Laboratories provide nearly 90% of all data that comprise a new drug application. Undetected error in test results can delay, complicate and even jeopardize your clinical trial. **How well are you managing testing quality?**

Summary

The expense and timelines of drug development are well understood – an estimated $802 million and 90.3 months are needed to take a drug from phase 1 trials to approval.[1] That translates into $8 million per month or $300,000 per day. The significant role of laboratory tests is also well understood – laboratories generate nearly 90% of all data that comprise a new drug application.[2] Serial test results of individual patients form the empirical basis of drug efficacy and safety. They are intended to demonstrate that experimental patients fare <u>measurably</u> better from baseline than control patients.

Given these economics and the importance of serial test results, one would expect that test results would have a rigorous, scientific underpinning that would substantiate their accuracy and precision and correlate them to real physiological change in patients. Unfortunately, that's simply not the case.

In fact, the management of testing quality in clinical trials usually begins and ends with the selection of accredited laboratories that pass proficiency testing and report reference intervals with their test results. While this may well meet prevailing standards for routine laboratory testing, it falls well short of what is required in clinical trials. Test results in clinical trials should be held to a much higher standard as they form the basis for assessing safety and efficacy, and support the drug approval process.

This White Paper briefly examines the sources of variability inherent in test results and how prevailing practices fall short in clinical trials. It then outlines a new approach to more rigorously manage the variability inherent in test results so that they correlate – with statistical significance – with real physiological change in patients. Efficacy and safety – or lack thereof - can be established faster, with greater certainty and potentially with fewer patients. This approach can significantly simplify, streamline and even shorten clinical trials – providing a compelling value proposition for stakeholders within the entire drug development continuum. **To learn more, see the Online Consultation form at the end of this White Paper.**

Serial Testing

A patient is tested at baseline and thereafter at regular intervals. The test results change over time. But do the changes in test results represent <u>actual</u> physiological change – in other words, is the patient getting better or worse? Not necessarily, as the observed changes may also be due to pre-analytical, analytical and biological variation. These sources of variation need to be rigorously managed and quantified to enhance the discriminatory value of the test result. Failure to do so introduces the potential for significant errors.

Consider a candidate drug whose posited therapeutic effect is to reduce Analyte A by 25% from baseline. However, associated with Analyte A there is pre-analytical variation of 15%, analytical variation of 25% and biological variation of 30%. As illustrated below, these sources of variation could be additive (70%), cancel each other out (30%), or anything in between.

Clearly, a 25% reduction in Analyte A can easily be obscured if test results vary by 30% to 70% for reasons independent of therapeutic effect. Unaccounted test result variation can complicate an already expensive and lengthy clinical trial process.
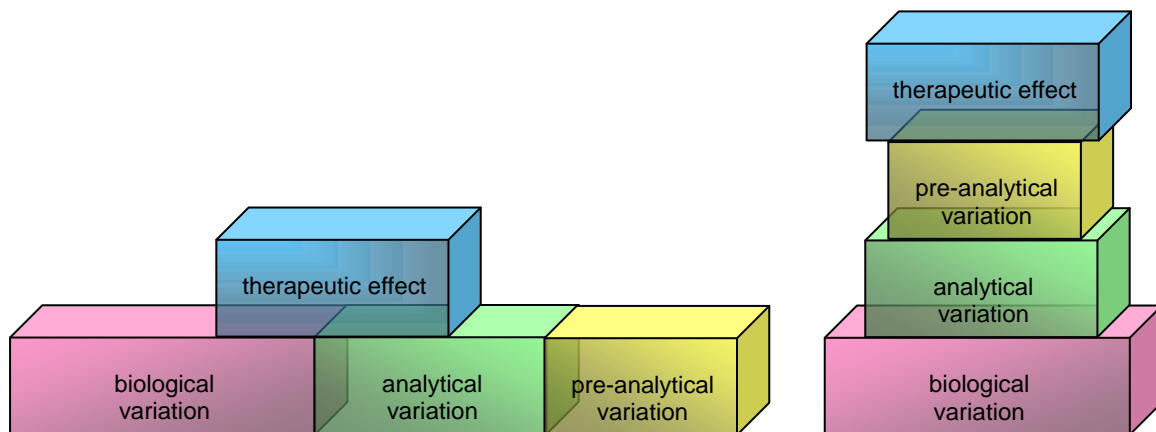
The posited therapeutic effect (20% reduction in Analyte A from baseline) is indicated in the blue box. Pre-analytical variation (15%) is indicated in the yellow box. Analytical variation (30%) is indicated in the green box. Biological variation is indicated in the red box. These sources of variation could be additive (left graphic), they could cancel each other out (right graphic) or combine in other ways. Failure to manage and quantify these sources of test variability introduces the potential for significant errors into clinical trial data.
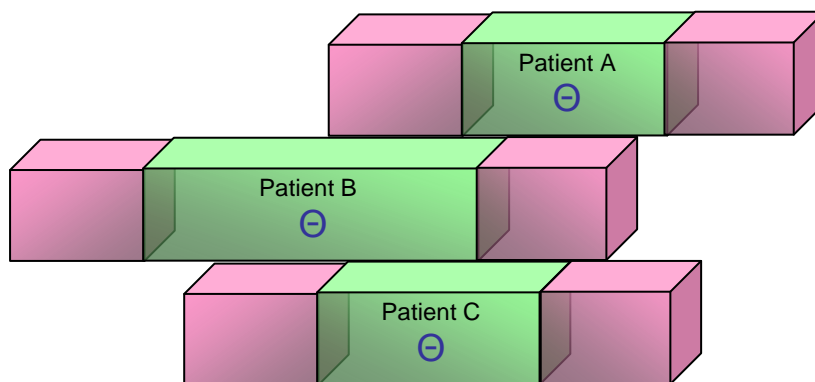
Biological Variation

