Clinical Chemistry 00:0
1–11 (2023)

Special Report

Recommendations for Setting a Criterion and Assessing Commutability of Sample Materials Used in External Quality Assessment/Proficiency Testing Schemes

Sverre Sandberg,[a,b,c,*] Pernille Fauskanger,[a] Jesper V. Johansen,[d] Thomas Keller,[e] Jeffrey Budd,[f] Neil Greenberg,[g] Robert Rej,[h] Mauro Panteghini,[i] Vincent Delatour,[j] Ferruccio Ceriotti,[k] Liesbet Deprez,[l] Johanna E. Camara,[m] Finlay MacKenzie,[n] Alicia N. Lyle,[o] Eline van der Hagen,[p] Chris Burns,[q] and W. Greg Miller;[r] for the IFCC Working Group on Commutability in Metrological Traceability

It is important for external quality assessment materials (EQAMs) to be commutable with clinical samples; i.e., they should behave like clinical samples when measured using end-user clinical laboratory in vitro diagnostic medical devices (IVD-MDs). Using commutable EQAMs makes it possible to evaluate metrological traceability and/or equivalence of results between IVD-MDs. The criterion for assessing commutability of an EQAM between 2 IVD-MDs is that its result should be within the prediction interval limits based on the statistical distribution of the clinical sample results from the 2 IVD-MDs being compared. The width of the prediction interval is, among other things, dependent on the analytical performance characteristics of the IVD-MDs. A presupposition for using this criterion is that the differences in nonselectivity between the 2 IVD-MDs being compared are acceptable. An acceptable difference in nonselectivity should be small relative to the analytical performance specifications used in the external quality assessment scheme. The acceptable difference in nonselectivity is used to modify the prediction interval criterion for commutability assessment.

The present report provides recommendations on how to establish a criterion for acceptable commutability for EQAMS, establish the difference in nonselectivity that can be accepted between IVD-MDs, and perform a commutability assessment. The report also contains examples for performing a commutability assessment of EQAMs.

Background

An important but often unrealized goal of external quality assessment (EQA)/proficiency testing schemes in laboratory medicine is to determine if different in vitro diagnostic medical devices (IVD-MDs), also called measuring systems, produce equivalent results with clinical samples (CSs) (1). To make this evaluation, it is obligatory to use an EQA material (EQAM) that is commutable with clinical samples (2, 3). Consequently, evaluating the commutability of an EQAM with CSs is required. Commutability is an important property of an EQAM, just as it is for certified reference materials (CRM) used as common calibrators in implementing metrological traceability to higher-order references. Recommendations for commutability assessment have been described in recent publications from this working group (4–6) and from the Clinical and Laboratory

[a]Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway; [b]Norwegian Porphyria Centre, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway; [c]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway; [d]Radiometer Medical ApS, Copenhagen, Denmark; [e]ACOMED statistic, Leipzig, Germany; [f]Jeff Budd Consulting, St. Paul, MN, United States; [g]Neil Greenberg Consulting, LLC, Rochester, NY, United States; [h]Department of Biomedical Sciences, School of Public Health, State University of New York at Albany, Albany, NY, United States; [i]Research Centre for Metrological Traceability in Laboratory Medicine, University of Milan, Milan, Italy; [j]Laboratoire national de métrologie et d'essais, Paris, France; [k]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; [l]European Commission, Joint Research Centre, Directorate F, Geel, Belgium; [m]National Institute of Standards and Technology, Gaithersburg, MD, United States; [n]Birmingham Quality/UK NEQAS, University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom; [o]Centers for Disease Control and Prevention, Atlanta, GA, United States; [p]Queen Beatrix Hospital, Winterswijk, the Netherlands; [q]National Institute for Biological Standards and Control, A Centre of the MHRA, Hertfordshire, United Kingdom; [r]Virginia Commonwealth University, Richmond, VA, United States.

*Address correspondence to this author at: Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Boks 6165, 5892 Bergen, Norway. E-mail sverre.sandberg@noklus.no.

Disclaimer: The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the Agency for Toxic Substances and Disease Registry. Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention, the Public Health Service, or the US Department of Health and Human Services.

Received June 20, 2023; accepted August 21, 2023.
https://doi.org/10.1093/clinchem/hvad135

Standards Institute (7). The criterion to be used in commutability assessment is different for a CRM and an EQAM because these materials are used for different purposes (Table 1). The criterion for commutability assessment of a CRM is described in another report from this working group (8).

A commutable EQAM is used to evaluate performance of individual medical laboratory IVD-MDs for their ability to provide results for CSs that are metrologically traceable to a higher-order reference and thus equivalent to results from other IVD-MDs calibrated to be traceable to the same or an equivalent higher-order reference. A commutable EQAM can also be used to assess and monitor the status of harmonization and standardization implementation according to International Organization for Standardization (ISO) 17511 and ISO 21151 standards (9, 10).

Definitions

Commutability is a property of a reference material that means results for a reference material, EQAM in this report, and for CSs have the same numeric relationship, within specified limits, across the measurement procedures (MPs) for which the reference material is intended to be used. Consequently, a commutable reference material produces a measurement result that is equivalent to the measurement result that would be obtained for a clinical sample with the same concentration of the measurand.

Equivalent results denote agreement of measured values among different IVD-MDs intended to measure the same measurand, where the differences in measured values on the same CSs do not affect clinical interpretation. A conclusion of equivalence of measured values for the same CSs among 2 or more IVD-MDs is based on the differences in measured values being within a predefined clinically acceptable margin or limit [adapted from ISO 17511:2020 (9)].

