

Opinion Paper

Obtaining reference intervals traceable to reference measurement systems: is it possible, who is responsible, what is the strategy?

Mauro Panteghini^{1,*} and Ferruccio Ceriotti²

¹Centro Interdipartimentale per la Riferibilità Metrologica in Medicina di Laboratorio (CIRME), Università degli Studi, Milano, Italy

²Diagnostica e Ricerca S. Raffaele, Istituto Scientifico Universitario S. Raffaele, Milano, Italy

Abstract

An issue associated with standardization efforts is the need to develop useful reference intervals (RI). Lack of proper RI may hamper the implementation of standardization in Laboratory Medicine as standardization can modify analyte results and, without adequate RI, this can impair the result interpretation. Once defined, RI obtained with analytical procedures that produce results traceable to the corresponding reference system can be transferred among laboratories, providing that they use commercial assays that produce results traceable to the same reference system and populations have the same characteristics. Multicenter studies are needed for a robust definition of traceable RI, using experimental protocols that include well defined prerequisites. Particularly, employed methods must produce results that are traceable to the reference system for that specific analyte. Thus, the trueness of laboratories producing reference values should be verified and, if necessary, experimental results corrected in accordance with correlation results with the selected reference. If requirements in the adoption of traceable RI are fulfilled, the possibility of providing RI that are applicable to any laboratory, able to produce results traceable to the reference system, is realistic. The definition of traceable RI should hopefully cause the disappearance of different RI employed for the same analyte, providing more effective information to clinicians.

Keywords: reference values; standardization; traceability.

*Corresponding author: Prof. Mauro Panteghini, Laboratorio Analisi Chimico Cliniche, Ospedale Luigi Sacco, Via GB Grassi 74, 20157 Milan, Italy
Phone: +39 02 3904 2806, Fax: +39 02 503 19835,
E-mail: mauro.panteghini@unimi.it

Received November 3, 2011; accepted November 22, 2011;
previously published online December 20, 2011

Introduction

The issue of standardization in Laboratory Medicine is a priority for public health as laboratory customers (i.e., doctors and patients) expect test results to be accurate and comparable and interpreted in a reliable and consistent manner (1). To be accurate and comparable, results must be traceable to high-order references (2–4). The objective of traceability implementation is to enable results obtained by the calibrated routine procedure to be expressed in terms of the values obtained at the highest available level of the calibration hierarchy (5). To be interpreted in a reliable and consistent manner, results should be compared with appropriate reference intervals or decision limits or with previous results from the same individual (6).

Previously, we already emphasized that the lack of proper reference intervals and/or decision limits for the newly standardized methods may markedly hamper their adoption (7–9). As the implementation of standardization can significantly modify the analyte results, without adequate reference intervals/decision limits this situation can impair the interpretation of the results and, paradoxically, worsen the patient's outcome. However, a single clinical laboratory has often not enough means to adequately produce reference limits and manufacturers have problems too (7). Available surveys performed in recent years on both sides of the Atlantic Ocean have shown that used reference intervals are often significantly different among laboratories even if, in the majority of cases, differences have no evident justification (10–12). Particularly, the source of reference intervals may be quite different and heterogeneous (manufacturer's recommendation, textbooks and literature, internal studies by healthy volunteers testing, other laboratories, undefined) (11, 12). To try to minimize these differences among laboratories, results are often expressed as multiples of the upper reference limit (URL). This approach has been used in international guidelines (13) and some authors in the past welcomed it as "physician liberation units" (14). Recently, it was proposed a more sophisticated version of that approach called "quantity quotient reporting" (15). Studies have demonstrated, however, that the expression of results as URL multiples paradoxically increases the scatter of results due to both inter-method differences and the different laboratory reference limits (16). Hence, this mode of result expression should be discouraged.

