

Reference Materials and Commutability

***Hubert W Vesper,¹ W Gregory Miller,² Gary L Myers¹**

¹Centers for Disease Control and Prevention, Division of Laboratory Sciences, Atlanta, GA 30341, ²Virginia Commonwealth University, Richmond, VA 23298 USA

*For correspondence: Dr Hubert Vesper e-mail: HVesper@cdc.gov

Abstract

Maintaining accurate laboratory measurements over time is crucial for assuring appropriate patient care and disease management. Accurate results over time and location are achieved by standardising measurements and establishing traceability to a reference system. Reference materials are key components of such reference systems and for establishing traceability. Commutability of reference materials is a critical property to ensure they are fit for use.

Commutability is defined as the equivalence of the mathematical relationships between the results of different measurement procedures for a reference material and for representative samples from healthy and diseased individuals. This material characteristic is of special importance for measurement procedures that are optimised for measuring analytes directly in patient samples. The commutability of a reference material is measurement procedure specific and its assessment requires special experimental designs.

This review explains the importance of commutability and summarises different experimental approaches described in the literature that have been used to assess the commutability of reference materials in clinical chemistry.

Introduction

The goal of standardisation in laboratory medicine is to assure that results from measurements in patients' samples are accurate, independent of the measurement procedure used or the location and time of testing. Accurate results are necessary to enable the use of laboratory measurements in clinical guidelines for patient care and disease management. Accurate results allow laboratory data to be collected and mined from different sources to identify public health needs, to monitor public health programs and to evaluate the effectiveness of these programs. Finally, accurate results produced in different research laboratories enable an efficient translation of research findings into information helpful in patient care.

A prominent example for the impact and importance of standardisation is cholesterol measurements. This effort started in the early 1960s when the lack of interlaboratory comparability of results became a major concern for clinical investigators conducting epidemiologic work.¹ These concerns led to the initiation of the first Cooperative Cholesterol Standardization Program. This program developed a hierarchy of approved methods and materials for the measurement of cholesterol. Standardised cholesterol measurements used in epidemiologic studies and clinical trials enabled identification

of risk factors associated with coronary heart disease and development of clinical practice guidelines for reducing the risk and incidence of heart disease. These trials led to national efforts for lowering cholesterol such as the National Cholesterol Education Program (NCEP), established in 1985 by the National Heart, Lung and Blood Institute in the USA.^{2,3} To ensure success of this effort, the NCEP recommended that all cholesterol measurements be standardised and traceable to a common reference system that had been established by the cholesterol standardisation program at the Centers for Disease Control and Prevention (CDC).⁴

The concepts established in these early clinical laboratory standardisation efforts created the basis for standardisation programs for a number of analytes. The key components of a standardisation program are reference measurement procedures, reference materials (RMs) and a system that enables measurement results obtained with the end-user's routine measurement procedure to be traceable to these RMs and reference measurement procedures.

While issues related to traceability and reference methods are discussed elsewhere in this issue, this review focuses on RMs with special emphasis on commutability, a characteristic of

particular importance for RMs intended for use with routine measurement procedures in laboratory medicine.

Reference Material, Definition and Use

Maintaining accuracy of laboratory measurement results over time is achieved by establishing and maintaining trueness of the measurement results i.e. traceability to a reference system, and a defined precision over time. The precision of a measurement result depends on the performance of the measurement procedure including the operator performance and is determined by assessing the results of repeated measurements performed under defined conditions. Procedures to assess the precision of measurement procedures are described in guidance documents such as Clinical and Laboratory Standards Institute (CLSI) EP5.⁵ The trueness of measurement results, defined as closeness of agreement between the average value obtained from a large series of test results and an accepted reference value⁶ is assessed, as the definition states, by comparing the measurement results obtained with the procedure in question with an established reference. The established reference is either a reference measurement procedure or a RM characterised with a reference measurement procedure. Thus, RMs are used to establish trueness of measurement procedures through calibration or to assess the trueness of the calibration of a measurement procedure.

Reference materials are defined as “materials, sufficiently homogeneous and stable with respect to one or more specified properties, which have been established to be fit for their intended use in a measurement process; **NOTE 1:** RM is a generic term; **NOTE 2:** Properties can be quantitative or qualitative, e.g. identity of substances or species; **NOTE 3:** Uses may include the calibration of a measurement system, assessment of a measurement procedure, assigning values to other materials, and quality control; **NOTE 4:** An RM can only be used for a single purpose in a given measurement”.⁷

As note 1 in this definition indicates, RM can be considered an umbrella term for all materials used to calibrate a measurement procedure or to assess the trueness of results obtained with measurement procedures. This umbrella would include materials such as method specific calibrators, trueness controls and certified RMs (CRMs).⁸ The main difference between the materials mentioned is in the uncertainty of the assigned value. The uncertainty of calibrators and trueness controls is typically larger than that of a CRM. A variety of naming systems have been used to describe RMs to imply different levels of uncertainty such as ‘primary RM’, ‘secondary RM’ or ‘higher order RM’, ‘lower order RM’, ‘primary calibrator’ and ‘secondary calibrator’. This imprecise nomenclature has resulted in a wide variety of terms used in

the current literature and in efforts by standards organisations to clarify the terminology.⁹⁻¹¹ The CLSI provides an online harmonised terminology database that contains a compilation of internationally accepted terminology.¹²

