CCQM-K80

Comparison of value-assigned CRMs and PT materials: creatinine in human serum

Final Report

Johanna E. Camara, David L. Duewer, Hugo Gasca-Aragon, Katrice A. Lippa
National Institute of Standards and Technology (NIST)
Gaithersburg, MD USA

Blaza Toman
Statistical Engineering Division, National Institute of Standards and Technology
Gaithersburg, MD USA

With contributions from:

Melina Pérez Urquiza
Centro Nacional de Metrologia (CENAM), Querétaro, México

Sang-Ryoul Park
Korea Research Institute of Standards and Science (KRISS), Taejon, Korea

Gavin O'Connor
LGC Limited (LGC), Teddington, Middlesex UK

Xinhua Dai
National Institute of Metrology (NIM), Beijing, China

William F. Guthrie, Reenie M. Parris, Karen W. Phinney, Antonio Possolo, Katherine E. Sharpless
National Institute of Standards and Technology (NIST), Gaithersburg, MD USA

Jolene D. Splett
National Institute of Standards and Technology (NIST), Boulder, CO USA

Ruediger Ohlendorf, André Henrion
Physikalisch-Technische Bundesanstalt (PTB), Braunschweig, Germany

Abstract

The 2009 CCQM-K80 "Comparison of value-assigned CRMs and PT materials: creatinine in human serum" is the first in a series of Key Comparisons directly testing the chemical measurement services provided to customers by National Metrology Institutes (NMIs) and Designated Institutes. CCQM-K80 compared the assigned serum creatinine values of certified reference materials (CRMs) using measurements made on these materials under repeatability conditions. Six NMIs submitted 17 CRM materials for evaluation, all intended for sale to customers. These materials represent nearly all of the higher-order CRMs then available for this clinically-important measurand.

The certified creatinine mass fraction in the materials ranged from 3 mg/kg to 57 mg/kg. All materials were stored and prepared according the specifications provided by each NMI. Samples were processed and analyzed under repeatability conditions by one analyst using isotope dilution liquid chromatography-mass spectrometry. The instrumental repeatability imprecision, expressed as a percent relative standard deviation, was 1.2 %.

Given the number of materials and the time required for each analysis, the measurements were made in two measurement campaigns ("runs"). In both campaigns, replicate analyses (two injections of one preparation separated in time) were made on each of two or three independently prepared aliquots from one randomly selected unit of each of the 17 materials. The mean value, between-campaign, between-aliquot, and between-replicate variance components, standard uncertainty of the mean value, and the number of degrees of freedom associated with the standard uncertainty were estimated using a linear mixed model. Since several of the uncertainties estimated using this traditional frequentist approach were associated with a single degree of freedom, Markov Chain Monte Carlo Bayesian analysis was used to estimate 95 % level-of-confidence coverage intervals, U_{95}. Uncertainty-weighted generalized distance regression was used to establish the Key Comparison Reference Function (KCRF) relating the assigned values to the repeatability measurements. Parametric Bootstrap Monte Carlo was used to estimate 95 % level-of-confidence coverage intervals for the degrees of equivalence of materials, $d \pm U_{95}(d)$, and of the participating NMIs, $D \pm U_{95}(D)$. Because of the wide range of creatinine mass fraction in the materials, these degrees of equivalence are expressed in percent relative form: $\%d \pm U_{95}(\%d)$ and $\%D \pm U_{95}(\%D)$.

On the basis of leave-one-out cross-validation, the assigned values for 16 of the 17 materials were deemed equivalent at the 95 % level of confidence. These materials were used to define the KCRF. The excluded material was identified as having a marginally underestimated assigned uncertainty, giving it large and potentially anomalous influence on the KCRF. However, this material's $\%d$ of 1.4±1.5 indicates that it is equivalent with the other materials at the 95 % level of confidence. The median $|\%d|$ for all 17 of the materials is 0.3 with a median $U_{95}(\%d)$ of 1.9. All of these higher-order CRMs for creatinine in human serum are equivalent within their assigned uncertainties.

The median $|\%D|$ for the participating NMIs is 0.3 with a median $U_{95}(\%D)$ of 2.1. These results demonstrate that all participating NMIs have the ability to correctly value-assign CRMs and proficiency test materials for creatinine in human serum and similar measurands.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

LIST OF FIGURES

0. INTRODUCTION

The Working Groups of the Consultative Committee for Amount of Substance – Metrology in Chemistry (CCQM) are responsible for selecting and overseeing the operation of Key Comparisons (KCs) that address chemical measurement-related issues important for international trade, environmental, health, and safety-related decision making. One objective of the Comité International des Poids et Mesures Mutual Recognition Arrangement (CIPM MRA) [1] is to establish the degree of equivalence of national measurement standards maintained by National Metrology Institutes (NMIs) and Designated Institutes (DIs). To more efficiently address this objective, the CCQM Organic Analysis Working Group (OAWG) has recently agreed to four basic study designs:

Track A: KCs that test core competencies for the delivery of measurement services to customers,

Track B: KCs that directly assess the equivalence of measurement services provided to customers,

Track C: KCs in emerging areas of global interest and importance with a potential parallel Pilot Study,

Track D: Capability assessment studies to allow assessment of measurement capabilities being established in a new area for NMIs and DIs.

0.1 Historical Background

At the April 2009 CCQM OAWG Meeting (Sévres, France), an experimental design was proposed for OAWG KCs that directly tests measurement services provided to customers through Certified Reference Materials (CRMs) and value-assigned Proficiency Testing (PT) materials. The premise of the Track B design is that these capabilities can be demonstrated when several different value-assigned materials are available that deliver a nominally identical measurand (i.e., a given analyte in a sufficiently similar matrix with the assigned quantity value expressed in given units). This can be accomplished by comparing the assigned values with measurements made on the entire group of materials under repeatability conditions. Such comparisons directly reflect the measurement capabilities of the organizations that value-assigned the materials when not constrained by sample amount or reporting deadline. Track B comparisons also address the material homogeneity and stability assessment capabilities of the participating institutions and potentially their packaging, storage, and shipping capabilities.

On the basis of the number of NMIs and DIs providing suitable materials, the number of available materials, and the intrinsic importance of the measurands, the OAWG authorized Track B KCs for ethanol in an aqueous matrix and creatinine in human serum. NIST volunteered to make the required repeatability-condition measurements for the creatinine KC and to coordinate the study. Given the potential utility of such KCs, all OAWG member institutions that deliver measurement services via one or more appropriately value-assigned materials were asked to participate.

0.2 Measurand Background

Creatinine is a very polar analyte that is a clinically important diagnostic marker for renal function. Routine clinical tests mostly based on enzymatic reactions are subject to interferences from various materials coexisting with creatinine in samples. Among these interferences is

creatine, which presents both separation and inter-conversion challenges. Use of different methods of analysis, different reagents, etc., may lead to significantly different results. Therefore, reference methods and reference materials are needed to maintain adequate accuracy in routine measurements for creatinine in blood or serum.

The OAWG first addressed creatinine measurement practice in 2000 with CCQM-P9 "Determination of creatinine in human serum." Based upon the excellent agreement among the one academic and four NMI participants, CCQM-K12 of the same title was conducted in 2002 with five NMIs reporting results. In 2004, two additional NMIs demonstrated their creatinine measurement capability in the subsequent study CCQM-K12.1.

0.3 Key Comparison Design Background

Since different CRM and PT materials for the same nominal measurand will likely deliver different analyte quantities and may have somewhat different matrices, the evaluation of degrees of equivalence among the participants requires a quite different approach than that of studies in which participants analyze nominally identical samples of the same material. The basic methodology for comparing two sets of measurement results (assigned values and repeatability measurements) for a given group of materials is well established and has been used with success in a number of studies conducted by the CCQM Gas Analysis Working Group (GAWG).

The relatively unfamiliar experimental design and methodology considerations for Track B KCs are described at some length in the companion report "Comparison of value-assigned CRMs and PT materials: experimental design and data evaluation" [2]. The design and analysis considerations and tasks can be divided into the following four steps:

Step 1) Design the study taking into account the number of candidate materials and their analyte levels and matrices together with the analytical capabilities and available resources of the measurement laboratory.

Step 2) Inform the participants of the materials and material quantities required and a target date for supplying those materials to the measurement laboratory. Collect all materials at the measurement laboratory. Store the materials under the conditions specified by the participants until such time as measurements are made. Make the measurements under repeatability conditions. The measurement procedure needs to provide results that are a simple function of analyte level (linear in mass fraction is best) but does not need to be calibrated. Summarize the measurement results for each material as a value and an uncertainty on that value.

Step 3) Establish a consensus model that relates the assigned and measured values, using a technique that takes into account the uncertainties on both the assigned and measured values.

Step 4) Estimate the difference between the assigned and measured value for each material and the value predicted from the consensus model, taking into account the uncertainties on the definition of the model as well as those on the observed values. Convert these differences into degrees of equivalence for each of the study materials. If there are two or more study materials from a given participant, combine the degrees of equivalence for those materials into a degree of equivalence for the participant.

This report presents the results of CCQM-K80 in the context of this four-step structure.

1. STEP 1: DESIGN OF THE STUDY

1.1 Timeline

Table 1: Timeline

Date	Action
21 Apr 2009	OAWG authorized K80 study and approved protocol
10 Aug 2009	Call for Participation emailed to OAWG members (see Appendix A)
1 Oct 2009	Sufficient documentation (see Appendix B) and study materials received at NIST
15 Oct 2009	Repeatability measurements complete
4 Nov 2009	Preliminary results presented to OAWG at Rio de Janeiro meeting
1 Apr 2010	Draft A report distributed to OAWG
14 Mar 2012	Draft B report distributed to OAWG
31 Jan 2013	Final report delivered to OAWG Chair

1.2 Participants

Table 2: Participating Institutions

Code	Institution	Country
CENAM	Centro Nacional de Metrología	México
KRISS	Korea Research Institute of Standards and Science	Korea
LGC	Laboratory of the Government Chemist	UK
NIM	National Institute of Metrology	China
NIST	National Institute of Standards and Technology	USA
PTB	Physikalisch-Technische Bundesanstalt	Germany

1.3 Materials

Only materials with certification values valid as of the 22 October 2009 measurement date were eligible for inclusion in CCQM-K80. Likewise, only materials either directly certified in units of mass fraction or that could be converted into units of mass fraction were eligible. Participants with materials certified in quantities other than mass fraction were asked to provide the converted values and the uncertainty on the converted values at a 95 % level of confidence. All but one of the OAWG members known to currently provide creatinine in human serum CRMs chose to participate in CCQM-K80; this one institution chose not to convert the mass/volume units of certification to the mass fraction quantity required for CCQM-K80.

To limit the number of materials in CCQM-K80 to a quantity that could be measured under repeatability conditions, each participating institution was asked to provide no more than four materials each. Any institution with more than four suitable materials was asked to provide the four that, in the institution's collective judgment, were deemed most representative (primarily with respect to matrix and analyte level) of its entire suite. Each participant was asked to provide three units of each value-assigned material. All materials were stored at the temperature

specified in the provided instructions from the time of receipt to the time when the material was prepared for analysis.

Table 3 summarizes the certification information as provide by the participants for the 17 materials submitted for inclusion in CCQM-K80. In addition to identifying the certifying institution, the certified value "*V*," the uncertainty on the certified value "$U_{95}(V)$" at a 95 % level of confidence, and the units of certification, Table 3 lists the auxiliary information deemed useful for evaluating the materials' suitability for inclusion in the KC, for storing and processing the materials, and for the measurement design: the form of the material, the amount of material available per unit, any specified minimum sample quantity per analysis, the recommended storage temperature, the original certification date, and the expiration date for the certification. Most of this information was available in the certification documents supplied by the participating institutions in response to the solicitation. When required information was not supplied in submitted documents, it was solicited via email. The repeatability measurements were not begun until all required information was compiled and the accuracy of the compilation confirmed by the participating institutions.

Table 3 also lists the basic analytical technique used within each institution for certification and the condition of the samples upon arrival at NIST. This information was recorded as a potential aid to the interpretation of results. The CENAM and LGC materials arrived well frozen on dry ice in completely intact packaging. The (lyophilized) PTB materials were shipped without temperature control and arrived in completely intact packaging. The KRISS materials arrived in intact packaging but, due to delay in Customs, at ambient temperature. Also due to Customs delay, the NIM materials were nearly thawed upon arrival at NIST; from the battered appearance of the packaging, they may well have gone through one or more freeze/thaw cycles. Transportation was not an issue for the NIST materials.

