

Reference Intervals: Strengths, Weaknesses and Challenges

Moderators: W. Greg Miller^{1*} and Gary L. Horowitz²

Experts: Ferruccio Ceriotti,³ James K. Fleming,⁴ Neil Greenberg,⁵ Alexander Katayev,⁶ Graham R.D. Jones,⁷ William Rosner,⁸ and Ian S. Young⁹

A reference interval (RI)¹⁰ is a standard component of reporting a laboratory result and is important to transform a numerical value into clinically meaningful information. An RI is intended to inform the clinical care provider that laboratory values within the interval indicate a nondiseased condition. The most common approach is to base an RI on the central 95% of laboratory test values observed for a reference population that is free of diseases that influence that laboratory test result. Because many diseases are asymptomatic, it becomes difficult to qualify people for a nondiseased condition, thus biasing the selection of reference individuals. Furthermore, information on the full complement of disease conditions that influence a laboratory test may be unknown. Thus, RIs may be influenced by inappropriately selected reference populations.

Another limitation in determining an RI is obtaining an adequate sample of a reference population to make an estimate of the central 95% of results with suitable uncertainty to be meaningful for interpreting a test result. The sample size requirement becomes even larger when partitioning by sex, age, ethnicity, menstrual cycle, and other parameters is necessary for meaningful RIs. CLSI Guideline EP28-A3c describes consensus approaches and some limitations for establishing and verifying RIs. However, some of the approaches in this guideline are statistically underpowered such that uncertainties in the RIs may not be appreciated.

A particularly challenging situation is the requirement for laboratories to establish an RI for a laboratory developed measurement procedure (MP) or to verify the RIs pro-

posed by the manufacturer of an in vitro diagnostic (IVD) MP. Identifying a suitable reference population for either requirement can be very challenging.

Evolution in laboratory practice is needed to enable appropriate RIs to be adopted by laboratories. We asked experts with different perspectives to address issues we all face in establishing or verifying RIs.

How can an IVD device manufacturer get a suitable sample of a reference population to establish an RI?



Ferruccio Ceriotti: “The RIs for the quantities being determined” are required from an IVD manufacturer to comply with the European Directive on in vitro diagnostic medical devices. Moreover, the RIs provided by the manufacturer are probably those most commonly used on clinical

laboratory reports. There is not a single answer to this critical question, it depends on the measurand. If a reference system exists for calibration traceability and there are published well-characterized RIs obtained with a well-standardized MP (“traceable RIs”), the manufacturer can provide those RIs without the need of a new RI study provided that its MP has the same level of standardization. For nonstandardized measurands, I do not see alternatives to an RI study performed according to CLSI

¹ Department of Pathology, Virginia Commonwealth University, Richmond, VA; ² Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA; ³ Former Chair of the IFCC Committee on Reference Intervals and Decision Limits (2006–2010) and collaborator in the CLSI document EP 28-A3, and Clinical Laboratory Service, San Raffaele Hospital, Milano, Italy; ⁴ Vice President and Director, Department of Science and Technology, Laboratory Corporation of America Holdings, Elon, NC; ⁵ Principal Consultant, Neil Greenberg Consulting, LLC, Rochester, NY, and Convenor, ISO/TC12 Working Group 2; ⁶ Associate Vice President and Director, Clinical Science Assessment, Department of Science and Technology, Laboratory Corporation of America Holdings, Elon, NC; ⁷ Chemical Pathologist, SydPath, Department of Chemical Pathology, St. Vincent's Hospital, Sydney, and Associate Professor, Faculty of Medicine, University of New South Wales, Sydney, Australia; ⁸ Professor of Medicine, Department of Medicine, Columbia

University, College of Physicians and Surgeons, New York, NY; ⁹ Professor of Medicine, Department of Medicine, Queen's University, Belfast, Wellcome Research Laboratories Mulhouse Building, Royal Victoria Hospital, Belfast, UK.

