

Establishing Reference Intervals for Clinical Laboratory Test Results

Is There a Better Way?

Alex Katayev, MD,¹ Claudiu Balciza,² and David W. Seccombe, MD, PhD³

Key Words: Reference; Interval; Estimation; Laboratory; Test; Result; Interpretation

DOI: 10.1309/AJCPN5BMTSF1CDYP

Abstract

Reference intervals are essential for clinical laboratory test interpretation and patient care. Methods for estimating them are expensive, difficult to perform, often inaccurate, and nonreproducible. A computerized indirect Hoffmann method was studied for accuracy and reproducibility. The study used data collected retrospectively for 5 analytes without exclusions and filtering from a nationwide chain of clinical reference laboratories in the United States. The accuracy was assessed by the comparability of reference intervals as calculated by the new method with published peer-reviewed studies, and reproducibility was assessed by the comparability of 2 sets of reference intervals derived from 2 different data sets. There was no statistically significant difference between the calculated and published reference intervals or between the 2 sets of intervals that were derived from different data sets. A computerized Hoffmann method for indirect estimation of reference intervals using stored test results is proved to be accurate and reproducible.

It is hard to underestimate the importance of clinical laboratory test results. Nearly 80% of physicians' medical decisions are based on information provided by laboratory reports.¹ A test result by itself is of little value unless it is reported with the appropriate information for its interpretation. Typically, this information is provided in the form of a reference interval (RI) or medical decision limit. An RI as defined by Ceriotti "is an interval that, when applied to the population serviced by the laboratory correctly includes most of the subjects with characteristics similar to the reference group and excludes the others."^{2(p115)} No RI is completely "right" or "wrong." The majority of RIs in use today refer to the central 95% of the reference population of subjects. By definition, 5% of all results from "healthy" people will fall outside of the reported RI and, as such, will be flagged as being "abnormal."

There are many problems associated with the calculation of RI. The latest edition of the Clinical and Laboratory Standards Institute–approved guideline, "Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory," recognizes the difficulties and controversies surrounding the establishment of RIs and the verification process: "...the working group recognizes the reality that, in practice, very few laboratories perform their own reference interval studies," "...instead of performing a new reference interval study, laboratories and manufacturers refer to studies done many decades ago, when both the methods and the population were very different."^{3(p1)}

It has been recommended that an RI be established by selecting a statistically sufficient group (a minimum of 120) of healthy reference subjects. However, it is noted in the guideline that "Health is a relative condition lacking a universal definition. Defining what is considered healthy becomes the

initial problem in any study....”^{3(p8)} In reality, there will always be some level of uncertainty with a given selection protocol not only because of the definition of health that was selected but also because of the very real possibility that some of the selected subjects may, in fact, have subclinical disease.

Recruiting a valid group of reference subjects and obtaining informed consent in today’s environment is costly, time-intensive, and virtually an impossible task for most laboratories. The challenge is further magnified in establishing RIs for different age groups (eg, pediatric patients and geriatric patients), uncommon sample types (eg, cerebrospinal fluid and aspirations), timed collections, challenge tests, and serial measurements.

In light of these difficulties, most laboratories elect not to establish their own RIs, but rather choose to verify RIs that have been reported by the manufacturer or as established by another laboratory. This is a relatively simple study and requires only 20 healthy subjects to recruit.³ The underlying assumption here is that the laboratory analytic system is calibrated and producing similar results as the method that was originally used in the published RIs. However, this may not be true because in many cases, the details of the reference study such as its design, the inclusion and exclusion criteria used for selecting the healthy recruits, preanalytic sources of variation, etc, are lacking.

A laboratory can elect to “transfer” the RIs that were in use with an older method (or from another laboratory) to a new method. To do this, the laboratory must first demonstrate that the 2 methods produce comparable results. It is well known that analytic systems drift over time, and there is no guarantee that the method of today is producing results that are comparable to those that were produced at the time of the original RI study. This technique is the main reason why many laboratories today are using RIs that were established decades ago and are out-of-date.

