Improved Monitoring of Differences in Serial Laboratory Results

Callum G. Fraser

The results of investigations in laboratory medicine are used for many purposes. Most are performed for monitoring individuals in acute situations and following the improvement or deterioration of chronic disease. In diagnosis, case-finding, and screening, conventional population-based reference intervals are relevant, and they are partitioned if necessary according to age, sex, or other important characteristics. In spite of the advantages of fixed decision limits, these intervals remain the mainstay of interpreting numerical results. Such tools are of very limited use, however, in evaluating serial results obtained for an individual. Within-individual biological variation is well known to be much smaller than between-individual variation for nearly all substances assayed in laboratory medicine (1). Each individual has a range of values that span only a part of the reference interval. In consequence, individuals can have important changes in results when they all lie within the reference interval. Such changes will usually be considered unremarkable by both laboratory professionals and clinicians, and thus ignored. In addition, results can change from inside to outside the interval (and vice versa) without having clinical importance. Laboratories conventionally flag results outside the reference limits, probably provoking some unnecessary follow-up activity, if only a repetition of the investigation (2).

Katzmann and coworkers in the current issue of the Journal have tackled the difficult problem of establishing criteria for monitoring changes in monoclonal protein concentrations (3). For some laboratory investigations, expert groups have recommended numerical criteria for interpreting changes. For the abnormal proteins in serum and urine investigated by Katzmann and colleagues (3), the guidelines from such expert groups state that reductions in the monoclonal protein in serum of at least 25% and 50% are considered minimal and partial responses, respectively, and the corresponding responses for urine monoclonal protein require at least 50% and 90% decreases (4, 5). As for many guidelines throughout medicine on the use of numerical data, these recommendations are round numbers. Guidelines on the imprecision required for investigations are replete with such numbers as 3%, 5%, 10%, and the like. Other examples from guidelines on clinical monitoring are statements that changes in serum troponin concentration of 20% (or 30%) are clinically important (6). Although such round-number guidelines may be based on considerable professional experience and may be easy to remember, they seem empirical. More objective strategies are required for interpreting test results in monitoring.

The use of the reference change value (RCV)2 has been advocated as a most appropriate tool for monitoring individuals (7). This strategy is based on the thesis that to decide whether test results have changed in an important way (perhaps informing the clinician that the individual is improving or deteriorating), they have to have changed by more than the difference expected from the inherent sources of variation. These sources are preanalytical variation, analytical imprecision (CV A), and within-individual biological imprecision (CVI). If preanalytical variation has been minimized (2), then the RCV can be calculated as: RCV = $2^{1/2}$ × Z × (CV A + CVI)$^{1/2}$, where Z is the number of SDs appropriate to the probability. Thus, RCVs are very simple to derive for commonly performed investigations, because every laboratory should know CV A from internal QC programs and CVI data are available (8). An additional argument for the use of the RCV in monitoring is that, for many analytes, the estimates of CVI are constant over time, geography, and analytical method (7), and in health and chronic, stable disease (9); thus, RCVs can be used by all, everywhere.

There are disadvantages to this approach, however, and they have been discussed recently by Cooper and colleagues (10). These disadvantages include the possibilities that (a) statistical information may overwhelm clinicians, (b) use of the Z score may deny clinical judgment, (c) RCV may be dependent on test frequency, (d) some biological variation may be depen-

---

2 Nonstandard abbreviations: RCV, reference change value; CV A, analytical imprecision; CVI, within-individual imprecision; M-spike, monoclonal protein spike.
dent on health status, (e) proper application requires a sophisticated laboratory information system, (f) education of laboratory staff and clinicians may be needed, and (g) terminology may be confusing.

Some of these problematic issues have been addressed by Katzmann and colleagues (3). They examined the analytical and biological variation of monoclonal protein spikes (M-spikes), urine M-spikes, and monoclonal serum free light chains. Monoclonal proteins do not exist in healthy people, and this study examined serial data from clinically stable patients with monoclonal gammapathies and then derived the RCVs in order to examine the validity of professional guidelines (4, 5). This approach is rarely done, but there would be no other means of examining the CV of these analytes except in a cohort of individuals with disease. To obtain good estimates requires that these analytes be stable. Although recognizing the possibility that some variation due to changes in clinical condition could confound the estimates of CV, Katzmann and colleagues (3) wisely applied preset exclusion criteria to minimize this possibility.