* Address correspondence to this author at: Virginia Commonwealth University, 403 N. 13th St., Rm. 501, Richmond, VA 23298. Fax 804-828-0353; e-mail greg.miller@vcuhealth.org.

Received April 28, 2016; accepted May 3, 2016.

© 2016 American Association for Clinical Chemistry

¹⁰ Nonstandard abbreviations: RI, reference interval; IVD, in vitro diagnostic; MP, measurement procedure, FDA, US Food and Drug Administration; RCV, reference change value.

EP28-A3c in collaboration with an academic or clinical center.



James K. Fleming & Alexander Katayev: Many quality-oriented IVD manufacturers are already performing RI studies using statistically significant numbers of reference subjects. It should be a responsibility of the US Food and Drug Administration (FDA) and other regulatory bodies to require IVD manufacturers to perform statistically powered RI studies and deny approvals if those studies are inadequate. It is also important that regulatory agencies require not only studies performed using disease-free subjects, but RI studies that include subjects with predefined conditions like pregnancy, diabetes, kidney disease, etc. that may affect the test result interpretation for those categories of patients. The common practice of literature reference citations should be avoided due to lack of standardization and MP comparability data, and unknown statistical power and assumptions used in those studies. Obtaining a suitable sample size for RI studies could potentially involve a collaborative effort between contract research organizations, biobanks, and medical societies.



Neil Greenberg: Obtaining suitable samples from a well-documented reference population for purposes of establishing an RI is challenging even in the most comprehensive clinical practice setting, such as an academic medical center. For an IVD manufacturer, often the best solution is to collaborate

with a clinical laboratory that has access to adequate numbers of samples from individuals fulfilling the specifications for the predetermined clinical and demographic characteristics of a reference population for the intended measurand. Depending on the sampling char-

acteristics that may be important (e.g., age, sex, race/ethnicity, disease history), the collaborating laboratory may also need to collect a number of different sample sets to establish specific ranges for multiple demographic and/or clinical reference populations.



Graham R.D. Jones: IVD device manufacturers face the same issues as any organization setting out to establish an RI, however, with greater responsibility since manufacturer provided RIs may be used in many laboratories. There are two important aspects. The first is doing the job well,

i.e., selecting an appropriate population, with age and sex ranges as needed for the analyte, managing the preanalytical and analytical issues correctly and then performing appropriate statistical analysis to derive the intervals. The second aspect, which is commonly not performed well, is to provide the end user with sufficient information about the population, the preanalytical and analytical processes, and the statistical processes so that the user can assess the suitability of the interval for use in their own laboratory. The uncertainties of the upper and lower reference limits should be provided to allow the user to compare the intervals with data from other sources.



William Rosner: Although it is in their best interests to do so, manufacturers probably are not ideally equipped to establish RIs properly, both in terms of the numbers required and the difficulty of deciding who is “normal.” Perhaps the best approach is to collaborate (both in defining and obtaining the reference population) with thought leaders to include statisticians as well as laboratorians and clinicians. A possible alternative would be to hire a contract firm with appropriate expertise to define and obtain the samples, e.g., such as the ones that supervise clinical trials. Both these solutions would be expensive and not trivial to initiate, but would yield the most reliable values. Of course, the usefulness of RIs would be greatly enhanced if MPs were harmonized and traceable to the same standard.



Ian S. Young: If the intention is to get a suitable sample to meet regulatory requirements, the easiest approach is probably to contract the selection of subjects and preanalytical aspects out to an appropriate third party supplier. More than one sample collection may be required for a manufacturer who is

marketing an MP globally. While a single source of reference individuals may be sufficient from a manufacturer perspective, it is of limited use to the end user clinical laboratory unless the manufacturer's selected sample is demographically similar to the patient population served by the clinical laboratory.

Is it necessary for every clinical laboratory to verify an RI included in an IVD device manufacturer's instructions for use? If so, what are practical approaches to use?

