



Abnormal laboratory results

Pitfalls in interpreting laboratory results

Pat Phillips, Senior Director, Endocrinology Unit, The Queen Elizabeth Hospital, Adelaide

Summary

The results of laboratory tests are affected by the collection and handling of the specimen, the particular laboratory and the method of analysis. They are also affected by variability within the individual and within the laboratory. Interpretation at one point in time should consider the position of the measurement within the laboratory reference range appropriate for the sample and the person being tested. Interpreting results over time should consider the likely variability of the measurement and the level of certainty required to identify a true change or absence of change. The more variable the measurement and the higher the required level of certainty, the larger the change between measurements needs to be before it can be considered clinically significant.

Key words: diagnostic tests.

(*Aust Prescr* 2009;32:43–6)

Introduction

Health professionals may find it hard to get clinically useful information from the barrage of figures, ranges, stars and comments in laboratory results. Some knowledge about the accuracy of laboratory results can help to sort out important clinical signals from the background 'noise'. The laboratory does not know all the patient's details. Clinicians should consider test results in the context of the clinical presentation and not rely completely on the laboratory's interpretation.

Reference ranges

Quoted reference ranges depend on the method used in the laboratory, and the population from which the reference range was derived. The results from one method may be systematically different from those of another and therefore the reference ranges will be different.

Some laboratories give the range quoted by the manufacturer of the test or derived from an easily accessible population such as blood donors. Others give ranges in terms of age, sex or biological phase. For example, the ranges quoted for female

sex hormones are related to pre- and post-menopausal status and the phase of menstrual cycle. Some important biological influences, such as seasonal effects on 25-hydroxyvitamin D, are often not included in the reference ranges. Perhaps this is because users would find it harder to interpret results if the reference ranges were changing all the time and because of the logistics and laboratory workload needed to derive such specific reference ranges.

The ideal reference range would relate to the individual being tested while healthy, at the same age, biological phase and in the same season. Clearly this is not possible, but sometimes one gets insights from looking back through previous results (ideally reported by the same laboratory using the same method).

By tradition, laboratories quote a reference range including 95% of the reference population. If results are normally distributed, this includes results within approximately two standard deviations above and two standard deviations below the mean value. The reference range therefore covers four standard deviations. Some results vary so much within the population that the laboratory may quote a reference range that includes a smaller proportion of the population. For example, the reference range commonly quoted for serum insulin may only include results within one standard deviation above and one standard deviation below the mean value. This includes 68% of the reference population. In this case, 16% of normal people will have 'abnormal' high insulin and 16% will have 'abnormal' low insulin according to the quoted reference range. Serum insulin is therefore not a useful test for assessing 'insulin resistance'.

Results have to be interpreted in terms of the particular laboratory reference range. When monitoring results over time, clinicians also need to be aware that different laboratories will have different reference ranges.

As reference ranges are population-based, a patient might have a result near the top or bottom of the normal range. Clinically significant changes could then occur, without the results moving out of the population reference range. For example, if an elderly patient's plasma creatinine concentration is usually near the bottom of the reference range but then rises to the upper end of that range, the patient may have had a significant deterioration in renal function. Similar considerations apply to a haemoglobin concentration falling from a high normal to a low normal value.

Table 1

Abnormal laboratory results caused by incorrect collection and handling †

Step	Mechanism	Result	Measurement affected
Sample	Incorrect sample	Incorrect results	For example, random spot urine calcium:creatinine ratio instead of first voided
Venepuncture	Prolonged venostasis Difficult venepuncture	Plasma filtration and concentration Haemolysis	Protein concentrations – globulins, albumins and lipoproteins and measurements affected by them (e.g. calcium) Red cell leakage with high potassium, phosphate and lactate dehydrogenase
Specimen tube	Incorrect collection tube	Incorrect results Assay added analyte Assay interference	Lithium heparin anticoagulant – lithium assay If potassium EDTA used for chemistry – potassium assay If potassium EDTA used for chemistry – assays for calcium and enzymes (calcium binding and enzyme inhibition)
Specimen handling	Delay in transport	Red cell use of glucose and leakage of contents	Blood glucose (if fluoride tube not used). Potassium, phosphate, lactate dehydrogenase
Laboratory	Specimen mislabelling Machine malfunction Transcription error	Incorrect results	Virtually everything

† Derived from reference 1

Specimen collection and handling

Laboratory results can be affected by the procedures for specimen collection and handling (Table 1).¹ If a result is a surprise, check the patient's name and date of birth on the result report. You can also contact the laboratory and ask if the specimen looked normal and consider repeating the test.

Why normal people often have abnormal results

A multiple biochemical analysis can be performed by one machine and produce 20 results. Assuming these results were all independent of each other (which they are not) and that results from the reference population are normally distributed (which they may not be), only 36% of normal people will have all 20 results in the reference range. There will be 64% with at least one abnormal result (Box 1). However, the more abnormal the result and the more related tests are abnormal, the more likely the abnormality is clinically significant.

