4 SD**
Probability of single test results outside range	32%	5%	0.27%	0.006%

Abbreviations: SD = standard deviation; RI = reference interval.

Checking change[2]

Results of tests repeated in a person vary considerably (intraindividual variability). Assessing whether an apparent change in the level of a particular analyte is a real 'signal' or the result of the variability of results (the background 'noise') requires information about the magnitude of the variability.

The coefficient of variation within an individual (CV_i) includes the individual's own biological variability (CV_b) and the laboratory variability for that analyte (CV_i); examples of desirable CV_is for several analytes are given in Table 2. The coefficient of variation (CV) is the ratio of the SD to the measured value, i.e. SD ÷ test result, expressed as a percentage.

If the observed difference between consecutive test results exceeds the least significant change (LSC), it is likely to be a signal of a real change rather than because of the background 'noise' of variability. To be 80% confident that an apparent change is a real change (the confidence interval usually used), the LSC is approximately twice the CV_i.

TABLE 2. DESIRABLE INTRAINDIVIDUAL COEFFICIENTS OF VARIATION (CV_i) FOR COMMON ANALYTES IN DIABETES CARE*

Analyte	CV_i (%)
Glucose	6.4
HbA_{1c}	3.8
HDL cholesterol	8.0
LDL cholesterol	9.3
Total cholesterol	6.0
Triglycerides	6.0
Creatinine (plasma)	5.9
Albuminuria (first morning)	40.2

Abbreviations: CV_i = intraindividual coefficient of variation; HbA_{1c} = glycosylated haemoglobin; HDL = high density cholesterol; LDL = low density lipoprotein.

* Derived from values given for within subject biological variability and laboratory variability (desirable specification for imprecision) in the database 'Desirable specifications for total error, imprecision, and bias, derived from intra- and inter-individual biologic variation' (www.westgard.com/biodatabase1.htm).[2]

Case continued

You arrange for David to have his fasting plasma glucose (FPG) measured. The result is 7.4 mmol/L (the diagnostic threshold for diabetes is 7.0 mmol/L or higher).

How likely is the result of a repeat FPG test to exceed the diagnostic threshold for diabetes?

From Box 1 ('Checking change'), assuming that 7.4 mmol/L (the original test result) is the mean value for David's FPG and using the intraindividual coefficient of variation (CV_i) for glucose of 6.4% (Table 2 in Box 1), the standard deviation (SD) of his FPG would be 0.47 mmol/L. (This is calculated by multiplying the mean value by the CV_i of the analyte, i.e. 7.4 mmol/L x 6.4%; Box 1, 'Checking change'). From Table 1 in Box 1, the probability that a single test result will lie outside ± 1 SD of the mean is 32%, which indicates that a repeat FPG measurement has at least a 16% chance of being 1 SD below 7.4 mmol/L – i.e. below 7.0 mmol/L (and an 84% chance of being above it).

1. TESTING TESTS – AN INTERPRETIVE TOOL KIT continued

Sensitivity, specificity and predictive values[3]

If a test is performed on a person with a problem, it can have a true-positive (TP) or a false-negative (FN) result. The sensitivity of the test is the proportion of tests that are positive in those with the problem, i.e. the TP rate, which is the number of TPs divided by the total number of people with the problem (i.e. the TPs plus the FNs).

Sensitivity = TP ÷ (TP + FN)

Similarly, in a person without the problem, the test result can be a true negative (TN) or a false positive (FP). The specificity of a test is the proportion of people without the problem who have a negative test, i.e. the TN rate, which is the number of TNs divided by the number of people without the problem (i.e. the TNs plus the FPs).

Specificity = TN ÷ (TN + FP)

The positive and negative predictive values (PPV and NPV, respectively) describe the performance of a diagnostic test. The PPV is the likelihood that a positive result for the test indicates the condition is present (the proportion of positive results that are TPs). The NPV is the likelihood that a negative result indicates the condition is not present (the proportion of negative results that are TNs).