Even if a laboratory was able to obtain a rigorously selected statistically sufficient number of healthy subjects and perform all the necessary testing, the next step would require statistical analysis of data. What statistical technique should be used: parametric, transformed parametric, non-parametric, or many others described? The judgment in most cases will be made subjectively because there are no clear guidelines, and the resultant intervals will differ depending on the method used.

Laboratories are often faced with test data that exhibit a multimodal or an asymmetric distribution. This may reflect a large prevalence of subclinical disease within the selected population or subgroup-related differences in normal ranges. The latter requires partitioning of test subjects by sex, age, race, and other factors. Partitioning by sex is relatively easy (select a minimum of 120 males and 120 females). However, partitioning by age groups is not a simple matter. What age

cutoffs should be used? How many groups should be studied? There are some complex statistical techniques available, but none seem ideal for solving partitioning problems.³

The last major challenge is cost. In the modern environment when laboratories are struggling to stay profitable, not everyone is willing to budget the appropriate resources for a lengthy and expensive RI study.

An alternative approach for establishing RIs is to do an indirect so-called a posteriori study of the patient data already collected and stored in the laboratory database. This is appealing because the data are readily available and will result in time and cost savings. A number of publications discuss this approach.⁴⁻⁸ Most of these studies were able to report clinically relevant and meaningful RIs. All of them used various sophisticated filters to exclude results from “unhealthy” subjects, and some used data from hospital laboratories and some from outpatient care settings or noninstitutionalized population study databases. Most of these studies used complex statistical algorithms to derive the final intervals. However, current guidelines do not endorse these methods as a primary approach for establishing RIs, mainly out of concern for the fact that most of the data may not come from reference or healthy subjects.³ This position may be justified for test results collected from hospitalized patients but is questionable when considering a very large number of results that have been collected in outpatient settings. Indeed, there is no disease with prevalence close to 50%. On the other hand, as discussed, the recommended direct sampling techniques are not without their own assumptions.

The reliability of an RI study should be a function of its accuracy and reproducibility and have a direct relationship with the number of observations used and method standardization. Statistically, it is more robust to analyze thousands of measurements that may include some unhealthy subjects than 120 measurements that are assumed to be from healthy subjects. The main problem with most of the reported indirect studies is that they used statistical analyses designed for a direct sampling technique. Hoffmann, in his classic *JAMA* article from 1963, described a technique designed for indirect estimation of RIs using all available test results from a laboratory’s database: “This statistical technique can be used for obtaining any normal values in medicine where a group of measurements are available and the mathematical assumptions are reasonable.”^{9(p868)} Although his work has been widely cited, few authors have actually applied the Hoffmann method in their calculations.

A notable exception is the manual of pediatric RIs by Soldin et al¹⁰ that is now in its sixth edition and published by American Association for Clinical Chemistry. This fundamental work was limited by the relatively small number of observations (typically 50-100) that were used and by the semimanual application of Hoffmann analysis of data, which added subjectivity to the calculations.¹⁰

Our goal in this study was to assess the reliability of the Hoffmann approach using a newly developed computer program designed to remove subjectivity from RI calculations.

Materials and Methods

Participants

The study was carried out using the data from Laboratory Corporation of America, one of the largest providers of laboratory testing in the United States. Data were collected from 6 regional laboratories: Burlington, NC; Columbus, OH; Houston, TX; Birmingham, AL; Raritan, NJ; and Tampa, FL. The average ordering source by test volume for this group of laboratories was as follows: 87% from outpatient general practice facilities and 13% from acute care facilities. The patient test results stored in the laboratory information system were queried for a given time frame without any filtering, no results were excluded, and all of these results constituted the original database for this study. The protocol for this study was determined to be exempt under existing regulations by the institutional review board.