Table 3: Materials

			Certified Value			Auxiliary Information [a]							
NMI	Material	Type	V_i	$U_{95}(V_i)$	Units	Matrix	mL	MinSam	°C	Year	Expires	Condition[b]	Method[c]
CENAM	DMR 263a	CRM	7.35	0.35	mg/kg	Frozen	1		-80	2004	3-Nov-09	frozen	ID-LC/MS
KRISS	111-01-01A	CRM	5.96	0.09	mg/kg	Frozen	3		-75	2007	31-Dec-12	thawed	ID-LC/MS
KRISS	111-01-02A	CRM	27.49	0.33	mg/kg	Frozen	3		-75	2007	31-Dec-12	thawed	ID-LC/MS
KRISS	111-01-03A	CRM	7.08	0.08	mg/kg	Lyoph	10		4	2007	31-Dec-17	ambient	ID-LC/MS
KRISS	111-01-04A	CRM	24.87	0.29	mg/kg	Lyoph	10		4	2007	31-Dec-17	ambient	ID-LC/MS
LGC	ERM-DA250a	CRM	39.	2.	mg/kg	Frozen	1	0.4 g	-20	2006	Shipment+3 mo	frozen	ID-LC/MSMS
LGC	ERM-DA251a	CRM	22.	2.	mg/kg	Frozen	1	0.4 g	-20	2006	Shipment+3 mo	frozen	ID-LC/MSMS
LGC	ERM-DA252a	CRM	3.1	0.2	mg/kg	Frozen	1	0.4 g	-20	2006	Shipment+3 mo	frozen	ID-LC/MSMS
LGC	ERM-DA253a	CRM	50.	2.	mg/kg	Frozen	1	0.4 g	-20	2006	Shipment+3 mo	frozen	ID-LC/MSMS
NIM	Creatinine-1	CRM	8.1	0.1	mg/kg	Frozen	1	0.011 mL	-70	2009	2013	thawing	ID-LC/MS
NIM	Creatinine-2	CRM	34.1	0.4	mg/kg	Frozen	1	0.011 mL	-70	2009	2013	thawing	ID-LC/MS
NIST	SRM 909b I	CRM	7.08	0.03	mg/kg/g	Lyoph	10		4	1996	Shipment+5 y	ambient	ID-GC/MS
NIST	SRM 909b II	CRM	33.93	0.16	mg/kg/g	Lyoph	10		4	1996	Shipment+5 y	ambient	ID-GC/MS
NIST	SRM 967a I	CRM	8.28	0.18	mg/kg	Frozen	1		<-60	2009	31-Dec-14	frozen	ID-LC/MS
NIST	SRM 967a II	CRM	37.90	0.80	mg/kg	Frozen	1		<-60	2009	31-Dec-14	frozen	ID-LC/MS
PTB	RELA 1/05 KS A	PT	44.89	0.92	mg/kg	Lyoph	5		4	2005		ambient	ID-GC/MS
PTB	RELA 1/05 KS B	PT	57.11	1.16	mg/kg	Lyoph	5		4	2005		ambient	ID-GC/MS

a **Matrix** is the form of the material, either liquid frozen or lyophilized; **mL** is the volume of material per unit (for lyophilized materials, the volume used for reconstitution), **MinSam** is the specified minimum amount of material per analysis, **°C** is the specified storage temperature; the **Year** the material was originally certified, and the **Expiration Date** of the certification.

b The **Condition** of liquid materials shipped frozen on “dry ice” upon arrival at NIST. The NIST materials were not shipped but taken from local storage.

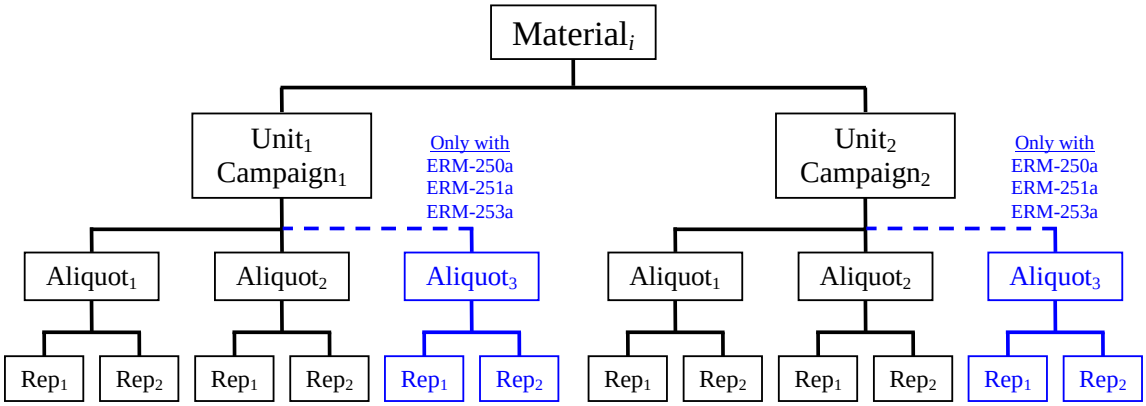
c The **Certification Method** used by the certifying institution to value assign the material: GC = gas chromatography, ID = isotope dilution, LC = liquid chromatography, MS = mass spectrometry, and MSMS= tandem mass spectrometry.

2. STEP 2: MEASURMENTS

2.1 Measurement Design

Participants provided the measurement laboratory with three units of each of their submitted materials, two to be analyzed and a spare in case of technical failure or to facilitate investigation of disputed results. Given the number of materials and the time required for each analysis, the measurements were made in two measurement campaigns ("runs"). In both campaigns, replicate analyses (two injections of one preparation separated in time) were made on each of two or three (see Section 2.1.1) independently prepared aliquots from one randomly selected unit of each of the 17 materials. This three-level nested design is summarized in Figure 1.

Figure 1: Repeatability Measurement Design



Measurements on the CCQM-K80 materials were performed following a randomized block design with blocking on aliquot and replicate. Control solution measurements were interspersed at regular intervals. All measurements within each campaign were made under repeatability conditions. The campaigns were separated in time by 5 days. No intentional changes were made to the equipment, reagents, instrumentation, or control solution between campaigns. The same analyst prepared all of the samples and made all of the measurements.

The above design confounds between-unit and between-campaign sources of measurement imprecision. The measurements made for this study thus cannot be used to estimate between-unit inhomogeneity for any of the study materials.

2.1.1 Modified measurement design for ERM-250a, -251a, and -253a

The specified minimum sample mass for ERM-250a, -251a, and -253a required use of a different sample preparation protocol than the routine protocol used for the other CCQM-K80 materials. For these three materials, Aliquot$_1$ and Aliquot$_2$ were prepared using the certificate-specified sample mass while Aliquot$_3$ was prepared following the routine protocol.

2.2 Analytical Method

All materials were analyzed under repeatability conditions using a definitive method recognized by the Joint Committee for Traceability in Laboratory Medicine (JCTLM). The method uses isotope dilution reversed-phase liquid chromatography with electrospray ionization mass spectrometric detection (ID-LC/MS) [3]. Quantitation was based on the relative peak areas for the creatinine target *m/z* 114 and the d_3-creatinine *m/z* 117. Table 4 lists all of the measurements for the CCQM-K80 materials. Experimental details are provided in Appendix C.

Table 4: Creatinine Measurements

All values are formally expressed in arbitrary units [a]

		Unit$_1$ (Campaign$_1$)						Unit$_2$ (Campaign$_2$)					
		Aliquot$_1$		Aliquot$_2$		Aliquot$_3$ [b]		Aliquot$_1$		Aliquot$_2$		Aliquot$_3$ [b]	
NMI	Material	Rep$_1$	Rep$_2$	Rep$_1$	Rep$_2$	Rep$_1$	Rep$_2$	Rep$_1$	Rep$_2$	Rep$_1$	Rep$_2$	Rep$_1$	Rep$_2$
CENAM	DMR 263a	7.105	6.983	7.067	7.042			7.132	7.096	7.172	7.422		
KRISS	111-01-01A	5.964	5.848	5.858	5.828			5.789	5.928	5.940	5.830		
KRISS	111-01-02A	27.370	27.436	27.170	27.467			27.508	27.429	27.431	27.651		
KRISS	111-01-03A	7.136	7.048	7.078	7.103			7.020	7.024	7.010	7.013		
KRISS	111-01-04A	24.825	25.083	25.220	25.278			25.163	25.387	25.406	25.171		
LGC	ERM-DA250a	39.911	39.023	40.485	39.723	40.168	39.628	39.400	39.251	40.362	39.826	40.759	41.267
LGC	ERM-DA251a	22.074	21.933	22.580	22.091	21.989	22.293	21.646	21.811	22.009	21.847	21.600	22.186
LGC	ERM-DA252a	2.986	3.098	3.146	3.128			3.121	3.125	3.083	3.124		
LGC	ERM-DA253a	50.868	50.225	49.429	49.346	50.042	49.163	51.468	50.045	50.621	50.521	50.618	50.478
NIM	Creatinine-1	7.959	8.098	8.143	7.996			8.001	8.152	8.104	8.157		
NIM	Creatinine-2	34.045	34.562	34.031	33.249			34.421	34.320	33.852	34.512		
NIST	SRM 909b I [c]	7.075	7.156	7.139	7.184			7.065	7.166	7.236	7.152		
NIST	SRM 909b II [c]	34.158	34.256	34.065	34.283			34.335	33.759	33.712	33.927		
NIST	SRM 967a I	8.347	8.116	8.270	8.161			8.265	8.218	8.293	8.294		
NIST	SRM 967a II	37.451	38.416	37.720	38.268			38.054	37.386	37.934	38.484		
PTB	RELA 1/05 KS A	45.180	46.153	45.425	44.829			45.792	45.188	45.009	45.174		
PTB	RELA 1/05 KS B	58.431	56.935	57.079	57.317			58.195	58.066	57.437	57.466		

a The results are listed with a uniform three digits to the right of the decimal for archival purposes.

b Results for Aliquot$_3$ were obtained using the "routine" sample preparation protocol used for all other materials; the Aliquot$_1$ and Aliquot$_2$ results for these three materials are for aliquots prepared using the certificate-specified minimum sample volume.

c Results adjusted for the measured fill masses of each unit.

2.2.1 Measurement quality assurance

In addition to the measurements made on the CCQM-K80 materials, a control solution was analyzed at regularly spaced intervals within each campaign. The relative standard measurement repeatability for these measurements is 1.2 %. Comparison of this relative variability with the range of measurement values observed for the very low creatinine level of ERM-DA252a and the lyophilized 111-010-03A, 111-010-03A, SRM 909b I, SRM 909b II, RELA 1/05 KS A, and RELA 1/05 KS B materials suggests that the modified preparation procedures required for these materials did not significantly affect measurement precision.

Since a significantly modified preparation procedure was required to meet the minimum sample mass requirements for ERM-250a, -251a, and -253a, a third aliquot prepared using the routine procedure provided both bias and precision control. Analysis of these data suggests that the modified preparation procedure used for these materials did not add bias or affect precision. Details of the measurement quality assurance analyses are provided in Appendix D.

2.3 Measurement Value, Variance Components, and Standard Uncertainty

The three-level nested measurement design for the CCQM-K80 materials addresses instrumental, sample preparation, and between-campaign sources of measurement variability by making two measurements on at least two independent aliquots of two different units of each material. The least complex analysis for these data is discussed at length in [2]. The repeatability measurement for each material, R_i, is the mean of the individual measurements

$$R_i = \sum_{1}^{N_c} \sum_{1}^{N_a} \sum_{1}^{N_r} R \Big/ \left(N_c \times N_a \times N_r \right)$$

where N_c is the number of measurement campaigns and is here always 2, N_a is the number of aliquots taken from each campaign and is here either 2 or 3, and N_r is the number of replicates of each aliquot and is here always 2.

The standard uncertainty of this mean is

$$u\left(R_i\right) = \sqrt{\frac{N_a \times N_r \times \sigma_{c,i}^2 + N_r \times \sigma_{a,i}^2 + \sigma_{r,i}^2}{N_c \times N_a \times N_r}}$$

where $\sigma_{c,i}^2$ is the between-campaign (confounded with the between-unit) variance, $\sigma_{a,i}^2$ is between-aliquot variance, and $\sigma_{r,i}^2$ is the between-replicate variance. These variances must be estimated from the data, most practically calculated with linear mixed model statistical analysis systems such as those discussed in [2, Appendix B]. Table 5 lists the estimated values.

2.3.1 Relative uncertainties

When expressed as estimated standard deviations, the non-zero estimates for $\hat{\sigma}_{c,i}$, $\hat{\sigma}_{a,i}$, and $\hat{\sigma}_{r,i}$ are approximately proportional to R_i. The pooled estimate for $\hat{\sigma}_{r,i} \big/ R_i$ is 0.0099 or 0.99 %. This is quite compatible with the 1.2 % estimated from the measurements on the control solution. The relationships between the standard deviations and the measurement means are documented in Appendix E.

2.4 95 % Coverage Intervals

Estimating defensible 95 % level-of-confidence coverage intervals for R_i from a small number of measurements can be much more complicated than multiplying $u(R_i)$ by a factor of 2 [2]. We use two approaches: a long-term frequency (frequentist) approach as recommended in the JCGM 100:2008 (GUM) [4] and a Bayesian approach that yields a probability interval interpretable as an uncertainty interval as defined in the JCGM 101:2008 (GUM-S1) [5].

2.4.1 Frequentist

Given $u(R_i)$ and its associated number of degrees of freedom, v_i, the usual frequentist approach to estimating a 95 % confidence coverage interval is to expand $u(R_i)$ using an appropriate two-tailed Student's *t* factor:

$$U_{95}\left(R_i\right)_{\mathrm{f}} = t_{\mathrm{s}}\left(1-0.95, v_i\right)\times u\left(R_i\right) \quad .$$

Depending on the relative magnitudes of the $\hat{\sigma}_{\mathrm{c},i}$, $\hat{\sigma}_{\mathrm{a},i}$, and $\hat{\sigma}_{\mathrm{r},i}$, v_i here ranges from 1 to 7. The corresponding Student's *t* factors range from 12.7 to 2.36. Table 5 lists the resulting $U_{95}(R_i)_{\mathrm{f}}$.