Dr. Callem G. Fraser (Clinical Director of Biochemical Medicine and Honorary Senior Lecturer at the Universities of Dundee and St. Andrews in Scotland) can be considered the dean of biological variation. Here's how he describes biological variation:[3]

Many of the quantities measured in laboratory medicine change over the span of life: one repercussion is that population-based reference values may need to be stratified into sub-groups according to age. A small number of quantities vary in predictable cyclical rhythms which may be daily, monthly, or seasonal in nature. Knowledge of these rhythms is important because it would be very difficult to create population-based reference values for each and every point in a cycle: this means that the appropriate times to investigate a cycle must be defined and reference values generated for these. Moreover, absence of predicted rhythms may give indications to the presence of disease. However, most quantities do not have predictable cyclical rhythms. It can be considered that, for each individual, the quantity varies round a homeostatic set point in a truly random manner (emphasis added).

In other words, each patient has a unique biological makeup for each analyte. This is reflected in a homeostatic set point ("Set Point") and a range of fluctuations around that set point ("Biological Variation"). Biological Variation for an individual patient remains constant over the patient's lifetime even if the Set Point changes.
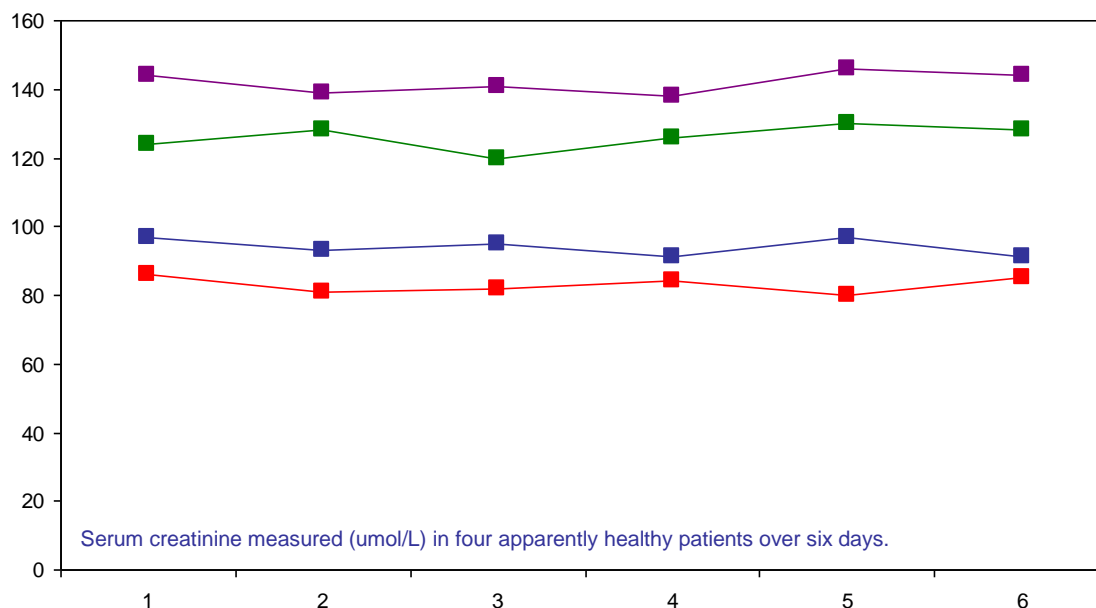

