



# IFCC Working Group Recommendations for Assessing Commutability Part 1: General Experimental Design

W. Greg Miller,<sup>1\*</sup> Heinz Schimmel,<sup>2</sup> Robert Rej,<sup>3</sup> Neil Greenberg,<sup>4</sup> Ferruccio Ceriotti,<sup>5</sup> Chris Burns,<sup>6</sup> Jeffrey R. Budd,<sup>7</sup> Cas Weykamp,<sup>8</sup> Vincent Delatour,<sup>9</sup> Göran Nilsson,<sup>10</sup> Finlay MacKenzie,<sup>11</sup> Mauro Panteghini,<sup>12</sup> Thomas Keller,<sup>13</sup> Johanna E. Camara,<sup>14</sup> Ingrid Zegers,<sup>2</sup> and Hubert W. Vesper,<sup>15</sup> for the IFCC Working Group on Commutability

Commutability is a property of a reference material (RM) that relates to the closeness of agreement between results for an RM and results for clinical samples (CSs) when measured by  $\geq 2$  measurement procedures (MPs). Commutability of RMs used in a calibration traceability scheme is an essential property for them to be fit for purpose. Similarly, commutability of trueness controls or external quality assessment samples is essential when those materials are used to assess trueness of results for CSs. This report is part 1 of a 3-part series describing how to assess commutability of RMs. Part 1 defines commutability and addresses critical components of the experimental design for commutability assessment, including selection of individual CSs, use of pooled CSs, qualification of MPs for inclusion, establishing criteria for the determination that an RM is commutable, generalization of commutability conclusions to future measurements made with the MPs included in the assessment, and information regarding commutability to be included in the certificate for an RM. Parts 2 and 3 in the series present 2 different statistical approaches to commutability assessment that use fixed criteria related to the medical decisions that will be made using the laboratory test results.

© 2017 American Association for Clinical Chemistry

## Background

A challenge in laboratory medicine is that lack of agreement between results among different measurement pro-

cedures (MPs)<sup>16</sup> means that results may not be suitable for medical decisions based on those laboratory results. One of the contributors to lack of agreement is using noncommutable reference materials (RMs) as calibrators in the calibration hierarchy for clinical laboratory MPs. In this series of articles, we consider how to decide whether an RM has suitable commutability to be used as a common calibrator in the calibration hierarchy of MPs, as a trueness control material provided by an MP manufacturer to verify calibration, or as an external quality assessment (EQA) or proficiency testing sample. In this part, we describe the common requirements for the experimental design. In parts 2 and 3, we present 2 different statistical approaches for assessment of commutability. Part 2 is suitable for use when an RM is intended for use as a calibrator, a trueness control, or an EQA material (1). Part 3 is suitable for use only when the RM is intended for use as a calibrator (2).

This series of articles recognizes the terms MP and measuring system (MS) as described in the International Vocabulary of Metrology (VIM) (3). An MP refers to a written specification for how a measurement is performed, including a technical description of reagents, calibrators, equipment, instrument, and other details necessary to create and operate an MS that implements those specifications. An MS is a physical in-vitro diagnostic (IVD) medical device manufactured according to the MP specifications and used to make measurements on clinical samples (CSs) to produce results (quantity values) that

<sup>1</sup> Department of Pathology, Virginia Commonwealth University, Richmond, VA; <sup>2</sup> European Commission, Joint Research Centre (JRC), Directorate F, Geel, Belgium; <sup>3</sup> Wadsworth Center for Laboratories and Research, New York State Department of Health, and School of Public Health, State University of New York at Albany, Albany, NY; <sup>4</sup> Neil Greenberg Consulting, LLC, Rochester, NY; <sup>5</sup> Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; <sup>6</sup> National Institute for Biological Standards and Control, A Centre of the MHRA, Hertfordshire, UK; <sup>7</sup> Beckman Coulter, Chaska, MN; <sup>8</sup> Queen Beatrix Hospital, Winterswijk, the Netherlands; <sup>9</sup> Laboratoire national de métrologie et d'essais (LNE), Paris, France; <sup>10</sup> Uppsala, Sweden; <sup>11</sup> Birmingham Quality/UK NEQAS, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK; <sup>12</sup> Research Centre for Metrological Traceability in Laboratory Medicine (CIRME), University of Milan, Milan, Italy; <sup>13</sup> ACOMED Statistic, Leipzig, Germany; <sup>14</sup> National Institute of Standards and Technology, Gaithersburg, MD; <sup>15</sup> Centers for Disease Control and Prevention, Atlanta, GA.