2.4.2 Bayesian

The frequentist approach may not be appropriate when there are constraints on the magnitudes of the measurement variances. It is unlikely that the σ_{c} of higher-order materials could be larger than the half-width of their certified uncertainty. However biased the V_i may be, their certified expanded uncertainties at least nominally include stability and heterogeneity uncertainty components in addition to measurement imprecision. This constrains each $\hat{\sigma}_{\mathrm{c},i}$ to be within the interval $0 < \hat{\sigma}_{\mathrm{c},i} \le U_{95}\left(V_i\right)/2$. Less importantly, the product of R_i and the pooled $\hat{\sigma}_{\mathrm{r},i}/R_i$ is a more robust estimate of the between-replicate variance component for each material than are the direct estimates. These types of prior information are difficult to include in frequentist analyses.

Prior information is easily included in Bayesian analyses. Under the Bayesian paradigm, parameters such as the measurand value and variance components have probability distributions that quantify our knowledge about them. The estimation process starts with quantification of prior knowledge about the parameters followed by specification of the statistical model that relates the parameters to the data. The priors are combined with the model and the data via Bayes Theorem to obtain posterior distributions for the parameters. These posterior distributions update our knowledge about the parameters based on the evidence provided by the data. For the CCQM-K80 data this analysis produces a probability distribution for each measurement mean that encompasses all of the information and variability present in the data but is confined by bounds based on prior knowledge. The 95 % coverage interval is then estimated from the percentiles of this distribution

$$U_{95}\left(R_i\right)_{\mathrm{B}} = \left(Ptile_{97.5} - Ptile_{2.5}\right)/2 \qquad [1]$$

where the "*Ptile*" are the 97.5 % and 2.5 % empirical percentile estimates from a suitable Markov Chain Monte Carlo analysis system such as discussed in [2, Appendix C].

Table 5 lists the $U_{95}(R_i)_{\mathrm{B}}$. The $U_{95}(R_i)_{\mathrm{B}}$ are smaller than the $U_{95}(R_i)_{\mathrm{f}}$ for materials with $u(R_i)$ associated with only 1 degree of freedom and somewhat larger than the $U_{95}(R_i)_{\mathrm{f}}$ for those associated with more than 1 degree of freedom.

Table 5: Variance Components, Means, Standard Uncertainties, and 95 % Coverage Intervals

		Numbers [a]				Estimates, a.u. [b]			Means and Uncertainties, a.u. [c]				
NMI	Material	N_t	N_r	N_a	N_c	$\hat{\sigma}_{r,i}$	$\hat{\sigma}_{a,i}$	$\hat{\sigma}_{c,i}$	R_i	$u(R_i)$	v_i	$U_{95}(R_i)_f$	$U_{95}(R_i)_B$
CENAM	DMR 263a	8	2	2	2	0.100	0.059	0.089	7.127	0.078	1	0.991	0.280
KRISS	111-01-01A	8	2	2	2	0.063	0	0	5.873	0.022	7	0.053	0.095
KRISS	111-01-02A	8	2	2	2	0.120	0	0.082	27.433	0.072	1	0.913	0.365
KRISS	111-01-03A	8	2	2	2	0.027	0	0.051	7.054	0.037	1	0.474	0.111
KRISS	111-01-04A	8	2	2	2	0.148	0.104	0.073	25.192	0.090	1	1.144	0.385
LGC	ERM-DA250a [d]	8	2	2	2	0.458	0.251	0	39.748	0.205	3	0.652	1.360
LGC	ERM-DA250a	12	2	3	2	0.431	0.516	0	39.983	0.245	5	0.629	1.130
LGC	ERM-DA251a [d]	8	2	2	2	0.198	0.134	0.199	21.999	0.171	1	2.171	0.990
LGC	ERM-DA251a	12	2	3	2	0.230	0	0.198	22.005	0.155	1	1.969	0.805
LGC	ERM-DA252a	8	2	2	2	0.043	0.029	0	3.101	0.021	3	0.067	0.137
LGC	ERM-DA253a [d]	8	2	2	2	0.554	0.437	0.266	50.315	0.348	1	4.429	1.635
LGC	ERM-DA253a	12	2	3	2	0.520	0.248	0.488	50.235	0.390	1	4.951	1.045
NIM	Creatinine-1	8	2	2	2	0.079	0	0	8.076	0.028	7	0.066	0.120
NIM	Creatinine-2	8	2	2	2	0.407	0.166	0	34.124	0.166	3	0.529	0.900
NIST	SRM 909b I	8	2	2	2	0.056	0	0	7.147	0.020	7	0.047	0.098
NIST	SRM 909b II	8	2	2	2	0.212	0	0.148	34.062	0.129	1	1.636	0.505
NIST	SRM 967a I	8	2	2	2	0.076	0	0	8.245	0.027	7	0.064	0.132
NIST	SRM 967a II	8	2	2	2	0.420	0	0	37.964	0.148	7	0.351	0.675
PTB	RELA 1/05 KS A	8	2	2	2	0.434	0	0	45.344	0.154	7	0.363	0.895
PTB	RELA 1/05 KS B	8	2	2	2	0.538	0.112	0	57.616	0.198	3	0.631	1.025

a N_r is the number of replicates per aliquot, N_a is the number of aliquots per campaign, and N_c is the number of campaigns. N_t is the total number of measurements, which is equal to $N_r \times N_a \times N_c$.

b $\hat{\sigma}_{r,i}$, $\hat{\sigma}_{a,i}$, and $\hat{\sigma}_{c,i}$ are the estimated between-replicate, between-aliquot, and between-campaign components of variance as standard deviations formally expressed in arbitrary units (a.u.).

c R is the mean of the measurements, $u(R_i)$ the standard uncertainty for R_i, v_i the number of degrees of freedom associated with $u(R_i)$, $U_{95}(R_i)_f$, a frequentist 95 % confidence estimate for R_i, and $U_{95}(R_i)_B$ a Bayesian 95 % confidence estimate for R_i. All values are in formally expressed in arbitrary units (a.u.).

d Estimated using only the Aliquot_1 and Aliquot_2 measurements.

2.5 Estimates for ERM-DA250a, -DA251a and -DA253a

Given the approximate equivalence of the estimated uncertainties for these materials with and without the Aliquot_3 data, all further analyses discussed in this Report are based on the eight Aliquot_1 and Aliquot_2 measurements for each material.

2.6 "Large Sample" Standard Uncertainties

Comparisons involving standard uncertainty estimates associated with different and sometimes small numbers of v_i can be misleading if the v_i are not taken into account. Since 95 % confidence intervals nominally provide consistent coverage of the true value, they can be used to estimate "large sample" standard uncertainties that formally are all associated with the same large number of degrees of freedom, $v_i \approx \infty$, and are suitable for use in further analysis [6]

$$u_\infty(.) = U_{95}(.)/2 \qquad [2]$$

where the "." represents a specified quantity such as V_i or R_i.

We symbolize the frequentist $u_\infty(R_i)$ as $u_\infty(R_i)_\text{f}$ and the Bayesian as $u_\infty(R_i)_\text{B}$. Estimated in this manner, these large-sample standard uncertainties define the dispersion of Gaussian kernels $N(V_i, u_\infty(V_i))$, $N(R_i, u_\infty(R_i)_\text{f})$, and $N(R_i, u_\infty(R_i)_\text{B})$ that have approximately 95 % of their density within the interval defined by the value and its 95 % confidence expanded uncertainty. They are listed in Table 6, with the materials sorted in order of increasing V_i. Each material is assigned a one-character identifying code to simplify graphical presentation.

Table 6: Data as Used in the Analysis

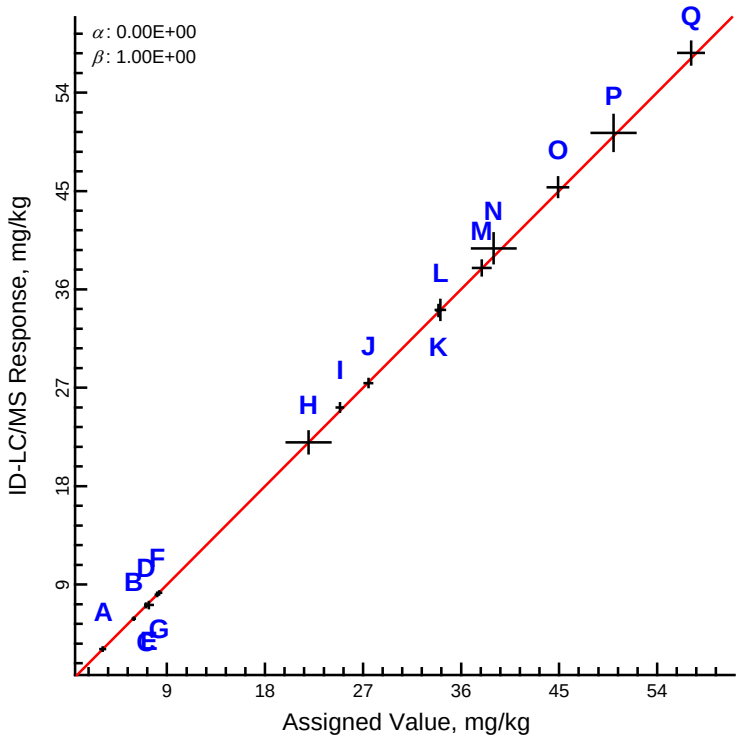
		mg/kg		Arbitrary Units		
Code	Material	V_i	$u_\infty(V_i)$	R_i	$u_\infty(R_i)_\text{f}$	$u_\infty(R_i)_\text{B}$
A	ERM-DA252a	3.10	0.100	3.101	0.033	0.068
B	111-01-01A	5.96	0.045	5.873	0.026	0.047
C	111-01-03A	7.08	0.040	7.054	0.237	0.055
D	SRM 909b I	7.08	0.015	7.147	0.023	0.049
E	DMR 263a	7.35	0.175	7.127	0.495	0.140
F	Creatinine-1	8.10	0.050	8.076	0.033	0.060
G	SRM 967a I	8.28	0.090	8.245	0.032	0.066
H	ERM-DA251a	22.00	1.000	22.005	1.085	0.495
I	111-01-04A	24.87	0.145	25.192	0.572	0.193
J	111-01-02A	27.49	0.165	27.433	0.456	0.183
K	SRM 909b II	33.93	0.080	34.062	0.818	0.253
L	Creatinine-2	34.10	0.200	34.124	0.264	0.450
M	SRM 967a II	37.90	0.400	37.964	0.176	0.337
N	ERM-DA250a	39.00	1.000	39.983	0.326	0.680
O	RELA 1/05 KS A	44.89	0.460	45.344	0.181	0.448
P	ERM-DA253a	50.00	1.000	50.235	2.214	0.818
Q	RELA 1/05 KS B	57.11	0.580	57.616	0.316	0.513

3. STEP 3: DEFINE A CONSENSUS MODEL

3.1 Key Comparison Reference Function, KCRF

Since a definitive method was used for the CCQM-K80 measurements, a linear relationship is expected between the certified and measured values. Figure 2 provides an overview of the relationship between the certified and measurement values that confirms this expectation.

Figure 2: Overview of the Relationship Between Certified and Measured Values



Each uncertainty cross denotes the $V \pm U_{95}(V)$, $R \pm U_{95}(R_i)_B$ for one of the 17 materials. The crosses are labeled in order of increasing V with the letters A to Q; see Table 6 for the association between the code and the material.

The red line represents exact equality between the assigned and measured values: $R = V$.

A linear relationship can be modeled as:

$$R = \alpha + \beta \times V + E \qquad [3]$$

where here α is the intercept, β is the slope, and E is the residual random error. In analogy to the "Key Comparison Reference Value (KCRV)" used with single-material comparisons, we term this function the "Key Comparison Reference Function (KCRF)" for CCQM-K80.

The issue is then how to appropriately parameterize Equation 3. While the CCQM-K80 measurements were made under repeatability conditions, no effort beyond that inherent to the method was made to characterize or correct for potential bias.

3.1.1 Generalized Distance Regression (GDR)

Ordinary least squares regression is not an appropriate approach to estimating the parameters of Equation 3 since both the certified values and the measurement results have known and non-negligible uncertainty [2]. However, generalized distance regression (GDR) provides appropriate parameters by iteratively minimizing the total uncertainty-scaled residual distances:

$$E = \sum_{i=1}^{N_m} \varepsilon_i^2; \quad \varepsilon_i^2 = \left(\frac{V_i - \hat{V}_i}{u_\infty(V_i)} \right)^2 + \left(\frac{R_i - \hat{R}_i}{u_\infty(R_i)} \right)^2; \quad \hat{R}_i = \hat{\alpha} + \hat{\beta}\hat{V}_i$$

where *i* indexes the materials, N_m is the number of materials, and $\hat{V}_i$, $\hat{R}_i$, $\hat{\alpha}$, and $\hat{\beta}$ are *predicted* estimates for the parameters. Note that the residual uncertainty-weighted distance for a given material, ε_i, is symmetric in V_i and R_i.