Traceable reference intervals as the fourth pillar of the reference measurement system

The reference measurement system represents a trueness-based approach in which different commercial methods that provide results traceable to the system are able to produce comparable results in clinical laboratories using these assays. Thus, reference intervals obtained with analytical procedures that produce results traceable to the corresponding reference system can be transferred between laboratories, providing that they use commercial assays that produce results traceable to the same reference system and served populations have the same characteristics (7, 8, 17). In fact, when the level of standardization is good, only population differences justify different reference intervals. There are therefore two main aspects in the applicability of traceable reference interval concept: an analytical and a biological one. From one hand, laboratories should verify and validate the correctness of traceability implementation; on the other hand, possible population differences (due to ethnicity, genetics, or environmental factors) should be clarified. In considering the latter issue, a sufficient number of individuals with possibility of comparing different populations/ethnic groups should therefore be evaluated.

Ways for obtaining suitable traceable reference intervals

There are two ways for obtaining suitable traceable reference intervals: a) verification, according to the theory, of the existing peer-reviewed published data on reference intervals to decide which one can be accepted and “recommended” (at least for some ethnic group); b) planning of the activity necessary to define, among the missing traceable reference intervals, those that are more urgently needed and organize appropriate experimental studies for their establishment.

A good example of the definition of a theoretical approach for the endorsement of already existing traceable reference intervals is represented by the study published in 2008 assessing reference intervals for serum creatinine concentrations for their global application (18). Particularly, the selection of studies to be included was guided by the following four criteria:

- criterion 1: a priori selection of reference individuals;
- criterion 2: adequate description of pre-analytical conditions;
- criterion 3: analytical correctness and demonstration of traceability of the produced results to the isotope dilution-mass spectrometry (IDMS) reference method;
- criterion 4: appropriate statistical treatment of data (selection of an adequate number of subjects and use of non-parametric statistical methods, with appropriate partitioning).

In comparison with the classical studies defining reference intervals, the newly introduced criterion was the need of incontrovertible demonstration of analytical correctness (i.e., result traceability) of the retrieved experimental studies (Figure 1). Using this approach, authors were able to

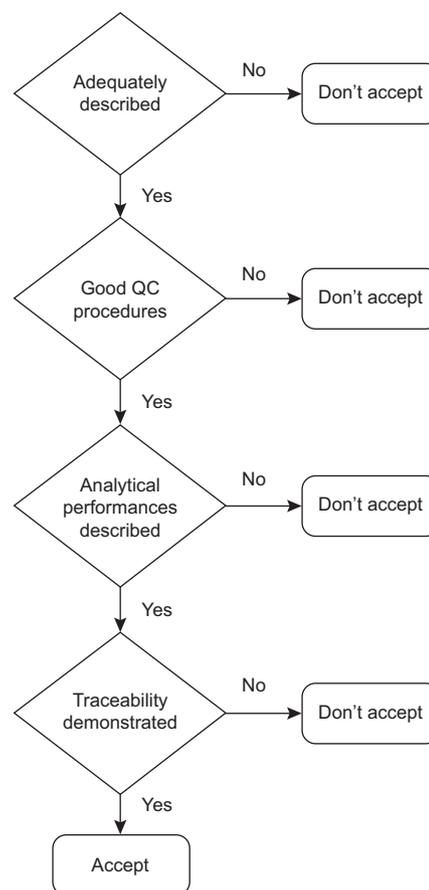


Figure 1 Recommended flow chart for checking the correctness of the analytical approach employed in the production of traceable reference intervals. QC, quality control.

recommend traceable reference intervals universally valid for white adults and children (18, 19).

General prerequisites for the organization of (multicenter) studies for defining traceable reference intervals have previously been described (8). There are two possible approaches: a) each participating center, properly standardized, analyzes its own fresh reference samples, or b) the reference samples are collected in different centers, frozen, and shipped to a central laboratory where all analyses are performed. The first approach requires a rather complex preliminary phase for verification of traceability of measurements, but uses native samples exactly as in clinical practice; the second option is simpler, allows a better control of the analytical phase, but uses frozen samples, thereby introducing a variable not typical for the routine laboratories (8). A recently published study has used a mix of both approaches to derive traceable reference intervals for aspartate aminotransferase (AST), alanine aminotransferase (ALT), and γ -glutamyl transferase (GGT) (20). Three clinical center laboratories were originally involved in the collection and measurements of reference samples. Additional samples were obtained frozen from a biobank [Nordic Reference Interval Project (NORIP)] and from an additional laboratory; these samples were then analyzed