\* Address correspondence to this author at: P.O. Box 980286, Richmond, VA 23298-0286. Fax 804-828-0375; e-mail gmillervcu.edu.

Disclaimer: The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official views or positions of the Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry.

Received June 2, 2017; accepted December 14, 2017.

Previously published online at DOI: 10.1373/clinchem.2017.277525

© 2017 American Association for Clinical Chemistry

<sup>16</sup> Nonstandard abbreviations: MP, measurement procedure; RM, reference material; EQA, external quality assessment; MS, measuring system; VIM, International Vocabulary of Metrology; IVD, in-vitro diagnostic; CS, clinical sample.

are used in research and to make medical decisions for patient care. For example, hundreds of the same MS can be manufactured as an implementation of a single MP; the MSs are used by different clinical laboratories to produce results for CSs. Another example of an MS is one developed by a clinical laboratory that is an implementation of an MP developed by that laboratory for its own use (frequently called a laboratory-developed test). Results for an RM and for CSs measured using different MSs are used to assess commutability of an RM. The conclusion regarding commutability is assumed to be applicable to all other MSs that are implementations of the same MP. For simplicity, in this series of reports we use the term MP when referring to either an MP or results from a specific MS that is an IVD medical device representative of the MP.

Differences in results among MPs may be because of the following sources of error: (a) variation within runs, including trends, caused by variation in performance conditions; (b) variation between runs caused by random error in establishing the calibration response function and interaction between performance conditions and calibrator properties; (c) errors in the assigned values of calibrators; (d) an unsuitable calibration model, for example, a linear model when the relationship between response and concentration (i.e., amount of substance present or other quantity value) is nonlinear; (e) difference in response to influence quantities between the calibrators and the CSs intended to be measured (this difference causes a different relationship between signal and concentration for calibrators than for CSs that is a systematic error referred to as noncommutability of the calibrator; see later definition); (f) differences in response to influence quantities that differ among CSs, referred to as sample-specific effects; and (g) differences in selectivity for the measurand.

When we compare results for CSs measured in 1 run with  $\geq 2$  MPs, as in a commutability assessment, we can only estimate the combined effect of the error sources *b* through *e*. The error sources *c* through *e* contribute to systematic differences in results for CSs between MPs. The systematic difference (bias) between 2 MPs can be expressed by a constant or a function of the concentration.

The effects of error sources *e*, *f*, and *g* above depend on the selectivity of the MPs for the measurand. Error source *e* contributes to the noncommutability of an RM intended for use as a common calibrator for a number of MPs and is the subject of this report. If an RM is commutable with CSs in the measurement responses of MPs, the RM can be used in the calibration traceability schemes for the MPs to reduce the systematic differences among the results for CSs to produce equivalent results for CSs, within stated limits, irrespective of the MP used.

The sample-specific effects in error source *f* above may be a confounding issue in assessing commutability of an RM. If we consider a specific CS, the sample-specific error can be considered as systematic (it cannot be reduced by repeated measurements), but in a population of CSs, the sample-specific errors of individual CSs appear as random variability to which we can assign an SD. Consequently, the magnitude of sample-specific effects can influence a commutability assessment and can be estimated in the experimental design in part 2 of this series (1).