If you consider the 99% reference range (approx. ± 2.6 standard deviations) and the 99.9% reference range (approx. ± 3.3 standard deviations), 82% and 98% of people will have all 20 tests within the reference range (0.99²⁰ and 0.999²⁰ respectively). These facts can be useful when interpreting an isolated abnormal result.

For example, the reference range of alkaline phosphatase is 30–110 U/L. This covers two standard deviations below the mean and two above the mean. One standard deviation is therefore 20 U/L [(110–30) ÷ 4]. A result of 150 U/L is two standard deviations above the upper limit of the reference range and therefore four standard deviations above the mean. This is very unlikely to occur in a normal individual. However, the result may be normal

Box 1

Normal results in normal people

If the reference range covers 95% of results for a normal population, the chance of a healthy individual having a certain number of normal tests is:

- Two out of two tests 90% (0.95 x 0.95 = 0.90)
- All 20 of 20 tests 36% (0.95²⁰)

if the quoted reference range is inappropriate. For example, in pregnancy and growing children alkaline phosphatase is produced by the placenta and bone. These are good examples of why it is important to consider whether the population reference range is appropriate for the individual being tested.

When deciding if a result is abnormal, look at related tests. Alkaline phosphatase is one of the 'liver function tests' (others are bilirubin, gamma glutamyl transferase, alanine aminotransferase, aspartate aminotransferase and lactate dehydrogenase). Abnormalities in the other tests would suggest that the abnormal alkaline phosphatase could be the result of liver disease. An elevated alkaline phosphatase in isolation may indicate another problem, such as bone disease.

Laboratory accuracy

We often know the within-laboratory, within-method variability as this is usually quoted by the laboratory. Modern laboratories provide remarkably consistent results for many analytes – typical coefficients of variation (see Box 2) are 1–6% for the components of multiple biochemical analysis, electrolytes, calcium and phosphorus, and renal and liver function tests.

Box 2

Coefficient of variation

The coefficient of variation (CV) is calculated as:

$$CV = \frac{\text{standard deviation of the measured value} \times 100}{\text{mean value}}$$

Variability is different at different absolute values of the measurement and is usually quoted at a specific clinically relevant value. For example:

CV for plasma sodium	0.8% at 139 mmol/L
CV for plasma bilirubin	6.1% at 10 micromol/L

The coefficient of variation is one way of expressing the variability of biological measurements. Laboratories sometimes also refer to the imprecision of a measurement.

National quality control programs monitor the accuracy and imprecision of different methods used in different laboratories. One result has been that the differences between laboratories for individual methods are now usually a small component of the overall variability of measurements.

Why values vary within one individual

In addition to the variations caused by specimen collection and handling and the differences within and between laboratories and their methods, there is intra-individual variation. Assuming specimen collection and processing errors do not occur, the largest source of variability is within the individual. Values vary by age, sex and within the menstrual, diurnal and seasonal cycles. Intra-individual biological variability for different analytes can range from very large to moderate, for example, 8% for total cholesterol² versus 40% for microalbuminuria³ assessed by the albumin:creatinine ratio. In addition, the longer the interval between tests, the greater the total intra-individual variability of the measure.

It is much more difficult for laboratories to provide information on the total intra-individual variability than for the within-laboratory, within-method variability which is automatically generated by their quality control programs. However, it is the total variability within an individual which is important when interpreting results.

Are changes in results caused by intra-individual variability or the effects of treatment?

One trap is the phenomenon of 'regression to the mean'.⁴ Results within an apparently homogeneous group of patients are likely to lie within the 95% reference range for that measurement. If the same patients are retested at a different time, the pattern of the overall results will look much the same. In a normal distribution, values are bunched around the group mean and progressively 'thin out' further from the mean. However, individual results are likely to have changed, particularly those at the extremes.

Box 3

Least significant change

1. The overall variability of the difference between two measurements is greater than the variability of the individual measurements: $\sqrt{2} CV_i$ *
2. The more confident one wishes to be that the change in a measurement is a signal rather than noise, the greater the change needs to be relative to this: $\sqrt{2} CV_i \times z$

The z value is used to refer to normally distributed values and describes the distance of a particular value from the mean in numbers of standard deviations (SD). The greater the distance from the mean (the z value) the less likely a result has occurred by chance.

z varies from 1.28 for 80% confidence to 2.6 for 99% confidence.

3. Generally 80% confidence is used (z = 1.28):

$$\text{Least significant change} = \sqrt{2} CV_i \times 1.28 = 1.8 CV_i$$

This approximates to $2CV_i$

CV_i Intra-individual coefficient of variation

* For explanation of the variability of the difference between two measurements, see extended Box 3 in this article online at www.australianprescriber.com/magazine/32/2/43/6.

The initial results at the extremes are the result of extreme random variability in one direction or the other. The same amount and direction of variability is unlikely to occur on the second measurement in the same individual. Subsequent measurements will therefore move closer to the middle (or 'regress to the mean'). Results from other individuals who initially were closer to the mean may now lie closer to the extremes of the distribution.

This phenomenon can be exploited intentionally or unintentionally in trials that select and treat individuals with high values of a measurement to demonstrate that a treatment is effective. 'Regression to the mean' is one reason why randomised placebo-controlled prospective trials are the gold standard for assessing treatments.