PPV = TP ÷ all positive tests (i.e. TP + FP)
NPV = TN ÷ all negative tests (i.e. TN + FN)

For example,

If the sensitivity (TP rate) of a test result for diabetes was 90% (0.9 as a proportion) and its specificity (TN rate) was 80% (0.8 as a proportion), 90% of those with diabetes would have a positive test and 80% of those without diabetes would have a negative test (Table 3).

The PPV of this test for diabetes in a population of 100 people with a pre-test probability of having diabetes of 4% – i.e. 4% (4) will have diabetes and 96% (96) will not – would be:

The number of TPs (i.e. 4 x the TP rate of 0.9) divided by the total number of positive tests – i.e. the TPs plus the FPs [96 x the FP rate of 0.2], the FP rate being (1 - TN rate)

PPV = (4 x 0.9) ÷ [(4 x 0.9) + (96 x 0.2)]
= 3.6 ÷ 22.8
= 16%

And the NPV would be:

The number of TNs (i.e. 96 x the TN rate of 0.8) divided by the total number of negative tests – (i.e. the TNs plus the FNs [4 x the FN rate of 0.1], the FN rate being (1 - TP rate)

NPV = (96 x 0.8) ÷ [(96 x 0.8) + (4 x 0.1)]
= 76.8 ÷ 77.2
= 99%

TABLE 3. RESULTS FOR A DIABETES TEST WITH 90% SENSITIVITY AND 80% SPECIFICITY

Patients	Positive test result	Negative test result
With problem	90% True positives	10% False negatives
Without problem	20% False positives	80% True negatives

Another way to explain this is that David's apparently positive test would be confirmed with a repeat test 84% of the time (the true-positive rate of the FPG test when performed on David) and not confirmed 16% of the time (the false-positive rate).

We don't actually know that David's mean FPG is 7.4 mmol/L so how likely is a repeat FPG test to be diagnostic of diabetes?

On the basis of his age alone David has a 0.6% of having undiagnosed diabetes; his other risk factors increase this to 7%. As shown in Box 1, 'Sensitivity, specificity and predictive values', combining these pre-test probabilities of diabetes (PD) and the true- and false-positive rates for the FPG test when performed give positive predictive values (PPV) of:

- For a pre-test probability of 0.6%,
  PPV = (0.6% x 0.84) ÷ [(0.6% x 0.84) + (99.4% x 0.16)]
  = 3%
- For a pre-test probability of 7%,
  PPV = (7% x 0.84) ÷ [(7% x 0.84) + (93% x 0.16)]
  = 28%

In this case, initially – when only David's age and gender were known, and these conferred a relatively low pre-test risk of him having diabetes – you were screening, knowing that false-positive test results are more likely than true-positive ones. Later, when the patient was known to have more risk factors for diabetes and therefore the pre-test risk of him having diabetes was high, you were case-finding, knowing that true-positive results are more likely than false-positive results.

The diagnostic criteria for type 2 diabetes are two abnormal blood glucose results or one abnormal laboratory test and

glycaemic or glycosuric symptoms of diabetes.[6] As it happens, Mark had nocturia because of glycosuric polyuria and satisfied these requirements for diagnosis after the first test.

How abnormal is abnormal?

Case continued

In reviewing David's biochemical profile from the laboratory you note an asterisked value of alanine aminotransferase (ALT) of 55 U/L (reference interval [RI], 0 to 40 U/L).

You wonder if this suggests David might have a fatty liver.

How abnormal is this test?

The SD of the RI for ALT (0 to 40 U/L) is 10 U/L and David's result lies 35 U/L above the midpoint of the RI – i.e. between 3 and 4 SDs higher. (As indicated in Box 1, 'The laboratory reference interval', the RI is 4 SDs, so for ALT results 1 SD is 25% of 40.) A result this far from the healthy population mean is highly unlikely to be an outlier of the healthy population (0.006 to 0.27% conservatively – Table 1 in Box 1). David's ALT result therefore strongly suggests some liver dysfunction. Fatty liver associated with David's diabetes and features of the metabolic syndrome is a likely diagnosis.[7]

At this stage, however, further investigation is not indicated because the abnormality may decrease or resolve as his lifestyle and glycaemic control improve with the initiation of type 2 diabetes management.

