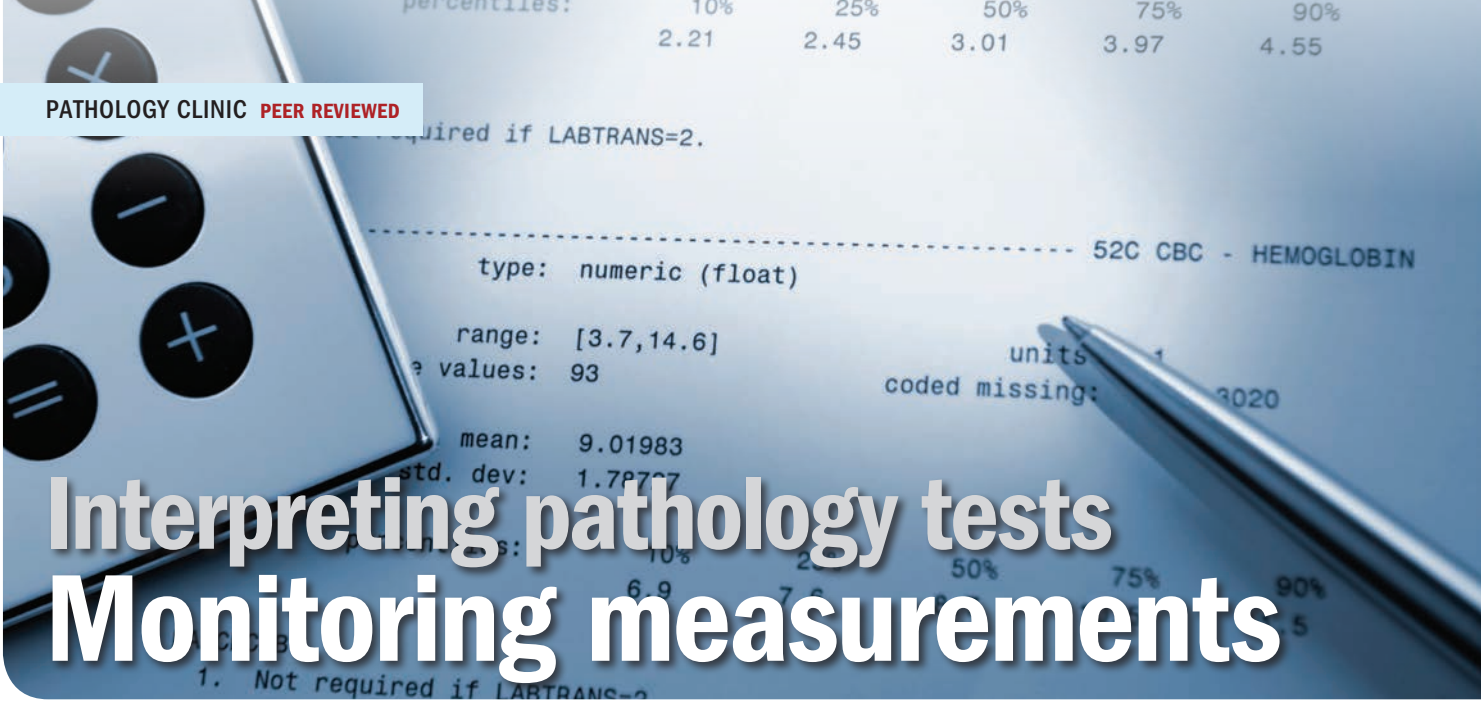

Interpreting pathology tests
Monitoring measurements

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

Variability and determining the significance of changes in measurements over time are discussed in this second article in a series of four aiming to provide a framework for interpreting test results.

Random variability in pathology results needs to be minimised so that real change can be identified when results of sequential tests differ. As well as a difference being statistically significant it also needs to be clinically significant to prompt medical intervention.

This second article in a series of four outlining a framework for interpreting laboratory results focuses on variability. It discusses the source of variability between sequential laboratory measurements of the same variable within the same individual and discusses ways to minimise the background 'noise' of random variability, maximise the 'signal' of any biological change and interpret the significance of changes in measurements over time.

MedicineToday 2015; 16(3): 70-74

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

Measuring variability

Variability in the results of laboratory tests is measured as standard deviations (SD) in absolute terms and as coefficients of variation (CV) in relative terms. The CV is the ratio of the SD to the measured value, i.e. SD ÷ test result, expressed as a percentage.[1]

Sources of variability

As noted in the first article in the series, reference intervals (RI) for laboratory test results include several sources of variability. These are:[2]

- biological variability between individuals
- biological variability within each individual
- variability introduced by specimen collection and transport
- variability introduced within the laboratory.