This figure illustrates three patients, each with a unique Set Point (Θ) and a unique Biological Variation (as indicated by the green box). Test results within the Biological Variation represent expected biological fluctuations. Test results outside the Biological Variation (red boxes) represent actual physiological change.
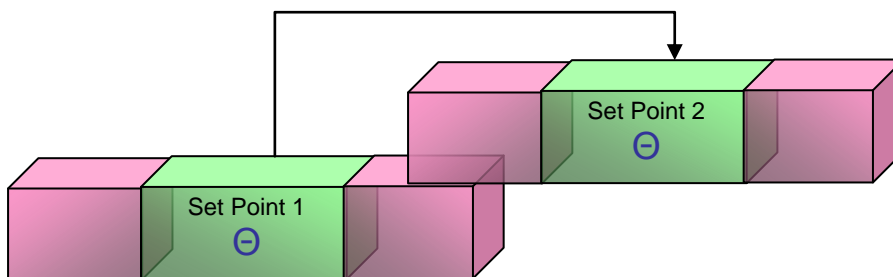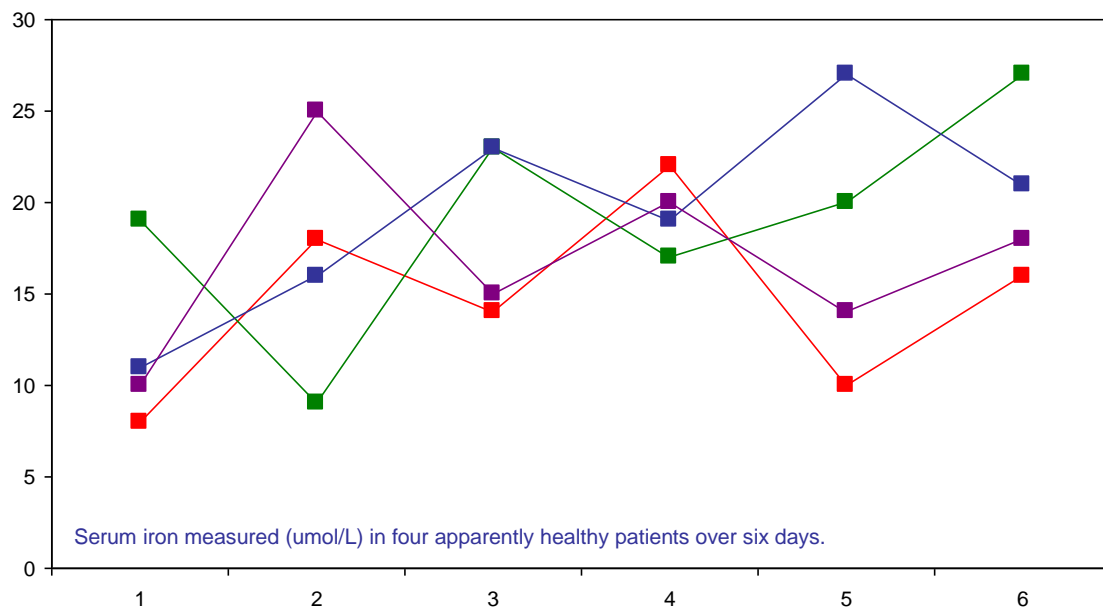
Actual physiological change in a patient can be viewed as a shift from an old Set Point to a new one. In a clinical trial, this evidence is used to validate or negate the hypothesis of the candidate drug – that the experimental group receiving the drug experiences positive physiological change compared with the control group.

Between Patient + Within-Patient Biological Variation

Biological variation differs by analyte and patient. Creatinine and iron are outlined below using data furnished by Dr. Fraser.[4] Creatinine has large between-patient variation and small within-patient variation. Iron has small between-patient variation and large within-patient variation.

Serum creatinine measured (umol/L) in four apparently healthy patients over six days.

Serum iron measured (umol/L) in four apparently healthy patients over six days.

What underlies within-patient biological variations? There is an inherent or true biological variation. As well, measuring this variation itself introduces fluctuations associated with obtaining and handling the test sample (pre-analytical variation) and in the actual testing (analytical variation).

Reference Intervals + Biological Variation

While the existence of biological variation has long been appreciated, this has not resulted in an operational framework of individual Set Points and Biological Variations. Rather what have evolved are reference intervals or ranges in which a patient's test result is compared with a "normal" reference interval for a larger peer group – or reference population. Laboratories generally report test results with a reference interval for the age and gender of the patient so the patient's result is compared to their reference group. Clinical judgment and personal knowledge of the patient is then used to interpret a test result within the context of a myriad of other factors known to affect test results (e.g. diet, medication, exercise, pregnancy, stress, anxiety).

Consider the following example.

Patient Biological Variation

❶ ❷ ❸

Population Reference Interval
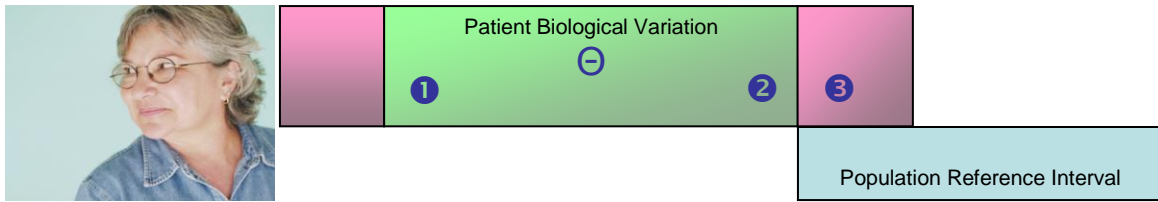
This patient is outside the Population Reference Interval. She is initially healthy, but progresses into a disease state. Her first two test results are within her Biological Variation. But her third is outside her Biological Variation indicating actual physiological change – in this case, the progression of disease. However, the Population Reference Interval supports the opposite – and erroneous – interpretation, namely that she is progressing from a disease state to a healthy one. Her first two test results are outside the Population Reference Interval (suggesting disease state), yet her third is within it (suggesting a healthy state). This erroneous interpretation is reinforced as her serial test results appear to form a trendline of incremental improvement to the "normal" values of the population reference interval.