A measurement procedure is a detailed description of a measurement according to one or more measurement principles, including a description of the logical organization of operations used in a measurement and any calculation to obtain a measurement result. An MP describes how to manufacture and operate a measuring system. A measuring system is manufactured according to an MP and in the field of laboratory medicine is called an IVD-MD at the metrological traceability level of the end-user clinical laboratory. A measuring system is a set of physical items including a measuring instrument, calibrators, reagents, and other devices needed for a measurement to generate a measured result for a measurand in a sample, e.g., a CS. Any performance estimate (including traceability, selectivity, or commutability assessment) can only be performed using measuring systems that are manufactured according to an MP. Definitions are adapted from ISO 17511:2020. 3.29 (9).

A measurand is the quantity intended to be measured (11). The chemical species being measured is an important consideration when assessing commutability of a CRM or an EQAM. In some cases, more than one chemical species may be measured, either intentionally, e.g., when a parent molecule and a metabolite are clinically meaningful, or due to poor selectivity of an MP.

Selectivity of an MP is a property whereby the measured value of a measurand is independent of other quantities in the sample (definition adapted from *Vocabulaire international de metrologie*) (11). Other quantities may be metabolites of the measurand, molecular forms of the measurand, other ions or molecules, or influences on the measurement from any source other than the measurand itself. Selectivity of an IVD-MD was formerly called "analytical specificity."

Sample-specific effects refer to the impact of influence quantities on measurement trueness that may be observed as variable magnitude errors (biases) among a set of CSs when measured using 2 or more IVD-MDs. Per ISO 17511:2020 (9): "when the selectivity of a measuring system is not fit-for-purpose, sample-specific influence quantities in human samples due to factors including disease, diet, drugs or other pathological conditions may lead to erroneous values for the intended measured quantity." Even with acceptable selectivity of a measuring system, if we consider a particular CS, a sample-specific error (bias) may be observed as a systematic bias because it cannot be reduced by repeated measurements. However, residual sample-specific errors (biases) may be observed as distributed in a statistically random fashion among a set of CSs.

Characteristics of EQAM

EQA is used for 3 main purposes: (*a*) to inform a laboratory of its results compared to other laboratories using the same or similar MPs, (*b*) to inform a laboratory of the suitability of its results vs medical requirements, and (*c*) to inform in vitro diagnostic manufacturers and the laboratory community regarding the metrological traceability of their MPs with the goal of obtaining equivalent results for CSs among different MPs. In addition, special EQA surveys are performed to inform laboratories and in vitro diagnostic manufacturers about the influence on results of interfering substances and selectivity for a given measurand. In all 4 cases, to use a common target value for the participating IVD-MDs, the EQAM used must be commutable with CSs when measured by the IVD-MDs participating in the scheme. If a reference measurement procedure (RMP) is used,

Table 1. Comparison of EQAM and CRM characteristics impacting their commutability assessment.

Consideration	EQAM	CRM
Intended use	To assess suitability of results by individual medical laboratories and from IVD-MDs.	To calibrate IVD-MDs.
Prerequisites for including end-user IVD-MDs in commutability assessment	None—include all that are in general use.	Some end-user IVD-MDs may be excluded due to poor performance.
Reasons for using commutable material	Without commutable material, it is not possible to assess the trueness or equivalence of end-user results.	Without commutable CRM, it is not possible to transfer values from the selected higher-order references to the end-user IVD-MDs.
Consequences of not having commutable material	Equivalence of results among individual laboratories and different IVD-MDs cannot be assessed. Difficult to identify and eliminate poor-quality IVD-MDs.	Not possible to transfer trueness from CRM to end-user IVD-MDs. As a result, different end-user IVD-MDs used by medical laboratories may not give equivalent results for CSs.
Principal differences for material that should be assessed for commutability	Usually produced and distributed several times a year with a short use cycle; also, may have short stability lifetime. May intentionally include potentially interfering substances.	Intended to be stored and used for years; long-term stability is required. Should not include potentially interfering substances.
Recommendation for commutability studies	Should be simplified to be practical for relatively frequent assessment. It is not possible to assess every batch of EQAM.	Commutability criteria should be sufficiently stringent to have low impact on the measurement uncertainty of CSs and thus suitability for use. Commutability assessment should be performed with every batch, unless scientifically valid rationale for exemption is provided.
What is the "state-of-the-art" concerning commutability?	Evidence for commutability is seldom provided. Commutability is often "assumed" based on EQAM preparation procedures.	Newer CRMs are assessed for commutability based on ISO 15194:2009 requirements. Most older CRMs have not been assessed for commutability.
Criterion for commutability	The criterion for commutability takes into account analytical performance of all IVD-MDs included in an EQA scheme.	The criterion for commutability assessment should be a fraction of the maximum allowable measurement uncertainty for CS results because its uncertainty is propagated through the calibration hierarchy to a CS result.
Using fixed criteria for commutability based on	Not suitable because even poorer-performing IVD-MDs (e.g., those with too high measurement	Suitable because poorer-performing MPs (i.e., those with poor performance or different selectivity)

Continued

Table 1. (continued)

Consideration	EQAM	CRM
analytical performance specifications for CS results	uncertainty) are included in EQA schemes.	can be excluded from the commutability assessment.
Using flexible criteria derived from analytical performance capability	Suitable because all IVD-MDs in use are included in an EQA scheme.	Not suitable. However, an allowance is possible to revise the fixed criteria at a less stringent level when no or few CRMs can meet it.

the EQAM must be commutable with CSs when measured by the IVD-MDs and the RMP.

