Traceability as a unique tool to improve standardization in laboratory medicine

Mauro Panteghini *

Centre for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Milan, Italy

Abstract

The standardization of measurements is of high priority in Laboratory Medicine, its purpose being to achieve closer comparability of results obtained using routine analytical systems. In order to achieve standardization, an approach is required that provides reliable transfer of the measurement values from the highest hierarchical level to methods which are routinely used in the clinical laboratories. Such a structure is presented by the reference measurement system (RS), based on the concepts of metrological traceability. Key elements of a comprehensive RS are the reference measurement procedure and reference materials. Other essential elements include the definition of the measurand in regards to the intended clinical use and the reference laboratories that may collaborate in a network. At present, there is international cooperation in developing RS for analytes of clinical significance. Thanks to the work of the Joint Committee on Traceability in Laboratory Medicine (JCTLM), a list of higher order reference materials and reference methods is now publicly available. JCTLM has also published the list of reference laboratories that are able to deliver a reference measurement service. As soon as a new RS is implemented, clinical validation of the correctly calibrated routine methods (the IVD products sold onto the market) should take place. Other important issues concerning the implementation of a metrologically-correct approach for result standardization are: 1) the clear definition of the clinically allowable error of measurements and 2) the post-market surveillance of the performance of IVD products. These are tasks of our profession through the organization of appropriate External Quality Assessment programs.

Introduction

The primary goal of Laboratory Medicine is to provide information that is useful to assist medical decision-making, allowing optimal healthcare. This can only be obtained by generating reliable analytical results on patient samples. Leading to incorrect interpretation by the clinician, inadequate laboratory performance may have extensive consequences for practical medicine, healthcare systems, and, last but not least, for the patient.

Foremost among the laboratory's problems is the poor comparability of analytical results that originate from different laboratories using different methods. Even today considerable differences can still be observed in the results obtained using different measurement procedures for the same analyte [1]. Such differences may cloud interpretations of reported data, creating problems for both clinician and laboratory communities.

Standardization of laboratory measurements would ensure the interchangeability of results over time and space and significantly contribute to improvements in healthcare by allowing results of clinical studies undertaken in different locations or times to be universally applied [2]. This will enable an effective application of evidence-based medicine and use of guidelines established by scientific or professional bodies which often advocate use of specific decision limits for diagnosis and therapeutic intervention [3].

The recognition that it is the standardization of results requiring improvement in Laboratory Medicine has raised questions about what contributes to the lack of standardization.

* Laboratorio Analisi Chimico Cliniche, Ospedale Luigi Sacco, Via GB Grassi 74, 20157 Milan, Italy. Fax: +39 02 503 19835.
  E-mail address: sd.chair@ifcc.org.

It was recognized that an insufficient calibration approach, due to the lack of result traceability to certified standards, is the major cause [4]. Consequently, an international agreement on the need to improve standardization through the implementation of metrologically-correct measurement systems has been reached.

The importance of the metrological principles has been described in two documents of the International Organization for Standardization (ISO), the ISO 17511 and the ISO 18153 [5,6]. In these documents, the traceability to internationally recognized and accepted reference materials and measurements is considered to be the key element in assuring the accuracy and comparability of clinical laboratory measurements. The directive of the European Union (EU) on in vitro diagnostic (IVD) devices supports these ISO standards and requests applications of the standards for all diagnostic reagents, the aim being to ensure that the use of IVDs do not compromise the health and safety of patients, users, and third parties and to attain the performance levels attributed to them by their manufacturer [7]. From a practical perspective, diagnostic manufacturers must ensure that the analytical systems they market have been calibrated against available certificate reference materials and reference measurement procedures and that uncertainty of their internal calibration procedures is quantified and documented.