Note 3 of the definition of RMs provides information on uses of RMs. The most common use is for calibration and as a trueness control to verify calibration. The goal of traceability is to have results obtained by a calibrated routine measurement procedure traceable to the highest available level of the calibration hierarchy.^{13,14} The highest available level is frequently a single-substance RM, or set of RMs. However, such RMs are frequently not suitable for use with routine measurement procedures, because those procedures are optimised for use with patient samples and thus require samples with the same or similar matrices as patient samples. It needs to be pointed out, however, that RMs with matrices similar to patient samples are not necessarily suitable for routine measurement procedures. Some of these matrix-based RMs are intended for use with higher order reference measurement procedures and not for use with routine measurement procedures (e.g. serum-based RM CRM 576, 577 and 578 for oestradiol¹⁵). In such situations, different matrix-based RMs are needed that are suitable for the different measurement procedures used in a traceability chain for the same measurand. Because the appropriate intended use for a RM may not be apparent, it is advisable to always assess the suitability of RMs for use with a particular measurement procedure before they are used for calibration or trueness assessment.

The use of RMs as trueness controls in external quality assessment (EQA) schemes, although desirable, is rather rare. However, trueness controls in EQA are very useful tools in assessing the impact of standardisation programs as demonstrated by Little for the National Glycohemoglobin Standardization Program (NGSP).¹⁶ In this report, trueness-based proficiency testing with native clinical samples was able to show marked improvements over the years in the comparability of HbA_{1c} results with most standardised measurement procedures being within 0.8% HbA_{1c} from the NGSP target values in 2002. The development of trueness controls for EQA programs is especially challenging, because these materials need to be suitable for a wide range of different measurement procedures and at the same time need to be commutable with native clinical samples. Use of trueness controls in EQA has been accomplished by using specimens prepared from pooled native patient samples with matrices that are essentially the same as those of patient samples.¹⁷⁻²¹

Whether RMs are used as calibrators in certain steps of the metrological traceability chain or as trueness controls, their fitness for the intended use, as stated in the definition for RMs,

needs to be established. One important criterion for establishing fitness for use is the commutability of the material.

Commutability, Definition and Impact

The term “commutability” was first used to describe the ability of a reference or control material for enzyme measurements to have interassay properties comparable to the properties demonstrated by authentic clinical samples when measured by more than one analytical method.^{22,23} This description was later expanded from enzymes to other analytes and commutability is now defined as the equivalence of the mathematical relationships between the results of different measurement procedures for a RM and for representative samples from healthy and diseased individuals.²⁴ A number of different definitions of commutability have been described in standards documents^{13,25} and scientific literature.²⁶ Though the basic principles are the same in all definitions, they differ in the description of the samples and materials used to assess commutability, as well as in the description of the relationship between the measurement procedures used for the commutability assessment.

The definition of commutability implies that it is a measurement procedure-specific characteristic and that any statements about the commutability of a RM require further information about the specific measurement procedures for which it was found to be commutable. Furthermore, since commutability is a method-specific characteristic, RMs can be commutable for some measurement procedures but may be non-commutable for others. The applicability of a RM for general use as a calibrator or a trueness control depends on the number of measurement procedures for which it was found commutable. Christenson et al. reported in a study assessing commutability of two cardiac troponin I materials among 15 measurement procedures that commutability was observed for 39% and 45% of measurement procedures, respectively. The authors concluded that the proportion of measurement procedures demonstrating commutability was too low for either of these materials to be used as a common calibrator.²⁷

A RM would be considered commutable when a measurement procedure produces the same result for a RM as it does for an authentic patient sample that contained the same analyte concentration. Measurement procedures calibrated with commutable RMs will produce results for clinical samples that are equivalent among all procedures, i.e. the results are traceable to the reference system and there is no calibration bias among the measurement procedures. In contrast, measurement procedures calibrated with materials that are non-commutable will show a measurement bias for clinical samples and results will not be equivalent among all procedures. Consequently, biases observed among measurement procedures calibrated

with materials of unknown commutability cannot be properly attributed to genuine measurement procedure problems or to problems related to the material used for calibration. The condition of unknown commutability is of special importance when RMs are used to assess acceptability of performance of measurement procedures. A study by Eckfeldt et al. investigated the effects of EQA materials on the measurements of cholesterol, uric acid, calcium and potassium measurements and showed that material-related biases lead to measurement procedures being considered non-acceptable, while data obtained with native patient samples concluded the agreement was acceptable.²⁸ Cattozzo et al. reported that recalibration with non-commutable RMs caused results for native clinical samples to change from having pathological values to non-pathological values and vice versa.²⁹