Efficient specimen collection and transport to the laboratory aims to eliminate the third component in the list above but variability between and within individuals and variability within the laboratory remain. Laboratory variability is generally low (e.g. 1 to 6% for common biochemical measurements) and much of the variability within the RI comes from between and within individuals.

The individual variability (intraindividual variability, CV_i) includes the individual's own biological variability (CV_b) and the laboratory variability (CV_l), as noted in Box 1. CV values for common analytes are published in the Desirable Biological Variation Database at http://www.westgard.com/biodatabase1.htm.[3]

When assessing sequential measurements within one individual, it is the variability within the individual and within the laboratory that matter. The random variability within an individual, in sample collection and in measurement is the background 'noise', and must be minimised as much as possible so that any 'signal' of real change can be distinguished.

Key points

- The variability of laboratory results is expressed in terms of SD or CV (SD ÷ test result, as %).
- Variability in results from an individual occurs because of the biological variability within that individual and the variability introduced before and after the specimen arrives in the laboratory (i.e. collection, storage, transport and laboratory analysis).

Minimising the background 'noise' of random variability

Background result variability (background noise) is reduced by minimising the variability introduced before the specimen reaches the laboratory and that contributed by the laboratory.

Variability before the laboratory

The variability contributed by specimen collection and transport to the laboratory can be reduced by:

- collecting blood specimens at the appropriate time (usually at 8.00 to 10.00 a.m. after fasting overnight for 10 hours) – this minimises any influence from diurnal rhythms and eating (e.g. triglyceride and testosterone concentration)
- collecting the blood with the patient seated and removing the tourniquet before withdrawing the blood sample – this minimises any influence on the specimen collection from venous stasis (e.g. calcium and protein concentrations)
- for urine collections, making sure the patient understands the arrangements for the timing of collection – this minimises errors such as mistiming the duration of collection (e.g. collecting a 24-hour specimen for more or less than 24 hours) or in the timing of a spot specimen (e.g. first voided morning urine for albumin to creatinine ratio rather than a random specimen)
- ideally using a nearby laboratory that has frequent collection services – this minimises any effects of delay in specimen processing (e.g. on plasma potassium level from erythrocyte leakage; plasma glucose level from utilisation by blood cells).

Variability within the laboratory

The variability contributed by the laboratory can be reduced by:

- using the same laboratory for the initial and subsequent tests – this removes the variability between laboratories and methods. Different laboratories and/or methods may give different results when testing the same specimen; using the same laboratory is very important for tests such as tumour markers
- if the result is unexpectedly abnormal, repeating the test while ensuring that all the above is considered – mistakes do happen, even in the most efficient laboratory, and the results from a specimen from patient A may be attributed to patient B.

1. COEFFICIENTS OF VARIATION AND CALCULATING INTRAINDIVIDUAL VALUES

Coefficient of variation (CV)

- The CV is the ratio of the SD to the measured value, i.e. SD ÷ test result, expressed as a percentage
- The author sometimes finds the following very rough rule useful for calculating the SD for a test result in individuals:*
  - for an individual, the SD = RI ÷ 10

Calculating the intraindividual CV

The variability of the results from an individual (the intraindividual CV, or CV_i) includes the biological variability within that individual (CV_b) and the laboratory variability (CV_l).

The sum of two components of variability is given by the square root of the sum of their squares, i.e.

$CV_i = \sqrt{(CV_b^2 + CV_l^2)}$

The values for biological variability and laboratory variability for different tests are available online – the Desirable Biological Variation Database at http://www.westgard.com/biodatabase1.htm (in this database CV_w is CV_i and I (%) is CV_l).[3]

For example:
For plasma creatinine (P-creat)

$CV_b = 5.3\%$
$CV_l = 2.7\%$
Therefore, $CV_i = \sqrt{(5.3^2 + 2.7^2)} = \sqrt{(28.1 + 7.3)} = 5.9\%$

* This rule has been derived empirically by the author by comparing published values of the cv of common analytes with their respective reference intervals.

Total variability

If a test result suggests therapy is needed, the test should be repeated before intervening because of variability within the individual. The overall CV_i of repeated tests in the same individual is the CV_i divided by the square root of the number of times the test is performed, i.e. $CV_i \div \sqrt{}$ number of tests, such that for two tests, the CV_i of the average of the two measurements is $CV_i \div \sqrt{2} = CV_i \div 1.4 = 0.7$.