## Definition of Commutability

The VIM defines commutability as a property of an RM, demonstrated by the closeness of agreement between the relation among the measurement results for a stated quantity in this material, obtained according to 2 given MPs, and the relation obtained among the measurement results for other specified materials (3). For medical laboratories, other specified materials are the CSs intended to be measured, and the quantity is usually referred to as the measurand. The CSs may come from healthy volunteers, patients visiting a clinic, or patients with disease. The definition of commutability of an RM concerns 2 MPs. When applied to a calibration traceability scheme as described in ISO 17511, *IVD medical devices—Measurement of quantities in biological samples—Metrological traceability of values assigned to calibrators and control materials*, the definition is applicable for each combination of 2 MPs used in the traceability scheme. However, when the definition is applied to an RM intended for use by several different MPs for a measurand, the experimental design for commutability assessment should include all MPs for which an RM is intended to be used.

Commutability can be stated as a property of an RM that indicates how well an RM mimics the characteristics of typical CSs in an MP for a stated measurand. Commutability is important for RMs used in a calibration hierarchy to ensure that the results for CSs measured in clinical laboratories will be equivalent irrespective of the MP used. Equivalent means within limits determined by the medical requirements for use of a laboratory test result in a patient care decision.

The VIM definition of measurand (the quantity intended to be measured) has limitations when applied to assessment of commutability. The chemical species being measured is the important consideration when selecting samples or qualifying MPs for inclusion in assessing commutability of an RM. In some cases, more than 1 chemical species may be measured either intentionally or because of poor selectivity of an MP. EQA data from commutable samples may identify MPs that do not measure the same quantity. Alternatively, the subgroups in an EQA scheme may represent an inadequately defined measurand

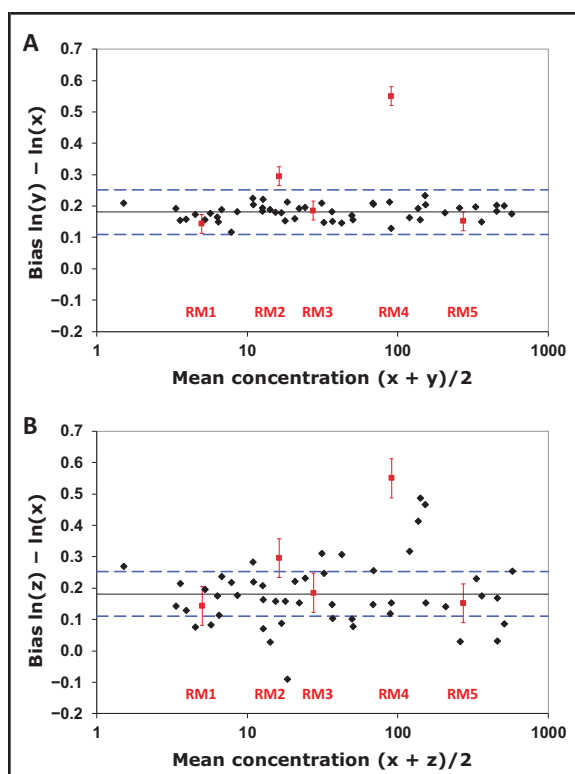
in the EQA material. In such situations, the quantity to be measured may need to be more clearly defined.

### Assessment of Commutability

Assessment of commutability requires the following steps: (a) obtain RM(s) to be evaluated; (b) obtain representative CSs; (c) measure the RM(s) and CSs using the MPs included in the commutability assessment; and (d) apply a procedure to evaluate the commutability of the results for RM(s) in relation to the results for the CSs. Currently available procedures to assess commutability use criteria based only on the statistical distribution of differences in results for CSs between 2 MPs; thus, the criteria can be different for different pairs of MPs with different precision performance (4, 5). Criteria for commutability assessment should be the same for all MPs in an assessment and should be based on the influence of differences between RM and CSs on medical decisions made using the laboratory test results. The statistical approaches in parts 2 and 3 in this series present commutability assessment procedures that can use criteria based on medically relevant differences in results between an RM and CSs (1, 2).

An example of commutability assessment based on the approach in part 2 of this series is shown in Fig. 1. The bias between 2 pairs of MPs ( $x$  and  $y$ ;  $x$  and  $z$ ) is shown for a panel of CSs and for 5 candidate RMs. The relative bias is constant over the concentration interval for the CSs; consequently, the average bias of the CSs is suitable for assessment at each of the RM concentrations examined. The error bars for each RM indicate the uncertainty of the difference in bias between that RM and the average bias of the CSs. The uncertainty consists of 2 components: the uncertainty of the estimate of bias for the CSs and the uncertainty of the estimate of bias for each RM.