The reference measurement system

In order to achieve standardization, an approach is required that provides reliable transfer of the measurement values from the highest hierarchical level to methods which are routinely used in the clinical laboratories. Such a structure is presented by the reference measurement system (RS), which is based on the concepts of metrological traceability (Fig. 1) [8]. Key elements of the system are the reference measurement procedure and reference materials. The reference procedure is used to assign a certified value to a given reference material. Once the appropriate reference material is certified, this material and the manufacturer's testing procedure can be used in industry to assign values to commercial calibrators. Clinical laboratories use routine procedures with validated calibrators, both from commercial sources, to measure human samples. In this way, the obtained value will be traceable to the reference procedure

and materials, and the standardization of measurement, that is, the process of realizing traceability, will be reached.

The commutability issue

However, it should be noted that the above statements are true only if the materials used to transfer trueness to the field methods are commutable [9]. Commutability is the ability of a reference or calibrator material to show inter-assay properties similar to those of human samples. In practical terms, the numerical ratio between the results determined by a given routine and a reference procedure found for the reference material must be the same as the average ratio found for patients' samples [10]. Only commutable materials can be used by industry for direct calibration of commercial methods to ensure there is an unbroken traceability chain. It is well known that purification procedures that are sometimes used in the preparation of reference materials may result in non-commutability of these materials with native samples. Pure compounds prepared by recombinant techniques may also have altered structures with the consequence that the derived materials have a high probability of matrix effects. A solution to the commutability problems can be the preparation and use of secondary reference materials as an intermediate step in the traceability chain. In their preparation, human serum is the preferred base matrix, the effects of the natural variation between donors being minimized by using pooled collections from a number of individuals. However, although matrix-based materials are desirable as they are more likely to behave in a similar fashion as human samples, this does not *a priori* eliminate the non-commutability problems. Thus, "patient-like" reference materials should be used for calibration of commercial methods only if their commutability has been proven experimentally.

If commutable reference materials suitable for direct use in the field method calibration are lacking, the only possible alternative for establishing traceability to a reference measurement procedure is for diagnostic manufacturers to split human fresh samples with a laboratory performing the reference measurement procedure. Calibration of the commercial system will be in accordance with correlation results obtained using the value-assigned samples [11].
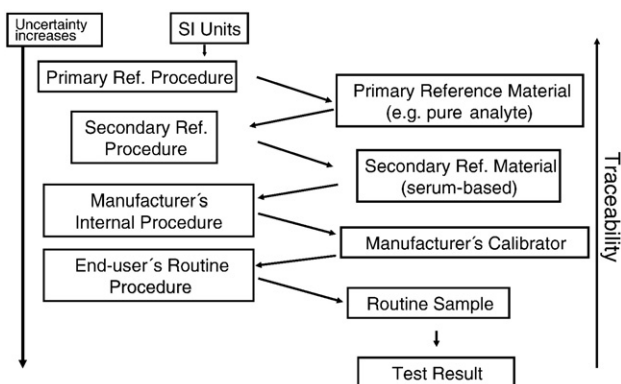
Definition of the "measurand"

In addition to reference procedures and materials, essential elements of a comprehensive RS also include the definition of the measurand in regards to the intended clinical use and the individuation of reference laboratories that may collaborate in a network. The main responsibility of reference laboratories is to assign target values to reference materials, using the reference measurement procedures. In addition, they may assist commercial companies in the validation of routine procedures through direct comparison of a routine analytical system with the reference procedure, using a number of appropriately selected, native human samples. Finally, reference laboratories may be regarded as a concerted means of supporting External Quality

Fig. 1. The reference measurement system (generic flowchart).

Assessment Schemes (EQAS) by setting up reference methods for their control materials in the post-market surveillance of clinical laboratory performance [12].