Some laboratories establish their own reference intervals to reflect the composition of their patient population. Many use the reference intervals suggested by instrument manufacturers (typically as "inserts" with their reagents) based on their reference studies done on their analytical platforms. Of course, these reference intervals will vary by manufacturer given analytical biases associated with their platforms. As a result, reference intervals vary among laboratories, depending on the analytical platforms they use and how they select a reference interval. Reference intervals can also vary within the same laboratory where the same test is performed on different instruments – even from the same manufacturer.

It turns out that what's "normal" can be highly variable from laboratory to laboratory.

Analytical Variation

How do clinical trial sponsors evaluate a laboratory's accuracy (ability to produce a true result) and precision (ability to repeat the same result)? Typically, they do not evaluate accuracy and precision at all. Rather accuracy and precision is <u>inferred</u> from the fact the laboratory is accredited by the College of American Pathologists (CAP) or some other reputable accreditor – and is required to pass proficiency testing from an approved provider.

Is this a reasonable inference?

The short answer is no. Proficiency testing typically uses <u>minimum</u> standards for <u>legal</u> oversight purposes – if a laboratory fails to meet those standards, then its performance justifies sanctions. Whereas, test results in clinical trials need to be accountable to a much higher standard – they need to meet a clinical need and demonstrate efficacy and safety.

There are significant differences between clinical versus legal need. Consider lipid testing and the contrast between clinical standards under the National Cholesterol Education Program (NCEP) of the National Heart, Lung, and Blood Institute versus legal standards under the Clinical Laboratory Improvements Amendments, 1988 (CLIA) administered by CMS.

Analyte	NCEP Performance Goals	CLIA Evaluation Criteria
Total Cholesterol	9% of reference value	10% of peer group mean
HDL Cholesterol	13% of reference value	30% of peer group mean
LDL Cholesterol	12% of reference value	Not regulated
Triglycerides	15% of reference value	25% of peer group mean

NCEP asks *how good does a test result have to be to support evidence-based medicine?* Laboratories are assessed against a reference value, typically established by a credentialed reference method using test samples which mimic patient samples (e.g. human samples). Moreover, the performance goals are based upon an allowable error budget that combines accuracy and precision that is traceable to a defined base of accuracy at the CDC.

CLIA asks a different question: *how bad does a test result have to be before sanctions are invoked?* Not surprisingly, CLIA's standards are considerably lower. Laboratories are assessed not against a single reference value, but against their instrument peers. Wide ranges from the peer group mean are permitted – up to 30% in the case of HDL cholesterol. LDL cholesterol is not even regulated under CLIA. Moreover, test samples for commercial reasons can be stabilized, preserved, lyophilized and augmented. Such samples introduce known matrix effects (they don't mimic patient samples) which precludes using them to assess accuracy.

So, in effect, CLIA only measures precision and not accuracy. This means that under CLIA there can be as many correct answers as there are instrument platforms. This also means that a laboratory could correct a calibration bias and provide an accurate test result, yet still fail CLIA because its peers continue with improper calibration. The net effect of CLIA is to reward peer group precision without consideration of accuracy.
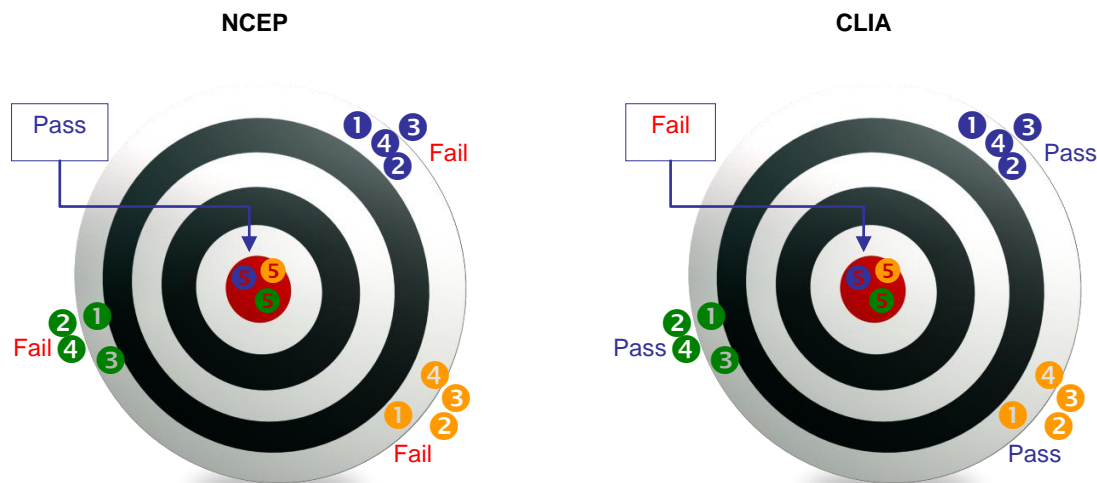

The bull's-eye represents an accurate or true test result. There are 3 different instrument groups (blue, green and amber) with 5 laboratories in each group. Each instrument is calibrated differently with significant bias as represented by the outer rings of the target. Laboratory 5 in each group, however, has corrected its respective instrument bias so that each provides traceably accurate test results. Under NCEP, laboratories 1 – 4 in each group would fail and laboratory 5 in each group would pass. Under CLIA, the opposite would occur.