Ferruccio Ceriotti: Yes, in most cases each clinical laboratory needs to verify the RI. The only exception is the case, indicated previously, when the MP in use is able to provide traceable results and "traceable RIs." In this case, it is sufficient to evaluate the trueness of the MP. If traceable RIs exist, but the population served by the laboratory has different characteristics, then verification is needed.

Besides the approaches proposed by CLSI EP23-A3c, a possible practical approach could be the use of data mining and suitable programs to analyze the data. Care is needed to select data from presumably healthy individuals (e.g., blood donors, presurgical screening for elective surgery, workplace screening programs where the prevalence of healthy individuals should be high). Comparison between the obtained RIs and those proposed by the manufacturer can be used for verification.

The situation becomes very complicated for esoteric tests, where the number of nondiseased patients included could be low. In this case, the binomial test proposed by EP23-A3c is the only practical verification, even with all the inherent statistical limitations.

James K. Fleming & Alexander Katayev: It may not be necessary to require verification of the manufacturer's instructions for use claims for RIs if the corresponding MP used by the laboratory is proven to be standardized/harmonized with that of the manufacturer. In addition, MP comparability studies should meet quality performance criteria for bias and ensure that the populations served by the laboratory are comparable to those populations used for the manufacturer's studies, or any other

donor study source. However, it is usually more propitious to perform a smaller verification study when performance bias approaches the limits of desirability and/or when study samples and other resources are readily available.

Neil Greenberg: Depending on the extent and quality of the effort undertaken by the IVD manufacturer to establish the RI, the individual clinical laboratory's role in verification of the RI can range from minimal obligation up to the opposite extreme of conducting a full, independent, and comprehensive RI study. If the IVD manufacturer states that the claimed RI was established in actual use of their MP in accord with an accepted national or international consensus standard (e.g., CLSI Guideline EP28-A3c), the requirements for each clinical laboratory implementing the measurement procedure should be limited to verification of the accuracy of the measurement procedure in their hands, with no need to conduct an RI verification.

Accuracy verification can often be accomplished by use of value assigned quality control materials provided by the manufacturer of the MPs, with data from commutable external quality assessment (proficiency testing) materials, or with commutable certified reference materials or "trueness controls" available for the measurand.

Where the quality of information from the IVD manufacturer is less robust, additional effort will be required to either verify or independently establish suitable RIs. Often this assessment will be dictated by the clinical needs of the patients (and clinicians) being served by the laboratory. Consultation and dialog with medical practitioners/specialists are often a useful place to begin the assessment.

Graham R.D. Jones: RIs are commonly used to flag results for further attention by the treating physician. Consequently, incorrect intervals have the potential to draw attention to results where there is no need for an action, or to miss drawing attention to results that may be of clinical interest. While RIs may be a "blunt tool" for identifying clinically important disease, their widespread use and unifying principles make the influence of incorrect intervals important.

RIs are one of the ways we turn numbers into information (e.g., "this result is twice the upper reference limit"). All laboratories spend considerable effort ensuring accuracy of their results through MP verification, internal quality control, and external quality assessment. An incorrect RI can have the same effect on the information a laboratory supplies as a bias in the MP itself; consequently, it is vital to assess and verify the suitability of an RI. RIs can be suboptimal by being too wide, too narrow, biased high or low, or not properly matched for age, sex, or some other aspect of the population. A rela-

tively simple way to assess an RI as suitable for use is to analyze data from the laboratory looking at any bias of the distribution relative to the distribution on which the RI is based, and to assess the flagging rate above and below the interval. This assessment can be done matched for age and sex as needed. This process can be made more sophisticated by selection of a more healthy subpopulation such as patients attending general practice locations.

William Rosner: The need to verify an RI depends on the situation. We live in an imperfect world. If the measurand in question does not yield the same values for the same clinical samples across manufacturers' MPs (a common enough situation), the solution is different from that entailed with harmonized measurands.