The blue dashed lines in Fig. 1 indicate the predetermined maximum allowable bias for the commutability of the RM to be considered sufficient to fulfill medical performance requirements for its intended use. The criterion is the same for all pairs of MPs examined. An RM is commutable when the bias of the RM including the error bars is completely within the blue dashed lines. In Fig. 1A, RM1, RM3, and RM5 are commutable with the CSs because their bias, including the error bars, is within the blue dashed lines. RM2 and RM4 are not commutable with the CSs because the error bars are completely outside the blue dashed lines. In Fig. 1B, the criterion is the same because it is based on medical use requirements. However, there is more scatter in the CS results, suggesting MP  $z$  has a larger random error component. The increased scatter is reflected in larger error bars that now make the commutability of RM1, RM2, and RM5 indeterminate because the error bars are outside of the blue



**Fig. 1.** The bias for logarithmic transformation of concentration [ $\ln(\text{concentration})$ ] between 2 measuring systems is shown for a panel of CSs (black diamonds) and for 5 candidate RMs (red squares) vs the mean concentration measured with each MP.

The error bars for each RM indicate the uncertainty of the difference in bias between each RM and the average bias for the CSs. The bias values for  $\ln$ -transformed results can be multiplied by 100 to give an approximate percent bias. The black solid line is the mean bias (18%) for all the CSs. The blue dashed lines indicate the maximum allowable commutability-related bias for the RM to be considered commutable with the CSs. (A) shows results for 2 MPs ( $x$  and  $y$ ) for which the random errors of the MPs and the sample-specific influences are reasonably small. (B) shows results for 2 MPs ( $x$  and  $z$ ) with more scatter for the CS results, suggesting that MP  $z$  has poorer precision and/or larger sample-specific influences in relation to MP  $x$ .

lines. The assessment of commutability illustrated in the Fig. 1 example reflects data from CSs and MPs that qualified to be included in the assessment as described in sections that follow.

### Individual CSs for Assessment of Commutability

Individual CSs are the matrix that a clinical laboratory MP is designed to measure. Consequently, CSs represent

ideal samples to establish the relationship between different MPs. A commutability assessment is intended to qualify an RM as suitable for use as a calibrator, trueness control material, or EQA sample. A commutability assessment is not intended to evaluate the selectivity of MPs for the measurand. In most cases, an RM is intended to simulate the types of CSs commonly encountered when measuring a given measurand.

CSs should be selected with consideration of MP selectivity limitations. The influence of interfering substances in an individual CS can make it unsuitable for use in a commutability assessment. CSs should be excluded that are known to contain interfering substances or unusual molecular forms, such as found in less common pathologic conditions, when these affect all or most of the MPs in a study. Sourcing more CSs than the minimum needed for statistical assessment is recommended to ensure that enough usable data will be available to meet the statistical requirements for a study. The presence of an interfering substance or unusual molecular form in a CS may not be known until identified as an outlier result in data analysis.

Note that qualification of MPs to be included in a study (see subsequent section of this report) should be done before setting requirements for exclusion of CSs so that MPs with inadequate selectivity will not compromise the usefulness of the evaluation of commutability of an RM. In some EQA applications, an RM may contain an uncommon molecular form for the purpose of challenging the selectivity of MPs. Such special cases are outside the scope of the usual purpose for a commutability assessment and are not addressed in these recommendations.

The interval of concentrations (quantity values) of the measurand in CSs must include that of the RM(s) but does not need to cover the entire measuring interval for the MPs included in the commutability assessment. The number of CSs needed will vary with the experimental design and the performance characteristics of the MPs in the commutability assessment. For measurands that have large differences in measuring interval for different clinical uses, for example, C-reactive protein, commutability assessment may be restricted to 1 of the intended use intervals or may require separate experiments for each interval. The concentrations for the RM(s) and CSs must be within the measuring intervals of the MPs included in an assessment.