Commutability assessment of an EQAM is challenging because EQAM suppliers prepare them frequently and at varying concentrations for a given measurand. The frequency of preparation means that commutability assessment of EQAM can only be done on a periodic basis. It is reasonable to assume that the property of commutability of EQAM is retained in the new batches prepared by a controlled process. In addition, EQAMs are intended for use with all end-user IVD-MDs in a scheme irrespective of the performance capability (trueness, precision, selectivity) of each IVD-MD.

Criteria for Assessing Commutability of EQAM

Because EQA is used to assess performance of all participants, the criterion for assessing commutability of EQAM is influenced by the analytical performance capability (trueness, precision, selectivity) of all end-user IVD-MDs in an EQA scheme. For EQAM, we recommend that the commutability criterion be based on the statistical distribution of the CS results of 2 IVD-MDs in a comparison, because IVD-MDs vary in performance for repeatability and for sample specific effects from differences in non-selectivity (DINS) for the measurand.

Commutability assessment for an EQAM is not influenced by bias in CS results that may be present between pairs of IVD-MDs in the assessment (Table 1). How an EQA provider sets target values and analytical performance specifications (APS) for participant performance are beyond the scope of this report. Note that commutability assessment is independent of, and not influenced by, any calibration bias that may be present for an IVD-MD. Commutability assessment demonstrates the closeness of agreement between results for an EQAM and for CSs when measured using pairs of IVD-MDs or between an IVD-MD and an RMP. If any IVD-MD in the paired comparisons demonstrates a systematic bias, the bias will have the same influence on results for CSs and an EQAM when an EQAM is commutable with CSs.

Assessment of Commutability of an EQAM

Figure 1 shows a flowchart for the assessment of EQAM commutability. The first 2 steps can in principle be done without EQAM as a comparison of results for CSs between pairs of IVD-MDs to determine if the DINS of these IVD-MDs for a measurand is suitable for commutability assessment of the EQAM.

Step 1: The data must be examined for outliers (12). An outlier may represent a sample specific effect in a CS with a particular IVD-MD in the comparison but could also be an error in measurement, transcription, or other type of error not associated with the selectivity characteristics of an IVD-MD. The number of outliers should be reported, and the reason for a decision to exclude them must be documented. If there are many outliers, this may suggest excessive DINS between 2 IVD-MDs. In this case, a commutability assessment cannot be performed. If in doubt, the analysis in step 2 can be performed with and without outliers to assess if the conclusion changes.

Step 2: After relevant outliers are excluded, the results from pairs of IVD-MDs are examined to determine if the IVD-MD selectivity for the measurand is suitable, thus permitting the inclusion of the IVD-MD pair in the assessment of commutability.

Step 3: A commutability assessment is performed between all combinations of 2 IVD-MDs (representing different MPs) whose selectivity performance in step 2 was suitable for commutability assessment. Because commutability assessment of EQAMs is influenced by analytical performance capability, the criterion is based on statistical distributions of results from CS pairs. A commutable EQAM is expected to have a result that falls within the prediction interval (PI) limits based on the statistical distribution of the CS results from the 2 IVD-MDs being compared.