Non-commutable RMs have a direct impact on the trueness of clinical measurements. Franzini and Ceriotti summarised different studies on the impact of non-commutable reference and control materials and found that up to 82% of commercial control materials were non-commutable for 10 common analytes.³⁰ Another study by Ross et al. using data obtained from the College of American Pathologists proficiency testing program found that 69% of material/method combinations for 11 analytes showed effects related to non-commutability. Because of non-commutability, reference method target values were only suitable for trueness evaluation for 32% of the analyte/routine method combinations.¹⁷ Similar findings were reported by Thienpont et al. in a study comparing results on serum glucose and cholesterol obtained with fresh-frozen single-patient specimens and lyophilised materials.³¹ In another study by Cattozzo et al. investigating the effects of myoglobin and creatine kinase MB, an increase in intermethod differences was observed when non-commutable materials were used as RMs.³²

Although the impact of non-commutable RMs is well documented and international standards and guidance documents require RMs to be validated for commutability,^{13,25} the assessment of commutability of RMs is still not routinely performed.^{24,33}

Non-commutability of RMs is commonly attributed to differences between the material's matrix and that of native patient samples.^{20,23,28,29,33} Therefore, the presence of matrix effects has frequently been associated with non-commutability of materials. The matrix, however, includes all components of a material except the analyte itself and a matrix effect is defined as the influence of a property of the sample, independent of the presence of the analyte, on the measurement and thereby on the value of the measurable quantity.¹³ This definition is very generic and applies to all materials including authentic

clinical specimens. Using this definition, interferences of other normally occurring compounds, such as bilirubin, would be considered matrix effects. Interferences from endogenous substances can occur in authentic samples as well as processed RMs. The so called ‘matrix-effects’ in reference or control materials that cause non-commutability of these materials refers to differences that are observed only in the RM but not in authentic clinical samples. Matrix-effects that cause non-commutability of a RM need to be distinguished from influences of the specificity of a measurement procedure for the measurand. For example, different measurement procedures may target different epitopes in a protein measurand in which case there are actually different analytes that are reported as the same measurand. In practice however, it can be difficult to distinguish whether non-commutability of a RM is caused by matrix-effects or lack of specificity of the measurement procedures.

The reasons for non-commutability of a material are not predictable and are highly specific to a measurement procedure. Material handling, such as time spent in contact with red blood cells or clot during blood collection, reconstitution of serum from plasma, dialysis, concentration, freeze-thaw cycles, filtration and lyophilisation, can affect the matrix of the material, for materials used for lipid and lipoprotein standardisation.²⁰ Another study by Thienpont et al. found that processing of native serum (e.g. sterile filtration, storage before aliquotting and freezing) may disturb the equilibrium between free and protein-bound thyroid hormone and hence jeopardise the commutability of a RM prepared from native sera with minimal processing.¹⁸ Other modifications that can compromise commutability of materials are supplementation with human or non-human analytes, as commonly performed to adjust the quantity present, because a non-native form of the analyte or impurities may be introduced.^{34,35} However, supplementation with human or non-human materials does not necessarily result in non-commutability of materials as shown in one study by Uldall et al. where supplemented materials were tested for 50 analytes and non-commutability was found for only eight analytes.³⁶

Careful preparation of materials can minimise matrix effects and thus non-commutability. A robust protocol to prepare an off-the-clot frozen serum pool is described in CLSI document C37-A.³⁷ Materials prepared according to this protocol have been evaluated by Cobbaert et al.¹⁹ for several lipoprotein and apolipoprotein measurands and were found to be superior in commutability to patient serum pools prepared by less-stringent processes and to commercially available human serum-based materials. This protocol was used to prepare frozen serum creatinine RMs that were validated for commutability.²¹

Assessment of Commutability

Different approaches for assessing the commutability of materials have been described in the literature. They were developed to address different situations such as the availability or lack of reference measurement procedures, and thus no single approach is currently recommended for the assessment of commutability. The existing approaches use descriptive statistics or some form of regression analysis to compare the numeric relationships among methods for authentic patient samples to those for RMs. All assessment procedures for commutability are based on determining the mathematical relationship and distribution of results observed for native patients’ samples measured by two or more measurement procedures, and determining if a reference material is a member of the same distribution.

The approaches using multivariate descriptive statistics utilise principal component or correspondence analysis to compare relationships for authentic patient samples to those for RMs.³⁸⁻⁴⁰ This approach has the advantage that it allows comparison of multiple materials with multiple methods in one graph. Figure 1 shows such a graph in which the results for authentic patient samples form a cluster of points. Points from commutable RMs would fall within the cluster of points for the patient samples, while non-commutable materials would be outside the distribution of patient sample points. The authors suggest correspondence analysis as the more appropriate of the two multivariate statistical techniques, because of its transformation of analytical data in terms of profiles (specimen profiles and method profiles) and the ability to separate specimen from method components. Interpretation is therefore independent of the magnitude of results in contrast to principal component analysis. The multivariate graphical approach does not provide clearly defined numeric criteria for distinguishing commutable from non-commutable materials.

