Demonstrating the extent-of-equivalence between multiple certified reference materials (CRMs) for the same measurand

1. Purpose

This document describes the JCTLM process which can be followed by the producers of Certified Reference Materials (CRMs) for demonstrating the extent-of-equivalence of the value assigned to CRMs that are nominally fit for the identical purpose(s). The process is intended to ensure that information to indicate the extent-of-equivalence is available for these CRMs. When at least one such CRM is listed by the JCTLM, the nominator is responsible for the submission of an extent-of-equivalence report that will be made available to the end users on the JCTLM website. The extent-of-equivalence evaluation allows demonstration of whether or not the certified values of CRMs are consistent within their stated uncertainties and thus are fit-for-purpose when used in a calibration hierarchy for end-user measurement procedures. When more than one CRM is listed by the JCTLM, potential users need this information to make an informed selection of the CRM best suited to their needs.

The extent-of-equivalence evaluation provides no information on the commutability of CRMs which must be assessed through commutability studies, nor on assessment that the uncertainty or other property is fit-for-purpose in a calibration hierarchy.

2. Contents

3. Scope

This procedure applies to all nominations of CRMs that are submitted for evaluation by JCTLM DB WG for inclusion in the JCTLM database where there are JCTLM listed CRMs, that are claimed to be fit for the same purpose and they have certified values for one or more of the same measurands. Replacement lots are deemed to be new CRMs and are thus subject to the extent-of-equivalence testing procedures described for new CRMs.

Not in scope for extent-of-equivalence is demonstration of commutability of CRMs, when applicable.

4. Acronyms and definitions

All acronyms and definitions employed in the procedures of the JCTLM DB WG Quality Manual are given in the procedure document JCTLM EXE-G01, Glossary of terms and definitions.

5. Responsibilities and Authorizations

5.1. The JCTLM Executive Committee (EC) has the following responsibilities under this procedure.

5.1.1. To accept or reject DB WG's recommendations for listing in the JCTLM Database extent-of-equivalence reports of CRMs evaluated for the same measurand.

5.1.2. To accept or reject DB WG's recommendation for follow up investigation of unacceptable extent-of-equivalence data for a nominated CRM when compared to one or more listed CRMs for the same measurand.

5.2. The JCTLM Secretariat has the following responsibilities under this procedure.

5.2.1. To forward DB WG's recommendations to the JCTLM EC and communicate the EC's decisions to the CRM nominator.

5.2.2. To appropriately list approved CRM extent-of-equivalence reports on the JCTLM database website.

5.3. The DBWG Vice-Chairs have the following responsibilities under this procedure.

5.3.1. To review RT recommendations

5.3.2. To communicate DB WG's recommendations to the JCTLM Secretariat for EC approval.

5.4. The RT leaders have the following responsibilities under this procedure.

5.4.1. To ensure that representative comparison data for material extent-of-equivalence demonstration are available for evaluation by the RT.

5.4.2. If necessary, to communicate with the JCTLM Secretariat for obtaining additional supporting evidence for extent-of-equivalence studies.

5.5. The RTs have the following responsibilities under this procedure.

5.5.1. To review the extent-of-equivalence data as part of the evaluation of a nominated CRM.

5.5.2. To determine if the extent-of-equivalence data demonstrates that the nominated CRM will produce fit-for-purpose results when used in the calibration hierarchies of end-user measurement procedures.

5.6. The CRM nominator has the following responsibilities under this procedure.

5.6.1. To initiate extent-of-equivalence evaluation when at least one JCTLM listed CRM is identified as being fit for the same purpose for the same measurand(s) as the nominated CRM.

5.6.2. To define and justify an appropriate measurement performance criterion for the particular measurand(s) in the given matrix.

5.6.3. To define an appropriate measurement protocol.

5.6.4. To perform the necessary measurements.

5.6.5. To prepare a report of analysis that suitably documents the performed measurements and the measurement systems used.

5.6.6. The final report documenting the extent-of-equivalence determination of the evaluated CRMs will be published by JCTLM as part of the listing of the CRM.

6. General Procedures

6.1. The process for demonstrating the extent-of-equivalence of CRMs will be initiated by CRM nominators whenever one or more materials that are nominally fit for the same purpose are identified in the JCTLM database. The extent-of-equivalence demonstration process should be completed when the producer of a CRM submits a nomination for evaluation by DB WG RT.