A detailed definition of the quantity to be measured (the "measurand") constitutes an indispensable part of any analytical RS. In Laboratory Medicine, many hundreds of different analytes are measured or determined. With regard to the implementation of traceability, it is, however, important to differentiate between analytes which are well defined chemical entities and are traceable to International System (SI) units (type A quantities) and analytes which are rather heterogeneous in human samples and are not directly traceable to SI units (type B quantities) [8]. Type A analytes represent a relatively small number of well defined compounds (approximately 65), which belong to 'classical' clinical chemistry, including electrolytes, minerals, metabolic products (such as cholesterol, creatinine, etc.), steroid hormones, and vitamins. Test results for these measurements are nowadays expressed in terms of moles per litre, which represent the accepted system of SI units. However, for many hundreds of measurable quantities, designated as type B analytes, e.g. most of the proteins – usually measured by some kind of immunochemical techniques –, test results are not expressed in terms of SI units, but in terms of arbitrary units, for example WHO international units or mass units of a preparation belonging to a manufacturer.

For type A analytes, reference materials containing the analyte as a pure compound can usually be prepared and reference measurement procedures which specifically measure the analyte and are independent of routine analytical principles can be developed. Consequently, for many of these analytes, RS are already available. An example of a RS for type A analytes is that for creatinine in blood serum [13]. Creatinine is a chemical substance whose entity can be unequivocally defined as a single species. The unit for the measurement of the amount-of-substance concentration of creatinine is mol/L and gravimetry can therefore be used for the value assignment of a primary reference material prepared with the pure substance. The reference measurement procedure for creatinine, applying the isotope dilution-mass spectrometry principle, is directly calibrated against this primary reference material. Using this reference procedure, reference laboratories working under well defined performance conditions are able to assign values to commutable secondary reference materials. Manufacturers then may use these materials for calibration of a routine method, leading to traceable results for the end user's routine method.

For type B analytes, the implementation of standardization is in general more difficult. Much scientific work still has to be done before RS for this type of quantities can be established. Since they are heterogeneous and their composition in human body fluids varies, all reference materials for type B analytes are, by definition, only surrogates for the analytes measured in patient samples. While such materials may resemble to some extent the typical heterogeneous mixture of the analyte present in the human fluids, they often may represent only an 'average' condition [14]. Furthermore, for type B analytes, reference measurement procedures independent of routinely employed analytical principles are currently lacking in majority of cases.

Thus, the value assignment of candidate secondary reference materials is frequently problematic [15]. As a consequence, manufacturers prepare their own calibrators that are often not available to other manufacturers and assign values to the selected preparation on a mass basis. This can lead to a disagreement between results from different commercial assays.

As many type B analytes are very important parameters in the medical field, such as in oncology, endocrinology, and virology, the establishment of a RS for these measurands is urgently needed. For these analytes in particular, the traceability model emphasizes the importance of the definition of the measurand. For complex substances the definition may not be so clear because of their potential intrinsic or acquired heterogeneity. One way to circumvent the issue of heterogeneity of type B analytes is to define the measurand as a unique, invariant part of the molecule that is common to all components of the mixture present in blood. Methods used for the development of commercial assays should, without distinction, recognize this common part with a consequent increase in the homogeneity of assay reactivity. Using this approach, a number of significant efforts have recently been initiated to standardize measurement results for type B analytes. An excellent example is the IFCC project for standardization of measurement of haemoglobin A_{1c} (HbA_{1c}). Firstly, HbA_{1c} was defined as hemoglobin molecules having in common a glycated amino-terminal hexapeptide of the hemoglobin β-chain. Two equivalent reference measurement procedures specifically measuring this hexapeptide were then developed, using either a combination of HPLC and electron-spray mass spectrometry or, alternatively, a two-dimensional approach using HPLC and capillary electrophoresis [16]. Finally, secondary reference materials have been prepared and their HbA_{1c} values certified by a network of reference laboratories, allowing the establishment of a complete RS [17].