in one of the laboratories that were collecting samples. Analytically, the traceability of enzyme results to the available reference measurement systems was verified through the use of trueness control materials that had values assigned by the reference procedures. For AST and ALT, the global implementation of traceable reference intervals appeared to be possible, because no differences between laboratories and centers collecting reference samples were observed (20). On the contrary, GGT values in serum appeared to be markedly different in the Scandinavian population vs. the other three regional populations (South Europe, Turkey, and China) for both males and females (20). More recently, a similar study deriving reference intervals for alkaline phosphatase (ALP) in adult males and premenopausal females, traceable to the newly IFCC recommended reference procedure, has been performed (21). Sample collection was done in four different European centers. The frozen aliquots were sent in dry ice to the reference laboratory in Germany where all the analyses were performed. The calibration of the employed assay was performed by using a set of pooled sera with assigned target values for ALP obtained by measurements using the primary IFCC reference measurement procedure (21). It is expected that the obtained traceable reference limits for ALP become usable once the new reference measurement procedure for ALP is approved by the Joint Committee on Traceability in Laboratory Medicine (JTLM) and used to standardize commercial assays.

Cautions needed in the adoption of traceable reference intervals

Boyd has cautioned against the uncritical use of “common” reference intervals, stressing the need to fully document the possible presence of differences in tests results across populations due to biological or environmental factors (22). Until it can be reasonably shown that no differences exist for a given test between the population served by individual laboratories and that used in defining traceable reference intervals, their adoption should indeed be discouraged.

The selection of individuals for the production of reference values needs much attention. Selection of “normal reference” subjects is from many years a key point in all recommendations for the development of reference intervals (23). The application of different models for selection of individuals may result in large variability between reference limits (24). We already mentioned the difference recently found between traceable reference intervals for GGT in individuals from Nordic countries and those of other parts of the world (20). In a sub-analysis of the NORIP, Alatalo et al. have explained this peculiarity of North European populations with factors reflecting life style, such as drinking habit or body weight (25). In their study, by applying strengthened inclusion criteria, the URL for GGT in NORIP males ≥40 years old (69 U/L) became indeed very similar to that found in the IFCC multicenter study in the same group of individuals after the exclusion of samples obtained from the NORIP group (68 U/L). Therefore, criteria used for acknowledging reference

population still maintain a central role also in the traceability era.

To specifically test the influence of regional variations in East and South-East Asia, a multicenter study was recently conducted to derive traceable reference intervals for 30 analytes for which reference measurement systems exist (26). Clinical centers located in six different countries took part in the study and more than 3500 well-defined healthy volunteers were recruited. All the specimens were sent to Japan at -80°C for centralized measurements, performed with assays of which the traceability was checked to the corresponding reference system using reference materials or reference measurement procedures available in the JCTLM database. Particularly, when reference laboratory services were involved (enzymes and steroid hormones), a set of fresh-frozen (-80°C) pooled sera for each enzyme or different levels of lyophilized specimens (for steroid hormones) were first value assigned by reference laboratories and then used to check alignment of systems used for obtaining reference values. Limiting the discussion to the three enzymes already evaluated in the IFCC study, no differences were found in traceable reference intervals for AST and GGT between the two multicenter studies, while ALT showed lower upper reference limits in the Asian study (Table 1). Once again, criteria used for selection of individuals may, at least partly, explain this difference. Tolerated body mass index was <30 kg/m² in the IFCC study and <26 kg/m² in the Asian study, respectively, and this may justify the higher ALT values in the population of the former study as this enzyme is no longer be considered solely as a marker of hepatocellular damage, but also a marker of ectopic fat deposition (27).

Table 1 Comparison between traceable reference intervals for three liver enzymes obtained in the IFCC multicenter study (20) and in the Asian-Pacific multicenter study (26).