Eckfeldt et al. used linear or polynomial regression to assess the commutability of materials used in EQA schemes.²⁸ This experimental design was adopted and refined as a consensus guideline available as CLSI document EP-14.⁴¹ In this approach, regression analysis is performed to establish the relationship between results obtained from authentic patient samples using two measurement procedures and the two-tailed 95% prediction interval was calculated for the distribution of patient results. Measurement results obtained with a RM are then compared against the 95% prediction interval. Figure 2 shows an example of the regression evaluation. Materials falling within the prediction interval have the same numeric relationship between the two measurement procedures as native patient samples, and those falling outside this interval are considered to have a matrix effect and therefore not be commutable with native patient samples.

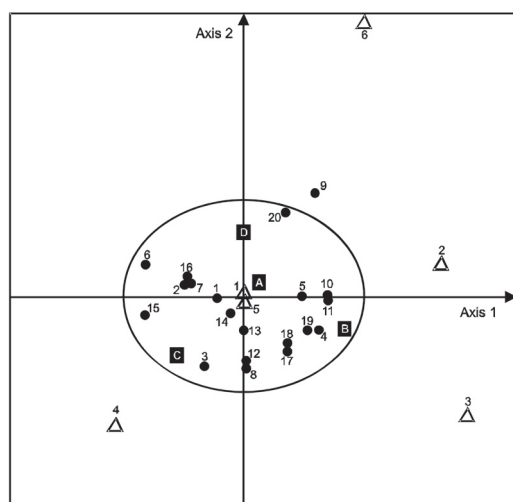


Figure 1. Correspondence analysis. Patient specimens (●) and RM (Δ). Lettered boxes indicate projections of four analytical methods. The axes represent the first two components of the correspondence analysis that account for most of the relationships observed between different specimens. Because correspondence analysis describes relationships between materials in terms of profiles and is independent of the magnitude of the result, the scale of the axis does not have units. The ellipse defines the 95% confidence area describing the multivariate variability in behaviour of patients' specimens with respect to the four methods.³⁸ Reprinted with permission from Rej R, 1993, Arch Pathol Lab Med, 117 pp. 352-64. Copyright © 1993, American Medical Association. All Rights reserved.

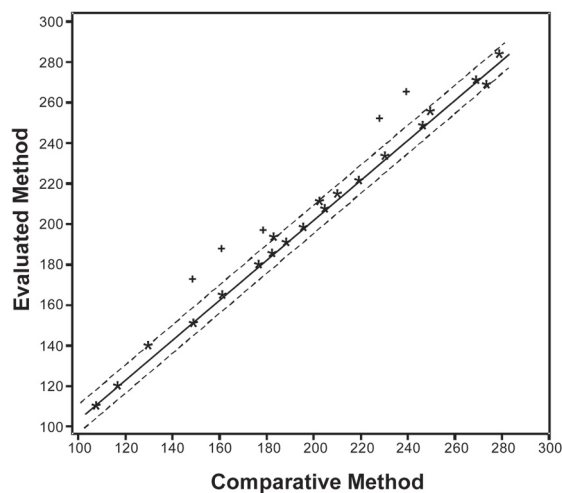


Figure 2. Assessment of reference materials (+) for agreement with patient samples (*) using linear regression with prediction limits (dashed line) according to EP14-A2.⁴¹

An approach based on evaluation of the residuals from regression analysis was introduced by Franzini to evaluate commutability.²⁴ In this approach, regression analysis is

performed between the results obtained with two measurement procedures using authentic clinical samples. Results for a RM measured by the same two procedures are evaluated by plotting its results and determining its residual vs. the regression line for the patient results. The residual is the difference between the value plotted for the RM and the calculated value predicted from the regression analysis for the patient results. The residual for the RM is normalised by dividing its value by the standardised residual ($S_{y,x}$) for the patient samples. Franzini used this approach to evaluate commutability of 27 RMs for 12 analytes in serum. An example is shown in Figure 3.

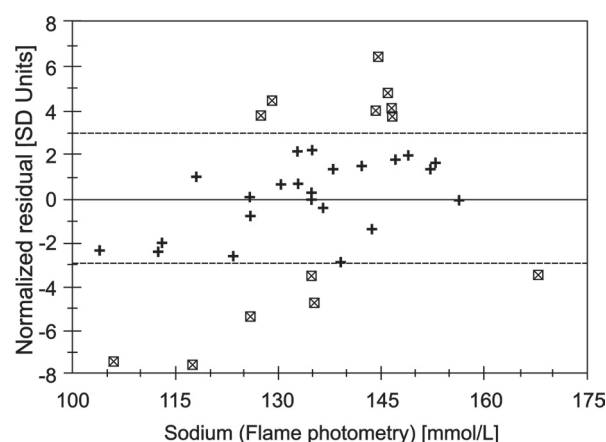


Figure 3. Assessment of commutability of control materials (squares) using normalised residuals and $\pm 3 S_{y,x}$ limits (dashed lines) calculated with patients samples (+) as described by Franzini et al.³⁰ Reprinted from Clinical Biochemistry, 31, Franzini C and Ceriotti F, Impact of reference materials on accuracy in clinical chemistry, pp. 449-57, Copyright 1998, with permission from Elsevier.