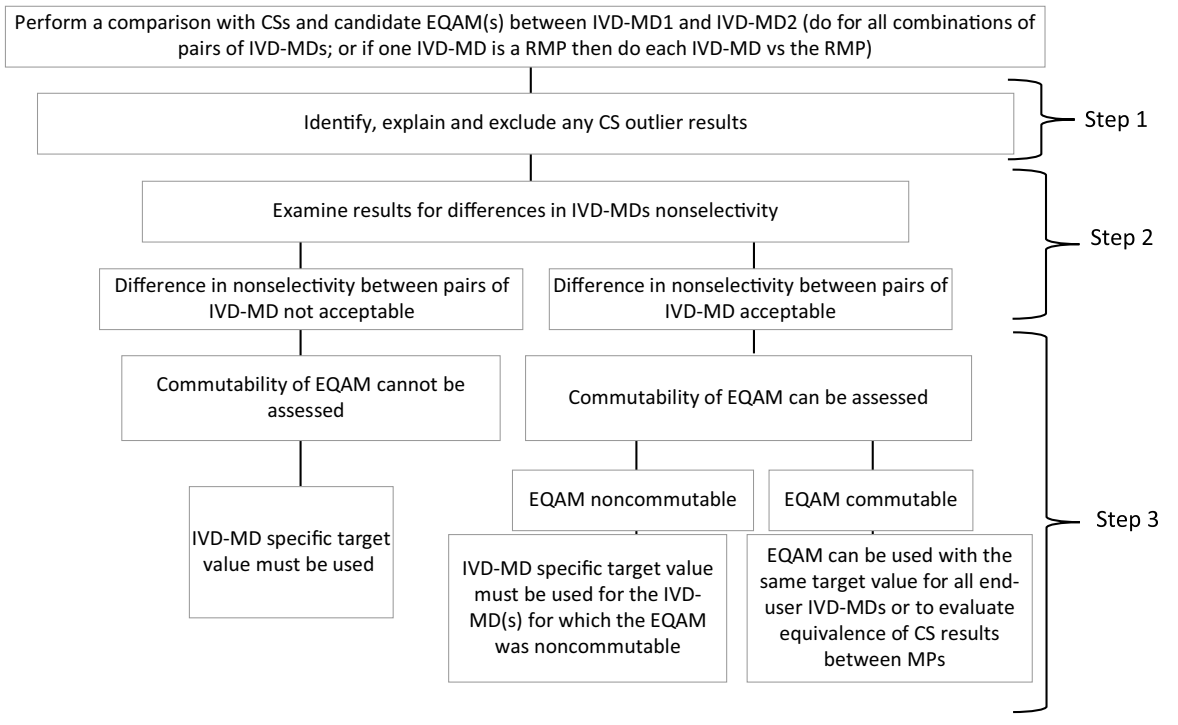

Fig. 1. Flow chart for commutability assessment of an EQAM. The commutability assessment of EQAMs will usually be performed by analyzing EQAMs and CSs in the same run, and the evaluation of the results consists of the 3 sequential steps.

Inclusion of a RMP for a measurand, when available, can simplify the commutability assessment because all end-user IVD-MDs are compared only to the RMP, and comparisons of all combinations of pairs of IVD-MDs are not needed.

Evaluation of Selectivity of Pairs of IVD-MDs as Suitable for Inclusion in a Commutability Assessment of an EQAM (Step 2 in Fig. 1)

In order to detect DINS for the measurand in the 2 IVD-MDs (IVD-MD1 and IVD-MD2), the replication variance of each of the IVD-MDs, $SD^2_{IVD\text{-}MD1}$ and $SD^2_{IVD\text{-}MD2}$, is estimated from the pooled variance of the individual variances for each CS analyzed in replicate. Considering the regression relationship between the 2 IVD-MDs, the average variance of the residuals of a fitted regression model, SD^2_R, is expected to be close to $SD^2_{IVD\text{-}MD1} + SD^2_{IVD\text{-}MD2}$ if the compared IVD-MDs have similar nonselectivity profiles. In cases when one IVD-MD has excessive nonselectivity relative to the other, SD^2_R will be increased compared to $SD^2_{IVD\text{-}MD1} + SD^2_{IVD\text{-}MD2}$. Consequently, the ratio of the calculated SD^2_R and $SD^2_{IVD\text{-}MD1} + SD^2_{IVD\text{-}MD2}$ will exceed 1. Let the ratio of SD^2_R and $SD^2_{MS_1} + SD^2_{MS_2}$ be denoted by the Greek letter ζ:

$$\zeta = \frac{SD^2_R}{SD^2_{IVD\text{-}MD1} + SD^2_{IVD\text{-}MD2}}. \quad (1)$$

Thus, ζ is a quantitative measure of the DINS between the 2 IVD-MDs. ζ should be reported for every study comparing pairs of IVD-MDs for the same measurand.

A calculated ζ value smaller or equal to an upper limit of ζ (denoted ζ_{upper}) refers to an acceptable DINS. Consequently, the acceptable SD^2_R can be stated as follows:

$$SD^2_R \leq \zeta_{upper} \cdot (SD^2_{IVD\text{-}MD1} + SD^2_{IVD\text{-}MD2}). \quad (2)$$

ζ_{upper} is chosen to be the 99th percentile of ζ obtained using Monte Carlo simulation based on measurement results from a pair of IVD-MDs assumed to have acceptable DINS. The approach for detecting excessive DINS described in this report, uses the proof-of-hazard approach (13), meaning a DINS is deemed acceptable if

there is not sufficient statistical evidence that it exceeds an acceptable magnitude. The benefits of the proof-of-hazard approach are that it requires smaller sample sizes and is less likely to yield erroneous conclusions of excessive DINS, making it more practical. The disadvantage is that small DINS between compared IVD-MDs may not be detected. This limitation occurs when we have excessive DINS but our estimated ζ value is small (i.e., close to 1). However, when ζ is small, the actual relative increase in PI width for regression analysis of CS results from the 2 IVD-MDs is also small, minimizing the potential undesirable consequences of having small DINS between compared IVD-MDs.

The PI around the regression line of paired CS results is used to evaluate the influence of DINS between 2 IVD-MDs. If both IVD-MDs in a regression pair have perfect selectivity or identical nonselectivity profiles for the measurand, the PI reflects the combined imprecision of the 2 IVD-MDs. When the 2 IVD-MDs have different nonselectivities for the measurand, the PI will be increased by the influence of DINS. Differences in nonselectivity between the pair of IVD-MDs may create a difference in bias for one or more CS results. However, for a set of CS results the DINS bias appears as a random imprecision contribution that causes the width of the PI to increase.

The relative PI width increase for IVD-MDs with DINS profiles compared to IVD-MDs with identical nonselectivity profiles is denoted by the relative prediction interval width increase (M). In other words, M is a function of DINS between 2 IVD-MDs. When a maximum acceptable M is established, it can be used to identify IVD-MD pairs with excessive DINS that must be excluded from a commutability assessment. Such an excluded IVD-MD pair would have its results evaluated vs a peer-group target because the DINS is too large for assessing commutability of the EQAM with CSs. The ζ_{upper} in Eq. 2 is determined from M, the number of clinical samples used, and the number of replicates of these samples (Table 2). Consequently, the value of ζ_{upper} in Eq. 2 is used as the criterion for acceptable DINS between IVD-MDs in a paired comparison.