A special class of analytes is represented by the enzymes, defined in terms of the so-called "catalytic amount", which is the amount of an agree-upon substrate converted to product in an agreed-upon measurement system. Theoretically, enzymes defined by substrate conversion do not belong to the SI category of analytes, even if the definition of "katal" may suggest so. Rather the measurand may well be part of a family and, in some cases, may be totally or partially unknown. Hence, the problems of mixture analysis and unknown entities, typical of type B analytes, may also apply for enzymes defined by substrate conversion. Compared with other analytes, the numerical results of catalytic activity measurements depend entirely on the experimental conditions under which the measurements are made [6]. Therefore, in the standardization of enzyme assays, a reference measurement procedure, which defines the conditions under which a given enzyme activity is measured, occupies the highest level of the traceability chain [18]. Complete RS, comprising reference measurement procedures, reference materials, and reference laboratories, are currently available for alanine aminotransferase (ALT), aspartate aminotransferase (AST), creatine kinase (CK), lactate dehydrogenase (LD), γ-glutamyltranspeptidase (GGT), and amylase [19].

The Joint Committee on Traceability in Laboratory Medicine

Since the development of metrologically sound RS is a complicated and expensive process, it is clear that the objective of improving standardization in Laboratory Medicine will only be achieved if the problems are dealt with not on a national level but through international cooperation. This was the reason for the creation of the Joint Committee for Traceability in Laboratory Medicine (JCTLM), established in 2002 under the auspices of the Bureau Internationale des Poids et Mesures (BIPM), the IFCC and the International Laboratory Accreditation Cooperation (ILAC). Thanks to its activity, a list of higher order reference materials and reference methods for analytes measured in Laboratory Medicine, identified by a thorough review process for conformity with appropriate ISO standards, is now publicly available. JCTLM has also published a list of reference laboratories, which fulfill the established selection criteria and are able to deliver a reference measurement service. Using these validated RS, industry can assign traceable values to commercial calibrators. Clinical laboratories, which will use routine procedures and these validated calibrators to measure human patient specimens, may finally obtain comparable results. Then, the traceability requirement, as formulated by the IVD directive of the EU and in the corresponding ISO standards, can finally be implemented in practice (Fig. 2).

Further issues in the implementation of traceability

As soon as a new RS is adopted and implemented, clinical validation of the correctly calibrated routine methods (the IVD products sold onto the market) should take place. In specific cases, in order to maintain the value of clinical experience, correlation of measurement results obtained with the new calibration to results of measurements obtained with the previous calibration should be established. Adjusting the decision-making criteria is of major importance since, even if from a metrological point of view the routine method was biased,

clinicians can still reach correct clinical decisions if the decision-making criteria they apply incorporate the same bias. In contrast, they could arrive at incorrect clinical decisions if patient results are true with regard to the RS, but the decision-making criteria are only valid by using the previous calibration for the test.

The prostate-specific antigen (PSA), one of the most common tumor markers, provides a practical example. Currently, two sources of calibration are in common use for PSA. One is based on the traditional calibration scheme that produces results consistent with the first PSA assay marketed by Hybritech, used to establish the clinically relevant PSA cutoff of 4.0 μg/L. The second calibration approach provides traceability to the WHO International Reference Preparation 96/670, thus fulfilling the IVD directive directions. Recalibrating a PSA assay from an original 'Hybritech' calibration to the WHO calibration results, however, in about 20% lower PSA values indicating that the 4.0 μg/L cutoff would not provide optimal clinical efficacy for the WHO standardized assays. However many clinicians are unaware that different PSA results are produced for the same patient specimen if tested by assays using different calibration schemes and this may result in potential variation of clinical interpretation, with adverse consequences for patients. For instance, a study has shown that use of the traditional 4.0 cutoff for a WHO calibrated assay potentially missed 19% of patients candidates for prostate biopsy because of suspected cancer [20].

In the case of HbA_{1c}, reliable, linear relationships between results traceable to the IFCC RS and previous routine methods were demonstrated allowing the conversion of analytical and clinical data from one system to another [21]. In practice it is therefore possible to translate target values generated in previous landmark clinical studies, using methods not traced to the IFCC system, in order to maintain the clinical experience. In addition, use of the SI unit as the unit of measurement for HbA_{1c}, namely "mmol/mol", can avoid confusion in the recalculation of the old HbA_{1c} targets to the new IFCC standardized results if clinical laboratories wish to implement HbA_{1c} results traceable to the IFCC RS [22]. Other advantages of this approach include a positive impact of a change of scale of reported HbA_{1c} results thereby allowing clinicians and diabetic patients to better understand HbA_{1c} changes and the increased potential for future use of HbA_{1c} as a diagnostic tool.