CSs must be collected and aliquots prepared, stored, and distributed such that no alteration of the measurand or matrix occurs. For practical reasons, it may be necessary to store samples before a commutability assessment. Use of preservatives, freezing or other storage conditions, pooling, or any modification to an individual CS may affect commutability and should be evaluated for suitability in a preliminary experiment. Some mea-

surands in CSs are not stable for prolonged times under defined storage conditions. Such situations need to be considered in the experimental design for a commutability assessment. Aliquots of individual CSs used for a commutability assessment should be retained when possible to use for commutability assessment of new MPs or MPs that were excluded because they required improvement in performance.

### **Pooled CSs for Assessment of Commutability**

Although individual CSs are preferred for a commutability assessment, sufficient volumes of individual CSs cannot always be obtained to enable aliquoting for distribution and measurement by all MPs in a commutability assessment. In this case, using pooled CSs instead of single donations is a practical solution that may reduce the cost and complexity of a commutability assessment. Pooling dilutes the influence of a substance present in an individual donor sample. Pools prepared from many single donations are more likely to reduce the influence of a few individual donors with sample-specific effects. Consequently, pooled samples cannot be used to assess sample-specific effects in the statistical analysis.

When preparing pools, the individual donors should meet the same requirements as described for individual CSs used alone. A preliminary experiment should be performed to demonstrate that pooled CSs and their storage conditions are suitable surrogates for individual CSs. Pooled CSs may have matrix alterations that cause them to be noncommutable with individual CSs. Consequently, pooled CSs should be validated to be commutable with individual CSs before being used for commutability assessment of RMs. Demonstrating that some pools are commutable does not guarantee that other pools prepared in the same manner will also be commutable, but it is a reasonable assumption. When sufficient volume of donor samples is available, Clinical and Laboratory Standards Institute document C37 (6) includes a validation scheme to demonstrate commutability of a pool based on recovery of a value expected from the proportion of each donor sample used to prepare the pool. When pooled CSs are used for commutability assessment, their preparation and qualification must be fully documented.

### **RM(s) to Include in an Assessment of Commutability**

Candidate RM(s) intended for use as calibrators in a calibration hierarchy (7), as trueness control materials, or as EQA samples are typically produced by an organization that assesses their commutability as part of the qualification for intended use. When RMs are intended to be used after dilution to obtain quantities within the measuring

interval of the MPs, the diluted RMs must be evaluated for commutability. The manufacturer of an RM that is intended to be diluted before use should provide instructions for making the dilutions that address compatible or incompatible diluents and other known influences that could affect the performance of the diluted RM.

RMs intended to be used as common calibrators by different MPs must be commutable. However, product calibrators (end-user calibrators) that are provided by an IVD manufacturer intended for use only with a specifically identified MP are not required to be commutable. When necessary, the value assignment of such product calibrators can compensate for a known and constant noncommutability bias such that patient results from that specific MP are traceable to higher order references. Such calibrators are validated and intended for use only with MPs specifically identified by the IVD manufacturer and are not intended to be used with any other MP.

### Qualification of MPs for Inclusion in a Commutability Assessment

Cooperation with manufacturers of MPs is important during RM development to anticipate and minimize possible constraints related to an assessment of commutability. An MP manufacturer may be an IVD medical device manufacturer or a laboratory that develops its own MP. It is desirable to include as many different MPs and analytical measurement principles as possible in a commutability assessment. However, it may not be possible to include all MPs in a commutability assessment. Including the most representative group of MPs will improve the likelihood of an RM being suitable for use with other MPs not included in the initial assessment or with a new MP that may enter the market. Considerations for inclusion include market share for commercially available MPs and types of analytical measurement principles. For laboratory-developed tests, inclusion can be based on methods used on a large scale or offered as a service to other laboratories. The manufacturer of an MP is responsible to ensure that an RM is commutable and suitable for use in a calibration hierarchy.

MPs to be included in a commutability assessment must have acceptable performance characteristics as described below. Improvement of some MPs may be a required precursor to inclusion in a commutability assessment. An MP manufacturer may require substantial time to improve an MP and may have to make provision for a follow-up commutability assessment of an RM when MP improvements have been addressed.