If the measurand results have been harmonized by having calibration traceable to a common reference system, then subscription to an accuracy-based quality control or proficiency testing program to verify traceability would be sufficient. This begs the question of local differences in the population on which the RIs are based, but is not an unreasonable short cut.

If the MPs in use yield different, nonlinearly related values and accuracy-based testing is not possible, the problem is large and the approach would be different in different settings. Large reference laboratories could work with specific manufacturers of MPs to establish RIs, as indicated earlier. For smaller laboratories, the simplest solution would be to ascertain which reference laboratory is using the same MP that they are, and adopt their RI, ensuring the preanalytical conditions (and populations) are comparable.

Ian S. Young: The need to verify RIs is likely to vary depending on the accreditation standard to which a clinical laboratory operates. However, in my view it is, at the very least, good practice to do this. There are several possible approaches. One is to develop a laboratory sample bank from relatively healthy subjects which can be used to verify RIs. An alternative approach, which will become more common, is to make use of the large dataset which most laboratories generate. By ensuring the dataset consists of relatively healthy individuals and by excluding outlying results, an RI can be determined from the remaining very large number of values.

Many IVD device manufacturers use RIs published in textbook or literature reports even when the measurement procedure used to establish the RIs may not be known. Why does this practice persist? How should a clinical laboratory verify the suitability of such RIs for a particular measurement procedure?

Ferruccio Ceriotti: As previously stated, RIs from the literature can be used if the MP's calibration traceability

requirements are fulfilled. The situation may improve in the future because large RIs studies have been recently performed by some manufacturers. In addition, the Committee on RIs and Decision Limits of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC C-RIDL) is concluding a large worldwide study that includes different manufacturers.

A main limitation in applying the manufacturer's RIs is that frequently a small number of individuals were included in these studies and the population does not match the one served by the laboratory. These are the main reasons why verifying the suitability of the proposed RI is a must for any laboratory.

James K. Fleming & Alexander Katayev: Using textbook or literature RIs persists because of the lack of regulatory oversight where approvals of submissions are granted without properly conducted RIs. In such situations it is difficult to determine what MP was used and the population used in the reference citation study. When a laboratory perform its own RI study, it should satisfy the established quality requirements, or at least verify that the proposed intervals are applicable to the laboratory's own patient's population and MPs.

Neil Greenberg: Because the performance of comprehensive de novo RI studies can be quite costly, some developers of new MPs prefer to quote published RIs from other sources, with the implication that the published RIs are applicable to their new MPs. Such RIs should be accepted by end users in clinical laboratories only under circumstances where the new MP has been demonstrated (preferably by the developers of the measurement procedure) to have calibration traceability and specificity that is equivalent to that of the MP used to establish the published RI.

Graham R.D. Jones: The main concern I have with the practice to provide literature-based RIs is that there is usually inadequate information provided, making it difficult for the laboratory to assess the quality of the supplied RI. There is obviously a clear need for a statement about any possible MP bias relative to that used to derive the RI as well as descriptive information about the population. In the same way that laboratories must verify an RI (and be ready to supply that information if requested), it seems unacceptable for a manufacturer not to undertake at least this level of verification.

William Rosner: The practice to provide literature-based RIs persists because of: inertia, e.g., getting over the energy hump to change, the cost of changing, the difficulty of promulgating and getting acceptance of changes, and the basic difficulty of establishing traceability and harmonization.

Ian S. Young: The practice of providing literature-based RIs persists because it is relatively easy and inexpensive for manufacturers, and because it is tolerated by regulators and users. As for verifying the suitability of such RIs, see response to the second question.

Activities are underway by several professional organizations to establish common RIs suitable for use by different measurement procedures. What are the strengths and limitations of common RIs?

Ferruccio Ceriotti: The concept of “common RIs” is based on three assumptions: *i*) use of a well-controlled MP able to provide results traceable to a reference measurement system; *ii*) existence of RIs obtained with such a MP; and *iii*) independence of the RIs from the type of population (or availability of suitable RIs for the different populations). The first condition should be true for many frequently ordered clinical chemistry tests and also for the frequently ordered hematological tests. However, even if traceability to a reference measurement system is theoretically present, the practical realization of the traceability principle is often quite poor. In addition, the other two conditions are not easily achievable.