Figure 2A shows the effect on the PI for different values of M. If we accept a difference in nonselectivity that can cause an increase in the PI width of, e.g., 50%, the maximum acceptable noncommutability bias of an EQAM will be increased by (0.5×PI width), assuming that the PI width is constructed based on identical nonselectivity profiles for the paired IVD-MDs. An acceptable M is established based on the intended use of the IVD-MD results for a given measurand. One of the most important aspects of having a commutable EQAM is to enable the evaluation of the equivalence of results from different IVD-MDs including their metrological traceability to an RMP when one is available. Consequently, the noncommutability bias of an EQAM must be small compared to the APS used when evaluating the IVD-MDs in an EQA.

Figure 2B shows the percentage maximum noncommutability bias for different imprecisions of 2 IVD-MDs and different M values. As an example, for glucose, if both CVs are 1% and M is 50%, the maximum noncommutability bias will be 1.2%, rather small compared to the APS for glucose in an EQA scheme that is typically ±6%. In EQA, the APS is rarely < 5% for any measurand. In general, when IVD-MD imprecision is small, we can allow for a larger M (Fig. 2B). The EQA provider must decide how big the noncommutability bias can be for a certain measurand in relation to the chosen APS. Since the imprecision of IVD-MDs often is <2% and the APS is usually >5%, an M of 50% will often be suitable (Fig. 2B).

When M has been determined, the ζ value is calculated based either on raw data, log-transformed data, or data transformed in another way [see software application (14)], whichever gives the ζ value closest to 1. For example, if the ζ resulting from log-transformation is closer to 1 than for nontransformation of the results, we will choose the log-transformation. The recommended upper limits of ζ for different study designs are based on Monte Carlo simulations using measurement results from 2 IVD-MDs that were simulated for study designs that measured 20, 25, 30, and 40 CSs, with 2, 3, and 4 replicates for each CS, using randomly sampled CVs between 0.1% and 10%. Based on each simulated set of measurement results, ζ was calculated (unpublished observations). The ζ_{upper} values were determined from the simulation studies that allowed a relative increase of the PI with M of 0%, 5%, 15%, 20%, 25%, 30%, 40%, 50%, or 100% for each unique study design. One hundred thousand ζ values were simulated for each of 108 combinations of number of CSs, number of replicates, and M. ζ_{upper} was then calculated as the 1000th (99th percentile) largest ζ value of the 100 000 simulated ζ values (Table 2).

If the ζ value in the experiment (Eq. 1) is smaller than the ζ_{upper} value selected from Table 2, we may include the IVD-MD comparison in the commutability assessment. If the ζ value in the experiment is larger than ζ_{upper}, we will not include the IVD-MD comparison in the commutability assessment due to excessive DINS. As an example, for glucose ζ is 3.50 and ζ_{upper} is 3.23 from Table 2 for 25 CSs, 3 replicates each, and M = 50%. In this case, when $\zeta > \zeta_{upper}$, excessive DINS between the 2 IVD-MDs is observed in the data, and commutability of EQAMs cannot be assessed for this pair of IVD-MDs. Testing EQAM for commutability with one of these IVD-MDs is possible. If the second IVD-MD in the comparison is an RMP, then the other IVD-MD (which is not an RMP) has inadequate

Table 2. Different upper limits of ζ depending on study design and relative increase in the prediction interval accepted (M).

Number of clinical samples	Number of replicates	ζ_{upper} values for different Ms								
		M = 0%	M = 5%	M = 15%	M = 20%	M = 25%	M = 30%	M = 40%	M = 50%	M = 100%
20	2	1.92	2.10	2.50	2.73	2.93	3.19	3.69	4.28	7.65
20	3	1.51	1.66	1.97	2.14	2.32	2.52	2.92	3.38	6.03
20	4	1.38	1.52	1.80	1.95	2.12	2.30	2.66	3.09	5.53
25	2	1.76	1.93	2.32	2.51	2.74	2.94	3.42	3.98	7.08
25	3	1.44	1.59	1.88	2.05	2.22	2.40	2.78	3.23	5.76
25	4	1.33	1.47	1.73	1.89	2.05	2.22	2.57	3.00	5.33
30	2	1.67	1.83	2.19	2.38	2.58	2.79	3.23	3.75	6.68
30	3	1.40	1.54	1.82	1.98	2.15	2.32	2.69	3.13	5.56
30	4	1.30	1.44	1.69	1.84	2.00	2.16	2.50	2.92	5.21
40	2	1.54	1.71	2.03	2.21	2.38	2.58	3.01	3.48	6.19
40	3	1.33	1.47	1.74	1.89	2.05	2.22	2.58	3.00	5.32
40	4	1.26	1.39	1.63	1.78	1.93	2.09	2.42	2.84	5.03

ζ_{upper} is the 99th percentile of ζ, which in practice means that for a chosen M, only 1% of the corresponding ζ values will exceed ζ_{upper}; M is the acceptable average increase in prediction interval width due to differences in nonselectivity between 2 IVD-MDs.

selectivity, and MP-specific target values are required for assessment of EQA participants who use that IVD-MD. When both IVD-MDs are end-user clinical laboratory IVD-MDs and one IVD-MD in this pair shows consistent nonselectivity when paired with the majority of other IVD-MDs, we conclude that the IVD-MD with consistent nonselectivity cannot be included in a commutability assessment and must use MP-specific target values.