#### MP PRECISION

MPs must have adequate precision because inadequate precision can inappropriately influence assessment of commutability. For example, a value for a measurand

from an MP with inadequate precision will have larger uncertainty that could compromise assessment of commutability for that MP. There are no fixed guidelines for adequate precision. Considerations for acceptable precision will be determined by the experimental design that in turn will be influenced by the criteria for acceptable commutability, availability of CSs with sufficient volume for replication, and cost for distributing materials and making measurements. The desired closeness of agreement between RM and CSs that needs to be identified in the commutability assessment should be determined, and the required precision of MPs can be estimated based on statistical power analysis. The replication of measurements is an experimental design detail that can be adjusted to reduce random error.

#### MP SELECTIVITY

MPs to be included in a commutability assessment must have adequate selectivity for the measurand. MPs with inadequate selectivity could inappropriately disqualify an RM that may be suitable for use with many MPs being used by clinical laboratories. Inadequate selectivity is identified as bias in an individual CS, called a sample-specific effect, that is greater than the typical bias for other CSs and is caused by the influence of substances other than the measurand on the measurement signal, or by differences in response to the measurand by different MPs. A sample-specific effect is a systematic error (bias) for an individual CS that can be considered a random component of bias within a group of CSs that cannot be reduced by replication or calibration.

The selectivity of MPs for which the RM will be used must be considered when determining whether adequate commutability can be demonstrated. It may not be possible to demonstrate adequate commutability of an RM with an MP that has excessive sample-specific effects. The MP manufacturer will need to improve the selectivity for the measurand to enable the RM to be suitable for use with that MP. Ideally, inadequate selectivity will be identified in a qualification assessment or a preliminary experiment. However, the nonselectivity may not be recognized until the commutability assessment is performed, in which case results for that MP should be excluded from a commutability assessment of an RM for other MPs for which the RM may be suitable for use.

Selectivity of an MP can be concentration-dependent, for example, if different molecular forms of an analyte are produced in disease conditions. An example of a selectivity limitation is the presence of molecular forms that may not be measured in a consistent ratio by different MPs. MPs that recover the same ratio can be included; others may need to be excluded until they can be modified to recover the same ratio. An expert group may need to determine which molecular forms and ratios are appropriate for the medical use of the laboratory test.

## Criteria to Determine That an RM Is Commutable

The intended use of an RM will influence the choice of a criterion for commutability. The criterion for commutability of an RM relates to the allowable bias for an individual CS result (8). Several approaches to determine the allowable bias for a CS result have been described that consider the risk of harm to a patient from medical decisions based on uncertainty in a measurand result for a CS (9). The criterion for commutability for an RM intended for use as a calibrator in a calibration hierarchy for an MP should be a fraction of the uncertainty required at the relevant position in the hierarchy to achieve the allowable bias in a CS result.

EQA or trueness control materials are usually intended to verify (not establish) that an individual result is within an acceptable measurement error. The criterion should consider that EQA and trueness control materials are usually measured in singlicate or a few replicate measurements. Consequently, both bias and imprecision can influence a measured value. The criterion for commutability should be a fraction of the bias component of the acceptance limits for evaluating an EQA or trueness control result.

Closeness of agreement is a relative term, and some RM results may have closer agreement with results for CSs and, thus, better commutability for some MPs than for others. A criterion based on the intended medical use of laboratory test results is preferred but needs to be established with consideration of MP performance capability. The criterion for a conclusion that an RM is commutable based on medical requirements should be established at the beginning of a commutability assessment. If no available RM can meet the criterion, the criterion may be reconsidered to permit production of an RM that when used in a calibration hierarchy substantially improves agreement of results for CSs among different MPs such that medical decisions are improved and the risk of harm to a patient is reduced.

### CRITERIA RELATED TO MP PERFORMANCE CHARACTERISTICS

Criteria based on statistical distribution of CS results between MPs are less desirable and not recommended because they can produce different criteria for different combinations of MPs for the same measurand. Criteria based on statistical distribution of CS results may be unreasonably small or large compared with the intended medical use of laboratory test results. Criteria based on statistical limits expect some fraction of RM–MP combinations to fail to meet the criteria that could cause an erroneous conclusion regarding the RM being fit for purpose. However, a criterion related to achievable performance of available MPs may be acceptable if the same

criterion is used for all MPs in a commutability assessment.