- For example:
  The intraindividual variability CV_i for HbA_{1c} is 3.8%, and includes the individual's biological variability (CV_b 3.4%) and the laboratory variability (1.7%).[3]
  - if the test were repeated once, the CV_i of the average of the two measurements would be $CV_i \div \sqrt{2} = 3.8\% \div 1.4 = 2.7\%$, rather than 3.8%
  - for the average of three measurements in one individual, the CV_i would be $CV_i \div \sqrt{3} = CV_i \div 1.73 = 2.2\%$, rather than 3.8%.

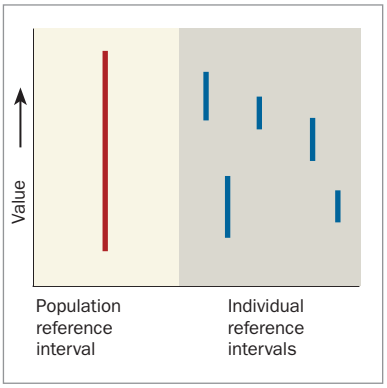

Figure 1. Population reference interval versus individual normal range for laboratory tests. Physiological ranges for individuals lie within the population reference interval but the mean values and the physiological ranges around the mean for individuals may be quite different from the population mean and reference interval and from each other.

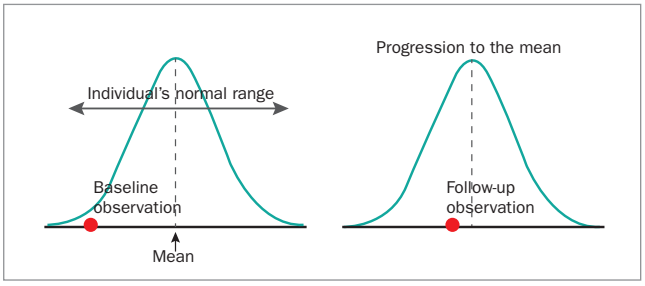

Figure 2. Regression to the mean. Initially high and low results usually move (regress) towards the mean when remeasured.

As discussed later, this repeating of a test sets a baseline and thus increases the ability to tell if an apparent change over time is a real 'signal' of change rather than reflecting the background 'noise' of the variability of the test results.

Key points

- To minimise the background 'noise' of variability introduced by collection and analysis, specimens should be collected correctly and transported promptly to the laboratory, and the same laboratory should be used for the initial and subsequent tests.
- If a test result is unexpectedly abnormal or if it would prompt medical intervention, it should be repeated to identify errors and to reduce the variability of what will become the baseline for future monitoring.

Regression to the mean

Case scenario

John is 64 years old and has had type 2 diabetes for eight years. Three months ago he came to see you as a new patient for a 'check-up' because his father died of a heart attack at age 57 years. You checked his cardiovascular risk factor profile and noted a triglyceride of 3.0 mmol/L (target <2.0 mmol/L).[4] After some discussion about cardiovascular risk associated with a reduced triglyceride level, John agreed to take fenofibrate 48 mg/day.

At a consultation with John about another problem a week ago, you took the opportunity to check his triglyceride level. The results show it is now 2.0 mmol/L and he has now come to discuss this with you. You comment on the effectiveness of the fenofibrate. John looks embarrassed and says that he stopped taking fenofibrate a week after starting when one of his friends told him that it could cause muscle problems.

The case is an example of 'regression to the mean'.[2] In the population, results at the extremes of the distribution are likely to reflect the combination of individuals' mean results being higher or lower than the population mean and the fact that these particular results lie at the extremes of the individuals' normal ranges (Figure 1).

Results of tests repeated in the same individual at different times lie within the person's normal range for that measurement, bunched around the mean value and progressively thinning further from the mean (i.e. the normal distribution). Initial results at the edges of the distribution result from 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 (Figure 2).

This phenomenon can be exploited intentionally or unintentionally in trials that select and treat individuals with high values of a measurement to check that a treatment is effective. 'Regression to the mean' is one reason why randomised placebo-controlled prospective trials are the gold standard to assess treatments.[2]

As mentioned earlier, before intervening because of an abnormal result it is wise to repeat the measurement. If the first value was at the edge of the individual's normal range, it is unlikely that the second will be equally extreme. If the second value is similar to the first, it is likely that both are representative of the individual's mean value. If the second result is very different from the first (e.g. by more than twice the CV_i), it may be wise to do a third test to determine that the intervention is justifiable and to establish a reliable baseline for monitoring after the intervention.

Identifying significant change

Statistically significant change – the least significant change

Having reduced the background variability as much as possible, some way is needed to determine the least significant change (LSC) that is a true signal of change; the LSC is also known as the 'significant difference'.