6.2. The extent-of-equivalence of materials must include considerations about the impact of a new CRM when used for its intended purpose in the calibration hierarchies of existing end-user measuring systems.

6.3. For CRMs (m.1, m.2 and m.3 in ISO 17511:2020), the following information is adequate to demonstrate extent-of-equivalence. Other experimental designs may also be acceptable.

6.3.1. Demonstrate the assigned value of a nominated CRM and a listed CRM with similar or smaller uncertainty than the nominated CRM is recovered within the stated uncertainties of both materials when measured using the same RMM/P in the same run.

6.3.2. If there is no listed RMM/P available, then measurements may be made using a representative number of well validated end-user MPs for which the listed CRM has been shown to be commutable with clinical samples.

6.4. Whenever possible, the CRM producer may rely on already available data to demonstrate extent-of-equivalence of their CRM (m.1, m.2 or m.3 in ISO 17511:2020). The following approaches may be used for demonstrating extent-of-equivalence of primary and secondary CRMs for well defined molecules. Alternate approaches may be needed for complex measurands such as proteins.

6.4.1. For CRMs, relevant published calibration and measurement capability claims (CMCs) within the BIPM Key Comparisons Database (KCDB) can be reported, as

these have been internationally accepted based on results of international comparisons.

6.4.2. The results of key comparisons where the producer's CRM has been measured at a central facility using an RMM/P at the same time as measurements were performed on nominally the same CRMs from other producers and all results compared. National Metrology Institutes (NMIs) and Designated Institutes (DIs) participating in the CIPM Mutual Recognition Arrangement (MRA) may use the results of key comparisons to demonstrate the extent-of-equivalence of their CRMs with others in the comparison exercise. This is possible for different CRMs compared directly at one NMI/DI (so called Model 2 comparison) or alternatively for comparisons of NMI/DI RMM/Ps on a common comparison material (so-called Model 1 comparison), where these RMM/Ps are the basis of value assignment of the different CRMs for which extent-of-equivalence information is needed.

6.4.3. As part of the producer's certification process of the CRM, where the RMM/P used to certify the producer's CRM has been validated by performing measurements in the same run on nominally the same CRM(s) already listed in the JCTLM database.

6.5. When more than one CRM exists within the JCTLM database for a measurand with extent-of-equivalence already demonstrated, the number of CRMs required to include in an extent-of-equivalence study will depend on the measurement uncertainty that the producer of the nominated CRM wishes to demonstrate equivalence for. The comparison may need to only include one already listed CRM with an uncertainty that is small enough that the maximum allowable combined uncertainty for the clinical sample is fulfilled.

6.6. If other laboratories than the nominator are involved in generating data for the extent of equivalence report, operation according to the requirements of the most recent editions of ISO/IEC 17025 and ISO 15195 is expected. The nominator will be responsible to select the appropriate laboratory.

6.7. For CRMs, other interested parties may have performed and published the needed information. The CRM nominator will investigate the availability of extent-of-equivalence information publicly available and, if it exists, evaluate the data against the criteria established above. Based on this evaluation, an extent-of-equivalence report will be prepared to be added to the JCTLM database with the CRM to be listed.

7. Relative measurements against JCTLM listed CRMs

7.1. Extent-of-equivalence evaluation can be performed with adequate measurements of measurand values in the different CRMs. The major considerations for selecting an appropriate measurement system for these relative measurements are sensitivity, selectivity, linearity, and repeatability. It is the responsibility of the CRM nominator to establish fit-for-purpose criteria for these performance metrics for each set of multiple materials, given the stated purpose of the materials and the certified values and uncertainties of the measurands.

The following apply separately to each measurand evaluated:

7.1.1. Sensitivity: the measurement system must provide an adequately sensitive signal at the lowest level certified in any of the CRMs.

7.1.2. Selectivity: the measurement system must provide an adequately selective signal given the materials' general matrix and any explicitly stated CRM components.

7.1.3. Linearity: the measurement system must provide an adequately linear signal to allow direct proportional comparison between the signals of the lowest and highest levels certified in any of the CRMs.