Accuracy Study

The following 5 analytes were evaluated: mean corpuscular volume (MCV), hemoglobin, creatinine, calcium, and thyroid-stimulating hormone (TSH). Each of these analytes are known to have peer-reviewed RIs as calculated from direct sampling of volunteers, multicentered studies, or national survey databases. The following numbers of test results were collected retrospectively from the laboratory information system: MCV, 40,744 results (men and women); hemoglobin, men, 16,463; hemoglobin, women, 24,809; creatinine, men, 24,068; creatinine, women, 29,471; calcium, 51,492 (men and women); and TSH, 129,443 (men and women). All results originated from tests performed in June 2008 and were from adult patients (18 years or older).

Reproducibility Study

The reproducibility of this method was assessed using TSH as an analyte that historically has been problematic for RI studies. Some authors have suggested that “TSH reference range cannot be determined from population data, because occult thyroid dysfunction skews the TSH upper limit.”^{11(p4239)} A comparison of 2 TSH RIs as calculated from 2 different sets of test results collected 6 months apart was considered to be a good challenge for testing the reproducibility of this method. The second set of TSH results (n = 151,235) from adult patients (18 years or older) originated from December 2008 was collected.

Technical Information

Standardization to the same methods, reagent lot numbers, calibrators, controls, and standard operating procedures significantly reduces the between-laboratory variability in test results. All 6 laboratories that participated in this study had been standardized in this manner. The methods were as follows: MCV and hemoglobin, LH750/GEN*S (Beckman Coulter, Miami, FL); creatinine and calcium, Roche/Hitachi MODULAR (Roche Diagnostics, Indianapolis, IN); TSH, ADVIA Centaur (Siemens Medical Solutions Diagnostics, Tarrytown, NY). All methods are US Food and Drug Administration–approved and were operated according to the manufacturer’s specifications. Centralized comparative method performance monitoring for each laboratory location was conducted on a monthly basis using quality control data and patient mean data.

Statistical Analyses

The Hoffmann indirect method for the derivation of RIs was programmed as originally described.⁹ Chauvenet criteria were used for the detection and elimination of outliers. With Chauvenet criteria, a measurement (result) is eliminated if the probability of its occurrence is less than $1/(2N)$, where N is the number of measurements in the data pool and is greater than 4.

Following the elimination of outliers, the cumulative frequency for each test result was determined. The frequency of a test result was taken as the number of times a result occurs in the data set divided by the total number of results times 100%.

$$F_{X_i} = \frac{\text{Count}_{X_i}}{\text{Count}_{\text{data-pool}}} \times 100\%$$

The cumulative frequency is $CF_{X_i} = \sum_{k=2}^i F_{X_k}$ ordered by X_i .

Once the outliers had been eliminated, the data were refined so that only values from the linear portion of the cumulative frequency graph were used.

Visual analysis provides a good approximation of the linear data so that only values from the linear portion of the cumulative frequency graph were used. By computing the maximum deviation of the data from the regression line in this interval, the acceptable linearity error may be defined. This approach was taken because it is the best way to translate human fuzzy logic into computer precision. Once the acceptable linearity error had been defined, the portion of the data pool that was deemed to be “linear” was selected and used in computing the associated regression line **Figure 1**.

Subsequently, the best-fitting linear regression ($y_i = \alpha * x_i + \beta + \epsilon_i$) equation was determined by least-squares analysis (α is the slope, β is the intercept of the line and ϵ_i is the error). The line with the minimum sum of square residual values was identified accordingly. A residual value (r_i) was taken as the difference between the measured value (y_i) and the approximated one as determined by the linear regression function $[f(x_i)]$:

$$r_i = y_i - f(x_i)$$

The RIs were then determined from the linear regression equation following extrapolation of the preceding curve. The RI was calculated (for x = 2.5% and 97.5%):

$$RI_{\min} = \alpha * 2.5 + \beta \quad RI_{\max} = \alpha * 97.5 + \beta$$

The reference change value (RCV) was selected for determining the statistical significance of observed differences between the calculated RI and the published RI. If the observed difference is less than the RCV, the difference is not significant.