Other important issues concerning the implementation of a metrologically-correct approach for result standardization are still to be improved. Firstly, a clear definition of the clinically allowable error of measurements is required. Since methods with a total error of zero do not exist, agreement is required as to what percentage of misclassification of patients is acceptable and whether it is preferable to avoid false positive or false negative classification. Even though statistical validation criteria for analytical performance can be easily defined, tolerable deviations for clinical use are often undefined. The scientific community has to be aware that the absence of specifications derived from clinical needs for validation of metrologically traceable calibrations might result in a large gray zone with respect to the extent of traceability expected from IVD manufacturers, partially or totally invalidating its theoretical advantages [23].

What has JCTLM delivered?

The World´s only quality-assured database of:

a) Higher Order Reference Materials
b) Higher Order Reference Measurement Procedures
c) Laboratory Reference Measurement Services

http://www.bipm.org/en/committees/jc/jctlm/jctlm-db/

For use by (primarily):

a) IVD industry (to ensure that results produced by IVDs are traceable to)
b) Regulators (to verify that results produced by IVDs are traceable to)

Fig. 2. The Joint Committee on Traceability in Laboratory Medicine (JCTLM) promoting traceability of the in vitro diagnostic (IVD) products and standardization in Laboratory Medicine.

In a recent study involving 70 laboratories from three European countries, enzyme assays from six major manufacturers were assessed for traceability to IFCC RS through a commutable serum-based material targeted with ALT, AST, CK, GGT, LD, and amylase reference methods [24]. Results from commercial methods were assessed by a system using a maximum allowable error derived from the desirable analytical performance that is based on the biological variation model. Of these enzyme measurements, ALT results were relatively good. For AST, CK and GGT not all manufacturers would fully comply. LD and amylase measurements have, however, major drawbacks. Collectively, these observations suggest that a number of routine analytical systems for serum enzymes are still significantly biased when compared to the reference methods, even if all manufacturers in the European market declare their traceability to the RS.

Another important issue relates to the post-market surveillance of the performance of IVD medical devices. This should be one of the major tasks of our profession through the organization of appropriate EQAS. However the applicability of the true value concept in EQAS requires the availability of control materials with target values assigned by laboratories using reference methods and that these materials behave exactly as human patient specimens [25]. True value assignment to commutable EQAS materials facilitates objective evaluation of the performance of IVD devices, together with an accuracy-based (instead of inferior consensus-based) grading of the competency of participating clinical laboratories. Using this approach, results of the 2007 German EQAS has shown that a large number of laboratories measuring serum creatinine using alkaline picrate-based assays are still significantly inaccurate, particularly at lower creatinine concentrations [26].

In conclusion, the time has come to provide compatible numerical results from all clinical laboratories in the world to permit common decision-making criteria. With this goal in mind, the scientific community in Laboratory Medicine is dedicated to the necessity of standardizing measurement results on the basis of trueness by consistent application of metrological concepts. A number of projects are currently underway using the RS approach to add quality and value to Laboratory Medicine.

References

[1] Örnemark U, Van Nevel L, Smeyers P, Harper C, Taylor PDP. The International Measurement Evaluation Program IMEP-17. Trace and minor constituents in human serum. EUR 20694 EN. Report to participants. Part 2: methodology and quantity specifications. *www.imep.ws*.

[2] Panteghini M, Forest JC. Standardization in laboratory medicine: new challenges. Clin Chim Acta 2005;355:1–12.

[3] Klee GG. Clinical interpretation of reference intervals and reference limits. A plea for assay harmonization. Clin Chem Lab Med 2004;42:752–7.

[4] Büttner J. The need for accuracy in Laboratory Medicine. Eur J Clin Chem Clin Biochem 1995;33:981–8.