If the aforementioned conditions are realized, the main strength of the common RI approach is to provide harmonized and comparable clinical interpretations regardless of where and when a result was generated. They improve the safety for the patient by minimizing the risk of erroneous evaluation by a clinician due to different RIs from different laboratories. Common RIs allow application of common guidelines and can be applied by the laboratory without the need of verification even when changing the MP, provided the new MP has the same level of standardization that can be easily verified with a simple MP comparison.

James K. Fleming & Alexander Katayev: Standardization and harmonization of MPs across different manufacturers and platforms are paramount for establishing common RIs. Multicenter laboratory networks may then be able to conduct RI studies that are of the highest quality and would apply to different, but standardized/harmonized, methods across different instruments. This will be the ultimate solution for the old and controversial RI dilemma. It is important to understand that common RIs may not necessarily be “one size fits all,” but be specifically established for the given ethnic or geographic groups or populations of subjects with predefined conditions. Nevertheless, they would be “common” for use in those populations across all laboratories that serve those populations. This approach has been successfully implemented in several countries and is in use by our laboratory network that is standardized to the same MPs and reagent systems.

Neil Greenberg: The problem of common/shared RIs is closely related to the problem of MP harmonization. As long as the various MPs intended to be included in an exercise for pooling RI data are harmonized from the standpoint of calibration traceability, this approach can theoretically be useful. However, it is incumbent upon the sponsors of these studies to ensure verification of the accuracy of the MPs included in the project are consistent with, representative of, and fully sustainable and supported by the claims of the manufacturers/developers of the measurement procedures being included.

Graham R.D. Jones: When surveys on RIs are undertaken a number of themes commonly arise. These include the use of “historical” intervals derived many years ago for which the source is not available, the uncritical use of manufacturer’s intervals, and intervals based on underpowered studies. When RIs in use are compared with MP performance, there is often little or no correlation between among-method MP bias and differences in the RIs. Thus the between-laboratory variation in RIs increases the noise in the information supplied by laboratories rather than reducing it.

The practice of setting common RIs also has the advantages of sharing the workload for establishing the intervals and brings more minds and more data to the process. Setting an RI can be seen as analogous to identifying truth based on experimental science. The more data we have supporting a conclusion the more confident we are in the result. When properly done, common intervals are likely to be of a higher quality than laboratory-specific intervals.

William Rosner: If common RIs include those appropriately corrected for age, sex, ethnicity, cyclicity of the measurand, etc., then their strength cannot be overstated. Normal would be normal from laboratory to laboratory, from country to country, and from time to time. The major dangers of common RIs would be a disregard of preanalytical handling of samples, and possible differences between the reference population and the local population served by a laboratory.

Ian S. Young: The idea of common RIs is attractive in a world where patients are mobile and increasingly obtain healthcare information from globally available information sources, and where guidelines with fixed decision limits are widely used. However, in my opinion common RIs can only be used where between-MP differences are modest (<10%, perhaps) and where there are not significant between-population differences as a consequence of genetic factors (including ethnicity), demographics, or environment. I think that these conditions are rarely met at present. When there are significant differences between MPs, a common RI needs to be unrealistically

wide to embrace results from all MPs, and can disguise the fact that methodological differences exist.

An RI may be used to compensate for lack of harmonization among results from different measurement procedures for the same measurand. What are the issues that face clinical care givers in interpreting laboratory results that have different RIs?

Ferruccio Ceriotti: MP-dependent results require MP-dependent RIs. These differences are also a patient safety issue; therefore standardization or harmonization of the results of different MPs is really an ethical imperative for our profession of clinical laboratory specialists.