This approach of using patients with stable disease as the reference sample group for generating RCVs of clinical utility has previously been used, but rarely. Although the results of many investigations are well known to vary with age and given that childhood and puberty may be periods of important changes, it would probably be difficult, if not unethical, to undertake traditional studies of biological variation with healthy children. Desmeules et al. (11) estimated the biological variation and RCVs for glycohemoglobin in children with cystic fibrosis who did not have evidence of diabetes mellitus. Similarly, it would be difficult to perform experimental work in healthy individuals for investigations of such variables as arterial pH, gases, and electrolytes. Cembrowski et al. (12) demonstrated that for many investigations done regularly in the intensive care unit, only patient results needed to be statistically analyzed to provide reliable estimates of CV. These examples show that it is not necessary to undertake the classic experimental work and statistical analyses, as documented by Fraser and Harris (13), for all analytes, although that might be the ideal. Flair and imagination can be used, however, to identify reference populations with available test results that could well be used to determine CV and then, together with estimates of CV generate clinically useful RCVs.

Katzmann and colleagues (3) note that the guidelines for monitoring monoclonal gammapathies have proposed that decreases in serum and urine analytes are to be evaluated to assess response. Although both the laboratory medicine and clinical communities are generally rather less than careful with the way words are used, semantics is very important in clinical decision-making and the way that RCVs are calculated. The RCV formula presented above and most publications on the generation and application of data for the analytical and biological components of variation have documented that the Z score to be used is 1.96 for P values <0.05; sometimes, a Z score of 2.58 is used for P values <0.01. It is vital to note, however, that these Z scores are 2-sided and can be used only when both an increase and a decrease are being considered—in other words, any change. If the real clinical requirement is the evaluation of a significant decrease or reduction, then 1-sided Z scores must be used: 1.65 for P values <0.05 and 2.33 for P values <0.01. This correct usage has been performed by Katzmann and colleagues (3).

Another example is the evaluation of serum troponin concentrations, and although “change” is the term most often used in the literature, what is usually actually being assessed clinically is the probability of an increase, not any change. Again, application of a 1-sided Z score is warranted (6). A better understanding of the clinical decision-making context and the marked difference in meaning between a change and an increase or decrease is required for the correct calculation of appropriate RCVs.

In addition, Katzmann and colleagues (3) provide an interesting table of data on the percentage decrease needed for statistical significance at various probability thresholds for each analyte studied. It is very easy to calculate the probability that any change (increase or decrease) has occurred by using a simple rearrangement of the RCV formula to make the Z score (and thus the probability) the unknown, namely: $Z = \frac{\text{change}}{\sqrt{\frac{\text{CV}_{\text{analyte}}}{{\text{P}}^2} + \frac{\text{CV}_{\text{analyte}}}{\text{P}^2}}}$ [1.645]. Not all decisions in medicine are made at statistically significant (P < 0.05) or highly significant (P < 0.01) levels of probability. Knowledge of CV and CV, and this equation can be used to create value-added information for the user of test results. A tabular approach as presented (3) might indeed be useful. A perhaps more user-friendly graphical approach to presenting probability vs change has also been used (7). This simple aid to interpretation has been adopted in the study of glycohemoglobin variation in children (11). A similar approach has been used in a very recent study on RCVs for dehydration markers (14). In addition to providing probabilities appropriate for change on the abscissas of the graphs of probability vs change for plasma osmolality, urine specific gravity, and body mass, this approach gave semantic interpretative anchors, namely that change was likely at P > 0.80, more likely at P > 0.90, very likely at P > 0.95, and virtually certain at P > 0.99.

Approaches such as those used by Katzmann and colleagues (3) negate some of the cogent concerns expressed about RCV (10). Others are encouraged to emulate and expand their work, especially on important
analytes for which the derivation of traditional estimates of biological variation is problematic.

**Author Contributions:** All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

**Authors’ Disclosures or Potential Conflicts of Interest:** No authors declared any potential conflicts of interest.

**References**