In some cases, practical limitations in study design (e.g., limited number of replicates or limited number of CSs) and/or the performance capability of all or many MPs (e.g., poor precision or high susceptibility to sample-specific effects) could produce large uncertainties that could cause high rates of inconclusive commutability decisions and limit the ability to make a decision about suitability of an RM (10, 11). In such situations, using a less stringent acceptance criterion accounting for the study design and performance capability of MPs can be considered. In this situation, the claimed use of the RM should be reconsidered accordingly.

### CRITERIA FOR A SET OF CALIBRATORS

Criteria for a set of calibrators intended to be used together to cover a measuring interval need to be considered as a group. When different concentrations of RMs are prepared independently, each RM should be independently evaluated for commutability. If RMs are prepared by a process such as admixing 2 concentrations or another approach that provides common characteristics between the materials, then each RM concentration may have the same or similar relationship (i.e., bias) when compared with the results from CSs. In this case, the commutability of the RM may be considered as a set. If the variability of results [i.e., SD, CV, SD(log)] for an MP is different at different measurand values (concentrations), then the data may need to be partitioned accordingly (readers may refer to Clinical and Laboratory Standards Institute document EP09 for procedural details on partitioning) (12). Criteria based on medical requirements may be different at different concentrations (quantity values).

### FRACTION OF MPs FOR WHICH A RM IS COMMUTABLE

A number of MPs will be included in the assessment of commutability of an RM. Ideally, an RM will be commutable for all MPs that it might be used with. However, this goal is not always achievable, and in many cases there will be some MPs for which the RM is not commutable. An RM intended for use as a common calibrator should be commutable for a sufficient fraction of MPs that an improvement in medical decisions will occur. Similarly, when an RM is intended to be used as a trueness control or an EQA sample, its commutability should be suitable to obtain useful information on trueness and performance for most of the laboratories and MPs that use the RM.

There are no simple recommendations for the fraction of available MPs for which an RM must be commutable that would qualify an RM as being fit for purpose. Considerations include the market share for MPs and consequently the number of tested individuals who may

be affected. Note that market share can be different in different regions of the world. Another consideration is the overall health improvement that would follow from use of an RM in a calibration hierarchy even if it failed the preestablished criteria for commutability for some fraction of MPs in use. In such a case, the criteria could be reconsidered based on the impact on medical decisions of using a less stringent criteria.

### Generalization of Commutability of an RM to Future Measurements

The conclusions from a commutability assessment are strictly applicable to only the MPs and measurement conditions (reagent lot and other parameters) used in the experiment. It may be possible to use several reagent lots, calibration events, and IVD medical devices of the same MPs in a commutability experiment, but in many cases such duplication is not realistically achievable. The experiment typically assumes that the MPs and measurement conditions are representative of those that will be encountered in medical laboratories and that the conclusions regarding commutability of an RM will be maintained when used with other IVD medical devices of the same MPs, as well as for other reagent lots and measuring conditions in the future. The assumption is accepted as reasonable in laboratory medicine, but users should keep in mind that changes in reagents or measuring conditions can occur that could make conclusions from a commutability assessment no longer applicable. The more substantial a change, such as a change in reagent formulation, the greater is the risk that commutability conclusions may no longer be valid. In such cases, a commutability assessment should be repeated to ensure that an RM remains suitable for its intended use.

### Information on Commutability to Be Provided to the User of an RM

Table 1 includes the information regarding commutability assessment that should be documented for a certified RM, trueness control material, or commutable EQA material. The certificate of analysis should include a summary of the Table 1 information with details at the discretion of the RM manufacturer. All the information must be provided on request to users of an RM. It is recommended that the results of a commutability assessment for an RM intended to be used in a calibration hierarchy or as a trueness control be published in a peer-reviewed journal.