3.1.2 Parametric Bootstrap Monte Carlo (PBMC) uncertainty evaluation

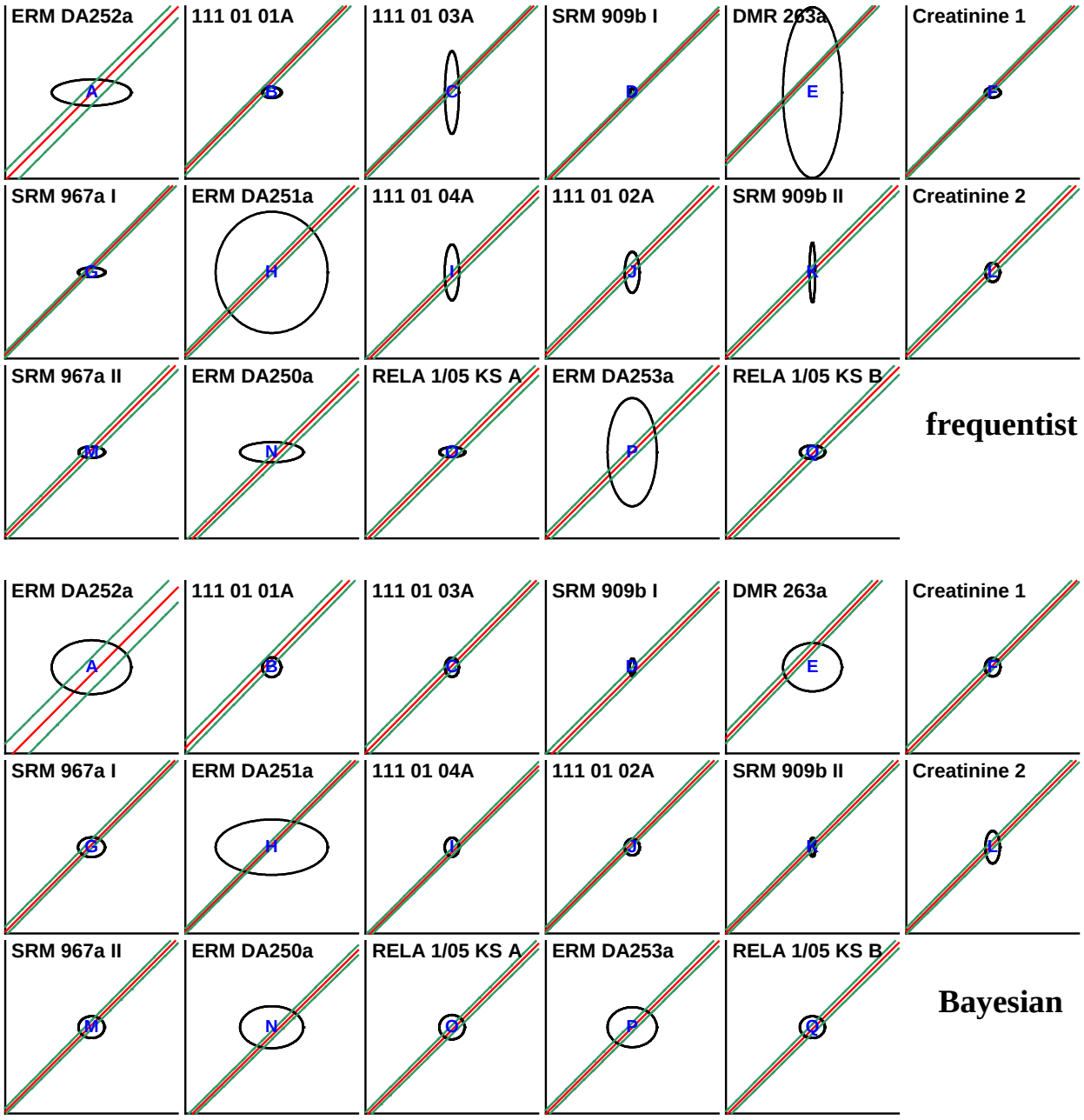
The variability for all quantities estimated with GDR is readily estimated using the parametric bootstrap Monte Carlo (PBMC) technique [2]. With PBMC, the entire set of V_i and R_i values used in the GDR analysis are repeatedly replaced with corresponding "pseudo-values" randomly drawn from each of the $N(V_i,u_\infty(V_i))$ and $N(R_i,u_\infty(R_i))$ normal kernels. The parameters and associated quantities are stored and, once a suitably large number have been generated, approximate 95 % expanded uncertainty intervals are estimated from the percentiles of the empirical distributions. Since only the central 95 % of the distributions are of interest, relatively few pseudo-sets are required for stable estimates.

3.1.3 Graphical results

Figure 3 displays GDR results using the frequentist $u_\infty(R_i)_f$ and Bayesian $u_\infty(R_i)_B$. The graphical resolution required for simultaneously displaying all materials in single scatterplot of Figure 2 is insufficient for adequately visualizing the analysis. The Figure 3 summaries therefore display each material in a series of individual "thumbnail" scatterplots. Each thumbnail displays the GDR analysis for one material and is centered at $\{V_i, R_i\}$. All of the thumbnails have the same relative scale. The red lines represent the candidate KCRF. The green lines are approximate 95 % level of confidence intervals on the candidate KCRF, U_{95}(KCRF). The ellipse bounds all points that are two large sample standard uncertainty units distant from $\{V_i, R_i\}$. The top set of thumbnails display results using $u_\infty(R_i)_f$, the bottom set display results using $u_\infty(R_i)_B$.

The square of the GDR uncertainty-weighted residuals, ε_i^2, are expected to be distributed as χ^2 with two degrees of freedom. Therefore, ε_i less than the value for the 95th percentile expected from this distribution, $\sqrt{5.99} \approx 2.45$, indicate that the uncertainties adequately cover the difference between the estimated and observed values at the 95 % level of confidence. Thus ellipses that overlap the KCRF line indicate the observed values are consistent with the KCRF and ellipses that do not enter the KCRF $\pm U_{95}$(KCRF) interval are not consistent. By this graphical test, all 17 of the materials are at least marginally consistent with the KCRF as defined from either $u_\infty(R_i)_f$ or $u_\infty(R_i)_B$.

Figure 3: High-Resolution Display of GDR Results



3.2 Identifying Strongly Influential Materials

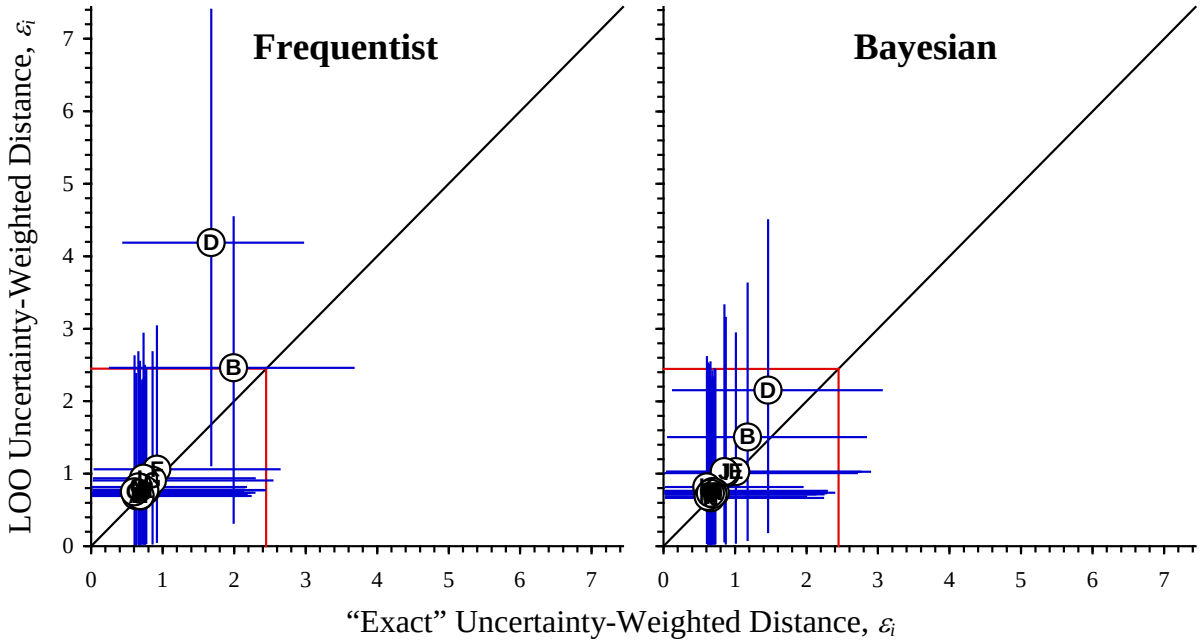
The GDR solution can be strongly influenced by materials having small $u_\infty(V_i)$ and/or $u_\infty(R_i)$ [2], e.g., 111 01 01A ("B"), SRM 909b I ("D"), and SRM 909b II ("K"). The magnitude of this influence depends not only on the magnitudes of the uncertainties but also on where the $\{V_i, R_i\}$ pair is located relative to the other materials.

Leave-one-out (LOO) validation is an efficient, if empirical, approach to establishing which, if any, materials are sufficiently influential to distort the consensus estimation of the KCRF [7]. A

LOO analysis proceeds by excluding each material in turn from its own evaluation. For the CCQM-K80 materials, this involves 18 GDR analyses: one solution with all 17 materials included in the analysis and 17 solutions each with one material excluded.

Figure 4 compares the "exact" ε_i, calculated using all materials with their LOO-estimated analogues. The panel to the left presents results using the frequentist $u_\infty(R_i)_f$, the panel to the right presents results using the Bayesian $u_\infty(R_i)_B$. The circles represent estimates for individual materials; the crosses represent the PBMC-estimated 95 % level of confidence intervals on the estimates. Results inside of the red lines indicate materials that are consistent with the consensus GDR solution whether or not they are included in the GDR model. Results far from the diagonal line indicate materials that strongly influence the consensus solution. SRM 909b I ("D") appears to be inconsistent and strongly influential when the $u_\infty(R_i)_f$ are used.

Figure 4: Identification of Strongly Influential Materials



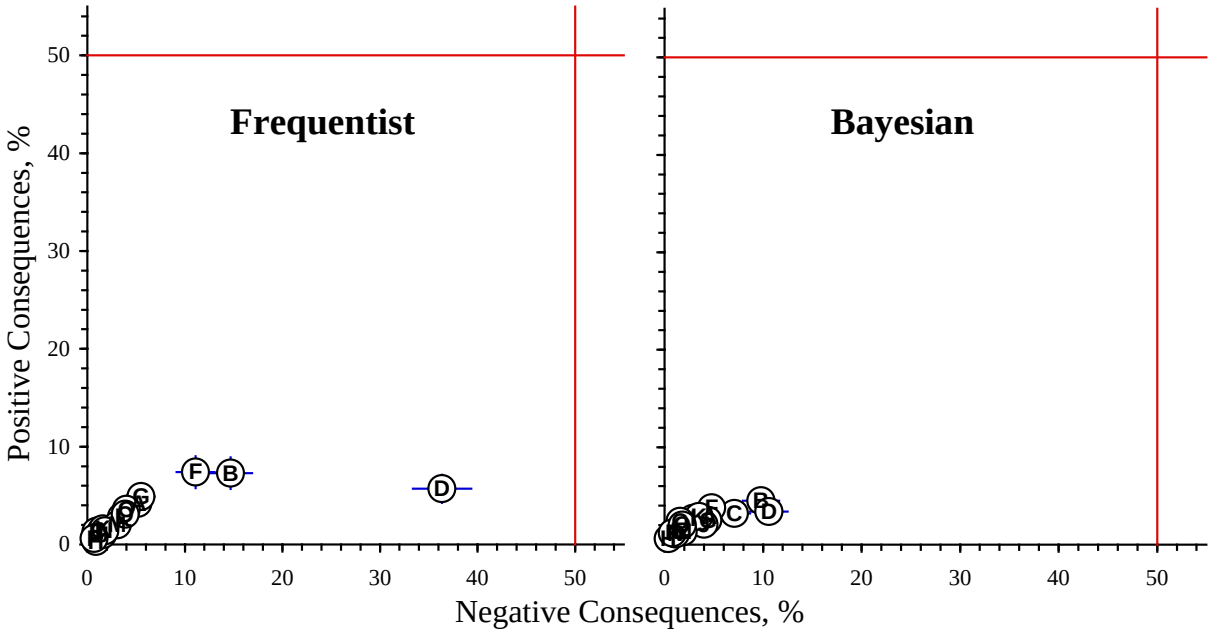
3.3 Identifying Strongly Consequential Materials

While the presence of a material in the GDR model may strongly influence the consensus solution, that influence may or may not have consequence for the other materials in the study. Figure 5 visualizes the consequences of including each CCQM-K80 material in the GDR model. Inclusion of a material has a negative consequence when its presence causes the ε_i of one or more other materials in a given PBMC iteration to change from being less than the critical distance of 2.45 to being greater than that value. Likewise, inclusion of the material has a positive consequence when its presence causes the ε_i of one or more other materials to change from being greater than 2.45 to being less than that value.

Figure 5 displays the negative and positive consequences for the frequentist and Bayesian $u_\infty(R_i)$, estimated from 1000 PBMC iterations. The 95 % error bars for most of these estimates are

covered by the circles. The red lines enclose materials whose presence in the GDR model have a very strong negative or positive consequence for any other material.

Figure 5: Identification of Strongly Consequential Materials



Although SRM 909b I ("D") is not strongly consequential, with the frequentist estimates it is considerably more consequential than the other materials. After careful consideration NIST chose to have this material excluded from the estimation of the KCRF parameters on the grounds that 1) the 1996 definition of this lyophilized material's fill-mass corrected $U_{95}(V_i)$ is suspiciously small, 2) the process used to establish the $U_{95}(V_i)$ does not reflect current practice, and 3) SRM 909b is completely sold out and no longer available for re-evaluation or sale.