	IFCC study		Asian study	
	Females	Males	Females	Males
AST, U/L	11–33	14–35	14–29	16–35
ALT, U/L	8–41	9–59	11–31	14–54
GGT, U/L	6–40	12–68	15–43	15–68

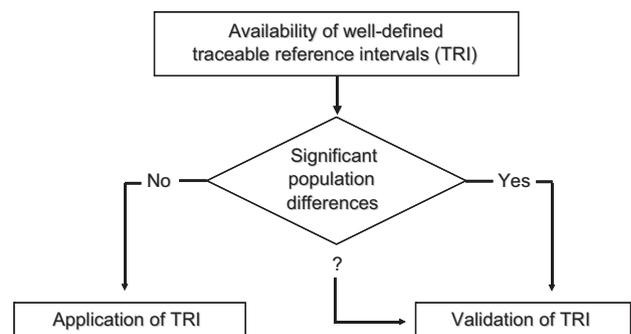


Figure 2 Steps for the implementation of traceable reference intervals in clinical practice.

Table 2 Summary of analytical and biological requirements in the adoption of traceable reference intervals [adapted from ref. (17)].

	Responsibility
Analytical requisite:	
• Existence of a reference system	• IFCC, JCTLM, National Metrology Institutes
• Availability of commercial assays producing traceable results	• IVD manufacturers
• Correct implementation of traceable assays in the clinical setting	• Clinical laboratories
• Control of the performance of commercial assays within stated limits of uncertainty	• Clinical laboratories, EQAS organizers
• Compatible pre-analytical conditions	• Clinical laboratories
Biological requisite:	
• Accurate definition of traceable reference intervals, providing information on the influence of biological and environmental factors	• Joint effort between laboratory profession and IVD manufactures
• Validation of the applicability of the traceable reference intervals to the laboratory's own population	• Clinical laboratories

EQAS, external quality assessment scheme; IVD, in vitro diagnostic.

Validation of available traceable reference intervals

Providing that there are no significant differences in the populations served by the laboratory, the existing traceable reference intervals can be applied without any further step. As described before, there are however many factors, varying according to different analytes, that may influence reference intervals. If doubts remain on the possibility that, for instance, different life habits may influence reference intervals, they should be validated before their clinical use (Figure 2). The validation can be done according to the CLSI document C28-A3, paragraph 11.2, by examining 20 reference individuals from a laboratory's own subject population (28). If no more than two (10%) of the 20 tested values fall outside the defined traceable reference interval, this can be adopted by the individual laboratory.

Concluding remarks

Table 2 summarizes the analytical and biological requirements in the adoption of traceable reference intervals. Providing that these prerequisites are fulfilled, the possibility of obtaining reference intervals that are applicable to any laboratory, able to produce results traceable to reference measurement systems, seems to be realistic, even if their establishment requires coordinated efforts at different levels and between different stakeholders. IFCC and its Committee on Reference Intervals and Decision Limits is putting a lot of efforts to accomplish the task of providing sound reference intervals. Manufacturers should also contribute by adopt the proposed reference intervals at the same time of the standardization of their assays. Last but not least, clinical laboratories should place in use traceable reference intervals without any further delay. Only the availability of traceable reference intervals will permit the definitive implementation of standardization projects, filling the gap between the production of traceable results and their correct use for patient care in order to eventually provide more effective information to clinicians.