James K. Fleming & Alexander Katayev: There are drawbacks to MP-specific RIs. First, physicians have to keep careful track of what laboratories/MPs are currently in use for their patients and adjust their decision-making process accordingly. Second, there is a relatively high chance for a medical decision error, especially when monitoring the same patient while using different laboratories or MPs.

Neil Greenberg: This scenario reflects the reality for a large portion of the measurands in laboratory medicine, where there is a lack of harmonization among MPs. Lack of harmonization can lead to dissociation between the values reported by a particular laboratory and any available clinical practice guidelines with regards to interpretation and appropriate clinical actions to be followed. Clinicians dealing with results for these types of measurands are commonly challenged to correctly interpret laboratory values in the face of different RIs for the same measurand, leading to higher risk for misinterpretations and diagnostic errors. Finding practical solutions to this situation is among the highest of priorities for the laboratory medicine profession.

Graham R.D. Jones: It is my experience that physicians commonly assume that differences in RIs between laboratories is an indicator that the results are similarly different, i.e., that the information supplied by the combination of results and interval is the same from different laboratories. When MP results are truly different, clearly a different RI is required. This process can be made clearer either by providing educational support, e.g., with a footnote or modified test name, and by avoiding unnecessary differences in RIs where the results are indeed comparable. In circumstances where a different RI is required, steps should be taken to ensure these results are not combined in electronic health records.

Ian S. Young: Caregivers need to be aware of local RIs and any related decision limits. This can be problematic

when caregivers move between facilities, or in the case of a mobile patient who has samples analyzed in different laboratories and who takes the results with them to another care provider. Misinterpretation of a result in the context of an inappropriate RI may give rise to a wrong clinical decision. For example, the Kidney Disease Improving Global Outcomes guidelines recommend decision limits based on a multiple of the upper RI for parathyroid hormone. If the wrong RI is used, an incorrect decision may be made.

How can RIs be derived from clinical outcomes rather than from a statistical distribution of results for a reference population? Is there a role for each approach?

Ferruccio Ceriotti: There is a terminology problem in this question. We can obtain outcome based references, but these are not “RIs” and we have to call them “decision limits.” RIs are a biological characteristic of a defined population and a specific statistical approach is used for their definition. Decision limits, in contrast, depend on the clinical outcome and define if a subject needs a specific medical intervention, so they may depend on the clinical condition or on the type of disease or clinical intervention, and the influence of other variables like age and sex can be absent or secondary. Theoretically, decision limits are what we really need and RIs are just a surrogate for them, but the definition of a decision limit implies clinical outcome studies and its use requires the fulfillment of the same high-level analytical requirements defined earlier for the use of common RIs. For these reasons, decision limits are not yet widely used. I imagine that in the future decision limits will gradually replace the present RIs.

James K. Fleming & Alexander Katayev: It is well accepted that setting analytical quality goals based on clinical outcomes may be one of the best mechanisms to establish those goals. However, it is also noted to be one of the most difficult approaches in day-to-day practice, especially in the referral laboratory setting where there are little or no data available on clinical outcomes. Deriving RIs (or medical decision points) using clinical outcomes may be easily accomplished in conjunction with the same data analysis as for quality goals. Linking laboratory test results with clinical outcomes to establish risk-based cutoffs has been reported in a number of recent publications, and has been successfully used in several large academic medical centers. Both the clinical outcomes/risk-based approach and the reference population-based approach have their respective roles in the RI establishment process.

Neil Greenberg: The conventional approach to derivation of RIs using reference populations is of course flawed

in that the use of predetermined and sometimes arbitrary selection criteria for the reference population will never completely rule out disease in the so-called nondiseased subjects. This approach has nevertheless served us well in laboratory medicine for a very long time. Use of clinical outcomes data in the derivation of RIs is potentially an elegant approach. This is essentially the approach that was taken for the establishment of clinical guidelines in the identification and management of patients for hyperlipidemia and in the diagnosis of Type 2 diabetes using hemoglobin A_{1c}. However, the clinical outcomes approach is usually much more costly and time-consuming, and meaningful data are often available only as part of a long-term prospective study of large populations. There is clearly a place for both approaches in the lifecycle of diagnostic tests, with the conventional reference population approach being the usual starting point for RIs, ultimately to be updated subject to the availability over time of outcomes data based on comprehensive prospective clinical studies.