Each RM was classified as commutable when its normalised residual was within a $\pm 3 S_{y,x}$ interval. The author points out that the evaluation procedure is sensitive to differences in imprecision of the methods compared. Larger variability between methods observed for patient sera causes a greater number of materials to appear commutable, while smaller variability between methods may cause more materials being considered non-commutable. For the 27 materials investigated, an average of 67% of method/material combinations were found to be commutable ranging from 7% to 100% for individual method/material combinations.

The same normalised residual approach using an acceptance criteria of $\pm 3 S_{y,x}$ has been used by several investigators. Cattozzo et al. applied this procedure to assess commutability of 29 commercial calibration and control materials for serum lipase and reported an overall non-commutability rate of 27% for liquid materials and 47% for lyophilised materials.²⁹

Mosca et al. assessed the commutability of control materials for glycohaemoglobin and reported the normalised residual exceeded $\pm 3 S_{y-x}$ for 2 out of 15 materials.⁴² Dominici et al. used normalised residuals to evaluate 12 control materials for measurement of serum carcinoembryonic antigen measurements and identified 7 materials as non-commutable.⁴³ The application of the normalised residual approach has varied in the number of native patient samples or pools used and is limited by the assumption that the average residual for the distribution of patient samples is appropriate over the concentration range of the RMs.

Baadenhuijsen et al. described an alternate study design that uses normalised residuals to assess commutability of the RMs but simplifies the native sample acquisition logistics by utilising a large number of laboratories organised into pairs.⁴⁴ In this “twin-study” design, locally acquired fresh patient samples were exchanged between each of two laboratories, forming a laboratory “twin” pairing. Each laboratory analysed the patient samples and the candidate RMs. Results were evaluated between the two laboratories using normalised residuals. The results for a number of laboratory pairs (“twins”) were combined to achieve adequate replication and coverage of all methods to be included. Results were aggregated to determine overall commutability among methods represented. With this approach, the investigators were able to assess commutability of 9 materials with 6 methods using 86 different laboratories. This approach overcomes problems related to limited availability of native patient samples when multiple methods need to be compared against each other. It is limited by the fact that different native patient samples are used between each pair of laboratories, and among method and among laboratory variability are included in the data. This approach has been primarily used to validate commutability of control materials for use in EQA schemes and requires substantial coordination among participating clinical laboratories.

Ricos et al. used regression analysis with an evaluation approach based on expressing the residual as a percent to identify non-commutable control materials for creatinine.⁴⁵ They used Passing-Bablok regression to determine the relationship between results for native patient samples. For each control material, a residual was determined as the difference from the value for that material predicted from the regression line for the native patient samples. The residual for the control material was expressed as a percent of the value predicted from the regression relationship and called a bias (in percent). The bias (in percent) for each control material was compared to three criteria to evaluate commutability: $\pm 2 S_{y-x}$ expressed as a percent, the 95% prediction interval expressed as a percent, or a value based on the biological variability. The authors found that the 95% prediction interval and the

biological variability criteria gave concordant conclusions throughout the concentration range, but the $\pm 2 S_{y-x}$ standardised residual criterion classified materials with large matrix bias as being commutable at lower concentrations where the measurement variability was larger. The observation that the $\pm 2 S_{y-x}$ standardised residual criterion was inappropriate can be explained by non-constant relative variance (CV) over the measurement range causing the S_{y-x} , determined for the midpoint of the concentration range of patient samples, to be an incorrect assessment of the distribution of patient results at lower concentrations.

Commutability evaluation based on parallelism and slope ratios have been reported for immunoassay measurement procedures.^{46,47} The assessment of parallelism of response to dilutions of RMs compared to dilutions of native sera is an approach to identify the presence of differences in reactivity with the antibody (differences in epitope). In this concept, the signal responses of a dilution series of native patient samples and RMs are plotted as the (y) variable vs the dilution ratios on the (x) axis. Reference materials with similar slopes for the regression of the measured values on the dilution ratios are said to have “parallel” dilution behaviour. Based on this observation, the diluted samples are considered to have equivalent reactivity with the antibody in the measurement systems evaluated. Non-parallel slopes indicate a difference in antibody response that may be caused by a matrix bias and/or antibody reactivities. When dilutions of a RM have a parallel response to that for dilutions of native clinical samples, the RM can be considered to have comparable response to the antibody used for that particular measurement system, and may be commutable with native patient samples for that measurement procedure. The ratio of the slope of the dilution response for a RM to that of native clinical sample(s) has been compared among different routine measurement procedures as a criterion that a RM was commutable among the different methods.⁴⁸⁻⁵⁰ The observation of parallelism provides an indication that matrix effects are not present. However, the assessment relies on results obtained by diluting and thus modifying the matrix of the RM. Thus, non-parallelism can be caused either by matrix differences among the RMs, by the diluents or by the influence of dilution on the analyte form in solution. Consequently, confirmatory evaluation using native patient samples and non-diluted RM is necessary to conclude the RM is in fact commutable.