3.4 KCRF Parameters and Predictions

In addition to identifying materials that could distort the consensus GDR solution, LOO-PBMC enables a more robust estimate of the variability of the GDR parameters. The LOO estimates are influenced by biases (systematic differences in the GDR solutions with-and-without each material in the models) that are not present when all materials are included ("Leave-All-In" or "LAI") in the model. Thus the LOO-PBMC parameter uncertainties are constrained to be somewhat larger than those determined with LAI-PBMC analysis.

Table 7 lists the GDR consensus solution parameters for the model with all 17 materials estimated for both the $u_\infty(R_i)_\mathrm{f}$ and $u_\infty(R_i)_\mathrm{B}$ definitions using data for the all materials except SRM 909b I. The slightly larger LOO-based asymptotic standard uncertainty estimates on the GDR parameters provide more conservative coverage than do the LAI. Table 8 lists estimates for the assigned values and repeatability measurements along with their LOO-estimated asymptotic standard uncertainties.

Table 7: GDR Model Parameter Estimates

$u_\infty(R_i)$	$\hat{\alpha}$ [a]	$u(\hat{\alpha})$ [a] LAI [d]	$u(\hat{\alpha})$ [a] LOO [e]	$\hat{\beta}$ [b]	$u(\hat{\beta})$ [b] LAI [d]	$u(\hat{\beta})$ [b] LOO [d]	$\rho(\hat{\alpha},\hat{\beta})$ [c] LAI [d]	$\rho(\hat{\alpha},\hat{\beta})$ [c] LOO [e]
frequentist	-0.102	0.056	0.059	1.0077	0.0059	0.0061	-0.80	-0.80
Bayesian	-0.100	0.055	0.058	1.0079	0.0051	0.0053	-0.77	-0.77

a Intercept and its uncertainty estimates are expressed in arbitrary units
b Slope and its uncertainty estimates are expressed in arbitrary units per mg/kg
c Correlation between the intercept and slope estimates
d Standard deviation of 1000 Leave-All-In PBMC parameter estimates where all eligible materials were included in model
e Standard deviation of 17 sets of 1000 Leave-One-Out PBMC parameter estimates, where one eligible material is in turn left out of the model in each set

Table 8: GDR Predicted Values and Responses

		frequentist, $u_\infty(R_i)_f$				Bayesian, $u_\infty(R_i)_B$			
		mg/kg		Arbitrary Units		mg/kg		Arbitrary Units	
Code	Material	$\hat{V}_i$	$u(\hat{V}_i)$	$\hat{R}_i$	$u(\hat{R}_i)$	$\hat{V}_i$	$u(\hat{V}_i)$	$\hat{R}_i$	$u(\hat{R}_i)$
A	ERM-DA252a	3.173	0.052	3.097	0.031	3.154	0.062	3.079	0.057
B	111-01-01A	5.935	0.036	5.880	0.025	5.944	0.037	5.890	0.038
C	111-01-03A	7.080	0.039	7.034	0.052	7.089	0.035	7.044	0.040
D	SRM 909b I	7.114	0.016	7.068	0.028	7.090	0.015	7.045	0.037
E	DMR 263a	7.334	0.159	7.289	0.164	7.237	0.108	7.193	0.110
F	Creatinine-1	8.112	0.037	8.073	0.030	8.105	0.040	8.067	0.044
G	SRM 967a I	8.283	0.043	8.246	0.030	8.276	0.059	8.240	0.054
H	ERM-DA251a	21.952	0.734	22.019	0.743	21.929	0.444	21.998	0.442
I	111-01-04A	24.887	0.139	24.976	0.173	24.952	0.118	25.044	0.127
J	111-01-02A	27.470	0.155	27.578	0.189	27.418	0.127	27.529	0.134
K	SRM 909b II	33.931	0.080	34.089	0.183	33.929	0.076	34.089	0.145
L	Creatinine-2	34.055	0.170	34.214	0.193	34.071	0.187	34.232	0.217
M	SRM 967a II	37.792	0.222	37.978	0.165	37.825	0.279	38.015	0.264
N	ERM-DA250a	39.473	0.351	39.673	0.305	39.375	0.593	39.577	0.587
O	RELA 1/05 KS A	45.074	0.265	45.316	0.171	45.000	0.338	45.245	0.335
P	ERM-DA253a	50.039	0.921	50.320	0.954	50.016	0.682	50.299	0.682
Q	RELA 1/05 KS B	57.243	0.368	57.578	0.278	57.213	0.400	57.551	0.401

Given that the two sets of GDR model parameters differ by much less than their standard uncertainties, combining the values provides a robust definition for the consensus KCRF:

$$\hat{R} = (-0.101 \pm 0.059) + (1.0078 \pm 0.0057)\hat{V}$$

4. STEP 4: DEGREES OF EQUIVALENCE

4.1 Degrees of Equivalence for Materials

An appropriate definition for the degrees of equivalence for materials in studies such as CCQM-K80 is the percent relative signed orthogonal distance [2]:

$$\%d_i = 100 \times \mathrm{SIGN}\left(V_i - \hat{V}\right) \times \frac{\sqrt{\left(V_i - \hat{V}\right)^2 + \left(\left(R_i - \hat{R}_i\right)/\hat{\beta}\right)^2}}{\left(V_i + \left(R_i - \hat{\alpha}\right)/\hat{\beta}\right)/2} \qquad [4]$$

where the measurement-related terms are transformed to have the same scale as the assigned values. The function SIGN returns the sign (±1) of its argument and defines whether the observed $\{V_i, R_i\}$ pair is "above" or "below" the KCRF. While not necessary for the nearly zero intercept and unit slope of the CCQM-K80 data, transforming the repeatability measurements through the KCRF ensures that they have the same origin and scale as the assigned values.

4.1.1 Definition of the uncertainty of the degree of equivalence for individual materials

Given the covariances among the $\hat{\alpha}$ and $\hat{\beta}$ parameters of the KCRF and the data used to define them, estimating appropriate uncertainties for the $\%d_i$ via the usual Taylor's series approximation for Equation 4 would be challenging even if the relative $u_\infty(V_i)$ and $u_\infty(R_i)$ were constant for all of the materials. However, empirical $U_{95}(\%d_i)$ are easily estimated by applying Equation 1 to $\%d_i$ values calculated for each set of PBMC pseudo-values. Large sample standard uncertainties, $u_\infty(\%d_i)$, can then estimated using Equation 2.

Further, LOO analysis enables making these empirical uncertainty estimates robust to each material's "self-referential" influence. The $U_{95}(\%d_i)$ for each material can be estimated from the distribution of the $\%d_i$ calculated when its own values are not used in the GDR solution. While requiring many more calculations, these LOO-PBMC estimates are free of correlation between each material's observed values and the KCRF.

4.1.2 Graphical representation of degrees of equivalence for materials

Figures 6 and 7 display the $\%d_i \pm U_{95}(\%d_i)$ in dot-and-bar format, with the horizontal axis used to display the V_i of each material. The red line denotes zero bias relative to the KCRF; the $\%d_i$ for materials with bars that cross this line are compatible with having a true value of zero with about a 95 % level of confidence. Figure 6 presents results obtained with the $u_\infty(R_i)_\mathrm{f}$, Figure 7 presents those for $u_\infty(R_i)_\mathrm{B}$.

4.2 Choice of Model for Degrees of Equivalence

The $\%d_i$ from the frequentist and Bayesian estimates of $u_\infty(R_i)$ are remarkably similar, reflecting the near-identity of the KCRF with the two approaches. However the $U_{95}(\%d_i)$ are considerably different for many of the materials associated with only one degree of freedom. The $k_{95} = 12.7$ frequentist coverage for those materials appears overly pessimistic. Given the similarity between the frequentist and Bayesian estimates for the $\%d_i$ and assuming that the true between-campaign variance estimates, $\hat{\sigma}_{\mathrm{c},i}$, are validly bounded by the uncertainties of the assigned values, the Bayesian approach provides the more realistic assessment of $\%d_i \pm U_{95}(\%d_i)$.

Table 9 lists the Bayesian degrees of equivalence for the materials and, for convenience, their assigned values, V_i, and percent relative expanded uncertainties, $100 \times U_{95}(V_i)/V_i$.