7.1.4. Repeatability: the measurement system must provide adequate repeatability precision over the time course of all measurements required for the comparison. This repeatability is relative to the certified uncertainties of the CRMs being compared.

7.1.5. Uncertainty: The measurement system used must provide results with adequate uncertainty to be fit-for-purpose to describe the extent-of-equivalence of the CRMs for the described intended use.

7.2. The CRM nominator will determine the extent-of-equivalence of the materials.

7.2.1. Given sufficiently unbiased measurements such as those summarized in Table 1 and Figures 1a and 1b of the potassium example in Attachment 1, extent-of-equivalence can be evaluated from the extent of overlap between the 95% confidence intervals of the measured value, $X_i \pm U_{95}(X_i)$, and the certified value, $C_i \pm U_{95}(C_i)$, of each material. The least squares regression of the measured X_i as a linear function of the certified values, $X_i = a + bC_i$, enables evaluation of bias: within their asymptotic uncertainties, the intercept is zero and the slope is one.

While Figure 1a displays both the measured and certified value intervals, it is difficult to visually compare them due to the range of measurand levels in the CRMs. Figure 1b provides a much more interpretable display by expressing the intervals relative to the certified values: $X_i/C_i \pm U_{95}(X_i)/C_i$ and $1 \pm U_{95}(C_i)/C_i$. For all materials, there is a nearly complete overlap between the measured and certified intervals. Consequently, these CRMs are fully and about equally equivalent for potassium.

7.2.2. Relative extent-of-equivalence can be evaluated using biased measurement systems. The measured and certified values summarized in Table 2 and Figures 2a and 2b of the cholesterol example in Attachment 2 are linearly related; however, while the intercept is zero to within model error, the slope is not unity, i.e., the measurement system is proportionally biased. The measurement and certified value intervals must be suitably adjusted before they can be compared.

Figure 2b displays the relative measurement interval, $X_i/C_i \pm U_{95}(X_i)/C_i$ and the slope-scaled certified intervals, $b \pm U_{95}(C_i)/C_i$. While the intervals for all materials do overlap, the overlap for two of the materials is marginal. Thus, while these CRMs are comparable for cholesterol, a few of the materials have poorer equivalence than the majority. The variability in the size and overlap of the measurement and certified value intervals in Figure 2b suggest that the certified

uncertainties (including the material homogeneity component) of several materials should be re-evaluated.

7.2.3. In general, extent-of-equivalence can be demonstrated whenever an adequately descriptive functional relationship between measured and certified values can be established. For complex relationships, one or more research chemometrics experts should be involved in the study.

8. Dealing with discordance between CRMs in extent-of-equivalence studies

8.1. If the results of an extent-of-equivalence study indicate discordance among the evaluated CRMs relative to fit-for-purpose criteria, the nominated CRM under review will not be accepted for listing in the JCTLM database until the discrepancy is explained. The DB WG Vice-Chair, in consultation with the DB WG Chair, will collaborate with the nominator and producers of listed CRMs to investigate the discrepancy. Producers of listed CRMs are expected to participate in the investigation.

9. Related documents

JCTLM DB WG-P-00	Quality Policy and Definitions
JCTLM DBWG-P-01A	Outline of JCTLM Procedures for Evaluating Materials and Methods to be Listed as Being of Higher Metrological Order
JCTLM DBWG-P-03A	Process for Review and Approval of Nominated Certified Reference Materials and Reference Measurement Methods

10. Revision History

Version number	Date of Issue/Review	Summary of change
1.0	09/15/2004	Initial issue of WG1 Quality System Procedures
2.0 2.1	10/15/2005 24/01/2006	Separation of comparison of reference measurement procedures using the same principle of measurement to WG1-P-04B Corrections to grammar and terms where identified. Resolution of inconsistencies. Correction of minor textual errors
3.0	27/01/2017	Document revised for implementing modifications requested by EC's at its 17th meeting
4.0	01/02/2023	Modifications discussed at the December 2022 Executive Meeting

11. Attachments

Attachment 1	Example, Comparison of Certified and Measured Values of Potassium in Human Serum CRMs
Attachment 2	Example, Comparison of Certified and Measured Values of Cholesterol in Human Serum CRMs

Attachment 1 Potassium Example

The following example tables and graphs represent possible ways to show extent-of-equivalence data. In many cases, comparison between a nominated and one listed CRM is sufficient. Other data display approaches may also be acceptable.