Evaluation of EQAMs for Commutability with CSs (Step 3 in Fig. 1)

When we have concluded from a series of IVD-MD pair-wise comparisons that a set of IVD-MDs can be included in a commutability assessment, Deming regression analysis with 99% PIs is recommended as the statistical acceptance limit for commutability of evaluated EQAMs. A suitable approach is to employ the 99% PI derived by regression analysis is described by Fuller and Gillard (15, 16). We do not recommend the CLSI EP14 approach (7) in this situation because it assumes no DINS between compared IVD-MDs. This assumption will result in underestimation of the PI width if there are acceptable DINS between the pairs of IVD-MDs being compared. Conversely, the Fuller–Gillard approach accounts for DINS making it more generally applicable. See the Supplemental Material for further explanation why the EP14 approach is not recommended.

An EQAM with measurement results inside the estimated PI is concluded to be commutable with that pair of IVD-MDs. In the software application for commutability (13), the "strength" of the conclusion regarding commutability is also given, taking the uncertainty of the conclusion into account. When there is no RMP and EQAM(s) of IVD-MD pairs are examined for commutability with CSs, a table can be constructed as shown in the Supplemental Material. Equivalence of EQA results can only be evaluated for IVD-MDs for which the EQAM is commutable with CSs and thus is suitable for use with these IVD-MDs.

In cases where the pairs of IVD-MDs being compared both have very small imprecision, the estimated PIs will be very narrow. In such scenarios, EQAMs with very small absolute noncommutability biases are more likely to be classified as noncommutable. The 99% PI is recommended, ensuring that only 1% of commutable EQAMs will be erroneously classified as noncommutable. If a 95% PI is used, 5% of commutable EQAMs would be erroneously classified as noncommutable. The amount of misclassification for commutable EQAMs for a 95% PI will increase the risk that an EQAM will be found to be noncommutable compared to a 99% PI, possibly causing rejection of that batch of EQAM, in turn increasing costs and making it less likely to achieve widespread implementation of routine commutability assessments of EQAMs. The 99% PI is

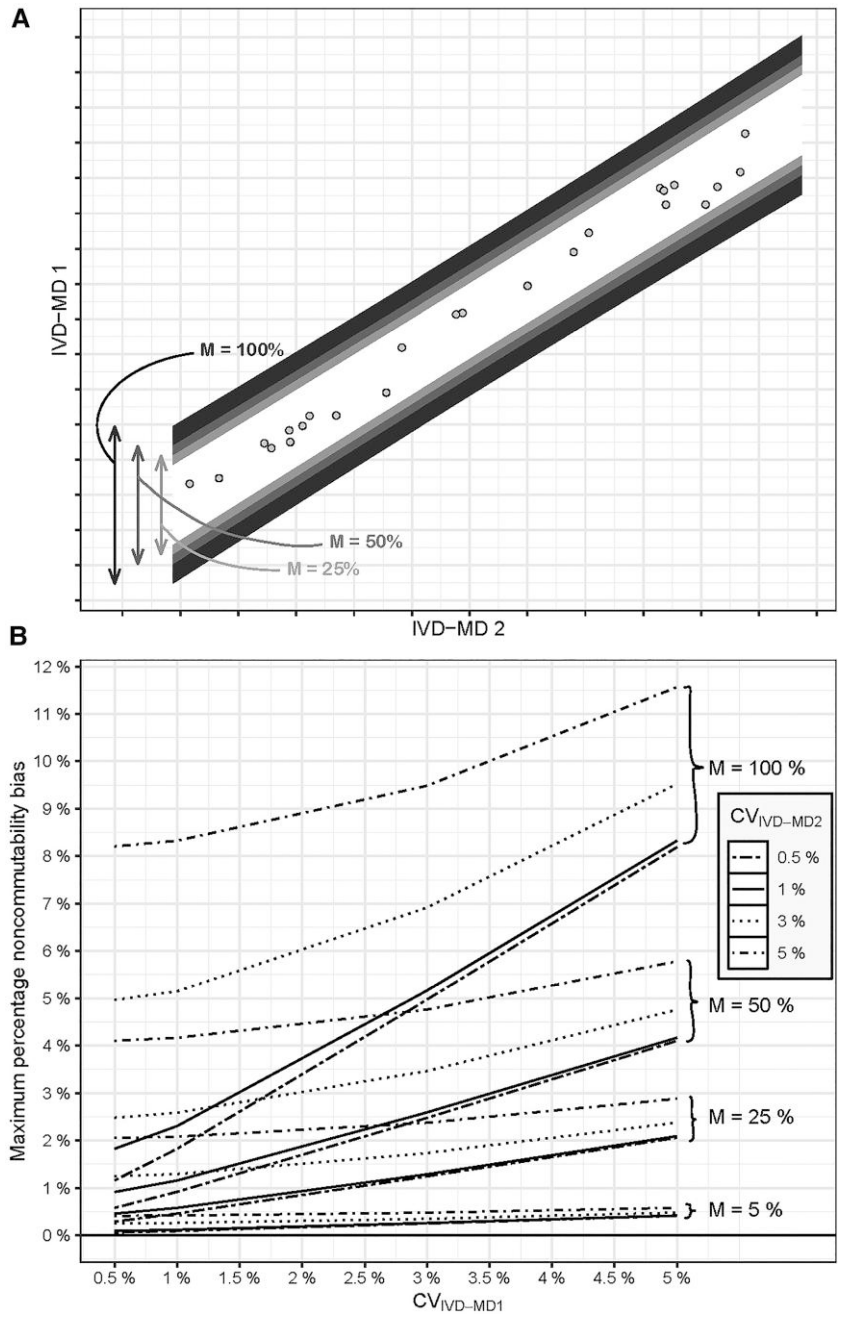

Fig. 2. (A), The effect of the relative increase in percentage (M) on the prediction intervals. M is the relative prediction interval width increase for IVD-MDs with DINS compared to IVD-MDs with identical nonselectivity profiles. The inner interval area (white) is the prediction interval when M is 0%. The light grey interval area illustrates the increase in width of the prediction intervals when M is 25% and the dark grey when the increase in width of the prediction intervals when M is 50%. The outer interval area (black) illustrates the additional increase in width of the prediction intervals when M is 100%; (B), The maximum percentage noncommutability bias as a function of imprecision (CV) of the compared IVD-MDs and different Ms.

recommended as a suitable choice that takes account of practical concerns and the ability to detect EQAMs with unacceptable noncommutability biases.