A large difference between two measurements is more likely to be a signal of a true change than the result of the background noise of measurement variability. Similarly, the smaller the total intra-individual variability, the more likely a specific absolute change is a signal. The less likely the observed change is caused by variability, the surer one can be that the change is real.

These three elements are brought together in the concept of the least significant change. To be 80% confident the observed change is real, the change should exceed approximately twice the intra-individual coefficient of variation (CV_i) (Box 3). For example:

- A total cholesterol which decreases from 7.0 to 5.6 mmol/L, after starting a statin, is a 20% fall from the initial value. The CV_i for total cholesterol is 8% so the least significant change

is approximately 16% ($2CV_i$). You can be 80% sure that the 20% change is real rather than apparent.

- A decrease in microalbuminuria from an albumin:creatinine ratio of 5.0 to 2.0 mg/mmol, after starting an ACE inhibitor, is a 60% fall. The total CV_i of the albumin:creatinine ratio is 40% so the least significant change is approximately 80% ($2CV_i$). It is likely that this 60% change is apparent rather than real.

The effects of treatment on measurements may be delayed

Laboratory results may take a long time to change after starting treatment. This may reflect pharmacokinetics, biology or a combination of the two.

The half-life of thyroxine in the body is approximately seven days. Testing after one week will only show half the expected total effect. (This may sometimes still be useful information.) By six weeks (six half-lives in this case) 98.4% of the effect will have occurred [$1 - (1/2)^6$].

When starting a thiazolidinedione (glitazone) the full effect on blood glucose requires a steady state of the glitazone (pharmacokinetic) but also requires the shift in fat metabolism which in turn causes the reduction in glucose (biologic). Finally, the glycated haemoglobin (HbA1c) reflects the average blood glucose over the preceding 4–6 weeks because of the slow turnover of the red cells (biologic and pharmacokinetic).⁵ The combination of these factors means that testing after one week of treatment may show little change in the HbA1c which may take 2–3 months to show the full effect of treatment.

Another glycated protein (albumin, which becomes fructosamine) has a much faster turnover. It therefore reflects the average glucose over a shorter period (2–3 weeks).

One can reduce the variability of the measurement change by reducing the variability of the baseline and final measurements (for example, the mean of two measurements for each). If both initial and final measurements were repeated the variability of the change would be reduced to CV_i (not $\sqrt{2} CV_i$).

Using the microalbuminuria example, with two measurements before and after the intervention, the least significant change would be 51% ($1.28 \times 40\%$). You could then be 80% sure that the 60% observed change was real and not apparent.

Recommendations

When interpreting laboratory results it is important to know that the sample was collected and handled correctly. The appropriate reference range for the test should be used. Different laboratories may report different results on the same specimen.

When comparing results over time, use the same laboratory and method for testing. Consider the variability of results within the individual and the least significant change. This is the amount of difference between measurements that is likely to be a real biological 'signal' instead of resulting from the noise

of biological variability within the individual and within the end measurement variability within the laboratory. As a rough rule, the least significant change is twice the intra-individual coefficient of variation ($2CV_i$).

If an important clinical decision depends on whether a change occurs with a particular treatment, consider making two (or more) measurements before and after starting treatment. This reduces the variability and the possibility of misinterpreting the regression to the mean of an initial high or low value. Monitoring trends with time involves more measurements and gives a more reliable indication of change than a single comparison at two points.

Remember, the more tests you do the more likely you are to get at least one 'false positive' outside the laboratory reference range. Aim to limit the number of tests to those that are relevant to the clinical situation rather than requesting a screening battery.

When assessing the effects of treatment, consider how long the treatment will take before the therapeutic effect reaches a steady state (e.g. 4–6 half-lives of a drug) and how long the biological response will take before the measurement you make reaches a steady state. Trying to assess therapeutic effects before treatment and response have reached a steady state can seriously underestimate the therapeutic effect.

References

1. Phillips P, Beng C. Electrolytes – 'fun with fluids'. Check (Continuous Home Evaluation of Clinical Knowledge) program of self assessment. No. 323. South Melbourne: Royal Australian College of General Practitioners; 1999.
2. Cooper GR, Myers GL, Smith SJ, Schlant RC. Blood lipid measurements. Variations and practical utility. *JAMA* 1992;267:1652-60.
3. Phillipou G, Phillips PJ. Variability of urinary albumin excretion in patients with microalbuminuria. *Diabetes Care* 1994;17:425-7.
4. Irwig L, Glasziou P, Wilson A, Macaskill P. Estimating an individual's true cholesterol level and response to intervention. *JAMA* 1991;266:1678-85.
5. Phillipov G, Phillips PJ. Components of total measurement error for haemoglobin A(1c) determination. *Clin Chem* 2001;47:1851-3.

Conflict of interest: none relevant to this article

Self-test questions

The following statements are either true or false (answers on page 55)

7. A laboratory result which is two standard deviations from the population mean is always abnormal.
8. If treatment reduces a patient's total cholesterol by 5% the change is significant.