Table 1: Protocol, Measurements, and Summary Statistics for the Comparison of Certified and Measured Values of Potassium in Human Serum CRMs
All values are in mmol/L.

	Certified [b]		Measured [c] Campaign 1		Measured [c] Campaign 2		Summary Statistics				
CRM [a]	Value	U_{95}	Value	U_{95}	Value	U_{95}	Mean	SD	R [d]	u_c [e]	U_{95} [f]
JCCRM 111-5, L	3.250	0.020	3.241	0.027	3.241	0.025	3.241	0.000	0.013	0.013	0.026
JCCRM 111-5, M	4.400	0.020	4.419	0.027	4.400	0.034	4.410	0.013	0.015	0.020	0.041
JCCRM 111-5, H	5.690	0.020	5.698	0.047	5.706	0.044	5.702	0.006	0.023	0.023	0.047
SRM 909b, Lv 1	3.424	0.025	3.437	0.028	3.425	0.027	3.431	0.008	0.014	0.016	0.032
SRM 909b, Lv 2	6.278	0.052	6.286	0.052	6.248	0.048	6.267	0.027	0.025	0.037	0.073
SRM 956a, Lv 1	6.008	0.020	5.983	0.049	6.008	0.046	5.996	0.018	0.024	0.030	0.059
SRM 956a, Lv 2	3.985	0.020	3.993	0.033	3.993	0.031	3.993	0.000	0.016	0.016	0.032
SRM 956a, Lv 3	2.025	0.008	2.022	0.017	2.016	0.016	2.019	0.004	0.008	0.009	0.019

a CRMs suitable for the calibration or verification of potassium in human serum proposed for listing by the JCTLM as of May, 2003. Materials with prefix 'JCCRM' are available from the HECTEF Standard Reference Center Foundation, Japan; those with prefix 'SRM' are available from the National Institute of Standards and Technology (NIST), USA.

b Certified values and 95% confidence level uncertainties as listed on the CRM Certificates.

c Measured values and fully evaluated 95% confidence level uncertainties. The determinations were performed at NIST using isotope dilution mass spectroscopy. The potassium level of each level of every CRM was fully evaluated in two separately analyzed vials of the material. The two sets of vials were analyzed in separate measurement campaigns.

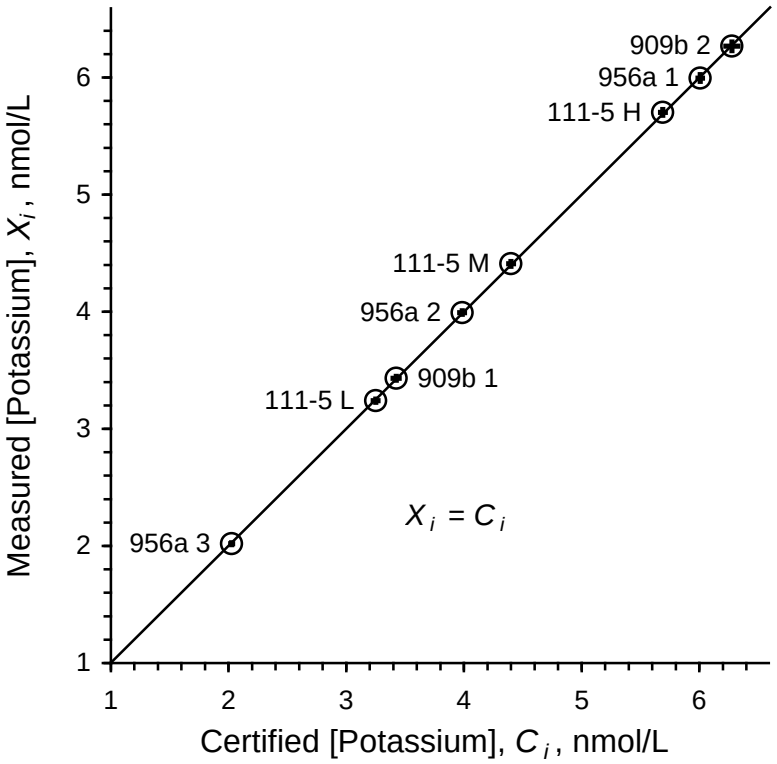
d Expected repeatability standard deviation of a single determination in this matrix in this laboratory using the given method, estimated as one-half of the pooled Campaign 1 and 2 U_{95} values:

$$R = 0.5\sqrt{\frac{U_{95,1}^2 + U_{95,2}^2}{2}}.$$

e Combined uncertainty, $u_c = \sqrt{SD^2 + R^2}$.