Figure 6: Frequentist Degrees of Equivalence for Materials

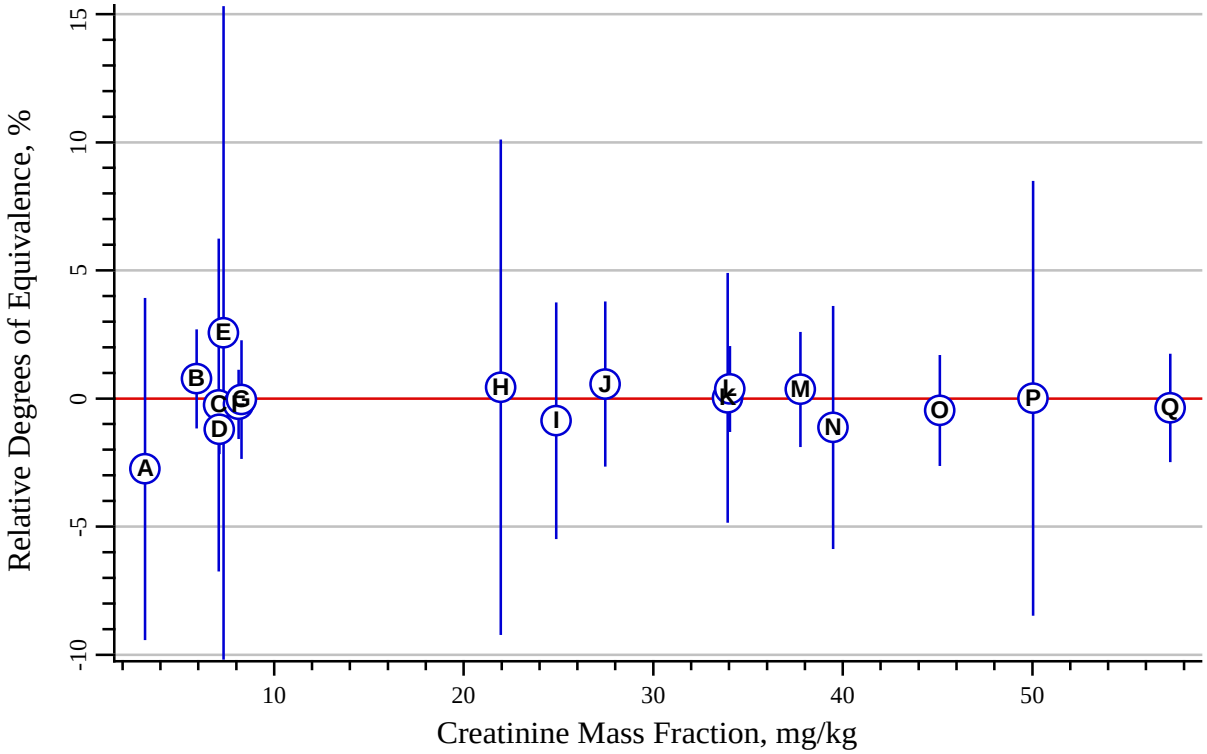


Figure 7: Bayesian Degrees of Equivalence for Materials

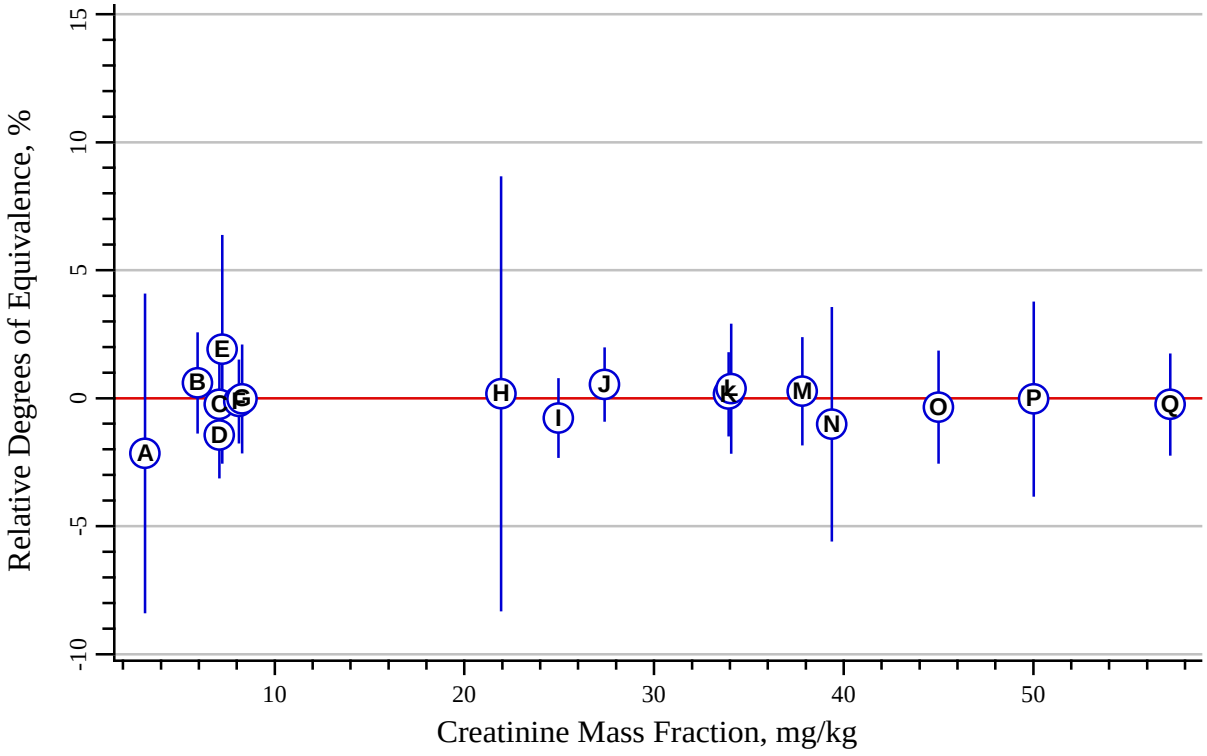


Table 9: Degrees of Equivalence

Participating NMIs				Materials						
	$\%D$, percent					$\%d_i$, percent			V_i	
NMI	Value	u_∞	U_{95}	Material	Code	Value	u_∞	U_{95}	mg/kg	$\%U_{95}$ [a]
CENAM	1.91	2.21	4.4	DMR 263a	E	1.91	2.21	4.4	7.35	4.8
KRISS	0.03	1.00	2.0	111 01 01A	B	0.60	0.96	1.9	5.96	1.5
				111 01 03A	C	-0.24	0.86	1.7	7.08	1.1
				111 01 04A	I	-0.78	0.75	1.5	24.87	1.2
				111 01 02A	J	0.54	0.70	1.4	27.49	1.2
LGC	-0.76	3.14	6.3	ERM DA252a	A	-2.16	3.10	6.2	3.1	6.5
				ERM DA251a	H	0.17	4.22	8.4	22	9.1
				ERM DA250a	N	-1.02	2.27	4.5	39	5.1
				ERM DA253a	P	-0.03	1.88	3.8	50	4.0
NIM	0.12	1.07	2.1	Creatinine 1	F	-0.12	0.79	1.6	8.1	1.2
				Creatinine 2	L	0.37	1.25	2.5	34.1	1.2
NIST	-0.26	1.16	2.3	SRM 909b I	D	-1.44	0.82	1.6	7.08	0.42
				SRM 967a I	G	-0.03	1.04	2.1	8.28	2.2
				SRM 909b II	K	0.16	0.80	1.6	33.93	0.47
				SRM 967a II	M	0.27	1.03	2.1	37.90	2.1
PTB	-0.30	1.03	2.1	RELA 1/05 KS A	O	-0.35	1.08	2.2	44.89	2.0
				RELA 1/05 KS B	Q	-0.24	0.97	1.9	57.11	2.0

a Percent relative expanded uncertainty, $100 \times U_{95}(V_i)/V_i$

4.2.1 Summary of degrees of equivalence for materials

For the 16 materials used to define the KCRF, the $U_{95}(\%d_i)$ estimated with LOO-PBMC range from 1.0 to 1.4 times larger than the simple PBMC results. The median enlargement factor is 1.07. Thus for most of the materials the LOO-PBMC estimates are only modestly more conservative.

Even without LOO enlargement, the interval $\%d_i \pm U_{95}(\%d_i)$ contains zero for all materials including SRM 909b I. This indicates that all of the materials evaluated in CCQM-K80 agree with the KCRF, and thus each other, within their assigned uncertainties with about a 95 % level of confidence. The median absolute value of the $\%d_i$, $|\%d_i|$ is 0.3; the median $U_{95}(\%d_i)$ is 2.0.

The assigned uncertainties, $U_{95}(V_i)$, are as large or larger than the $U_{95}(\%d_i)$ for several materials, suggesting that some of the $U_{95}(V_i)$ may be somewhat overestimated.

4.3 Degrees of Equivalence for Participating NMIs

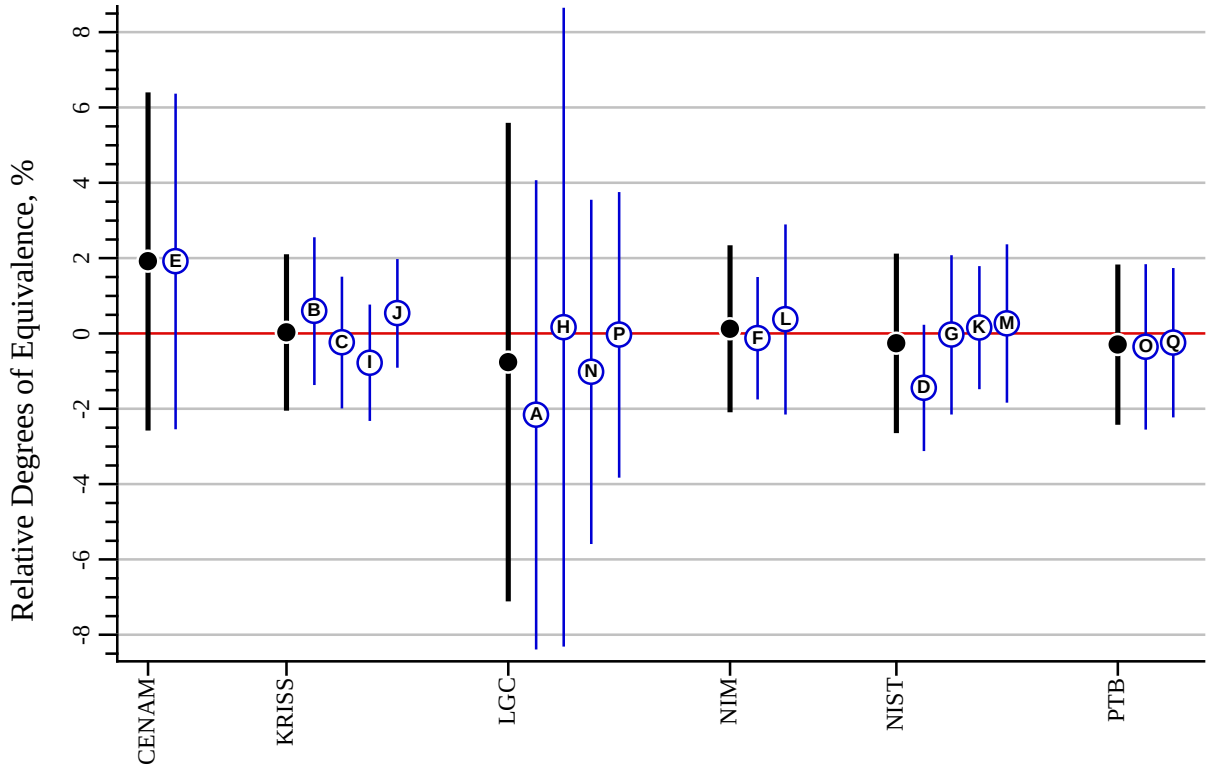
All of the CCQM-K80 participants except CENAM are represented by more than one material. The results for all of the materials from each participant contributing more than one material must therefore be combined in some way to provide the desired goal of the KC: the expected degrees of equivalence of the participating NMIs, $\%D$. There are two convenient ways to estimate the $\%D$ [2]: from the $\%d_i \pm u_\infty(\%d_i)$ values of all materials submitted by a given NMI or from the PBMC pseudo-values used to estimate the $U_{95}(\%d_i)$. For the CCQM-K80 data, the $\%D$ from two approaches differ by no more than 0.1 and the $u_\infty(\%D)$ differ by no more than 0.2.

Table 9 lists the degrees of equivalence for both participating NMIs and materials, with the %D estimated from the $\%d_i \pm u_\infty(\%d_i)$.

4.3.1 Graphical representation of degrees of equivalence for participating NMIs

Figure 8 displays the $\%D \pm U_{95}(\%D)$ and the $\%d_i \pm U_{95}(\%d_i)$ in dot-and-bar format, with the thick black bars and open dots representing the %D and thin blue bars and solid dots the $\%d_i$. The NMIs are arranged in alphabetical order.

Figure 8: Degrees of Equivalence for Participating NMIs



4.3.2 Summary of degrees of equivalence for participating NMIs

The median absolute value of the *%D*, |*%D*|, is the same 0.3 as for the constituent materials. The median $U_{95}(\%D)$ is 2.2, slightly larger than for the materials and reflecting the between-$\%d_i$ variability possible when an NMI submits more than one material. These values suggest a median within-NMI relative standard deviation of the $\%d_i$ $\sqrt{(2.2/2)^2 - (2.0/2)^2} = 0.5$. The "adjusted median absolute deviation from the median (MAD_e)" robust estimate of the between-NMI standard deviation of the *%D* is 0.3. This suggests that, within their assigned uncertainties, materials value-assigned by different NMIs are likely to be about as self-consistent as materials from the same NMI.

5. RECOMMENDATIONS FOR USE OF CCQM-K80 IN EVALUATING CIPM MRA CALIBRATION AND MEASUREMENT CAPABILITY (CMC) CLAIMS

5.1 How Far Does the Light Shine?

CCQM-K80 demonstrated that as a group the participating NMIs have the capability to successfully assign creatinine values from 3 mg/kg to more than 57 mg/kg with a relative bias of ±2 % or less in frozen or lyophilized human serum using ID/MS methods with GC or LC separation prior to analysis. The range covered by the CCQM-K80 study materials extends from low-normal to elevated levels of creatinine.

Creatinine is a relatively stable analyte in serum, but care is required to prevent inter-conversion between creatinine and creatine during separation and MS detection. The creatinine levels measured in this study are in the mid to high range compared to levels of other commonly measured serum analytes. The results for creatinine therefore suggest that the participating NMIs have the capability to value assign other well-defined, small, polar, non-protein analytes that are stable to somewhat sensitive to separation and MS detection conditions and that are present in a fairly homogeneous human body fluid matrix (serum, plasma, urine, etc.) at low mg/kg levels with relative uncertainties of a few percent at the 95 % level of confidence.

5.2 Demonstrated Competencies

Table 10 documents the “Core Competencies” that each of the participating NMIs assert that they demonstrated by participation in CCQM-K80. The scope and nature of these measurement process competencies include competencies embedded in CRM and PT material value assignment and the delivery of these materials to customers.

Table 10a: Core Competencies Demonstrated in CCQM-K80 by CENAM

Scope

Analyte	Small, polar, relatively stable non-protein organic compounds
Matrix	Frozen relatively homogenous complex body fluids such as human serum, plasma, and urine
Mass fraction	3 mg/kg to 10 mg/kg
Demonstrated U_{95} capability	4.4 % of certified values

Calibration Competencies

Competency	Specifications / Comments
Identity verification	A primary reference material (PRM) characterized at CENAM
Purity assessment	Karl Fischer titration, LC/MS, LC/Diode Array Detection (LC/DAD), and other techniques
Calibrant preparation	Gravimetric dilution of the solid PRM with water

Sample Analysis Competencies

Competency	Specifications / Comments
Measurand identification	Chromatographic retention time, mass profile
Extraction	Protein precipitation with water acetone
Cleanup	Filtration
Transformation	None
Method validation	Matrix-matched CRM control from NIST SRM 909b, comparison of results from independent methods
Use of analytical techniques	ID-LC/MS and ID-LC/DAD
Quantification mode(s), data analysis, and uncertainty evaluation	Interpolation in a calibration curve, calculations performed using purpose-built spreadsheet software; uncertainties combine instrumental imprecision, material heterogeneity and stability; gravimetric preparation of standard solutions

Material Delivery Competencies

Competency	Specifications / Comments
Homogeneity assessment	Appropriate to uncertainty claimed in Scope
Stability assessment	Appropriate to uncertainty claimed in Scope
Packaging	Appropriate for at least six years storage
Shipping	Appropriate for national and international customers

Table 10b: Core Competencies Demonstrated in CCQM-K80 by KRISS

Scope

Analyte	Small, polar, organic molecule of relatively high stability
Matrix	Frozen or lyophilized biological fluids of high homogeneity such as serum, urine, and cerebrospinal fluid
Mass fraction	6 mg/kg to 27 mg/kg
Demonstrated U_{95} capability	1.5 % of certified value

Calibration Competencies

Competency	Specifications / Comments
Identity verification	NIST SRM 914a
Purity assessment	For this particular analyte, creatinine, the purity claimed in the certificate of NIST SRM 914a was adopted. However, the KRISS team is now actively building capability to assess purities of pure organic compounds, which has been partially proved through CCQM-P117a and CCQM-K55b.
Calibrant preparation	Gravimetric dilution of NIST SRM 914a in a buffer solution
Other: Calibrant verification	Multiple calibration standards are cross-checked by ID-LC/MS for verification of their correct preparations.

Sample Analysis Competencies

Competency	Specifications / Comments
Measurand identification	Chromatographic retention time, mass profile using selected ion monitoring (SIM) and multiple reaction monitoring (MRM)
Extraction	Deproteinization using ultrafiltration after complete equilibration of the spiked isotopic analogue in the sample matrix
Cleanup	Ultrafiltration described above (molecular weight cut-off: 3 kDa)
Transformation	None
Method validation	Simultaneous measurement of creatinine in NIST SRM 909b
Use of analytical techniques	ID-LC/MS
Quantification mode(s), data analysis, and uncertainty evaluation	Use of the internally established uncertainty evaluation scheme that combines calibration uncertainty, weighing and sample preparation uncertainty, and scattering of data of instrumental analysis

Table 10b (KRISS), Continued

Material Delivery Competencies

Competency	Specifications / Comments
Homogeneity assessment	The claimed uncertainty covers the results of analysis of 10 of randomly selected samples
Stability assessment	Both short term (1 yr) and long term (5 yr) stability data are within the claimed uncertainty
Packaging	Appropriate for up to 20 years (glass vials with clamped silicone rubber caps: frozen serum kept at -20 ^{o}C, lyophilized serum kept at 5 ^{o}C). Absence of microbiological contamination was confirmed by observing no change after holding the materials at room temperature for a week.
Shipping	Successful in domestic shipping to 150 laboratories
Other: Material Handling	A good deal of experiences in handling serum materials in a controlled manner including compounds-fortification and sterilization Serum materials
Other: Use in PT	The CRM candidates had been successfully used for a PT of Korean medical diagnostic laboratories (150 labs) in year 2005.