Calibration Bias

The graphs below illustrate significant bias in laboratories measuring creatinine and HDL cholesterol. In both instances, calibration bias was by far the largest component of testing error.
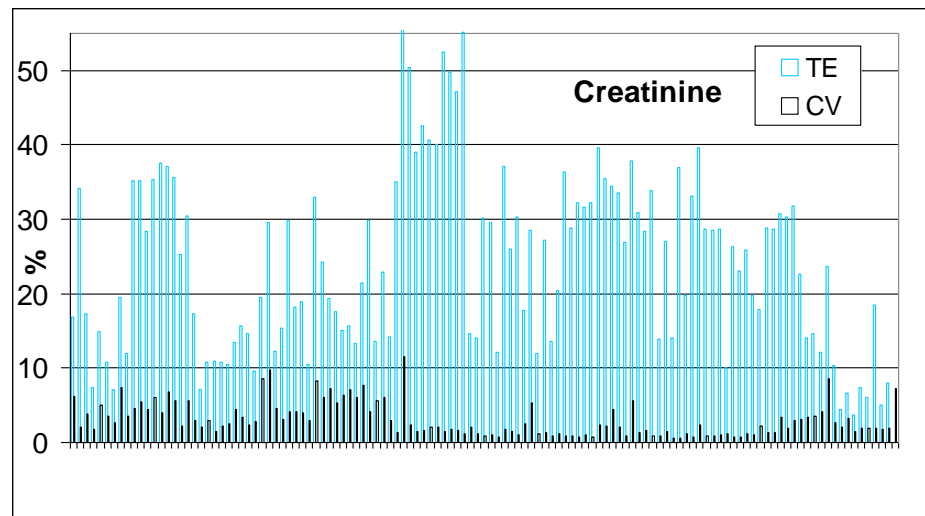
Laboratories in British Columbia standardized their creatinine measurements to enable them to report an estimate of kidney function (eGFR or estimated glomerular filtration rate) with every adult creatinine test result.[5] This graph summarizes each laboratory's total error (TE), imprecision (CV) and bias (TE less CV) at baseline.
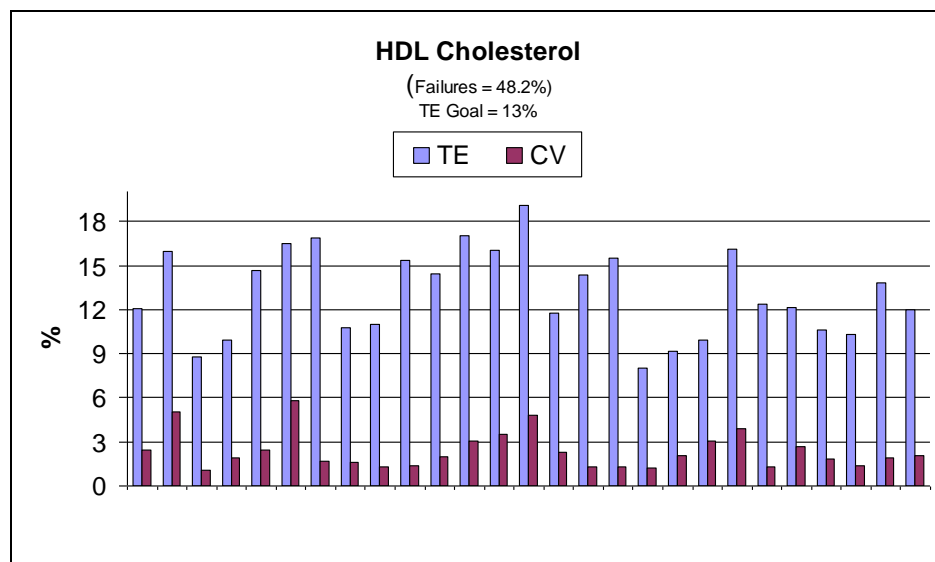

27 laboratories reputed to be the "best" in their country were recruited for a clinical trial involving a statin drug.[6] This graph summarizes each laboratory's total error, imprecision and bias in measuring HDL cholesterol. Nearly half had total error in excess of the NCEP performance goal of 13%.

Some have argued that calibration bias has minimal or no impact in a clinical trial. Their reasoning is that calibration (and associated bias) remains constant over a clinical trial so it has no impact on serial test results. In other words, serial test results detect incremental changes, which are independent of calibration. But just how sound is the assumption that calibration remains constant?

Certainly, that assumption should not be taken as an article of faith in a clinical trial. In fact, evidence suggests that there may be significant variation in calibration bias from one reagent lot number to the next.