RCV was calculated as described by Fraser et al¹²:

$$RCV = 2^{1/2} * Z * (CV_a^2 + CV_i^2)^{1/2}$$

where Z is the probability selected for significance (a Z value of 1.96 was selected for 95% probability corresponding to a significant change), CV_a is the analytic variation, and CV_i is the within-subject biologic variation. The CV_a and CV_i values were derived from Ricos et al.¹³

Results

The representative output graph is given in **Figure 2**, and the derived regression functions and calculated RIs for

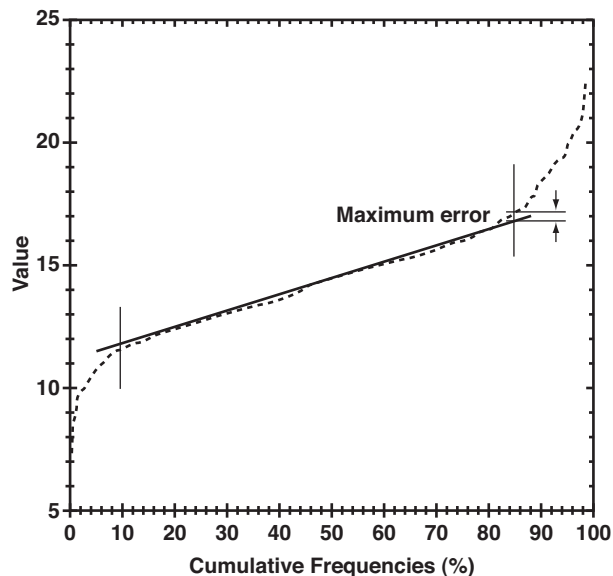
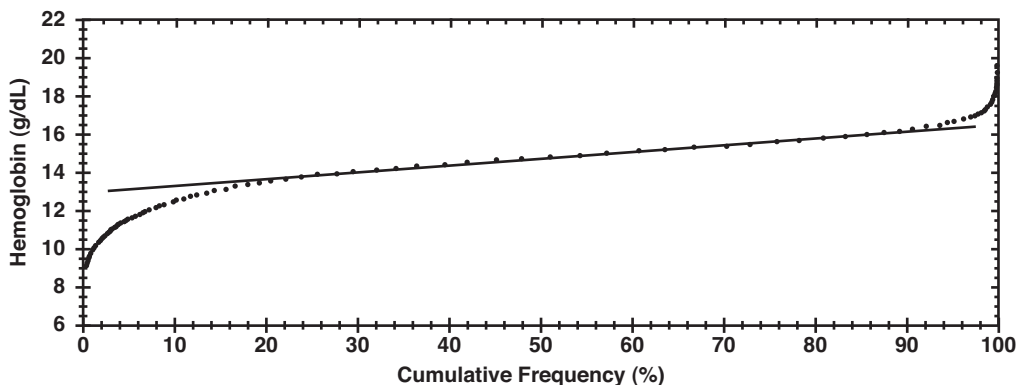


Figure 1 A typical plot of cumulative frequencies (hatched line) with the linear portion of the cumulative frequency graph (solid line).