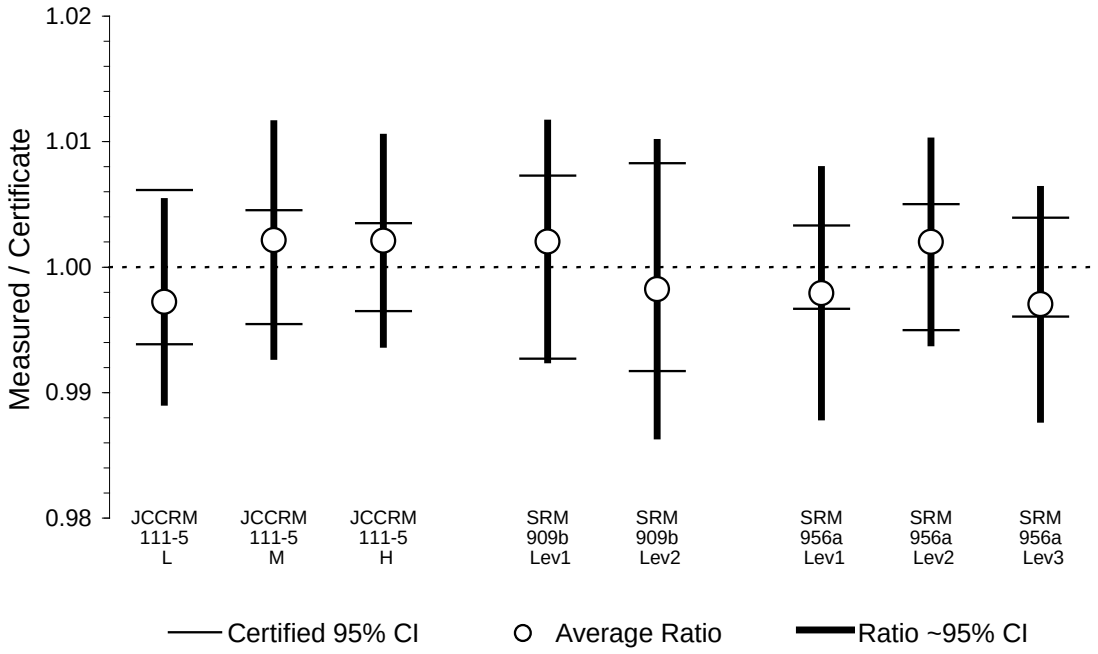
f Approximate 95% confidence expanded uncertainty, $2u_c$.

Figure 1a: Scattergram Display of the
Extent-of-equivalence of Potassium in Human Serum CRMs



The horizontal axis reports the certified values, C_i; the vertical reports the average measured values, X_i. Each level of each CRM is displayed as approximate 95% uncertainty intervals along both axes. The intersection of these intervals is bounded by an open circle to aid visual inspection. The line denotes the identify function: $X_i = C_i$. This model was chosen after finding that the intercept of the linear model was not significantly different from zero and the slope was not significantly different from unity:
$X_i = (0.003 \pm 0.013) + (0.9993 \pm 0.0028) \times C_i$.

Figure 1b: Ratio Display of the
Extent-of-equivalence of Potassium in Human Serum CRMs



The horizontal axis reports the CRMs evaluated. The vertical axis reports the ratio between the average measured and certified values of each CRM, X_i/C_i. The open circles denote these average ratios; the dark vertical lines represent the approximate 95% uncertainty interval on these averages. The light horizontal lines represent the certified 95% confidence intervals. The dotted line represents the expected ratio for the suite of all materials given the observed identity between the measured and certified values.

Attachment 2 Cholesterol Example

The following example tables and graphs represent possible ways to show extent-of-equivalence data. In many cases, comparison between a nominated and one listed CRM is sufficient. Other data display approaches may also be acceptable.