Reagent Lot Variation in Measuring Cholesterol
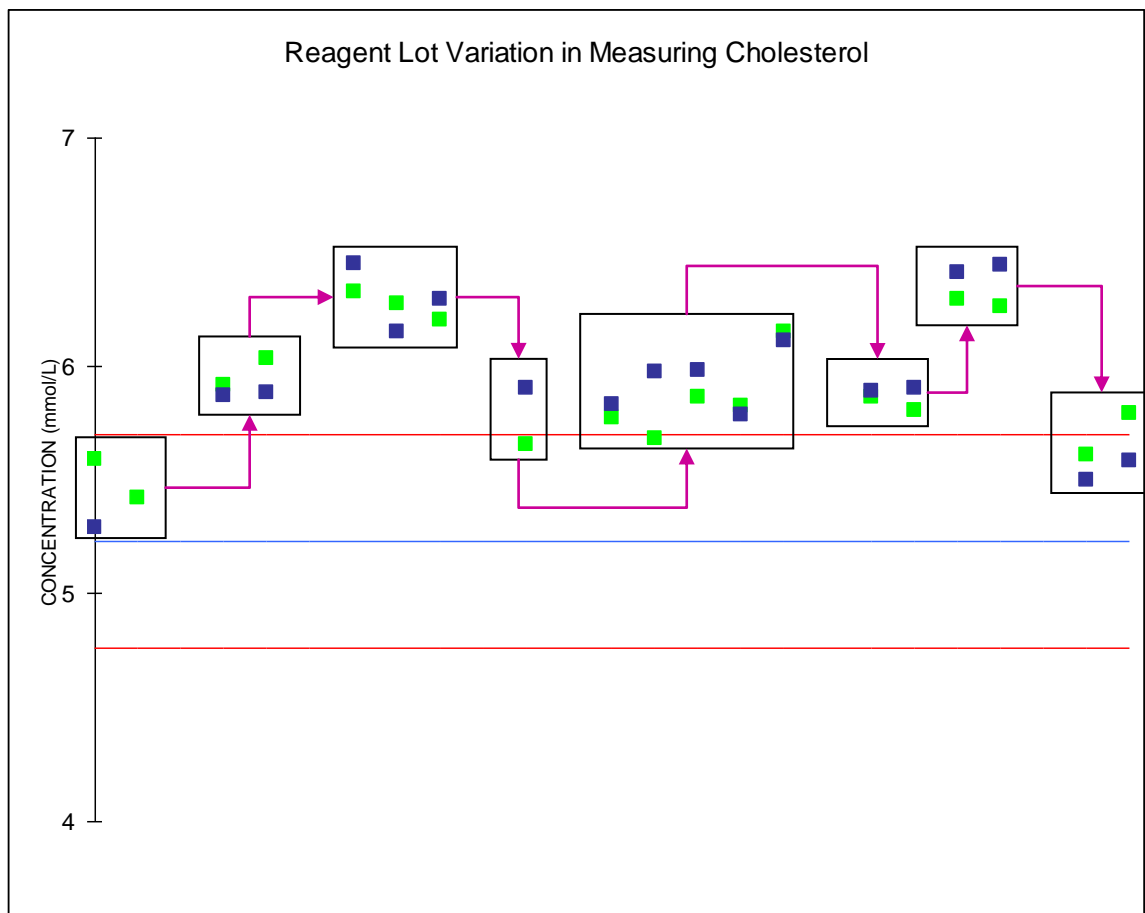
The same sample was tested repeatedly for cholesterol on the same analyzer using eight reagent lots from the instrument manufacturer.[7] The reference value for the sample is indicated by the blue line. The NCEP total error goal (+/- 9%) is indicated by the red lines. The black boxes encompass the test results for each reagent lot. The first reagent lot meets the NCEP performance goal. There is considerable variability in subsequent reagent lots and the overall trendline indicates a positive bias.
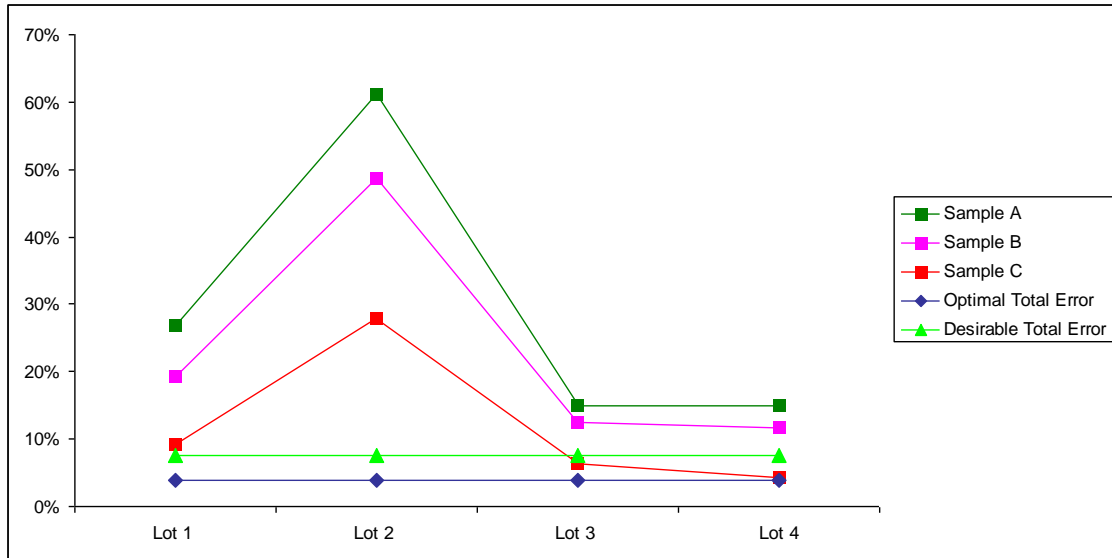
Three creatinine samples A, B, C with known reference values were tested on four sequential reagent lots from the same manufacturer.[8] The manufacturer claimed that the reagents were traceable to the same accuracy base used to establish the reference value. The graph measures error from the reference value. Optimal total error is 3.8% from the reference value and desirable total error is 7.6%.[9] This illustrates significant lot-to-lot calibration bias which increases across the concentration range for creatinine.