All evaluation procedures based on regression analysis are subject to limitations of regression analysis. The regression analysis must be appropriate for the relationship between measurement procedures. For example, if linear regression is used when the relationship is non-linear, then an inappropriate measure of dispersion for the patient sample results will be

observed that will cause incorrect acceptance criteria for the mathematical relationship between measurement procedures. When the relationship is non-linear or the CV is not constant, the results can be partitioned into segments over which the linearity or constant CV assumption is valid, or non-linear regression or weighted regression procedures can be used. Failure to recognise the influence of non-linearity or non-constant CV can cause incorrect assessment that a RM is a member of the distribution of patient results and therefore tests for linearity and constant CV should always be performed.

Ordinary linear regression, frequently used in commutability studies, has limitations because it assumes variation only in Y values, constant CV over the measurement range and is affected by the range and magnitude of numeric values with a few large numbers having disproportionate influence on the statistical coefficients. Due to these limitations ordinary linear regression is not always suitable for comparisons between laboratory measurements that have variability in both X and Y variables and may have non-constant CV and/or a relatively small range of numeric values.

Deming regression has the advantage of allowing variability in the values for both X and Y variables, is less sensitive to a small range of numeric values, but requires constant CV over the measurement range.⁵¹ Weighted Deming regression can be used to overcome the effect of non-constant CVs on the regression parameters. Non-parametric linear regression analysis as described by Passing and Bablok has been suggested to determine the relationship between two methods and has the advantage of being insensitive to extreme values and does not require constant CVs over the measurement range.⁵² However, Passing-Bablok regression analysis gives larger confidence intervals than parametric procedures and may result in inappropriate acceptance criteria based on the distribution of patient results. Consequently, RMs may be considered commutable using non-parametric procedures and non-commutable using parametric procedures. Studies assessing the effect of different regression procedures on the outcome of commutability evaluation have not been described in the literature.

An important consideration when determining acceptance criteria is the intended use of the RM. The uncertainty in a commutability decision should be smaller when a RM is intended to be used for calibration of a measurement procedure than when it is intended to be used as a trueness control or in an EQA program. The uncertainty can be reduced by using a statistical description of the distribution of patient sample results based on 80 or 90% probability rather than the commonly used 95.5% (± 2 SD). There are no published evaluations of the influence of acceptance criteria on use of

RMs as method calibrators vs as trueness controls or EQA samples.

Summary

The appropriate characterisation of RMs, especially those materials intended to be used with routine measurement procedures, must carefully address fitness-for-use for all methods for which the material is intended to be used. Commutability is a critical requirement to avoid introducing unintended, and sometimes undetected, bias in patients' results when using a RM.

Different approaches to assess the commutability of a RM have been described. All are based on determining the mathematical relationship and distribution of results observed for native patients' samples that have been measured by two or more measurement procedures, and determining if a reference material is a member of the same distribution. Different statistical procedures have been used to establish the mathematical relationship and distribution for patients' results and to define acceptance criteria for the RM results to be commutable. The procedure described by CLSI document EP14⁴¹ can be considered the most thoroughly evaluated approach, and is particularly suitable for situations when a reference, or designated comparison, measurement procedure is available. The procedure described by Ricos et al.⁴⁵ has been used for situations when no reference method is available. Any approach that requires separate paired comparisons between all combinations of measurement procedures may become very tedious when several materials need to be screened with many methods, as is frequently the case in EQA programs. For these situations, multivariate approaches allow faster identification of materials that may not be commutable for all the methods evaluated. However, multivariate methods lack defined acceptance criteria to make determination about the commutability of a particular material.

Although valid procedures to evaluate commutability have been described in the literature, they differ markedly in the number of native clinical samples used to define the mathematical relationships among methods and the statistical criteria to define acceptance criteria. There is a need for consensus guidelines to enable consistent assessment of commutability of RMs. The Clinical and Laboratory Standards Institute is currently developing a guideline addressing this need.⁵³ Having consistent procedures is of special importance considering the increasing number of RMs that have been introduced with the intended use to establish or to verify trueness for routine measurement procedures.

Competing Interests: None declared.