The size of the LSC is determined by the background variability of the test results and the desired level of certainty that an apparent

'signal' of change is real. To be 80% confident that an apparent change is a real one (the confidence level usually used), the LSC is approximately twice the CV_i – as explained in Box 2.[2] In some clinical circumstances, one might wish to be more than 80% certain, and then the LSC would increase (Box 2).

An example of calculating statistical significance is given below, using the case mentioned earlier.

John's HbA_{1c} was 7.8% (62 mmol/mol) several months ago. Now the result has increased to 8.2% (66 mmol/mol). Is this a significant change?

- From earlier, the total individual variability (CV_i) for HbA_{1c} is 3.8%
- The LSC = 2 x CV_i = 2 x 3.8% = 7.6%.

The actual HbA_{1c} % change is 0.4% above the initial 7.8%, i.e. a 5.1% increase. This is less than the LSC and the increase is quite likely to be a result of background 'noise' rather than a real change in the level of John's overall glycaemia.

Clinically significant vs statistically significant change

Clinical significance

To be considered clinically significant, a change must be statistically significant and exceed the LSC. However, a statistically significant change may not be clinically significant and suggest intervention is needed. An example is given below, again using the case mentioned earlier.

You are concerned John's blood pressure is not well controlled and you start an ACE inhibitor (or angiotensin receptor antagonist). As recommended on the prescribing information you recheck John's renal function two weeks later. His plasma creatinine (P-creat) has increased from 104 to 128 µmol/L. Is this clinically significant?

- The first step is to check that the change exceeds the LSC:
  - the RI for P-creat given by the laboratory is 50 to 120 µmol/L
  - the CV_i for P-creat is 5.9%, as calculated in Box 1
  - the LSC in John = 2 CV_i = 11.8%.

The actual change is an increase of 24 from 104 µmol/L, i.e. an increase of 23%. This is more than the LSC and is therefore statistically significant.

- The second step is to check if this statistically significant change is clinically significant:
  - the change is not unexpected. ACE inhibitors and angiotensin receptor antagonists constrict the afferent arterioles to the glomerulus and dilate the efferent arterioles, thereby decreasing glomerular pressure and filtration and increasing the P-creat level.

An evidence-based threshold recommended by respected authorities usually guides the decision as to the clinical significance of a change. In John's case, the Australian consensus suggests that a clinically significant increase in P-creat level after starting an ACE inhibitor or angiotensin receptor antagonist is 30% or greater in those with a P-creat level of 120 µmol/L or lower before starting the ACE inhibitor or angiotensin receptor antagonist.[5] Clinical studies suggest that this level of change may indicate significant renal artery disease. The 23% increase in John's P-creat level does not exceed this threshold. (A 20% or greater increase is clinically significant in those starting with a P-creat level above 120 µmol/L.[5])

Thresholds

Many thresholds need to be interpreted in the light of the clinical situation, as illustrated below.

- The general target for HbA_{1c} is 7% (53 mmol/mol) or lower according to the RACGP and Diabetes Australia guidelines (*General Practice Management of Type 2 Diabetes – 2004-15*) but the Australian Diabetes Society recognises that there are risks as well as benefits in improving glycaemic control and recommends different levels of glycaemic control for different groups of people, for example:[4,6]
  - 6% (42 mmol/mol) or lower in those using lifestyle

2. LEAST SIGNIFICANT CHANGE

- The overall variability of the difference between two measurements in an individual is greater than the variability of the individual measurements:
  - the CV of the difference between two measurements = $\sqrt{2} \times CV_i$*
- The more confident one wishes to be that the change in a measurement is a 'signal' rather than a 'noise', the greater the change needs to be relative to this – this is the least significant change (LSC):
  - $\sqrt{2} \times CV_i \times z$ (where the z value varies from 1.28 for 80% confidence to 2.6 for 99% confidence)†
    Note that this is the standardised one-sided probability where the difference is not the result of chance (the background 'noise' of test variability)
  - generally, 80% confidence is used: (z = 1.28):
    LSC = $\sqrt{2} \times CV_i \times 1.28 = 1.4 \times CV_i \times 1.28 = 1.8 \times CV_i$
    which approximates to 2 x CV_i
  - for 99% confidence (z = 2.6):
    LSC = $\sqrt{2} \times CV_i \times 2.6 = 1.4 \times CV_i \times 2.6 = 3.6 \times CV_i$

ABBREVIATIONS: CV = coefficient of variation; CV_i = total intraindividual coefficient of variation; SD = standard deviation.