Relative Change Index

The CEQAL Clinical Trials Program is an external, independent program that comprehensively manages test result variability so that changes in serial test results can be assessed - with defined statistical certainty. This increases the discriminatory value of the test and enhances identification of real physiological change in the experimental and control groups of patients. This means that efficacy and safety – or lack thereof – can be established faster, with greater certainty and potentially with fewer patients.

The program is operated on a secure, online informatics platform. The program is customized to the needs of each clinical trial. Sponsors and stakeholders receive third party, independent performance reports on a timely basis monitoring the quality of testing throughout the clinical trial together with a comprehensive set of standardized patient data from baseline to support an objective assessment of efficacy and safety.

Each laboratory is "standardized" to a common quality management program so that their total error, calibration bias and imprecision are assessed at baseline and monitored throughout the clinical trial. This means that their testing can be traced to an appropriate accuracy base (for instance, to the reference methods accepted by the Cholesterol Reference Method Laboratory Network under the aegis of the CDC).

Prior to randomization, patients are given a baseline assessment to establish their unique Set Point and Biological Variation for each target analyte. Changes from baseline and from last test result are calculated using a Relative Change Index or RCI. The RCI calculates whether a change is statistically significant by assessing the relative contribution of method imprecision (CVa) and biological variation (BVi) to the reported test result. The RCI assumes calibration remains constant, hence the need to monitor calibration at baseline and going forward for the duration of the clinical trial. Relative Change Index can be described as $RCI = 2^{\frac{1}{2}} \times Z \times (CVa^2 + BVi^2)^{\frac{1}{2}}$. An RCI > 1 is statistically significant. An RCI < 1 is not.

Here is a summary of the program referenced to the sources of testing variation:

Pre-Analytical Variation

- Each laboratory's SOP's (standard operating procedures) related to the collection, handling and accessioning of patient test samples are reviewed to assess whether they are standardized inter se and properly documented in the clinical trial protocol. The goal is to standardize these SOP's procedures among laboratories and throughout the clinical trial to minimize the impact of pre-analytical variation. This assessment has particular relevance for newer markers of disease.

Analytical Variation

- Performance goals are established for measuring target analytes. Wherever possible, an appropriate accuracy base is established for each target analyte with the goal to make test results traceable to a defined accuracy base for the duration of the clinical trial.

- A detailed analytical profile is developed for each laboratory for each target analyte. This profile includes method, instrument model, reagent, reagent lot, reportable range, reference intervals, calibrator, calibrator lot, critical limits and interpretive comments. As well, data are collected to assess the between-day and within-day imprecision for each test method.

- Sample sets are developed or procured whose target values can be traced to the defined accuracy base. These sample sets are used for laboratory baseline assessment, ongoing monitoring and internal quality control.

- Each laboratory receives a baseline assessment to measure their respective total error, imprecision and bias in measuring each target analyte against the stated performance goals.

- If needed, custom bias correction formulae may be developed for each laboratory for each target analyte (or other remedial strategy as appropriate). These formulae may be used post-analytically to normalize test results.

- Each laboratory receives on-going monitoring of testing (external quality assessment) using blinded samples. This monitoring is provided at quarterly, monthly or other intervals as appropriate to ensure continued compliance with the stated performance goals.

- Each laboratory receives custom internal quality control (IQC) materials for each target analyte which are traceable to the defined accuracy base. These IQC materials are used to bracket the testing of patient samples and are submitted online as an on-going monitor of precision and constancy of calibration bias.

Biological Variation

- Each patient in the clinical trial receives a baseline assessment for each target analyte. This consists of testing each patient daily under standardized conditions for a specified number of days (typically 3 or 4), depending upon the underlying biology of the analyte. These data are used to generate patient-specific BVi (Biological Variation).

- The test results are submitted online for analysis. The resulting baseline assessment details the Set Point and Biological Variation for each patient for each target analyte.

- After randomization, patients are tested at the intervals indicated in the clinical trial protocol. Patient samples are bracketed by IQC samples.

- The patient test results and IQC results are submitted online for analysis. The resulting report details the RCI (Relative Change Index) for each patient for each analyte – and compares the current test result against those prior and at baseline.

About CEQAL

CEQAL, based in Vancouver BC, specializes in the design and implementation of innovative, clinically-effective programs to monitor and standardize laboratory testing. Clients include clinical trial organizations, diagnostic manufacturers, government agencies, laboratory networks and proficiency testing providers – including an international EQA collaboration which currently numbers 20 Ministries of Health and other national-scale proficiency testing providers. CEQAL's reference method laboratory operates a number of credentialed reference methods that are used to assign target values within its programs. CEQAL is a member of the Cholesterol Reference Method Laboratory Network that operates under the aegis of the US Centers for Disease Control and Prevention.