References

1. Rifai N, Cooper GR, Brown WV, Friedewald W, Havel RJ, Myers GL, et al. Clinical Chemistry journal has contributed to progress in lipid and lipoprotein testing for fifty years. *Clin Chem* 2004;50:1861-70.
2. Lenfant C. A new challenge for America: the National Cholesterol Education Program. *Circulation*. 1986;73:855-6.
3. Report of the National Cholesterol Education Program Expert Panel on detection, evaluation and treatment of high blood cholesterol in adults. The Expert Panel. *Arch Intern Med* 1988;148:36-9.
4. Current status of blood cholesterol measurement in clinical laboratories in the United States: a report from the Laboratory Standardization Panel of the National Cholesterol Education Program. *Clin Chem* 1988;34:193-201.
5. NCCLS. Evaluation of precision performance of quantitative measurement methods; Approved Guideline – Second Edition. NCCLS document EP5-A2. Wayne, PA USA: NCCLS; 2004.
6. ISO 3534-1:2006. Statistics – vocabulary and symbols – Part 1: General statistical terms and terms used in probability. Geneva, Switzerland: ISO; 2006.
7. ISO Guide 35. Reference Materials – General and statistical principles for certification. 3rd ed. Geneva, Switzerland: ISO; 2006.
8. Emons H. The ‘RM family’-Identification of all of its members. *Accred Qual Assur* 2006;10:690-1.
9. Emons H, Linsinger TPJ, Gawlik BM. Reference materials: terminology and use. Can’t one see the forest for the trees? *Trends Analyt Chem* 2004;23:442-9.
10. ISO. International vocabulary of basic and general terms in metrology (VIM). 3rd Ed. Geneva, Switzerland: ISO; 2004.
11. Dybkaer R. Vocabulary for use in measurement procedures and description of reference materials in laboratory medicine. *Eur J Clin Chem Clin Biochem* 1997;35:141-73.
12. CLSI Harmonized Terminology Database. http://www.clsi.org/AM/Template.cfm?Section=Harmonized_Terminology_Database (Accessed 23 April 2007).
13. 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.
14. CLSI document X5-R. Metrological traceability and its implementation; A report. Wayne, PA USA: CLSI; 2006.
15. European Commission Community Bureau of Reference. EUR 17540 EN. The certification of estradiol-17 β in three lyophilized serum materials. Luxembourg: Office for Official Publications of the European Communities; 1997.
16. Little RR. Glycated hemoglobin standardization - National Glycohemoglobin Standardization Program (NGSP) perspective. *Clin Chem Lab Med* 2003;41:1191-8.
17. Ross JW, Miller WG, Myers GL, Praestgaard J. The accuracy of laboratory measurements in clinical chemistry: a study of 11 routine chemistry analytes in the College of American Pathologists Chemistry Survey with fresh frozen serum, definitive methods, and reference methods. *Arch Pathol Lab Med* 1998;122:587-608.
18. Thienpont LM, Van Uytvanghe K, Marriott J, Stokes P, Siekmann L, Kessler A, et al. Feasibility study of the use of frozen human sera in split-sample comparison of immunoassays with candidate reference measurement procedures for total thyroxine and total triiodothyronine measurements. *Clin Chem* 2005;51:2303-11.
19. Cobbaert C, Weykamp C, Baadenhuijsen H, Kuypers A, Lindemans J, Jansen R. Selection, preparation, and characterization of commutable frozen human serum pools as potential secondary reference materials for lipid and apolipoprotein measurements: study within the framework of the Dutch project “Calibration 2000”. *Clin Chem* 2002;48:1526-38.
20. Miller WG. Matrix effects in the measurement and standardization of lipids and lipoproteins. In: Rifai N, Warnick GR, Dominiczak MH, eds. *Handbook of Lipoprotein Testing*. 2nd ed. Washington, DC USA: AACC Press; 2000. p. 695-716.
21. NKDEP Creatinine Standardization Program. <http://nkdep.nih.gov/labprofessionals/commutabilitystudy.htm> (Accessed 6 May 2007).
22. Fasce CF Jr, Rej R, Copeland WH, Vanderlinde RE. A discussion of enzyme reference materials: applications and specifications. *Clin Chem* 1973;19:5-9.
23. Rej R, Jenny RW, Bretauiere JP. Quality control in clinical chemistry: characterization of reference materials. *Talanta* 1984;31:851-62.
24. Franzini C. Commutability of reference materials in clinical chemistry. *J Int Fed Clin Chem* 1993;5:169-73.
25. ISO 15194:2002. In vitro diagnostic medical devices – Measurement of quantities in samples of biological origin – Description of reference materials. Geneva, Switzerland: ISO; 2002.
26. Ferard G, Edwards J, Kanno T, Lessinger JM, Moss DW, Schiele F, et al. Validation of an enzyme calibrator-an IFCC guideline. *Clin Biochem* 1998;31:495-500.
27. Christenson RH, Duh SH, Apple FS, Bodor GS, Bunk DM, Panteghini M, et al. Toward standardization of cardiac troponin I measurements part II: assessing commutability of candidate reference materials and harmonization of cardiac troponin I assays. *Clin Chem* 2006;52:1685-92.
28. Eckfeldt JH, Copeland KR. Accuracy verification and identification of matrix effects. *Arch Pathol Lab Med* 1993;117:381-6.
29. Cattozzo G, Franzini C, Melzi D’Eril G. Commutability of calibration and control materials for serum lipase. *Clin Chem* 2001;47:2108-13.