Graham R.D. Jones: The starting point with this question is one of terminology addressing the difference between a population RI and a clinical decision point. In general, a clinical decision point should be a more powerful tool than an RI since it will have been set with appropriate clinical sensitivity and specificity for a defined clinical decision. A further refinement is where a decision point has been validated not just to predict an outcome, but to predict benefit from a specified treatment. Limitations to clinical decision points include the lack of available robust data and that they are commonly related to a single clinical scenario. By contrast, a population RI can be of value for analytes that are affected by a range of different pathologies. Another difference between clinical decision points and RIs is that the latter can, and should, be verified locally. By contrast, clinical decision points based on longitudinal or interventional studies can rarely be repeated locally, which places a higher importance for laboratories in ensuring traceability of their MPs relative to the methods used in the clinical trials.

William Rosner: The good clinician adjusts diagnostic judgments arising from statistical RIs based on knowledge of the significance of the observed deviation from normal in the context of the clinical situation. At the end of the day, the only thing important about abnormal laboratory values is their relationship to clinical outcomes. But a context is necessary when instituting such RIs. In the absence of some information about the distribution of values, it would be difficult to judge what constitutes a minor, a moderate, or a serious risk.

Ian S. Young: Large epidemiological studies can be used to establish the relationship between values of a measurand and outcomes. A good analogy might be body mass index, where

19–25 kg/m² is considered ideal because it is associated with lowest overall mortality. A similar approach could be taken and used to derive an RI for a laboratory test. This would differ significantly from the traditional RI, but arguably might have greater clinical utility.

In the UK, the consensus common RI for serum sodium, agreed through the pathology harmonization process, is 133–146 mmol/L. However, some international guidelines define hyponatremia as serum Na <135 mmol/L, overlapping with the RI. This overlap potentially causes confusion for clinicians and has led to at least one region declining to adopt the harmonized RI due to legal concerns. In this case, an outcome-based RI would make more sense.

Laboratory reports typically highlight results that are “abnormal,” but they rarely indicate the degree of abnormality. For example, given reference limits of 8.4–10.2 mg/dL (2.10–2.55 mmol/L) for total calcium, a report will highlight both 10.3 mg/dL and 11.4 mg/dL (2.57 mmol/L and 2.84 mmol/L) as abnormally high. As a result, clinicians may, in both cases, order a series of additional laboratory tests and other diagnostic studies. What can laboratories do to help clinicians not overreact to a value just outside, but close to, a reference limit?

Ferruccio Ceriotti: The problem can be seen also from the other side: underreacting to a value just inside, but close to the limit. Besides generating additional laboratory tests, results that are highlighted as abnormal may produce a great and unjustified patient anxiety.

A drastic approach could be eliminating any highlighting system that would oblige the clinician to look more carefully at the laboratory results. But I am not in favor of such a drastic position because it puts the patient in danger if the clinician misses some relevant abnormal values. One possibility would be to use different symbols, e.g., a single asterisk for the values close to the limit and a double asterisk for the highly pathological. A second possibility is a graphical representation of the value, maybe with the grey zone of the measurement uncertainty, in relation to the RI. A third possibility is an automatic comment occurring in any situation in which the uncertainty around the result value overlaps the upper or lower reference limit or the decision limit.

A long-term solution is education starting with the medical students. Education on the theory of RIs, that they include only the central 95% of the results of the reference population and that there is a confidence interval around each limit. In addition, biological variation and measurement uncertainty influence interpretation of laboratory results.