Table 10c: Core Competencies Demonstrated in CCQM-K80 by LGC

Scope

Analyte	Small polar organic molecule
Matrix	Fresh frozen or lyophilized biological fluid such as serum, plasma, whole blood, and urine
Mass fraction	2 µg/g to 50 µg/g
Demonstrated U_{95} capability	From 4 % to 16 % which includes uncertainty contributions from characterization, homogeneity, and prediction of long-term stability.

Calibration Competencies

Competency	Specifications / Comments
Identity verification	In this instance the material was sourced from another NMI so identity was only verified by precursor and product ion MS(/MS).
Purity assessment	High-purity substance sourced from another MRA signatory with a CMC for high-purity creatinine.
Calibrant preparation	Preparation of multiple calibration solutions by mass. To include multiple analysts, dissolving different amounts of solid in a solvent, and assessment of calibration solution stability in the chosen solvent.

Sample Analysis Competencies

Competency	Specifications / Comments
Measurand identification	Comparison of liquid chromatography retention times and precursor and product ion mass spectra with those of a known standard
Extraction	Protein precipitation
Cleanup	Centrifugation and filtration
Transformation	none
Method validation	Assessment of isotope equilibration, gravimetric spiked recovery experiments and inter-run QC with a CRM. Method used in CCQM-K12 and results were assessed in terms of accuracy and uncertainty for the production of a CRM.
Use of analytical techniques	LC/MS
Quantification mode(s), data analysis, and uncertainty evaluation	Exact matching ID-MS. Sample injection was bracketed by standards immediately before and after each sample. The uncertainty of each sample was derived from the standard uncertainty of the individual constituents of the ID-MS equation.
Other: Training protocols and SOPs.	The overall uncertainty of the reference material combines the standard uncertainties from the characterization data with an uncertainty from the homogeneity and stability data. These materials were prepared and characterized many years after CCQM-K12 using different analysts and instrumentation.

Table 10c (LGC), Continued

Material Delivery Competencies

Competency	Specifications / Comments
Homogeneity assessment	A number of vials, 10 in this instance, were randomly selected from the CRM production batch. The exact number depends on the number of units in the batch. Each unity was analyzed in duplicate. The same method that was used for characterization was used in the homogeneity assessment.
Stability assessment	Stability assessed two years after production and is routinely monitored while the CRM is available for sale.
Packaging	Screw cap plastic vials were assessed for ruggedness and to ensure long-term storage stability.
Shipping	Shipping was via LGC Standards using the standard procedures and practices for shipping reference materials ensuring the materials are shipped under appropriate conditions to maintain CRM stability in transit. Shipping is done in accordance with current labeling and transport restrictions.
Other: Use in PT	The CRMs were produced in accordance with LGCs SOPs for the production of CRMs under ISO-17025 and ISO-Guide 34 accreditation. The CRMs were used as PT samples for a UK WEQAS trail with more than 300 participants.

Table 10d: Core Competencies Demonstrated in CCQM-K80 by: NIM

Scope

Analyte	Small, polar, relatively stable non-protein organic compounds
Matrix	Frozen relatively homogenous complex bio-fluids such as human serum, plasma, and urine
Mass fraction	8 mg/kg and above
Demonstrated U_{95} capability	1.2 % of certified value

Calibration Competencies

Competency	Specifications / Comments
Identity verification	SRM 914a Creatinine (0.997 g/g±0.003 g/g, purchased from NIST) used as calibrant
Purity assessment	Purity certified by NIST and confirmed by our NMR and LC
Calibrant preparation	gravimetric method with acidified milli-Q water

Sample Analysis Competencies

Competency	Specifications / Comments
Measurand identification	Chromatographic retention time, standard stock solution, mass profile
Extraction	Protein precipitation
Cleanup	Filtration
Transformation	None
Method validation	SRM 909b for method validation, and comparison of results from independent scientists and different laboratories
Use of analytical techniques	ID-LC/MS
Quantification mode(s), data analysis, and uncertainty evaluation	The single-point calibration method; uncertainties combined the certification of reference values, the heterogeneity of materials, and the instability of the CRM

Material Delivery Competencies

Competency	Specifications / Comments
Homogeneity assessment	Appropriate to uncertainty claimed in Scope
Stability assessment	Appropriate to uncertainty claimed in Scope
Packaging	Appropriate for at least 1 yr storage
Shipping	Appropriate for national distribution

Table 10e: Core Competencies Demonstrated in CCQM-K80 by: NIST

Scope

Analyte	Small, polar, relatively stable non-protein organic compounds
Matrix	Frozen and lyophilized relatively homogenous complex body fluids such as human serum, plasma, and urine
Mass fraction	7 mg/kg and above on basis of our own materials, 3 mg/kg and above on basis of our successful analysis of all materials in study
Demonstrated U_{95} capability	2.2 % of certified value

Calibration Competencies

Competency	Description / Comments
Identity verification	SRM 914a Creatinine (0.997 g/g±0.003 g/g, certified in 1987, identity established with NMR and GC-MS) used as calibrant
Purity assessment	SRM 914a was value-assigned using acidimetric titration, Karl Fischer titration, ion chromatography, NMR, and NAA
Calibrant preparation	Gravimetric dilution of the solid SRM 914a with acidified water

Sample Analysis Competencies

Competency	Description / Comments
Measurand identification	Chromatographic retention time, mass profile
Extraction	Protein precipitation (LC/MS); solid phase extraction (GC/MS)
Cleanup	Filtration
Transformation	Post-separation esterification
Method validation	Matrix-matched CRM controls, comparison of results from independent methods
Use of analytical techniques	ID-GCMS and ID-LC/MS
Quantification mode(s), data analysis, and uncertainty evaluation	Quantification by ID-MS. Calculations performed using purpose-built spreadsheet software; uncertainties combine instrumental imprecision, material heterogeneity, and between-method bias.
Other (specify)	

Material Delivery Competencies

Competency	Description / Comments
Homogeneity assessment	Appropriate to uncertainty claimed in Scope
Stability assessment	Appropriate to uncertainty claimed in Scope
Packaging and storage	Appropriate for up to 15 yrs storage
Shipping	Not applicable (we did not ship materials to ourselves)

Table 10f: Core Competencies Demonstrated in CCQM-K80 by: PTB

Scope

Analyte	Small, polar, relatively stable molecules (i.e. M<= 1000 g/mol)
Matrix	Body fluids (e.g., serum, plasma, saliva, urine)
Mass fraction	5 mg/kg to 60 mg/kg (45 µmol/L to 550 µmol/L) and above after dilution
Demonstrated U_{95} capability	1.5 %

Calibration Competencies

Competency	Specifications / Comments
Identity verification	NIST-SRM 914a (identity by certificate)
Purity assessment	NIST-SRM 914a (purity and its uncertainty by certificate)
Calibrant preparation	Gravimetric dilution

Sample Analysis Competencies

Competency	Specifications / Comments
Measurand identification	Mass+ retention time in comparison to the pertaining data of SRM 914a
Extraction	LC-type extraction (here: strong cation exchange chromatography)
Cleanup	Same as above
Transformation	Chemical derivatization (here: silylation)
Method validation	SRM 909b; comparison to LC/MS results as alternative method; numerous intercomparisons including those on CCQM level
Use of analytical techniques	ID-GC/MS; ID-LC/MS
Quantification mode(s), data analysis, and uncertainty evaluation	Exact matching ID-LC/MS. Data analysis by use of a validated Excel-spreadsheet with LC/MS peak areas as input. GUM-compliant uncertainty evaluation using repeatability of measurement as type A input and uncertainties on purity (RM used for calibration) and possible nonlinearity of balance (data by manufacturer) as type B contributions.

Material Delivery Competencies

Competency	Specifications / Comments
Homogeneity assessment	Not applicable as PTB does not provide RMs in the organic area
Stability assessment	As above
Packaging	As above
Shipping	PTB re-shipped the samples (obtained from DGKL), thus demonstrating international shipping capability.

6. REFERENCES

1 CIPM. Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes. Comité international des poids et mesures. Paris, 14 October 1999. http://www.bipm.org/en/cipm-mra/documents/

2 Duewer DL, Gasca-Aragon H, Lippa KA, Toman B. Experimental design and data evaluation considerations for comparisons of reference materials. Accred Qual Assur 2012;17:567–588. DOI 10.1007/s00769-012-0920-4

3 Stokes P, O'Connor G. Development of a liquid chromatography-mass spectrometry method for the high-accuracy determination of creatinine in serum. J Chromatography B 2003;794:125-36. DOI: 10.1016/S1570-0232(03)00424-0

4 JCGM 100:2008. Guide to the Expression of Uncertainty in Measurement (GUM). BIPM, Sèvres, France. http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

5 JCGM 101:2008. Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement" — Propagation of distributions using a Monte Carlo method. BIPM, Sèvres, France. http://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf.

6 Ciarlini P, Cox MG, Pavese F, Regoliosi G. The use of a mixture of probability distributions in temperature interlaboratory comparisons. Metrologia 2004;41:116-121. DOI:10.1088/0026-1394/41/3/002

7 Picard RR, Cook RD. Cross-validation of regression models. J Amer Statist Assoc 1984;79(387):575-583. DOI: 10.1080/01621459.1984.10478083. See also "Cross-validation (Statistics)" http://en.wikipedia.org/wiki/Cross-validation_%28statistics%29#Leave-one-out_cross-validation

APPENDIX A
CALL FOR PARTICIPANTS

From: May, Willie E. Dr.
Sent: Monday, August 10, 2009
To: <list>

CCQM Call for Participants for CCQM-K80: value-assignment of CRMs and Proficiency Testing materials for Creatinine in Serum

All NMIs that deliver measurement services for Creatinine in human serum through one or more value-assigned CRMs or PT materials should participate in CCQM-K80.

> If your Institute cannot participate in CCQM-K80 but has Calibration and Measurement Claims (CMCs) in the CIPM MRA Key Comparison Database for creatinine in serum where CRMs or PT value assignment is one of the delivery mechanisms, this may result in your CMCs for this being deleted from the Database

Unlike previous OAWG studies, participation in this KC is accomplished by providing the Measurement Laboratory with creatinine in serum CRM and/or PT materials that your institution has value-assigned, i.e., established a value and the uncertainty on that value. **All of the comparison measurements will be made at the Measurement Laboratory** under repeatability conditions. See the attached study proposal/project plan for details[*]. Institutes that currently distribute from one to four different creatinine in serum materials are asked to provide of **all** of their materials; institutes that distribute five or more materials are asked to provide **four** materials that are representative of their complete suite.

Coordinating Laboratory:	NIST
Contact at the Coordinating Laboratory:	david.duewer@nist.gov
Measurement Laboratory:	NIST
Contact at the Measurement Laboratory:	karen.phinney@nist.gov
CRM/PT information due to coordinators:	**ASAP**
CRM / PT materials due to coordinators:	**7 Sep 2009**
Discussion of results	Nov 2009 meeting, if materials are received in time to complete the measurements.

To participate, please:

1) E-mail david.duewer@nist.gov to indicate your intention to participate and provide the name and contact information for your institute. Please also provide a listing of your relevant CRM and/or PT materials and for each of these, a copy of 1) the CRM Certificate or the PT material value-assignment report and 2) the handling information provided to the CRM customer or PT participants for the materials you believe suitable for inclusion in this KC.

---

[*] *The study proposal/project plan was a trivial adaptation of the* "Proposal for demonstrating NMI Measurement Capability from analysis of CRMs" *presented at the April 2009 OAWG Meeting. This OAWG working document is available as OAWG/09-10 (http://www.bipm.org/wg/CCQM/OAWG/Restricted/April_2009/OAWG_0910.pdf) on the BIPM-hosted website.*

2) Upon notification from David Duewer to you that your materials are suitable for inclusion in the KC, ship **three** units of each of the materials to:

   Karen Phinney
   NIST
   100 Bureau Drive Stop 8392
   Gaithersburg, MD 20899 USA
   301-975-4457

   Please clearly indicate on the shipping label that these materials are for an interlaboratory study and provide any other information that you believe will facilitate shipping.

   To ensure that the measurements can be completed in time for CCQM-K80 to be discussed at the November 2009 meeting, NIST *must* receive all materials by *7 Sep 2009*.

3) E-mail the shipping date and other shipping details to karen.phinney@nist.gov as soon as possible after the materials are shipped. An acknowledgement of receipt of materials will be sent upon their arrival at the Measurement Laboratory.

If you have any comments or concerns regarding the design or proposed conduct of this KC, please contact david.duewer@nist.gov with a cc to willie.may@nist.gov.