30. Franzini C, Ceriotti F. Impact of reference materials on accuracy in clinical chemistry. *Clin Biochem* 1998;31:449-57.
31. Thienpont LM, Stockl D, Friedecky B, Kratochvila J, Budina M. Trueness verification in European external quality assessment schemes: time to care about the quality of the samples. *Scand J Clin Lab Invest* 2003;63:195-201.
32. Cattozzo G, Franzini C, Melzi D'Eril G. Myoglobin and creatine kinase isoenzyme MB mass assays: intermethod behaviour of patient sera and commercially available control materials. *Clin Chim Acta* 2001;303:55-60.
33. Miller WG, Myers GL, Rej R. Why Commutability Matters. *Clin Chem* 2006;52:553-4.
34. Howanitz JH. Review of the influence of polypeptide hormone forms on immunoassay results. *Arch Pathol Lab Med* 1993;117:369-72.
35. Satterfield MB, Welch MJ. Comparison by LC-MS and MALDI-MS of prostate-specific antigen from five commercial sources with certified reference material 613. *Clin Biochem* 2005;38:166-74.
36. Uldall A, Glavind-Kristensen S, Bak S. Preparation of fresh frozen human sera for external quality assessment. *Scand J Clin Lab Invest* 1989;49:11-4.
37. NCCLS document C37-A. Preparation and validation of commutable frozen human serum pools as secondary reference materials for cholesterol measurement procedures; Approved guideline. Wayne, PA USA: NCCLS; 1999.
38. Rej R. Accurate enzyme activity measurements. Two decades of development in the commutability of enzyme quality control materials. *Arch Pathol Lab Med* 1993;117:352-64.
39. Bretaudiere JP, Dumont G, Rej R, Bailly M. Suitability of control materials. General principles and methods of investigation. *Clin Chem* 1981;27:798-805.
40. Bretaudiere JP, Rej R, Drake P, Vassault A, Bailly M. Suitability of control materials for determination of alpha-amylase activity. *Clin Chem* 1981;27:806-15.
41. CLSI document EP14-A2. Evaluation of matrix effects; Approved Guideline. Wayne, PA USA: CLSI; 2005.
42. Mosca A, Paleari R, Made A, Ferrero C, Locatelli M, Ceriotti F. Commutability of control materials in glycohemoglobin determinations. *Clin Chem* 1998;44:632-8.
43. Dominici R, Cabrini E, Cattozzo G, Ceriotti F, Grazioli V, Scapellato L, et al. Intermethod variation in serum carcinoembryonic antigen (CEA) measurement. Fresh serum pools and control materials compared. *Clin Chem Lab Med* 2002;40:167-73.
44. Baadenhuijsen H, Steigstra H, Cobbaert C, Kuypers A, Weykamp C, Jansen R. Commutability assessment of potential reference materials using a multicenter split-patient-sample between-field-methods (twin-study) design: study within the framework of the Dutch project "Calibration 2000". *Clin Chem* 2002;48:1520-5.
45. Ricos C, Juvany R, Alvarez V, Jimenez CV, Perich C, Minchinela J, et al. Commutability between stabilized materials and fresh human serum to improve laboratory performance. *Clin Chim Acta* 1997;263:225-38.
46. Blirup-Jensen S, Johnson AM, Larsen M. Protein Standardization IV: Value transfer procedure for the assignment of serum protein values from a reference preparation to a target material. *Clin Chem Lab Med* 2001;39:1110-22.
47. Johnson AM, Sampson EJ, Blirup-Jensen S, Svendsen PJ. Recommendations for selection and use of protocols for assignment of values to reference materials. *Eur J Clin Chem Clin Biochem* 1996;34:279-85.
48. Panteghini M, Linsinger T, Wu AHB, Dati S, Apple FS, Christenson RH, et al. Standardization of immunoassays for measurement of myoglobin in serum. Phase I: evaluation of candidate secondary reference materials. *Clin Chim Acta* 2004;341:65-72.
49. Kimberly MM, Vesper HW, Caudill SP, Cooper GR, Rifai N, Dati F, et al. Standardization of immunoassays for measurement of high-sensitivity C-reactive protein. Phase I: evaluation of secondary reference materials. *Clin Chem* 2003;49:611-6.
50. Tate JR, Rifai N, Berg K, Couderc R, Dati F, Kostner GM, et al. International Federation of Clinical Chemistry standardization project for the measurement of lipoprotein(a). Phase I. Evaluation of analytical performance of lipoprotein(a) assay systems and commercial calibrators. *Clin Chem* 1998;44:1629-40.
51. Linnet K. Evaluation of regression procedures for methods comparison studies. *Clin Chem* 1993; 39: 424-32.
52. Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. *J Clin Chem Clin Biochem* 1983;21:709-20.
53. CLSI Projects in Development. <http://www.clsi.org/Content/NavigationMenu/CommitteesandProjects/AreaCommitteeonClinicalChemistryandToxicology/ProjectsInDevelopment/default.htm> (Accessed 23 April 2007).