[5] ISO 17511:2003. In vitro diagnostic medical devices — measurement of quantities in biological samples — metrological traceability of values assigned to calibrators and control materials. ISO, Geneva, Switzerland.

[6] ISO 18153:2003. In vitro diagnostic medical devices — measurement of quantities in biological samples — metrological traceability of values for catalytic concentration of enzymes assigned to calibrators and control materials. ISO, Geneva, Switzerland.

[7] Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. Official Journal of the European Communities 1998(Dec 7);L331:1–37.

[8] Panteghini M. Traceability, reference systems and result comparability. Clin Biochem Rev 2007;28:97–104.

[9] Franzini C, Ceriotti F. Impact of reference materials on accuracy in clinical chemistry. Clin Biochem 1998;31:449–57.

[10] Vesper HW, Miller WG, Myers GL. Reference materials and commutability. Clin Biochem Rev 2007;28:139–47.

[11] Clinical and Laboratory Standards Institute. Metrological traceability and its implementation; A report. CLSI document X5-R. Wayne, PA: CLSI; 2006.

[12] Miller WG. Specimen materials, target values and commutability for external quality assessment (proficiency testing) schemes. Clin Chim Acta 2003;327:25–37.

[13] Panteghini M, Myers GL, Miller WG, Greenberg N. The importance of metrological traceability on the validity of creatinine measurement as an index of renal function. Clin Chem Lab Med 2006;44:1187–92.

[14] Panteghini M. Current concepts in standardization of cardiac marker immunoassays. Clin Chem Lab Med 2004;42:3–8.

[15] Stenman UH. Immunoassay standardization: is it possible, who is responsible, who is capable? Clin Chem 2001;47:815–20.

[16] Jeppsson JO, Kobold U, Barr J, et al. Approved IFCC reference method for the measurement of HbA1c in human blood. Clin Chem Lab Med 2002;40:78–89.

[17] Weykamp C, John WG, Mosca A, et al. The IFCC reference measurement system for HbA1c: a 6-year progress report. Clin Chem 2008;54:240–8.

[18] Panteghini M, Ceriotti F, Schumann G, Siekmann L. Establishing a reference system in clinical enzymology. Clin Chem Lab Med 2001;39:795–800.

[19] Infusino I, Bonora R, Panteghini M. Traceability in clinical enzymology. Clin Biochem Rev 2007;28:155–61.

[20] Link RE, Shariat SF, Nguyen CV, et al. Variation in prostate specific antigen results from 2 different assay platforms: clinical impact on 2,304 patients undergoing prostate cancer screening. J Urol 2004;171:2234–8.

[21] Hoelzel W, Weykamp C, Jeppsson JO, et al. IFCC reference system for measurement of hemoglobin A1c in human blood and the national standardization schemes in the United States, Japan, and Sweden: a method-comparison study. Clin Chem 2004;50:166–74.

[22] Panteghini M, John WG. Implementation of haemoglobin A1c results traceable to the IFCC reference system: the way forward. Clin Chem Lab Med 2007;45:942–4.

[23] Thienpont LM, Van Uytfanghe K, Rodriguez Cabaleiro D. Metrological traceability of calibration in the estimation and use of common medical decision-making criteria. Clin Chem Lab Med 2004;42:842–50.

[24] Jansen R, Schumann G, Baadenhuijsen H, et al. Trueness verification and traceability assessment of results from commercial systems for measurement of six enzyme activities in serum. An international study in the EC4 framework of the Calibration 2000 project. Clin Chim Acta 2006;368:160–7.

[25] Thienpont LM, Stöckl D, Friedecky B, Kratochvila J, Budina M. Trueness verification in European external quality assessment schemes: time to care about the quality of samples. Scand J Clin Lab Invest 2003;63:195–202.

[26] Panteghini M on behalf of the IFCC Scientific Division. Enzymatic assays for creatinine: time for action. Clin Chem Lab Med 2008;46:567–72.