A



B

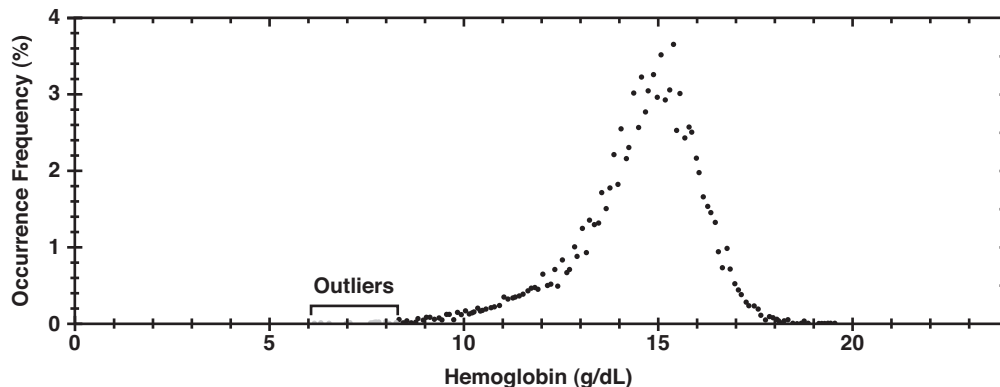


Figure 2 A representative output graph for hemoglobin (g/dL) for men. **A**, Cumulative frequencies (dots) and regression line. $y = 0.03594x + 12.93457$. **B**, Scatter graph showing good data (black dots) and outliers (gray dots). Linear range (%), 23.69302-90.67008; N value, 16,457 (without outliers); regression function, $y = 0.03594x + 12.93457$; reference interval, 13.02442-16.43872; maximum error, 0.05000. To convert hemoglobin values to Système International units (g/L), multiply by 10.0.

all 5 analytes are given in **Table 1**. The comparison of calculated RIs using the computerized Hoffmann method with peer-reviewed published intervals is given in **Table 2** (accuracy study).¹⁴⁻¹⁸ The comparison of 2 sets of TSH RIs derived from 2 separate data sets collected 6 months apart are given in the **Table 3** (reproducibility study).

Discussion

The computerized Hoffmann method for the indirect determination of RIs produced intervals that were remarkably similar to peer-reviewed RIs (Table 2).¹⁴⁻¹⁸ None of the calculated lower or upper limits displayed statistically significant differences from the corresponding published limits.

Table 1
Regression Functions and Calculated Reference Intervals for Five Analytes as Determined by the Computerization of the Hoffmann Method for the Indirect Estimation of Reference Intervals

Test/Sex	Linear Range (%)	Total No. of Results	Regression Function	Reference Intervals
Mean corpuscular volume, μm^3 (fL) Males and females	23.38-77.16	40,744	$y = 0.1267x + 83.18$	83.5-95.5 (83.5-95.5)
Hemoglobin, g/dL (g/L) Males	23.69-90.67	16,463	$y = 0.0359x + 12.93$	13.0-16.4 (130-164)
Females	21.40-89.48	24,809	$y = 0.0339x + 11.35$	11.4-14.7 (114-147)
Creatinine, mg/dL ($\mu\text{mol/L}$) Males	9.88-81.74	24,068	$y = 0.0053x + 0.75$	0.76-1.27 (67.2-112.3)
Females	8.83-75.93	29,471	$y = 0.0043x + 0.56$	0.58-0.99 (51.3-87.5)
Calcium, mg/dL (mmol/L) Males and females	9.40-88.03	51,492	$y = 0.0117x + 8.83$	8.86-9.98 (2.22-2.50)
Thyroid-stimulating hormone, mIU/mL (mIU/L)* Males and females (a)	6.09-72.22	129,443	$y = 0.0274x + 0.36$	0.44-3.05 (0.44-3.05)
Males and females (b)	7.06-71.18	151,235	$y = 0.0288x + 0.38$	0.45-3.19 (0.45-3.19)

* The a group, accuracy study; the b group, reproducibility study.

