PRACTICE

PRACTICE POINTER

Your results may vary: the imprecision of medical measurements

James P McCormack *professor*[1], Daniel T Holmes *clinical professor*[2][3]

[1]Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada; [2]St Paul's Hospital, Department of Pathology and Laboratory Medicine, Vancouver, BC, Canada; [3]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada.

What you need to know

- Inherent in every medical measurement is a degree of uncertainty: you must have a rough idea of the magnitude of that uncertainty to correctly interpret any reported measurement

- The greater the uncertainty, the greater the difference that needs to be observed between two measurements before you can be confident that a true change has occurred

- The "reference change value" (RCV) allows you to decide whether a change in two serial lab results is likely due to chance alone. The required change may be as small as 2% and as large as 50% depending on the test

- Biological variation is typically the largest contributor to the RCV. For analytical variation, your local lab director can tell you the measurement error of any test you are interested in

---

How this article was made

This article was based on a review of the available biological variation data for select routine clinical chemistry measurements as collated by the European Federation of Clinical Chemistry and Laboratory Medicine (https://biologicalvariation.eu/) and in select cases identified by PubMed search. These data were combined with directly calculated analytical variability measurements from a tertiary care hospital (see online supplemental table). Using these estimates, we calculated reference change value (RCV) estimates for the select analytes in order to assist clinicians in deciding whether changes seen in serial blood measurements are more likely due to random biological and analytical fluctuations or due to a true change in physiology. These results and calculations were incorporated into an interactive infographic for use as a clinical decision support and patient education tool.

Why knowing the uncertainty of a measurement matters

Measurement is core to the practice of medicine and drives much of the day-to-day decision making. Unfortunately, because test results are typically reported as a single static number without any statement of uncertainty, and often to more decimal places than appropriate, clinicians may fall into the erroneous assumption that laboratory results are exact. This can lead to over-interpretation of an apparent change in what was measured, which, in turn, leads to unwarranted intervention and feelings of fear, happiness, frustration, and confusion—both for patients and healthcare providers. With increasing numbers of patients having access to their laboratory reports, erroneous interpretation is becoming a more pressing issue.

Consider the following questions:

- Does the serum cholesterol level falling from 5 mmol/L to 4.5 mmol/L mean that lifestyle changes were successful?

- Is a rise in the vitamin D level proof that supplementation worked?

- Does an increase in HbA$_{1c}$ from 45 to 49 mmol/mol (6.3% to 6.6%) justify a lifelong label of diabetes?

Clinicians and patients need to interpret a multitude of medical measurements. These are often central to monitoring health and informed decision making. Has the serum cholesterol concentration come down since starting a statin? Have vitamin D levels gone up? Is the dose of thyroid medication correct? An understanding of the imprecision of medical measurements is essential to answer any of these questions. Even when laboratory and industry scientists have optimised their diagnostic testing processes to minimise inaccuracies, there always remains an error in any clinical measurement due to unavoidable, naturally occurring variability.

This practice pointer explains the nature of measurement errors and offers a practical guide to both estimating the confidence interval of a single result and deciding if changes between serial laboratory tests reflect true changes or simply fluctuations based on analytical or biological variation.

Correspondence to: J P McCormack jmccorma@interchange.ubc.ca

The answer to all these questions may well be "no" because of the various sources of uncertainty.

Uncertainty is always present and comes in different forms

The uncertainty arises from different causes of variation involved in measurement. There are human and other "preanalytical" sources of variation (the manner in which a specimen is collected, handled, or shipped and the storage conditions to which it was exposed), and these can lead to either random or systematic variation. Even when these are eliminated, there is always persistent analytical variation or imprecision ("instrument" variation). Finally, and most importantly, there is biological variation.