### Conclusion

This report provides recommendations for critical components of the experimental design for commutability

**Table 1. Information to be included in the documentation of an RM that is commutable for a stated number of MPs.**

• Selection criteria for individuals from whom CSs were obtained for the commutability assessment
• Number of CSs used in the commutability assessment and their collection, processing, storage, and distribution conditions
• Description of the experimental design used to assess commutability; state the reference MP if included in the experimental design
• Criteria used to conclude that an RM was commutable with clinical samples
• Summary of the results of the commutability assessment in sufficient detail that the conclusions can be verified; complete experimental results and data analysis must be available to a user on request
• MPs for which commutability was demonstrated, including the specific models of instruments and the part numbers and lot numbers of reagents, calibrators, and calibration confirmation materials

assessment, including selection of individual CSs, use of pooled CSs, qualification of MPs for inclusion in the assessment, establishing a criterion for the determination that an RM is commutable that is the same for all MPs and is related to how a test result is used in medical decisions, generalization of commutability conclusions to future measurements made with the MPs included in the assessment, and information regarding commutability to be made available to a user of the RM and included in the certificate for an RM.

**Author Contributions:** *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

**Authors' Disclosures or Potential Conflicts of Interest:** *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

**Employment or Leadership:** W.G. Miller, *Clinical Chemistry*, AACC; R. Rej, *Clinical Chemistry*, AACC; C.J. Burns, National Institute for Biological Standards and Control.

**Consultant or Advisory Role:** None declared.

**Stock Ownership:** None declared.

**Honoraria:** None declared.

**Research Funding:** None declared.

**Expert Testimony:** None declared.

**Patents:** None declared.

**Other Remuneration:** N. Greenberg, College of American Pathologists.

**Role of Sponsor:** No sponsor was declared.

## References

1. Nilsson G, Budd JR, Greenberg N, Delatour V, Rej R, Panteghini M, et al. IFCC working group recommendations for assessing commutability part 2: based on the difference in bias between a reference material and clinical samples. *Clin Chem* 2018;64:455–64.
2. Budd JR, Weykamp C, Rej R, MacKenzie F, Ceriotti F, Greenberg N, et al. IFCC working group recommendations for assessing commutability part 3: based on the calibration effectiveness of a reference material. *Clin Chem* 2018;64:465–74.
3. International vocabulary of metrology—basic and general concepts and associated terms (VIM). 3rd Ed. JCGM; 2012. Sevres, France: International Bureau of Weights and Measures.
4. Evaluation of commutability of processed samples; approved guideline. 3rd Ed. CLSI document EP14-A3. Wayne (PA): Clinical and Laboratory Standards Institute; 2014.
5. Characterization and qualification of commutable reference materials for laboratory medicine; approved guideline. CLSI document EP30-A. Wayne (PA): Clinical and Laboratory Standards Institute; 2010.
6. Preparation and validation of commutable frozen human serum pools as secondary reference materials for cholesterol measurement procedures; approved guideline. CLSI document C37-A. Wayne (PA): Clinical and Laboratory Standards Institute; 1999.
7. 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.
8. Braga F, Infusino I, Panteghini M. Performance criteria for combined uncertainty budget in the implementation of metrological traceability. *Clin Chem Lab Med* 2015;53:905–12.
9. Panteghini M, Sandberg S, editors. First EFLM strategic conference defining analytical performance goals—15 years after the Stockholm Conference. *Clin Chem Lab Med* 2015;53(6, special issue):829–958.
10. Korzun WJ, Nilsson G, Bachmann LM, Myers GL, Sakurabayashi I, Nakajima K, et al. Difference in bias approach for commutability assessment: application to frozen pools of human serum measured by 8 direct methods for HDL cholesterol and LDL cholesterol. *Clin Chem* 2015;61:1107–13.
11. Delatour V, Liu Q, Vesper HW. Commutability assessment of external quality assessment materials with the difference in bias approach: are acceptance criteria based on medical requirements too strict? *Clin Chem* 2016;62:1670–1.
12. Measurement procedure comparison and bias estimation using patient samples; approved guideline EP09-A3. Wayne (PA): Clinical and Laboratory Standards Institute; 2013.