Table 2
Comparison of Reference Intervals as Calculated by the Hoffmann Method With RIs as Reported in the Literature (Accuracy Study)

Analyte	Calculated RI		Reported RI		Absolute Difference (%)		RCV (%)*
	Lower-Upper		Lower-Upper		Lower-Upper		
Calcium, mg/dL (mmol/L)	8.86-9.98 (2.22-2.50)		8.60-10.04 (2.15-2.51) ^{14(p273)}		2.93-0.60		5.95
Creatinine, mg/dL ($\mu\text{mol/L}$) Males	0.76-1.27 (67.2-112.3)		0.72-1.18 (63.6-104.3) ^{15(p561)} ; 0.68-1.13 (60-100) ^{14(p273)}		5.26-7.09; 10.53-11.02		16.50
Females	0.58-0.99 (51.3-87.5)		0.55-1.02 (48.6-90.2) ^{15(p561)} ; 0.57-1.02 (50-90) ^{14(p273)}		5.17-3.03; 1.72-3.03		
MCV, μm^3 (fL)	83.5-95.5 (83.5-95.5)		82.0-98.0 (82.0-98.0) ^{16(p21)}		1.80-2.62		4.09
Hemoglobin, g/dL (g/L) Males	13.0-16.4 (130-164)		13.4-17.0 (134-170) ^{16(p21)}		3.08-3.66		8.67
Females	11.4-14.7 (114-147)		11.7-15.3 (117-153) ^{16(p21)}		2.63-4.08		
TSH, mIU/mL (mIU/L)	0.44-3.05 (0.44-3.05)		0.45-4.12 (0.45-4.12) ^{17(p495)} ; 0.54-3.37 (0.54-3.37) ^{18(p1227)}		2.27-35.08; 22.73-10.49		59.86

MCV, mean corpuscular volume; RI, reference interval; TSH, thyroid-stimulating hormone.

* Reference change value (95% probability) as calculated from Ricos et al.¹³

Table 3
Comparison of RIs as Calculated by the Hoffmann Method for Two Sets of TSH Data Collected 6 Months Apart (Reproducibility Study)

Analyte	First Data Set		Second Data Set		Absolute Difference (%)		RCV (%)*
	Lower-Upper		Lower-Upper		Lower-Upper		
TSH, mIU/mL (mIU/L)	0.44-3.05 (0.44-3.05)		0.45-3.19 (0.45-3.19)		2.27-4.59		59.86

RI, reference interval; TSH, thyroid-stimulating hormone.

* Reference change value (95% probability) as calculated from Ricos et al.¹³

The calculated ranges in most cases were slightly narrower than the ranges obtained from direct sampling techniques. A concern with the use of indirect methods has been that the resulting interval may be wider than it should be or skewed because of inclusion of unhealthy subjects. This was not observed in the present study. It is noteworthy that the choice of the statistical method and the large number of observations are critical components in maintaining the accuracy of RIs derived from indirect methods. The described technique may be the method of choice for this purpose.

It is well recognized that test results may be impacted by a host of different factors, some of which are known and some are not (preanalytic sources of variation). Preanalytic sources of variation should be taken into consideration and controlled when selecting the healthy subjects who will be used for establishing an RI by traditional techniques. Despite this, preanalytic sources of variation that may have been controlled for in establishing the RI are often neglected when it comes to the routine testing of patients. It may be argued that the Hoffmann indirect technique as applied in the present study should provide a more robust estimation of the RI than traditional methods in that the majority (if not all) of the preanalytic sources of variation impacting the test result will be reflected in the outcome.

Another interesting aspect from this study relates to estimates of disease prevalence. It is reasonable to suggest that the percentage of test results above and below the upper and lower limits in the entire data set used for RI calculation for a given analyte may correlate with a prevalence of conditions that cause its concentration to be abnormal. This assumption is less reasonable while using smaller sets of patient results or results from hospital settings because the results distribution will be skewed by the significant number of repeated tests from the same patient and results from patients with serious conditions. The large number of observations from outpatient settings will negate those factors. A recent study of the prevalence of chronic kidney disease in the United States reported the prevalence in 1999 to 2004 as 13.1% and raised concerns about its future increase.¹⁹ In the present study, the filtering technique excluded from the healthy subgroup a total of 14.4% of test results for creatinine as being above the calculated upper range (men and women). This mirrors the reported prevalence of kidney disease in the United States.