Biological variation: the noise that cannot be eliminated

Generally, the largest contributor to variability in laboratory results is biological variation. Biological variation is the "noise" attributable to normal physiological processes when repeated measurements are made over time in an individual. Biological variation comes about through the convergence of the innumerable perturbations of our physiology, state of health, living environment, diet, activity, stress, mood, the weather and climate, and environmental exposures. In the absence of an intercurrent illness or other physiological disruption (such as puberty or menopause), there still remains a random fluctuation (around an individual's true underlying mean value or "set point") in the concentration of ions, metabolites, and proteins in our tissues. The amount of fluctuation is dependent on human physiology, so the magnitude of biological variation varies considerably among the different things we measure. For example, sodium is very tightly regulated, whereas ions such as magnesium show larger fluctuations on a percentage basis. Blood glucose, even in people without diabetes, is tied closely to meals and their content, and some analytes vary with climate and exposure to sunlight, such as vitamin D and its precursors and metabolites. For example, the biological variation in cholesterol measurements is roughly 10-fold higher than the variation seen with sodium measurements (see online supplemental table and interactive infographic).

Biological variation is always present even when analytical imprecision and other forms of laboratory error have been minimised. When analytical variation is kept at a level less than half the biological variation (which is typically achieved in most laboratories), its effect on the total variation is less than about 12%—the rest is biological (see online appendix). Although this is non-intuitive, it is because these types of statistical errors combine in the same manner as the hypotenuse of a right-angled triangle grows when one or both of the other sides is lengthened. While a laboratory can seek to minimise analytical variation, biological variation cannot be eliminated but must definitely be understood.

Estimating variation in routinely ordered medical measurements

The table and the infographic provide estimates of variability for a variety of routinely ordered medical measurements. A detailed description of how these ranges were calculated is provided in the online appendix.

The first data column of the table conveys the actual analytical confidence interval or error due to the measurement process alone, without any consideration of biological effects. This column answers an immediate question, "What is the 95% confidence interval of a numerical result reported from the lab for a single analysis?"

The second data column provides the combined analytical and biological variation. It answers the question, "What is the 95% confidence interval of a single measurement for estimating the long term biological set point of the blood test?"

The last column of the table provides information to help you decide if a difference between two serial measurements reflects a real change or just fluctuation due to biological and analytical processes. It answers the question, "By how much does a test result need to change from the prior measurement in order to be considered different with 95% confidence?"

HbA$_{1c}$ for diagnosing diabetes

Suppose a single HbA$_{1c}$ measurement from a patient is 44 mmol/mol (6.2% NGSP), which is in the middle of the "prediabetic" range (42-47 mmol/mol; 6.0-6.5% NGSP).[1] Because of analytical error of this single measurement, the laboratory could have reported a result anywhere from 41 to 47 mmol/mol (5.9-6.5% NGSP). However, if we are trying to estimate the patient's long term biological HbA$_{1c}$ setpoint to establish what diagnostic category they would fall in over time, we should consider the combined analytical and biological variation. With this amount of variation, a measurement in the middle of the prediabetic range could really be anywhere from 39 to 49 mmol/mol (5.7-6.6% NGSP) over time. With this degree of variation, depending on when the patient had their test, the patient could fall into either of the adjacent diagnostic categories of "normal" and "diabetic." This shows how measurement variability can be a fundamental problem when it comes to rigid diagnostic thresholds.

Estimating whether the difference between two measurements is real or due to random variation

If we know the analytical and biological variation of a specific test, we can mathematically estimate the minimum difference between two consecutive results (each of which have measurement uncertainty) which must be exceeded before the change is considered statistically significant (a significance level of 0.05 to achieve 95% confidence is typically used). In other words, when the difference between two serial measurements exceeds the threshold for statistical significance, we can conclude that the difference is not wholly attributable to chance and we can have reasonable confidence that a bona fide change has occurred. This statistically significant difference is known as the reference change value (RCV).[2] RCVs for common measurements are given in the table and infographic. Naturally, if a significance level larger than 0.05 is chosen, the required change will be smaller (see appendix).

Even if we use the RCV to dismiss the possibility that the calculated difference is attributable to random biological and analytical fluctuations, this does not mean that the numerical value of that difference is perfectly accurate because it is still affected by the analytical errors of the first and second measurements from which it is calculated.