Table 2: Protocol, Measurements, and Summary Statistics for the Comparison of Certified and Measured Values of Cholesterol in Human Serum CRMs
All values are in mg/dL.

	Certified [b]		Measured [c]		Summary Statistics				
CRM [a]	Value	U_{95}	Set 1	Set 2	Mean	SD	R [d]	u_c [e]	U_{95} [f]
JCCRM 211-1, H	205.20	0.80	203.80	204.20	204.00	0.29	0.72	0.78	1.55
JCCRM 211-1, M	262.40	1.00	259.02	260.35	259.68	0.94	0.92	1.32	2.63
SRM 909b, Lv 1	146.40	1.11	146.21	146.40	146.31	0.13	0.52	0.54	1.08
SRM 909b, Lv 2	235.30	1.85	233.45	233.80	233.63	0.24	0.83	0.87	1.73
SRM 968c, Lv 1	133.50	1.30	132.08	132.40	132.24	0.22	0.47	0.52	1.04
SRM 968c, Lv 2	166.90	1.70	166.11	166.20	166.15	0.06	0.59	0.60	1.19
SRM 1589a	157.76	0.37	154.97	155.50	155.24	0.37	0.55	0.67	1.34
SRM 1951a, Lv 1	182.15	0.45	181.63	182.10	181.86	0.34	0.64	0.73	1.45
SRM 1951a, Lv 2	276.67	0.55	277.14	276.95	277.05	0.14	0.98	0.99	1.98
SRM 1952a, Lv 1	147.50	0.96	147.04	147.20	147.12	0.11	0.52	0.53	1.07
SRM 1952a, Lv 2	233.40	0.96	231.68	229.25	230.47	1.72	0.82	1.90	3.81
SRM 1952a, Lv 3	333.00	1.65	327.28	330.45	328.87	2.24	1.19	2.53	5.07

a CRMs suitable for the calibration or verification of cholesterol in human serum proposed for listing by the JCTLM as of May, 2003. Materials with prefix 'JCCRM' are available from the HECTEF Standard Reference Center Foundation, Japan; those with prefix 'SRM' are available from the National Institute of Standards and Technology (NIST), USA.

b Certified values and 95% confidence level uncertainties as listed on the CRM Certificates.

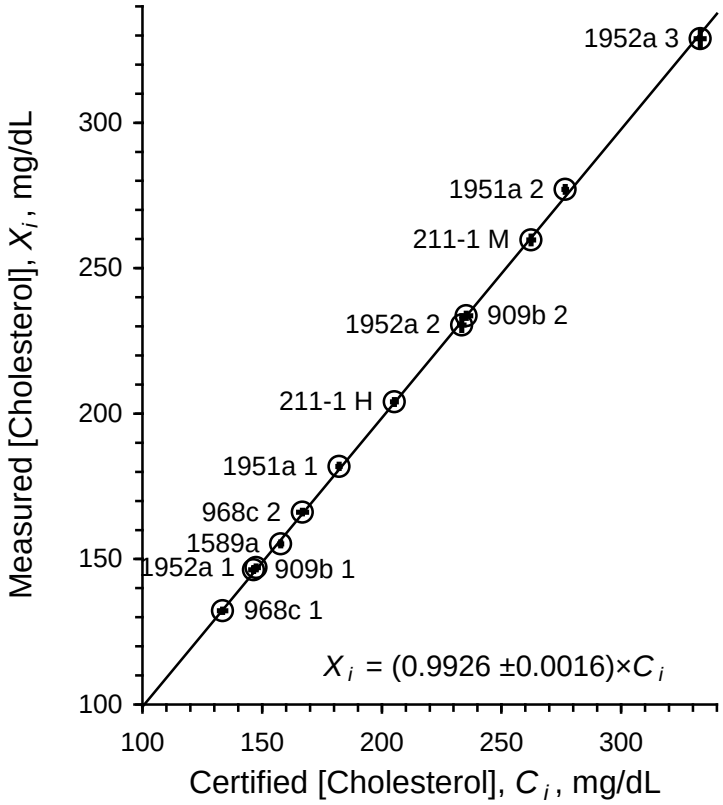
c Measurements performed at NIST using isotope dilution/gas chromatography/mass spectroscopy. The cholesterol level of each level of every CRM was determined in two separately analyzed aliquots. Each set of aliquots was analyzed during a single 24-hour period.

d Expected repeatability standard deviation of a single determination in this matrix in this laboratory using the given method, 0.35% of the measured value.

e Combined uncertainty, $u_c = \sqrt{SD^2 + R^2}$.