Recommendation for Experimental Setup

A worked example of a practical approach to evaluate DINS of IVD-MDs and commutability of an EQAM using a software application is given in an online resource (14).

The CSs and the EQAM should be analyzed under similar measurement conditions e.g., if 2 out of 3 replicates for CSs are analyzed in one run and the last replicate in another run, EQAM must be analyzed in a similar manner. The EQAM must be interspersed among the CSs in a balanced sequence for measurement. The EQAMs and CSs must be analyzed in a manner that ensures that different sources of variability; e.g., repeatability and between-run imprecision, if applicable, affect the results of both materials equally. A consequence of ignoring this assumption may be overestimation of ζ, which may result in falsely concluding that excessive DINS exist between the compared IVD-MDs.

The commutability of EQAMs should be examined after the same period of days an EQAM is expected to be measured by participants in cases when measurand changes are likely to occur as, for example, with hematology surveys. In such conditions measurements of the CSs are analyzed fresh and the analysis of EQAM(s) are coordinated to mimic typical time since production when participants are expected to make measurements of the EQAM(s). For example, the EQAM(s) are analyzed 5 days after preparation together with fresh CSs. This approach allows the commutability of the EQAM(s) to be assessed vs unaltered CS in the condition they are intended to be measured for clinical use.

EQA schemes usually include EQAMs at more than one concentration to examine participant performance. Each EQAM concentration should be examined for commutability. Different preparation approaches will likely be used to achieve different concentrations, and commutability is a unique property of each preparation of an EQAM. Consequently, whenever possible, each preparation to achieve different concentrations of a measurand in EQAMs should be examined for commutability.

It is not feasible for an EQAM provider to undertake a commutability assessment on every batch produced. EQAM providers should have in place manufacturing process control and risk management procedures to ensure consistency in preparation and the reasonable likelihood that future batches maintain commutability with CSs as observed in formal commutability assessment of the initial batch. Manufacturing controls to manage changes should be according to published international guidance on risk management for medical devices such as World Health Organization TGS-7 (17) and ISO 14 971 (18). Examples of questions to be addressed in a risk management process for a new batch of EQAM are:

1. What is the history of verified (non)commutability for a given EQAM?
2. Are there known issues concerning instability of any components and/or measurands of interest that may have changed in the new batch of the EQAM?
3. Are EQAMs with different levels of the target measurand prepared using different protocols, different amounts or sources of raw materials, or different donor populations?
4. Have there been any intentional or otherwise known changes in the ongoing production of the EQAM? Are subsequent batches of the EQAM prepared using different protocols, different equipment, different raw materials, different storage conditions, different donor populations, or different product containers (e.g., vials) for the EQAM?
5. Since the most recent commutability assessment of the EQAM was performed, are other newer MPs available for the target measurand(s) or have new measurands been added in the EQAM material that have not been examined for commutability or influence on the commutability of other measurands?

Summary of Recommendations

Commutability assessment of EQAMs will usually be performed by analyzing EQAMs and CS in the same run, and the evaluation of the results consists of 3 sequential steps (Fig. 1):

1. Examine results for outliers by measuring CSs using pairs of IVD-MDs; remove outliers as applicable.
2. Inspect CS results from pairs of IVD-MDs for DINS. Decide an acceptable M based on the DINS for the commutability study (Fig. 2B). In many cases, an M of 50% can be used.
3. Perform a commutability assessment of the EQAM including IVD-MDs not excluded by steps 1 and 2. An EQAM with measurement results inside the estimated PI is concluded to be commutable with CSs for use with that pair of IVD-MDs.

In addition, we recommend that commutability of an EQAM is examined at approximately the same time interval after preparation as the participants are expected to analyze the samples when stability of the measurand and thus commutability is influenced by such time intervals.

Since it is not possible to perform commutability assessment of each batch of EQAM, we recommend that EQAM providers have in place manufacturing process controls and risk management procedures to manage changes that can potentially influence the

commutability of a new batch of EQAM. If there is a reasonable risk that the new batch will have different commutability properties compared to earlier batches, a new commutability assessment should be performed.

Conclusion

We recommend EQA providers use commutable EQAM whenever possible. Commutable EQAM makes it possible to fulfill one of the most important goals of EQA, which is to examine if IVD-MDs for the same measurand give equivalent results for CSs. When an RMP is available for assigning a value to an EQAM that is commutable with CSs for use with a group of IVD-MDs, the EQA results can help assess whether successful implementation of metrological traceability to the corresponding reference measurement system has been achieved for a given IVD-MD. When an RMP is not available, the results for a commutable EQAM can be examined for equivalence among IVD-MDs to determine if metrological traceability to a CRM or to a harmonization protocol has been correctly implemented. One or more IVD-MDs that do not demonstrate equivalent results should be further examined to identify causes for the lack of agreement. In these ways, EQA schemes can be used to monitor and support harmonization and standardization of measurand results.