There are contexts where non-urgent decisions are made around chronic care using longitudinal data (for example, HbA$_{1c}$, cholesterol, vitamin D). When it comes to decisions for these measurements, clinicians may want to consider the measurement in the context not of just the analytic variation but also its expected biological variation occurring over the coming days,

weeks, or months. For instance, while a single cholesterol measurement has relatively low analytical variation, the biological variation is about threefold higher, meaning a cholesterol level reported two or three months later could be quite different from the original measurement on the basis of an individual's physiology and not because of any intervention or treatment. Given this, the last column of the table provides the combined variation that needs to be considered in these types of scenarios. Measurements of sodium, chloride and osmolality have relatively small measurement uncertainty and similarly small biological variation so changes in these lab results are most often real. However for measurements of liver enzymes (ALT, AST, GGT), vitamins and minerals (vitamin B12,vitamin D, iron) clinicians need to be aware of the total variability so they can ascertain whether a real change has occurred. If the combined effects of measurement variation and biological variation are not taken into account, interpretive errors can easily be made.

Serial low density lipoprotein cholesterol (LDL-C) measurements after starting a statin

Consider a person who has a baseline LDL-C measurement, starts a statin and subsequently has their LDL-C measured two months later. Using ballpark estimates of analytical and biological variation, we can calculate the RCV and estimate that the LDL-C needs to change by at least 21-30% (see table and infographic) before we can be confident a change has truly occurred. Let's say the patient's initial LDL-C was 3 mmol/L. On repeat measurement two months after starting the statin, the LDL-C level would need to fall to about 2.2 mmol/L for the clinician to be confident the LDL-C level has actually decreased. To appreciate the clinical relevance of this change, it must be put into context with the effects one might expect to see with treatment. A statin at a dose of 10-20 mg will lower LDL-C by roughly 30-35%.[3] In view of this relatively large change, a single follow-up measurement could probably be used to show whether the LDL-C has decreased because the expected change exceeds the RCV (~25%).

However, increasing the statin dose from 10-20 mg to 40 or 80 mg only lowers LDL-C by roughly a further 10%[3]—for example, from 2.2 to 2.0 mmol/L. This change is smaller than the RCV, and the small change we seek to detect in a repeat LDL-C measurement cannot be differentiated from the background biological variation. Repeat measurements after a statin dose change are therefore of limited or no benefit and can be misleading. Measurements to detect small changes are often hampered by variation—in extreme situations they can be akin to trying to hear a sparrow's call at a heavy metal concert.

Bone density with bisphosphonate treatment

Bisphosphonates given over two to three years produce an average increase of ~3-5% in bone density compared with placebo.[4] Despite bone density measurements having relatively small analytical variation and very little biological variation, this 3-5% is below the approximate minimal change (6-10%) required for us to be confident the measured change in femoral neck bone density isn't due to expected variation. For this and other reasons, the 2017 American College of Physicians osteoporosis guidelines state: "The data do not support monitoring [bone mineral density] during the initial 5 years of treatment in patients receiving pharmacologic agents to treat osteoporosis."[5]

HbA$_{1c}$ for monitoring diabetes

Typically, medications for diabetes will lower a baseline HbA$_{1c}$ level of 64 mmol/mol (~8.0% in National Glycated Haemoglobin Standardization (NGSP) units) by about 8 mmol/mol (~0.7% NGSP) to 56 mmol/mol (~7.3% NGSP).[6] This ~12% drop (from 64 to 56 mmol/mol) in HbA1c (which corresponds to an ~9% drop in NGSP units from 8.0% to 7.3%) is less than the calculated RCV—about 17% for IFCC measurements and 12% for NGSP measurements—required to confidently exclude the combined effects of analytical and biological variation in diabetic patients (see table and infographic). Readers may be surprised that RCVs are different for the IFCC and NGSP reporting systems. This is discussed elsewhere[7] and further explained in the appendix.

Given the RCV for HbA$_{1c}$, serial measurements of a patient's HbA$_{1c}$ while receiving treatment can be tricky to interpret, and it may be difficult to be confident that the treatment is working as intended. This exemplifies why we should not over-interpret changes that are small relative to a test's RCV.

Multiple measurements to reduce variation

In individuals, if one was to take two measurements before and two measurements after an intervention this reduces the change required (RCV) by about 30%: four measurements before and after reduce it by 50%.[8] However, the process of taking multiple measurements in short succession in routine clinical practice is impractical, and, even if it was done, clinicians and patients are still left with considerable variation.