The American Thyroid Association has estimated the prevalence of thyroid dysfunction in the United States for subclinical hypothyroidism to be up to 17% and for clinical hypothyroidism to be up to 2%, which makes a combined prevalence of both hypothyroid conditions of up to 19%. The prevalence of subclinical hyperthyroidism was reported to be up to 6% and clinical hyperthyroidism to be up to 0.2%, which makes a combined prevalence of both hyperthyroid conditions of up to approximately 6%.²⁰ The described method excluded

from the healthy subgroup a total 18.6% of test results for TSH that were above the calculated upper limit and 5.9% of test results for TSH that were below the calculated lower limit of the RI. Again, these values are very close to the reported prevalence of thyroid diseases.

As expected, the reported RIs for TSH showed the greatest degree of variation with the reference studies. However, the observed difference was well below the RCV. It is interesting that both of the reference studies for TSH reported higher upper limits than the upper limit that was generated in the present study. This may be explained in part by the observations of Spencer et al¹¹ regarding the difficulties in the estimation of an accurate TSH upper limit using the existing techniques. The presented method seems to accurately reflect the actual distribution of TSH in the thyroid disease-free population, and the new calculated TSH upper limit is in complete agreement with a recently suggested upper limit of 3.0 mIU/L and a treatment goal for hypothyroidism.^{21,22}

The limitations of this method for the use in many clinical laboratories may apply when one of the following is present: a large prevalence of results from hospitalized patients; a limited number of observations, especially for subject groups like pediatric or geriatric patients or rare sample types like synovial fluid; and lack of standardization between the methods in use.

This limitation could be minimized by linking laboratories that are operating similar instrumentation and methods into peer group–based operational networks. The performance of the peer group within the network would be monitored to ensure that standardization goals are being met. Under such circumstances, each of the participating laboratories could contribute its patient test results to a central network data bank from which the RIs for the network could be established.

Conclusions

The described computerized Hoffmann method for the indirect estimation of the RIs from the existing laboratory database of test results has proven to be reliable and reproducible. The method produced RIs that were statistically not different from the ones reported in peer-reviewed publications. This method was able to derive RIs for a problematic analyte (TSH), producing a result that was in strong agreement with recent scientific observations and clinical recommendations. In addition to its reliability and reproducibility, this technique has a number of advantages over the conventional direct sampling techniques. Notably, it provides a mechanism for deriving RIs for difficult-to-study populations like pediatric and geriatric, as well as for rare sample types like aspirations, for calculating RIs for timed collections, and for challenge tests while providing overall laboratory resources savings. In

addition, it may be used on an ongoing basis as part of a quality management program by providing an auditing tool for confirming the appropriateness of the current RIs for the existing methods. The caveat for this method is that it may be used only with large numbers of observations, with test results that are mostly from outpatient settings, and when standardization of methods has been implemented and monitored.

From the ¹Department of Science and Technology, Laboratory Corporation of America (LabCorp), Burlington, NC; ²Canadian External Quality Assessment Laboratory, Vancouver; and ³Department of Pathology and Laboratory Medicine, University of British Columbia and Canadian External Quality Assessment Laboratory, Vancouver.

Address reprint requests to Dr Katayev: Dept of Science and Technology, Laboratory Corporation of America, 112 Orange Dr, Elon, NC 27244.

Authors' contributions were as follows: study concept, design, and drafting the manuscript, Drs Katayev and Seccombe; software programming, Mr Balciza; statistical analysis, Drs Katayev and Seccombe and Mr Balciza. Drs Katayev and Seccombe had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Acknowledgment: We thank Mark Sharp for the collection of data and Mark Brecher, MD, for review of the manuscript.