A tool for the practical calculations of DINS and commutability is given in an online resource (14).

Supplemental Material

Supplemental material is available at *Clinical Chemistry* online.

Nonstandard Abbreviations: EQA, external quality assessment; IVD-MD, in vitro diagnostic medical device; CS, clinical sample; EQAM, external quality assessment control material; CRM, certified reference material; ISO, International Organization for Standardization; MP, measurement procedure; RMP reference measurement procedure; DINS, difference in nonselectivity; APS, analytical performance specifications; M, relative prediction interval width increase; PI, prediction interval.

Author Contributions: *The corresponding author takes full responsibility that all authors on this publication have met the following required criteria of eligibility for authorship: (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; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved. Nobody who qualifies for authorship has been omitted from the list.*

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

Research Funding: None declared.

Disclosures: C. Burns, IFCC Scientific Division—MHRA representative and WHO Expert committee for Biological Standardisation (both unpaid advisory roles). W.G. Miller, Chair, IFCC Working Group on Commutability in Metrological Traceability. F. MacKenzie, UK NEQAS Charity Board Member. N. Greenberg has received consulting fees from Ortho Clinical Diagnostics, paid to Neil Greenberg Consulting Services and is a member, College of American Pathologists—Accuracy-Based Programs Committee (unpaid). S. Sandberg, payment or honoraria for lectures for Technopath and BioRAD (paid to institution); participation with monitoring trial of Disc Medicine on drug for eryhtropoietic protoporphyria; Chair ICHCLR, www.harmonization.net; President Ipnet, www.porphyrianet.org.

Role of Sponsor: No sponsor was declared.

References

1. Miller WG, Sandberg S. Quality control of the analytical examination process. In: Rifai N Chiu RWK Young I Burnham C-AD and Wittwer CT, editors. Tietz textbook of laboratory medicine. 7th ed. St Louis (MO): Elsevier; 2022. p. 129–63.
2. Braga F, Pasqualetti S, Panteghini M. The role of external quality assessment in the verification of in vitro medical diagnostics in the traceability era. Clin Biochem 2018;57:23–8.
3. Badrick T, Miller WG, Panteghini M, Delatour V, Berghall H, MacKenzie F, et al. Interpreting EQA—understanding why commutability of materials matters. Clin Chem 2022;68:494–500.
4. Miller WG, Schimmel H, Rej R, Greenberg N, Ceriotti F, Burns C, et al. IFCC Working group recommendations for assessing commutability part 1: general experimental design. Clin Chem 2018;64:447–54.
5. Nilsson G, Budd JR, Greenberg N, Delatour V, Rej R, Panteghini M, et al. IFCC working group recommendations for assessing commutability part 2: using the difference in bias between a reference material and clinical samples. Clin Chem 2018;64:455–64.
6. Budd JR, Weykamp C, Rej R, MacKenzie F, Ceriotti F, Greenberg N, et al. IFCC working group recommendations for assessing commutability part 3: using the calibration effectiveness of a reference material. Clin Chem 2018;64:465–74.
7. CLSI. Evaluation of commutability of processed samples; approved guideline. CLSI document EP14 –A3. Wayne (PA): CLSI; 2014.
8. Miller WG, Keller T, Budd J, Johansen JV, Panteghini M, Greensberg N, et al. Recommendations for setting a criterion for assessing commutability of secondary calibrator certified reference materials. [Epub ahead of print] Clin Chem August 11, 2023 as doi: 10.1093/clinchem/hvad104.
9. International Organization for Standardization. ISO 17511. In vitro diagnostic medical devices—metrological traceability of values assigned to calibrators and control materials.pdf. Geneva (Switzerland): ISO; 2020.
10. International Organization for Standardization. ISO 21151. In vitro diagnostic medical devices—requirements for international harmonisation protocols establishing metrological traceability of values assigned to calibrators and human samples. Geneva (Switzerland): ISO; 2020.
11. Joint Committee for Guides in Metrology (JCGM). International vocabulary of metrology—basic and general concepts and associated terms (VIM) (2008 version with minor corrections). 3rd ed. Sevre (France): International Bureau of Weights and Measures; 2012.
12. Rosner B. Percentage points for a generalized ESD many-outlier procedure. Technometrics 1983;25:165–72.
13. Millard SP. Proof of safety vs proof of hazard. Biometrics 1987;43:719–25.
14. Sandberg S, Fauskanger P, Johansen J, Keller T, Budd J, Greenberg N, et al. Commutability assessment of external quality assessment material. https://www.noklus.no/en/a-practical-tool-for-commutability-

evaluation-of-external-quality-assessment-material/ (Accessed August 17, 2023).
15. Fuller WA. Measurement error models. New York (NY): John Wiley & Sons; 1987.
16. Gillard JW. Errors in variables regression: what is the appropriate model? Cardiff (United Kingdom): Cardiff University; 2007.
17. WHO. Technical guidance series (TGS) for WHO prequalification—diagnostic assessment. Risk management for manufacturers of in vitro diagnostic medical devices. TGS-7. Geneva (Switzerland): World Health Organization; 2018.
18. International Organization for Standardization. ISO 14197. Medical devices—application of risk management to medical devices. Geneva (Switzerland): ISO; 2019.