Biological variation in clinical trials

In contrast to measurements taken on an individual patient, clinical trials of multiple patients are often undertaken to estimate the impact of medications on various medical markers such as cholesterol or glucose. The average changes observed overall can often be smaller, relatively speaking, than the biological variation. However, in clinical trials, outcome analyses are less affected by biological variation because multiple measurements are performed in multiple people at multiple time points, causing the effect of analytical and biological variability to be attenuated. Therefore, the problem or evaluation of serial measurements and change over time is primarily an issue for measurements in individual patients not when these measurements are made in clinical trials.

Suggestions for practice

Much of the uncertainty in clinical measurements is not a fixable problem but only a knowable problem. We suggest that clinicians "bookmark" the interactive infographic on their web browser or use the table in their consulting room to help them interpret laboratory results in the context of the measurement uncertainty, because it is not typically provided in laboratory reports at present. Accreditation systems now mandate that the laboratories have established their analytical measurement uncertainty and can at least provide this data upon request.

Patients increasingly have access to their own laboratory results. We are not aware of any electronic records system that conveys uncertainty in measurement to patients accessing their own records. In our view, this needs to change. In the meantime, clinicians could consider discussing this measurement issue with patients who routinely follow their own laboratory results and potentially provide them with a ballpark estimate of what changes in reported results need to occur to be confident that a real change has occurred.

In addition, the concept of uncertainty in measurements also helps explain that, although serial measurements (such as at annual checkups) may seem intuitively useful, they may be of limited value and could potentially lead to confusion, inappropriate reassurance, or over-investigation. For instance, because the RCV of cholesterol is close to 25%, the typical average yearly increase in cholesterol (~0.5-1%) cannot be identified and therefore remeasurements more often than even every 3-5 years could be misleading.[9]

A problem we can't fix, but need to know about

Modern laboratory science has minimised analytical variation for many routine analytes to the point of manageability. However, biological variation is not something that can be minimised, managed, or fixed. The only answer is that clinicians must understand the concept and effect of biological variation and explain these to patients who access their results. Armed with the ballpark estimates provided in this article, clinicians can more appropriately interpret test results and empower their patients to understand what their results mean while avoiding misinterpretation.

Caveats and considerations

- Biological variations used in our calculations were derived from healthy cohorts. In states of illness or treatment, biological variation may differ by disease and analyte. For instance, the RCV of HbA_{1c} is provided in the table for both individuals with and without diabetes.

- If a significance level greater than 0.05 (95% confidence) were used, such as 0.1 (90% confidence) or 0.2 (80% confidence), the exceeded change required (the RCV) would be smaller. Refer to infographic to calculate these.

- Point of care testing devices typically have larger (sometimes much larger) analytical variation than instruments used by accredited clinical laboratories.

- Should readers wish to perform their own RCV calculations, they should be aware that analytical variations in their local laboratory may be larger or smaller than those presented in this paper based on local instrumentation. However, the information provided in the table and infographic is likely representative of a typical high-volume clinical laboratory.

Education into practice

- Consider how you typically explain blood test results to patients. Do you think you have overinterpreted changes in blood test results in the past?

- How can you use the concept of the reference change value (RCV) to help your patients interpret lab results they have seen online?

How patients were involved in this article

No patients were involved in the creation of this article.