Checking change

Case continued

At the time of diagnosis of type 2 diabetes, David's glycosylated haemoglobin (HbA_{1c}) level was 7.2% (55 mmol/mol). Three months later, after attending a series of group education sessions and trying to make healthy changes to his lifestyle, David's HbA_{1c} is 6.7% (50 mmol/mol).

How likely is the change to be a signal of real change (rather than because of the 'noise' of variability)?

Whether the decrease in David's HbA_{1c} level is a signal of real change or because of the 'noise' of variability requires a calculation of the least significant change (LSC) to see if this is less than the observed change (Box 1, 'Checking change'). David's HbA_{1c} level (in %) decreased from 7.2 to 6.7 – i.e. by 0.5, which when expressed as a percentage of the initial value of 7.2 is 6.9%. At a confidence of 80%, the LSC is approximately twice the CV_i, which for HbA_{1c} is 3.8% (Table 2 in Box 1):

LSC = 2 x 3.8% = 7.6%.

The observed change of 6.9% in David's HbA_{1c} is less than the LSC (7.6%) and is therefore quite likely to reflect background variability rather than a real change in overall glycaemic control.

Nonetheless, the change is in the right direction, possibly indicating David has improved his lifestyle. Given the progressive nature of type 2 diabetes, another measurement in the next six to 12 months would be useful to review the progression of David's diabetes.

Microalbuminuria – here today, gone tomorrow

Case continued

As part of the annual cycle of care under the diabetes incentive of the Australian Federal Government's Practice Incentives Program you tested for microalbuminuria by checking David's first-voided urine albumin to creatinine ratio (ACR). The result seems positive at 5.3 mg/mmol (microalbuminuria, males, ACR level 2.5 to 25 mg/mmol).

Should David start taking a blocker of the renin–angiotensin system?

It is recommended that if a first test for microalbuminuria is positive then a second test be performed to confirm the diagnosis, and if this is negative, a third test be performed. The results of this third test will decide whether microalbuminuria is present. The reason for this recommendation is that the variability (the background 'noise' – i.e. the biological and laboratory variabilities) of microalbuminuria is very high (CV_i for microalbuminuria is 40.2% [Table 2 in Box 1] and the LSC (approximately 2 CV_i) is therefore very high (80%).

If the mean value for many microalbuminuria tests in an individual was an ACR of 5.3 mg/mmol, the LSC would be 80% of 5.3 – i.e. 4.2 mg/mmol. A range of ACR values from 1.1 to 9.5 mg/mmol (5.3 ± 4.2 mg/mmol) could be attributed to 'noise' and not be a real 'signal' of biological difference. It would therefore be no surprise if the second 'confirmatory' test were negative. In general, urine tests of any sort have high CV_is, indicating that repeat testing may give different results.

The microalbuminuria that is here today could indeed be gone tomorrow.

Statistical versus clinical significance

Case continued

David's repeated microalbuminuria test gave a positive result of an ACR level of 7.2 mg/mmol, confirming the diagnosis. As recommended by the RACGP,[6] you started David on the renin–angiotensin system blocker irbesartan 150 mg/day, with the intention of increasing the dose to 300 mg/day if there are no significant changes in David's potassium or creatinine levels.

Two weeks later, David's potassium level was unchanged but his creatinine level had increased from 72 µmol/L to 83 µmol/L.

Should you stop the irbesartan?

Two questions should be asked when considering whether to stop treating David with irbesartan:

- is the change resulting from the treatment statistically

2. PRACTICE POINTS: INTERPRETING PATHOLOGY TEST RESULTS

- If a test for a condition is performed on a person with the condition, the result can be a true positive (TP) or a false negative (FN). The sensitivity of a test is the proportion of test results that are positive for the problem (the TPs) divided by the total number of people with the problem (the TPs plus the FNs), i.e. TP ÷ (TP + FN).
- Similarly, in a person free of the problem the test result can be a true negative (TN) or false positive (FP). The specificity of a test is the proportion of people without the problem who have a negative test (the TNs) divided by the number of people without the problem (the TNs and FPs), i.e. TN ÷ (TN + FP).
- The positive predictive value of a test is the proportion of all positive tests occurring in those with the problem, i.e. TP ÷ (TP + FP).
- Similarly, the negative predictive value of a test is the proportion of all the negative tests that occur in those without the problem, i.e. TN ÷ (TN + FN).
- The reference interval (RI) usually covers the span ± 2 standard deviations (SDs) around the healthy population mean (i.e. 4 SDs). The distance from the population mean in terms of SDs gives an estimate of how likely a result is to be a statistical outlier of the healthy population. Results lying 3 or 4 SDs from the population mean are unlikely to be outliers from the healthy population (likelihood 0.27% and 0.006%, respectively).
- To check the statistical significance of an apparent change in a particular analyte over time requires knowledge of the coefficient of variation within individuals (CV_i) for the analyte. The least significant change (LSC) is approximately twice the CV_i (for 80% confidence) and a greater change is likely to be a real 'signal' rather than a result of the background 'noise' of variability of the analyte.

  However, a statistically significant change (a change exceeding the LSC) may not be clinically significant (i.e. prompting therapeutic intervention recommended by best practice clinical guidelines).

significant (greater than the LSC) – i.e. is it a real signal of change?

- is the change clinically significant – i.e. a change greater than expected because of the therapeutic effect of blockers of the renin angiotensin system (decreasing glomerular pressure and reducing albumin filtration but also decreasing glomerular filtration and glomerular filtration rate)? A change of 20% or greater is likely to indicate an important clinical problem such as renal artery stenosis, where use of blockers of the renin–angiotensin system could greatly decrease the glomerular filtration rate.[8]

Given that the CV_i of P-creatinine is 5.9% (Table 2 in Box 1), the LSC is 11.8%. The change in David's creatinine level from 72 µmol/L to 83 µmol/L (an increase of 15%) exceeds this and is likely to be a real signal of change. But the change of 15% is not clinically significant (which would need to be a change of 20% or greater) and is well within the range expected because of the therapeutic effect of irbesartan.

Statistically significant change is not always clinically significant: the increase in David's creatinine level was statistically but not clinically significant and did not justify stopping the irbesartan.

Conclusion

There is more to interpreting results of pathology tests than the listing on the pathology report of whether a test is positive or negative and which results are abnormal. Clinical presentation should, of course, be taken into account, as should the intrinsic variability of measurement between and within individuals and from the collection and analysis of laboratory samples. Interpretation of diagnostic tests requires consideration of the test's sensitivity and specificity and the pre-test probability of the condition so the predictive power of positive and negative test results can be calculated. The interpretation of changes in laboratory measurements requires consideration of the statistical significance of the changes (i.e. the LSC) and their clinical significance (as outlined in best practice guidelines).

Practice points are listed in Box 2. MT

References

1. Phillips P. Interpreting pathology tests: results and reference intervals. Med Today 2014; 15(11): 64-68.
2. Phillips P. Interpreting pathology tests: monitoring measurements. Med Today 2015; 16(3): 70-74.
3. Phillips P. Interpreting pathology tests: screening versus diagnostic testing. Med Today 2015; 16(9): 67-71.
4. Phillips P. Diana's diabetes. Med Today 2011; 12(3): 57-64.
5. Chen L, Magliano DJ, Balkau B, et al. AUSDRISK: an Australian type 2 diabetes risk assessment tool based on demographic, lifestyle and simple anthropometric measures. Med J Aust 2010; 192: 197-202.
6. Royal Australian College of General Practitioners (RACGP), Diabetes Australia. General practice management of type 2 diabetes – 2014-15. Melbourne: RACGP and Diabetes Australia; 2014. Available online at: www.racgp.org.au/your-practice/guidelines/diabetes (accessed September 2015).
7. Harris MF. The metabolic syndrome. Aust Fam Physician 2013; 42: 524-527.
8. Mount P, Subramanian KK, MacGinley R; Mantha M, Roberts M, Mangos G; CARI. The CARI guidelines. Inhibition of the renin–angiotensin system. Nephrology 2010; 15: S234-S243.

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.