References

- Utah Governor Huntsman recognizes medical laboratory week. *Clinical Lab Products*. Issue Stories. May 2006. http://www.clpmag.com/issues/articles/2006-05_03.asp. Accessed January 1, 2009.
- Cerriotti F. Prerequisites for use of common reference intervals. *Clin Biochem Rev*. 2007;28:115-121.
- Clinical and Laboratory Standards Institute. Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline. Third Edition. CLSI document C28-A3. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.
- Grossi E, Colombo R, Cavuto S, et al. The RELAB project: a new method for the formulation of reference intervals based on current data. *Clin Chem*. 2005;51:1232-1240.
- Bock BJ, Dolan CT, Miller GC, et al. The data warehouse as a foundation for population-based reference intervals. *Am J Clin Pathol*. 2003;120:662-670.
- Surks MI, Hollowell JG. Age-specific distribution of serum thyrotropin and antithyroid antibodies in the US population: implications for the prevalence of subclinical hypothyroidism. *J Clin Endocrinol Metab*. 2007;92:4575-4582.
- Kouri T, Kairisto V, Virtanen A, et al. Reference intervals developed from data for hospitalized patients: computerized method based on combination of laboratory and diagnostic data. *Clin Chem*. 1994;40:2209-2215.
- Jones GR, Barker A, Tate J, et al. The case of common reference intervals. *Clin Biochem Rev*. 2004;25:99-104.
- Hoffmann RG. Statistics in the practice of medicine. *JAMA*. 1963;185:864-873.
- Soldin SJ, Brugnara C, Wong EC. *Pediatric Reference Intervals*. 6th ed. Washington, DC: AACC Press; 2007.
- Spencer CA, Hollowell JG, Kazarosyan M, et al. National Health and Nutrition Examination Survey III thyroid-stimulating hormone (TSH)-thyroperoxidase antibody relationships demonstrate that TSH upper reference limits may be skewed by occult thyroid dysfunction. *J Clin Endocrinol Metab*. 2007;92:4236-4240.
- Fraser CG, Hyltoft Petersen P, Lytken Larsen M. Setting analytical goals for random analytical error in specific clinical monitoring situations. *Clin Chem*. 1990;36:1625-1628.
- Ricos C, Alvarez V, Cava F, et al. Desirable specifications for total error, imprecision, and bias, derived from biologic variation. Westgard QC Web page. <http://www.westgard.com/biodatabase1.htm>. Accessed March 5, 2008.
- Rustad P, Felding P, Franzon L, et al. The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties. *Scand J Clin Lab Invest*. 2004;64:271-284.
- Cerriotti F, Boyd JC, Klein G, et al. Reference intervals for serum creatinine concentrations: assessment of available data for global application. *Clin Chem*. 2008;54:559-566.
- Martensson A. Reference intervals for haematology analytes: preliminary results from the Nordic Reference Interval Project (NORIP). *Klinisk Biokemi I Norden*. 2003;2:20-21.
- Hollowell JG, Staehling NW, Flanders WD, et al. Serum TSH, T4, and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab*. 2002;87:489-499.
- Hamilton TE, Davis S, Onstad L, et al. Thyrotropin levels in a population with no clinical, autoantibody, or ultrasonographic evidence of thyroid disease: implications for the diagnosis of subclinical hypothyroidism. *J Clin Endocrinol Metab*. 2008;93:1224-1230.
- Coresh J, Selvin E, Stevens LA, et al. Prevalence of chronic kidney disease in the United States. *JAMA*. 2007;298:2038-2047.
- Ladenson PW, Singer PA, Ain KB, et al. American Thyroid Association Guidelines for detection of thyroid dysfunction. *Arch Intern Med*. 2000;160:1573-1575.
- Spencer C. Clinical implications of the new TSH reference range. *AACC Expert Access*. August 15, 2006. http://www.aacc.org/events/expert_access/2006/tshrange/pages/default.aspx. Accessed May 5, 2008.
- American Association of Clinical Endocrinologists medical guidelines for clinical practice for the evaluation and treatment of hyperthyroidism and hypothyroidism. *Endocr Pract*. 2002;8:457-469.