1   Diabetes Canada Clinical Practice Guidelines Expert Committee. Diabetes Canada 2018 clinical practice guidelines for the prevention and management of diabetes in Canada. *Can J Diabetes* 2018;42(suppl 1):S1-325.
2   Fraser CG. Reference change values. *Clin Chem Lab Med* 2011;50:807-12.21958344
3   Law MR, Wald NJ, Rudnicka AR. Quantifying effect of statins on low density lipoprotein cholesterol, ischaemic heart disease, and stroke: systematic review and meta-analysis. *BMJ* 2003;326:1423. 10.1136/bmj.326.7404.1423 12829554
4   Sanderson J, Martyn-St James M, Stevens J, etal . Clinical effectiveness of bisphosphonates for the prevention of fragility fractures: a systematic review and network meta-analysis. *Bone* 2016;89:52-8. 10.1016/j.bone.2016.05.013 27262775
5   Qaseem A, Forciea MA, McLean RM, Denberg TDClinical Guidelines Committee of the American College of Physicians. Treatment of low bone density or osteoporosis to prevent fractures in men and women: a clinical practice guideline update from the American College of Physicians. *Ann Intern Med* 2017;166:818-39. 10.7326/M15-1361 28492856
6   CADTH. New drugs for type 2 diabetes: second-line therapy—recommendations report. 2017. https://www.cadth.ca/dv/new-drugs-type-2-diabetes-second-line-therapy-recommendations-report.
7   Weykamp CW, Mosca A, Gillery P, Panteghini M. The analytical goals for hemoglobin A(1c) measurement in IFCC units and National Glycohemoglobin Standardization Program units are different. *Clin Chem* 2011;57:1204-6. 10.1373/clinchem.2011.162719 21571810
8   Jones GR, Chung JZ. Reference change values using more than two results. *Ann Clin Biochem* 2016;53:413-4. 10.1177/0004563215613110 26628683
9   Glasziou PP, Irwig L, Heritier S, Simes RJ, Tonkin ALIPID Study Investigators. Monitoring cholesterol levels: measurement error or true change?*Ann Intern Med* 2008;148:656-61. 10.7326/0003-4819-148-9-200805060-00005 18458278

Figure

Approximate variability estimates for routine medical measurements

Test	Single measurement variability		Serial measurement variability
	Analytical variation†	Analytical and biological variation‡	Reference change value*
Bone density (spine, total hip)	<2%	2-5%	2-5%
Chloride	<2%	2-5%	2-5%
Osmolality	<2%	2-5%	2-5%
Sodium	<2%	<2%	2-5%
Bone density (femoral neck)	<2%	2-5%	6-10%
Albumin	2-5%	6-10%	6-10%
Calcium	2-5%	2-5%	6-10%
HbA$_{1c}$ NGSP (%)	2-5%	6-10%	6-10%
Haemoglobin	2-5%	6-10%	6-10%
Total protein	2-5%	6-10%	6-10%
Transferrin	2-5%	6-10%	6-10%
Creatinine	2-5%	6-10%	11-20%
Glucose	2-5%	6-10%	11-20%
HbA$_{1c}$ (diabetics) IFCC (mmol/mol)	2-5%	11-20%	11-20%
HbA$_{1c}$ (diabetics) NGSP (%)	2-5%	6-10%	11-20%
HbA$_{1c}$ IFCC (mmol/mol)	6-10%	6-10%	11-20%
Lactate dehydrogenase	2-5%	11-20%	11-20%
Magnesium	2-5%	6-10%	11-20%
PCO$_2$	2-5%	6-10%	11-20%
Potassium	2-5%	6-10%	11-20%
Total cholesterol	2-5%	11-20%	11-20%
Alanine aminotransferase	6-10%	11-20%	21-30%
Alkaline phosphatase	11-20%	11-20%	21-30%
Aspartate aminotransferase	6-10%	11-20%	21-30%
Gamma glutamyltransferase	6-10%	11-20%	21-30%
HDL cholesterol	2-5%	11-20%	21-30%
LDL cholesterol	2-5%	11-20%	21-30%
Phosphate	2-5%	11-20%	21-30%
Rheumatoid factor	11-20%	21-30%	21-30%
Uric acid	2-5%	11-20%	21-30%
Vitamin B$_{12}$	11-20%	11-20%	21-30%
25-hydroxy-vitamin D	6-10%	21-30%	31-40%
Holotranscobalamin	6-10%	21-30%	31-40%
Total testosterone (male)	6-10%	21-30%	31-40%
Urea	2-5%	21-30%	31-40%
Thyroid stimulating hormone	6-10%	31-40%	41-50%
Iron	2-5%	>50%	>50%
Lactate	2-5%	>50%	>50%
Total bilirubin	2-5%	41-50%	>50%
Triglycerides	2-5%	31-40%	>50%

*Change (%) required for two serial measurements to be considered different.
†Confidence interval (%) around a single measurement when only considering analytical variation.
‡Confidence interval (%) around a single measurement when considering both analytical and biological variation.
NGSP= National Glycated Haemoglobin Standardization. IFCC= International Federation of Clinical Chemistry.
HDL= high density lipoprotein. LDL= low density lipoprotein.