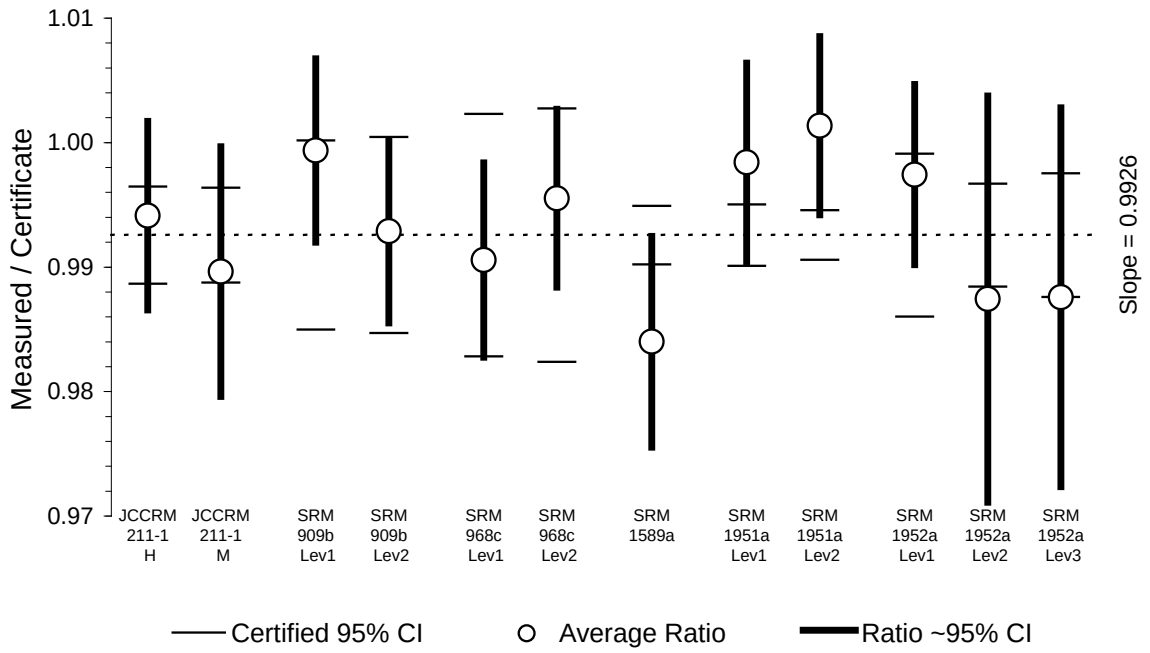
f Approximate 95% confidence expanded uncertainty, $2u_c$.

Figure 2a: Scattergram Display of the
Extent-of-equivalence of Cholesterol in Human Serum CRMs



The horizontal axis reports the certified values, C_i; the vertical reports the average measured values, X_i. Each level of each CRM is displayed as approximate 95% uncertainty intervals along both axes. The intersection of these intervals is bounded by an open circle to aid visual inspection. The line denotes the best linear model for the relationship between the measured values and the certified values: $X_i = (0.9926 \pm 0.0016) \times C_i$. This proportional model was chosen after finding that the intercept of the linear model was not significantly different from zero: $X_i = (0.92 \pm 1.27) + (0.9885 \pm 0.0059) \times C_i$.

Figure 2b: Ratio Display of the
Extent-of-equivalence of Cholesterol in Human Serum CRMs



The horizontal axis reports the CRMs evaluated. The vertical axis reports the ratio between the average measured and certified values of each CRM, X_i/C_i. The open circles denote these average ratios; the dark vertical lines represent the approximate 95% uncertainty interval on these averages. The light horizontal lines represent the certified 95% confidence intervals, scaled by the proportional model for the relationship between the measured values and the certified values: $X_i = (0.9926 \pm 0.0016) \times C_i$. The dotted line represents the expected ratio for the suite of all materials given the observed proportionality between the measured and certified values.