* The variance (S_1^2 and S_2^2) of the two measurements (M_1 and M_2) is added and the SD of the difference between the measurements = $\sqrt{}$ total variance:
$\sqrt{S_1^2 + S_2^2} = \sqrt{2S^2}$ (since $S_1 = S_2 = S$) = $S \times \sqrt{2}$.
The CV of the difference = SD ÷ mean value of the two measurements.

† The z value is a measure of the distance of a particular normally distributed value from the mean in terms of numbers of SDs; the further from the mean (i.e. the greater the z value), the less likely a result has occurred by chance.

3. PRACTICE POINTS: MONITORING PATHOLOGY TEST MEASUREMENTS

- The variability of laboratory results is expressed in terms of SDs in absolute terms or as the CV in relative terms (the ratio of the SD to the measured value, i.e. SD ÷ result, as a percentage).
- Variability in results from an individual occurs because of the biological variability within that individual and the variability introduced before and after the specimen arrives in the laboratory (i.e. collection, storage, transport and by laboratory analysis).
- To minimise the background 'noise' of variability introduced by collection and analysis, specimens should be collected correctly and transported promptly to the laboratory and the same laboratory should be used for the initial and subsequent tests.
- If a test is unexpectedly abnormal or if it would prompt medical intervention, it should be repeated to identify errors and to reduce the variability of what will become the baseline for future monitoring.
- The least significant change (LSC) indicating a real 'signal' of change within an individual is approximately 2 x CV_i (%).
- When monitoring results, a change must be clinically as well as statistically significant to prompt medical intervention.
- Different levels of results may be appropriate for intervention in different individuals and these levels should be set using evidence-based recommendations and clinical judgement.

ABBREVIATIONS: CV = coefficient of variation; CV_i = total intraindividual coefficient of variation; SD = standard deviation.

and metformin for glycaemic control
  - 8% (64 mmol/mol) or lower for those with recurrent severe hypoglycaemia.

Ongoing monitoring

Summarising the above section, ongoing laboratory monitoring – as occurs in diabetes – requires some assessment of the statistical and clinical significance of changes.

The statistical significance of a change in an individual is easily assessed by comparing the actual change (as a %) with the LSC. The assessment of clinical significance is guided by any evidence-based recommendation and by clinical judgement.

Key points

- The least significant change (LSC) indicating a real 'signal' of change is approximately 2 x CV_i (%).
- When monitoring results, a change must be clinically as well as statistically significant to prompt medical intervention.
- Different levels of results may be appropriate for intervention in different individuals and these levels should be set using evidence-based recommendations and clinical judgement.

Conclusion

Ongoing laboratory monitoring requires some assessment of the statistical and clinical significance of changes in measurements. Variability between sequential laboratory measurements of the same variable within the same individual occurs because of the biological variability within that individual and the variability introduced through the collection, storage and transport of the specimen to the laboratory and the laboratory analysis itself. This background 'noise' of random variability may be minimised by collecting specimens appropriately, transporting them promptly to the laboratory and using the same laboratory for the initial and subsequent tests.

Repeating a test that has a result that is unexpectedly abnormal or would prompt medical intervention enables any errors to be identified and takes into account variability within the individual, thereby reducing the variability of what will become a baseline for future monitoring.

To be considered clinically significant, a change must be statistically significant and meet the evidence-based threshold for clinical significance recommended for that test by respected authorities, as well as being interpreted in the light of the clinical situation. Practice points are listed in Box 3. MT

References

1. Phillips P. Pitfalls in interpreting laboratory results. Aust Prescr 2009; 32: 43-46.
2. Phillips P. Interpreting pathology tests: results and reference intervals. Med Today 2014; 15(11): 64-68.
3. Ricos C, Alvarez V, Cave F, et al. Desirable specifications for total error, imprecision, and bias, derived from intra-and inter-individual biologic variation. Desirable biological variation database (hosted by Westgard QC, Madison, WI, USA). Available online at: http://www.westgard.com/biodatabase1.htm (accessed March 2015).
4. Mathew TH, Australian Creatinine Consensus Working Group. Chronic kidney disease and automatic reporting of estimated glomerular filtration rate: a position statement. Med J Aust 2005; 183: 138-141.
5. The Royal Australian College of General Practitioners (RACGP) and Diabetes Australia. General practice management of type 2 diabetes – 2014-15. Melbourne: RACGP and Diabetes Australia; 2014. Available online at: https://diabetessociety.com.au/downloads/positionstatements/HbA1ctargets.pdf (accessed March 2015).
6. Cheung NW, Conn JJ, d'Embden MC, et al. Australian Diabetes Society Position Statement: Individualization of HbA1c targets for adults with diabetes mellitus. Sydney: ADS; 2009. 222.diabetessociety.org.au.

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.