David W. Seccombe MD PhD FRCPC | Managing Director, CEQAL • MD (University of Calgary), PhD (University of British Columbia) • Specialty training in Pathology (Royal College of Physicians & Surgeons Canada) • Past Director, Division of Clinical Chemistry, Department of Pathology and Laboratory Medicine, Vancouver General Hospital & Health Sciences Centre • Associate Professor (Tenured), Department of Pathology and Laboratory Medicine, University of British Columbia • Clinical Practice (Lipidologist) • Managing Director, CEQAL, a member of the CDC Cholesterol Reference Method Laboratory Network (CRMLN) • Member, NIH-NKDEP Laboratory Working Group on Creatinine Standardization • Member, IFCC Working Group on Creatinine / eGFR • Member, Guidelines Subcommittee of the Canadian Society of Nephrology • Fellow, College of American Pathologists • Member, Canadian Association of Pathologists • Member, Canadian Society of Clinical Chemists

More Information

To learn more, please complete the Online Consultation Form or contact Katherine Ray BSc MLT MBA, Business Development Manager | kray@digitalpt.com | k_ray.skype | 800.665.2575 | www.ceqal.com.

Endnotes

[1] DiMasi, Hansen, Grabowski (2003) "The price of innovation: new estimates of drug development costs," Journal of Health Economics; Vol 22, pp. 151-185.

[2] Drug Delivery Technology, May 2007.

[3, 4] Fraser CG, Stevenson HP. Production and use of data on biological variation in laboratory medicine. CPD Bulletin: Clinical Biochemistry 1998;1:5-8.

[5] Komenda P, Beaulieu M, Seccombe D, Levin A. Regional implementation of creatinine measurement standardization. J Am Soc Nephrol. 2008 Jan;19(1):164-9.

[6] Donohue WR, Seccombe DW, Jacobson BE, Ho An International study examining the "trueness" and precision of lipid measurements in 27 countries. Poster - XIX International Congress of Clinical Chemistry IFCC/AACC Annual Meeting, Orlando, 2005.

[7, 8] CEQAL data presented by David W. Seccombe, at Central Labs West Conference, San Francisco, February 3 – 5, 2010, www.centrallabswest.com.

[9] Recommended total error performance goals for measuring creatinine as extracted from Clin Chem 52:5-18 (2006).

References

Katayev A, Balciza C, Seccombe DW. Establishing reference intervals for clinical laboratory test results: is there a better way? Am J Clin Pathol. 2010 Feb;133(2):180-6. Feb 2010.

Miller WG, Bruns DE, Hortin GL, Sandberg S, Aakre KM, McQueen MJ, Itoh Y, Lieske JC, Seccombe DW, Jones G, Bunk DM, Curhan GC, Narva AS; National Kidney Disease Education Program-IFCC Working Group on Standardization of Albumin in Urine. Current issues in measurement and reporting of urinary albumin excretion. Clin Chem. 2009 Jan;55(1):24-38.

Komenda, P, Beaulieu, M, Seccombe D and Levin A. Standardization of creatinine measurement substantially improves GFR estimation. Nature Clinical Practice (Nephrology) (2008); 4(3):120.

David W. Seccombe, Beryl E. Jacobson and Amy Ho. A comparison of Proficiency Testing (PT) Data from a Fresh Human Serum Lipid PT program for the direct measurement of LDL, Apo A-1, Apo B. The Fats of Life; 21:22-24, 2007.

Mattman A, Eintracht S, Mock T, Schick G, Seccombe DW, Hurley RM, White CT. Estimating pediatric glomerular filtration rates in the era of chronic kidney disease staging. J Am Soc Nephrol. 2006 Feb;17(2):487-96.

Haag MD, Kelly JR, Ho A, Seccombe DW. A study to examine the accuracy of potassium measurements in clinical laboratories across Canada. Clin Biochem. 2000 Aug;33(6):449-56.

McGuinness C, Seccombe DW, Frohlich JJ, Ehnholm C, Sundvall J, Steiner G. Laboratory standardization of a large international clinical trial: the DAIS experience. DAIS Project Group. Diabetes Atherosclerosis Intervention Study. Clin Biochem. 2000 Feb;33(1):15-24.

Mary M. Kimberly, Elizabeth T Leary, Thomas G Cole and Parvin Waymack for the Cholesterol Reference Method Laboratory Network. Selection, validation, standardization, and performance of a designated comparison method for HDL-cholesterol for use in the Cholesterol Reference Method Laboratory Network. Clin Chem 1999; 45:1803-12.

Seccombe DW, Tholen DW, Jacobson BE. Standardization of creatinine: a pre-requisite for implementing the MDRD formula for the estimation of glomerular filtration rate (eGFR). Poster - XIX International Congress of Clinical Chemistry IFCC/AACC Annual Meeting, Orlando, 2005.

Levin A, Seccombe DW, Jacobson BE. A provincial program for standardization of creatinine measurements and the reporting of estimated GFR (eGFR). Poster - XIX International Congress of Clinical Chemistry IFCC/AACC Annual Meeting, Orlando, 2005. (NACB Distinguished Abstract Award).