References

- McLawnon RW. Patient safety and clinical effectiveness as imperatives for achieving harmonization inside and outside the clinical laboratory. *Clin Chem* 2011;57:936–8.
- Panteghini M. Traceability, reference systems and result comparability. *Clin Biochem Rev* 2007;28:97–104.
- Panteghini M. Traceability as a unique tool to improve standardization in laboratory medicine. *Clin Biochem* 2009;42:236–40.
- Panteghini M. Application of traceability concepts to analytical quality control may reconcile total error with uncertainty of measurement. *Clin Chem Lab Med* 2010;48:7–10.
- ISO 17511:2003. In vitro diagnostic medical devices – measurement of quantities in biological samples – Metrological traceability of values assigned to calibrators and control materials. Geneva, Switzerland: ISO, 2003.
- Büttner J. Laboratory findings: structure, validity and significance for medical cognitive processes. *Eur J Clin Chem Clin Biochem* 1991;29:507–19.
- Ceriotti F, Hinzmann R, Panteghini M. Reference intervals: the way forward. *Ann Clin Biochem* 2009;46:8–17.
- Infusino I, Schumann G, Ceriotti F, Panteghini M. Standardization in clinical enzymology: a challenge for the theory of metrological traceability. *Clin Chem Lab Med* 2010;48:301–7.
- Panteghini M. Implementation of standardization in clinical practice: not always an easy task. *Clin Chem Lab Med* 2012; DOI 10.1515/CCLM.2011.791.
- Ceriotti F. Reference intervals in the new millenium. *Biochim Clin* 2007;31:254–66.
- Dutta A, Saha C, Johnson CS, Chalasani N. Variability in the upper limit of normal for serum alanine aminotransferase levels: a statewide study. *Hepatology* 2009;50:1957–62.
- Friedberg RC, Souers R, Wagar EA, Stankovic AK, Valenstein PN. The origin of reference intervals. A College of American Pathologists Q-Probes study of “normal ranges” used in 163 clinical laboratories. *Arch Pathol Lab Med* 2007;131:348–57.
- AGA Institute Medical Position Statement on Acute Pancreatitis. *Gastroenterology* 2007;132:2019–21.
- Langdon DE, Leeper CK. Physician liberation units system. *N Engl J Med* 1970;283:1413.
- Haecckel R, Wosniok W. Quantity quotient reporting. A proposal for a standardized presentation of laboratory results. *Clin Chem Lab Med* 2009;47:1203–6.

16. Férard G, Imbert-Bismut F, Messous D, Piton A, Ueda S, Poynard T, et al. A reference material for traceability of aspartate aminotransferase (AST) results. *Clin Chem Lab Med* 2005;43:549–53.
17. Ceriotti F. Prerequisites for use of common reference intervals. *Clin Biochem Rev* 2007;28:115–21.
18. Ceriotti F, Boyd JC, Klein G, Henny J, Queraltó J, Kairisto V, et al. On behalf of the IFCC Committee on Reference Intervals Decision Limits (C-RIDL). Reference intervals for serum creatinine concentrations: assessment of available data for global application. *Clin Chem* 2008;54:559–66.
19. Panteghini M. Enzymatic assays for creatinine: time for action. *Clin Chem Lab Med* 2008;46:567–72.
20. Ceriotti F, Henny J, Queraltó J, Ziyu S, Özarda Y, Chen B, et al. Common reference intervals for aspartate aminotransferase (AST), alanine aminotransferase (ALT) and γ -glutamyl transferase (GGT) in serum: results from an IFCC multicenter study. *Clin Chem Lab Med* 2010;48:1593–601.
21. Schumann G, Klauke R, Canalias F, Bossert-Reuther S, Franck PF, Gella FJ, et al. IFCC primary reference procedures for the measurement of catalytic activity concentrations of enzymes at 37°C. Part 9: reference procedure for the measurement of catalytic concentration of alkaline phosphatase. *Clin Chem Lab Med* 2011;49:1439–46.
22. Boyd JC. Cautions in the adoption of common reference intervals. *Clin Chem* 2008;54:238–9.
23. PetitClerc C, Solberg HE. International Federation of Clinical Chemistry (IFCC). Approved recommendation (1987) on the theory of reference values. Part 2. Selection of individuals for the production of reference values. *J Clin Chem Clin Biochem* 1987;25:639–44.
24. Panteghini M. A critical appraisal of experimental factors influencing the definition of the 99th percentile limit for cardiac troponins. *Clin Chem Lab Med* 2009;47:1179–82.
25. Alatalo P, Koivisto H, Kultti J, Bloigu R, Niemela O. Evaluation of reference intervals for biomarkers sensitive to alcohol consumption, excess body weight and oxidative stress. *Scand J Clin Lab Invest* 2010;70:104–11.
26. Ichihara K. Study design: strategy for collaboration in the definition of reference intervals. *Clin Chem Lab Med* 2011;49:S69.
27. Ioannou GN. Implications of elevated serum alanine aminotransferase levels: think outside the liver. *Gastroenterology* 2008;135:1851–4.
28. CLSI. Defining, establishing, and verifying reference intervals in the clinical laboratory; Approved guideline, 3rd ed. CLSI document C28-A3c. Wayne, PA: Clinical and Laboratory Standards Institute, 2008.