James K. Fleming & Alexander Katayev: Clinicians

frequently view laboratory test results in terms of a high or low. This is directly tied to the historical paradigm of how test results have been reported by laboratories for decades. Laboratorians and astute clinicians know the fallacy of such reporting. Some laboratories have attempted to change the paradigm through the use of colorful graphic displays of RIs while inserting the test value within the linear graph. Others have attempted to show the current test result against the previous test result. However, most have ignored the variability inherent in the measurement itself. Any test result is a product of biologic variability, the analytical bias, and the analytical imprecision, i.e., the reported result is actually a subrange of values within and/or outside the continuum of the RI. Callum Fraser and colleagues have effectively applied these principles into the concept of reporting the reference change value (RCV), which our laboratory is about to pilot. We are developing a graphical representation of the RI in a linear fashion and will display the RCV range around the test result. Such a visual reference should provide an indication that not all abnormal test results are “abnormal” and not all normal test results are “normal.” Ideally, we would like to get to the point where we are looking more at the test result in the context of the patient’s biology and their homeostatic set points as opposed to a number referenced to a larger population. Reporting RCV should get us a little closer to the concept of “individuality” for each measurand’s variability and the understanding that many test results should not be evaluated against population-based RIs, but rather against the individual patient’s homeostatic thresholds.

Neil Greenberg: The issue regarding values near the limits of an RI is confounded by the question of what is “normal” for the individual patient vs what comprises the distribution of values observed in a population of “normal” subjects. What is meaningful for the individual in many cases may have little to do with how their laboratory value compares to values in other individuals, but rather how their value compares to their own personalized baseline value for the measurand. While a calcium value of 10.3 mg/dL (2.57 mmol/L) may appear to be very close to the high end of the distribution of values in a normal population, if that patient’s baseline or steady state value is 9.2 mg/dL (2.30 mmol/L), the clinician needs to consider that there may be a significant change in this particular patient in light of a new observed value of 10.3 mg/dL (2.57 mmol/L), regardless of where the population RI ends. I would rather see effort expended on developing clinical guidelines for what constitutes a significant change (δ value) in the laboratory values for individuals.

Graham R.D. Jones: This “problem” is caused by a number of factors. These include the common use of the

central 95th percentile, excluding 2.5% of the healthy or reference population at both ends. This practice causes 5% of results from the reference population to be flagged. A different percentile could be used, for example ± 3 SD, which would reduce the flagging but also reduce the sensitivity to identify results different from a reference population.

A benefit of the RI concept is that the same approach is applied to all results. However, there are limitations to the concept of “treating all results the same.” For example, a serum calcium value just above the RI has a reasonable chance of representing disease (e.g., hyperparathyroidism) but a serum lipase may need to be several times the upper reference limit to indicate a high clinical likelihood of disease (e.g., pancreatitis). A rule of thumb may be that a result close to the reference limits, from patients without a risk of relevant disease, has a high likelihood of representing a “normal” result. Pathology reports may be able to assist with variable flags for the degree of abnormality or clinical importance, or a graphical demonstration of these factors. Difficulties are conveying the meaning behind the additional flags, and deciding where to set them to support this meaning.

Ian S. Young: There is little doubt that clinicians often place undue weight on test values just outside an RI. They tend to view these values as “abnormal” and to forget that around one result in twenty by chance will be outside the RI. If the uncertainty of a result was included in a test report it would tend to reduce over-investigation in these circumstances, though considerable education of users would be required to assist them in interpreting the more complex reports.

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: Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:

Employment or Leadership: W.G. Miller, CLSI and *Clinical Chemistry*, AACC; G.L. Horowitz, *Clinical Chemistry*, AACC; J.K. Fleming, Laboratory Corporation of America, Holdings; I.S. Young, IFCC, ACB (UK), and *Clinical Chemistry*, AACC.

Consultant or Advisory Role: None declared.

Stock Ownership: None declared.

Honoraria: None declared.

Research Funding: None declared.

Expert Testimony: None declared.

Patents: None declared.

Previously published online at DOI: 10.1373/clinchem.2016.256511