APPENDIX B
SOURCES OF INFORMATION

B.1 CENAM: DMR 263a

Information extracted from documents “Certif_DMR-263_inglés_12_Aug_09.docx” and “MATERIAL SAFETY DATA SHEET SERUM.docx” emailed by Melina Pérez Urquiza on 12-Aug-2009 with the exception of: 1) conversion of units from mg/dL to mg/kg and 2) clarification of what is meant by “room temperature” emailed by Melina Pérez Urquiza on 18-Aug-2009.

B.2 KRISS: 111-01-01A, 111-01-02A, 111-01-03A, and 111-01-04A

Information extracted from “certificate-CRM-111-01-01A.doc,” “certificate-CRM-111-01-01A.doc,” “certificate-CRM-111-01-01A.doc,” and “certificate-CRM-111-01-01A.doc” emailed by Sang-Ryoul Park on 14-Aug-2009.

B.3 LGC: ERM-DA250a, ERM-DA251a, ERM-DA252a, and ERM-DA253a

Information extracted from email from Gavin O’Connor on 13-Aug-2009 and “Certificate of Analysis ERM® -DA250a”, “Certificate of Analysis ERM® -DA251a”, “Certificate of Analysis ERM® -DA252a”, and “Certificate of Analysis ERM® -DA253a” downloaded from http://www.lgcstandards.com/ShowProduct.aspx?productCode=ERM-DA250, –DA251a, -DA252a, and –DA253a on 13-Aug-2009.

B.4 NIM: Creatinine-1 and Creatinine-2

Information extracted from “To participate K80-Xinhua090819.doc” sent by Xinhua Dai on 19-Aug-2009 and draft certificate “GBW-090927.pdf” sent by Xinhua Dai on 27-Sep-2009.

B.5 NIST: SRM 909b I and SRM 909b II

Information extracted from “Certificate of Analysis Standard Reference Material® 909b Human Serum” downloaded from https://www-s.nist.gov/srmors/view_cert.cfm?srm=909B except for fill-mass data provided by Johanna Camara as part of the measurement report on 22-Oct-2009.

B.6 NIST: SRM 967a I and SRM 967a II

Information extracted from “SRM 967a ROA-090409.doc” emailed by Johanna Camara 22-Sep-2009 and “Summary on statistical analysis on SRM 967a” emailed by Nien-Fan Zhang on 1-Oct-2009.

B.7 PTB: RELA 1/05 KS-A and RELA 1/05 KS-B

Information extracted from email from Ruediger Ohlendorf on 24-Aug-2009 and documents “Certificate creatinin samples PTB.pdf” and “GUM consideration creatinin measurements PTB.pdf” emailed by André Henrion on 10-Sep-2009.

APPENDIX C
REPEATABILITY MEASUREMENT EXPERIMENTAL DETAILS

C.1 Reagents and Materials

SRM 914a Creatinine, with a purity of 99.7 % ± 0.3 %, was obtained from NIST. The stable isotope labeled internal standard material, d_3-creatinine, was commercially obtained. All solutions, LC mobile phase, and reconstitution of sera (with the exception of the lyophilized SRM 909b materials, see below) were prepared using LC-grade water. High-purity (>99 %) ammonium acetate was used to prepare the mobile phase.

C.2 Control Solution

Stock solutions of creatinine and d_3-creatinine were prepared gravimetrically and stored at -20 ºC when not in use. A 100-mL amber bottle was used to prepare the solutions. Approximately 1 mg of creatinine or d_3-creatinine was weighed into an aluminum boat and transferred to the pre-weighed storage bottle. Approximately 80 mL of water was added to the bottle and the final mass was determined. This yielded solutions of creatinine and d_3-creatinine with nominal mass fractions of 10.0465 µg/g and 17.3287 µg/g, respectively. These solutions were combined to yield a control solution having an approximate 1:1 mass ratio (≈3.5 µg of each component). All solutions were stored at -20 ºC when not in use.

C.3 Reconstitution of Lyophilized Materials

Each lyophilized sample was reconstituted according the directions provided by the NMI submitting the material. All samples were removed from storage at 4 ºC and allowed to equilibrate at room temperature. For KRISS samples 111-01-03A and 111-01-04A, vials were tapped to dislodge particles from the caps. After carefully removing the metal seals and caps, 10.00 mL ±0.02 mL water was added to each vial using a Type I Class A volumetric pipette. The stoppers were replaced and the vials were swirled to mix the contents. PTB samples RELA 1/05 KS-A and RELA 1/05 KS-B were reconstituted by the addition of 5.00 ±0.02 mL of water using a Type I Class A volumetric pipette. The samples were allowed to stand for 30 min protected from light. The lyophilized serum was dissolved by careful shaking.

For NIST SRM 909b I and II, the diluent water provided with the SRM was used after equilibrating to room temperature. Instructions for fill-mass correction were followed, which involved scraping off the labels and wiping the bottles with ethanol. The metal closure was removed and the bottom was lightly tapped to remove particles adhering to the stopper. The stopper was removed to equalize air pressure and then replaced. The bottle containing the lyophilized serum was weighed and the sample was then reconstituted with 10.00 mL ± 0.02 mL of the provided water using a Type I Class A volumetric pipette. This serum was swirled 2 to 3 times and allowed to stand for 10 min. The contents were again swirled, and the samples allowed to stand for an additional 30 min. This process was repeated with a final 10 min incubation period. Finally, the bottles were inverted several times to ensure mixing. Once the serum was removed, the SRM 909b I and II bottles were cleaned, dried, and weighed to determine accurate fill-masses: 0.87997 g and 0.87844 g for units 1 and 2 of SRM 909b I and 1.51158 g and 1.51119 g for units 1 and 2 of SRM 909b II. These masses are accurate to approximately ±0.00005 g. All materials remained at room temperature until all samples were reconstituted. The reconstituted samples were stored at 4 ºC until further processing (≈2 h).

C.4 Sample Preparation

Serum materials to be analyzed were removed from -80 ºC (frozen) or 4 ºC (reconstituted lyophilized) storage and allowed to equilibrate to room temperature. Based on provided assigned creatinine values, samples were prepared gravimetrically using the exact matching technique by dilution with d_3-creatinine internal standard solution, resulting in approximately equal masses of creatinine and d_3-creatinine (≈2.6 µg each). In both measurement campaigns, duplicate aliquots of each material were prepared and processed independently.

For most of the samples, ≈150 mg of d_3-creatinine internal standard solution was weighed into 15-mL plastic centrifuge tubes with screw caps. The appropriate amount of serum was then added and weighed accurately. All samples were vortex-mixed and allowed to equilibrate overnight at 4 ºC.

All LGC materials came with the stipulation that a minimum serum amount of 0.4 g be processed per analysis. Therefore, ERM-DA250a, -DA-251a, and -DA-253a were prepared by diluting ≈0.4 g aliquots with matching levels of d_3-creatinine internal standard (8.8 µg/g to 19.9 µg/g). To enable determination of whether this increased sample size introduced bias, normal-scale samples were also prepared from the same materials according to the above protocol. As an additional exception, ERM-DA252a is a low-level material (3.1 µg/g creatinine) provided in bottles of 1.0 mL each. For this material, 0.5 g aliquots were removed and combined with matching levels of d_3-creatinine internal standard (≈1.5 µg). These materials were mixed and equilibrated as stated above.

Following equilibration, three volumes relative to total sample volume of ice-cold ethanol were added to each of the samples, which were then vortex-mixed and allowed to stand for 5 min to precipitate proteins. Samples were then centrifuged at 314 rad/s (3,000 rpm) for 20 min at room temperature. The supernatant from each sample was transferred via a plastic pipette to a 5-mL amber glass vial. The supernatants from LGC materials with larger volumes (ERM-DA250a, -DA251a, and -DA253a) were transferred to 200 mL evaporation tubes. All samples were evaporated to dryness under nitrogen at 40 ºC. Dried residues were reconstituted in 500 µL water and vortex-mixed. Samples were then filtered through 0.45 µm polyvinylidene fluoride (PVDF) syringe filters into 2-mL screw-cap plastic tubes. The high-level LGC samples were diluted 1→10 (v→v) with water after filtration. Once the control solution was removed from -20 ºC and thawed at room temperature, 200 µL aliquots of the control solution and samples were transferred to amber glass LC vials with conical inserts for analysis.

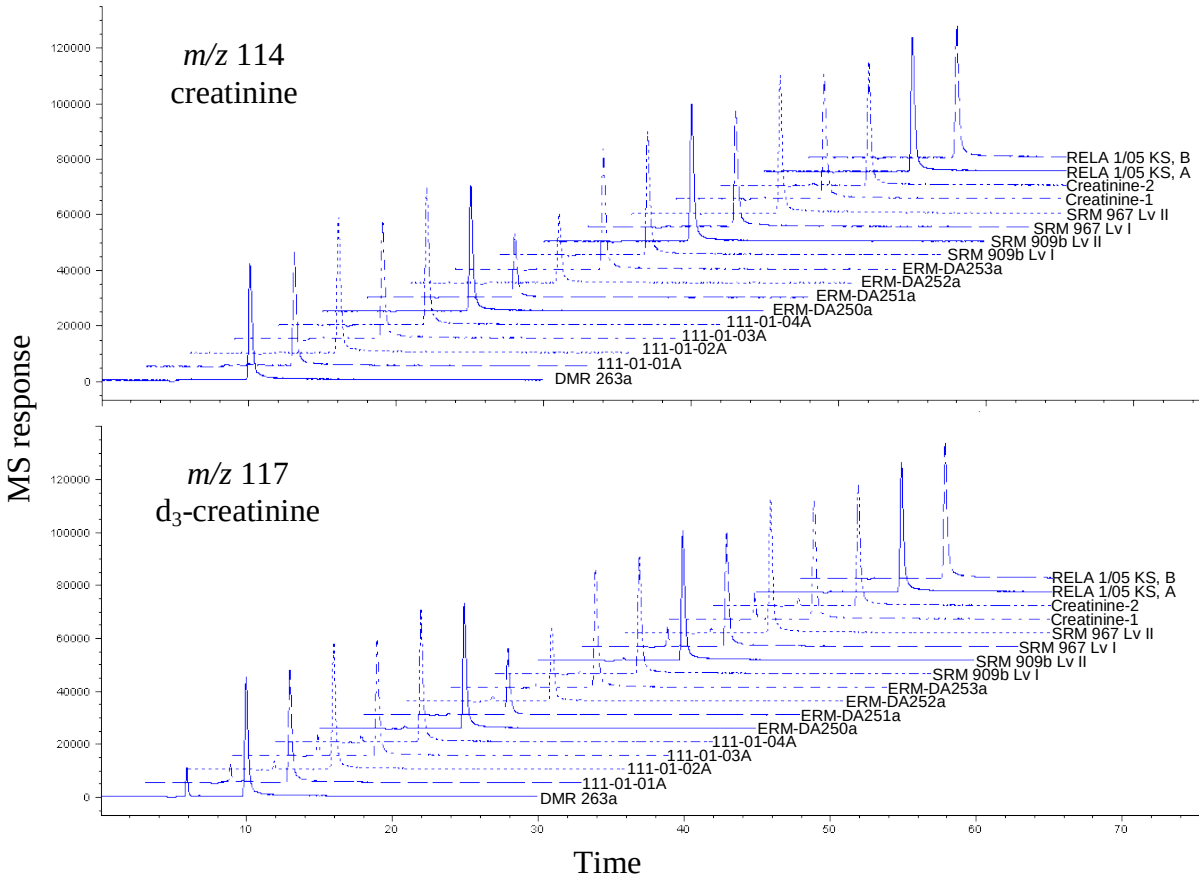
C.5 Instrumentation

A liquid chromatograph-mass spectrometric detector (LC/MS) was used to analyze all samples. The column utilized was a Phenomenex (Torrance, CA USA) Luna C18(2), 250 mm × 4.6 mm, 5 µm particle. The LC parameters were: mobile phase, 10 mmol/L ammonium acetate in water; flow rate, 0.5 mL/min; gradient, isocratic; column temperature, 22 ºC; injection volume, 5 µL. The MS detection parameters were: positive-mode electrospray ionization; gas temperature, 350 ºC; vaporizer temperature, 150 ºC; drying gas, 12.0 L/min; nebulizer pressure, 345 kPa (50 psig); capillary, 1500 V; charge, 2000 V. Selected ion monitoring (SIM) was used to detect creatinine at *m/z* 114 and d_3-creatinine at *m/z* 117.

C.6 Chromatograms

None of the chromatographic peaks used for quantitation appeared irregular on visual inspection. While *m/z* 117 peaks preceding the d_3-creatinine target peak were present in several of the materials, all such peaks were well separated from the target peak. Figure C1 presents representative chromatograms.

Figure C1: Representative Chromatograms



C.7 Measured Quantity

Equation 1 was used to transform the observed chromatographic peak areas and gravimetrically-determined masses into response measurements, *R*:

$$R = \frac{Area_{114} \times Mass_{IS}}{Area_{117} \times Mass_{material}} \qquad [C1]$$

where $Area_{114}$ is the *m/z* 114 peak area for the target creatinine peak, $Area_{117}$ is the *m/z* 117 peak area for the target d_3-creatinine internal standard peak, $Mass_{material}$ is the mass of sample, and $Mass_{IS}$ is the mass of the d_3-creatinine internal standard. While nominally having units of mg/kg, for the purposes of this study the *R* are formally expressed in arbitrary units (a.u.).

APPENDIX D
MEASUREMENT QUALITY ASSURANCE

D.1 Within- and Between-Campaign Measurement Quality Assurance

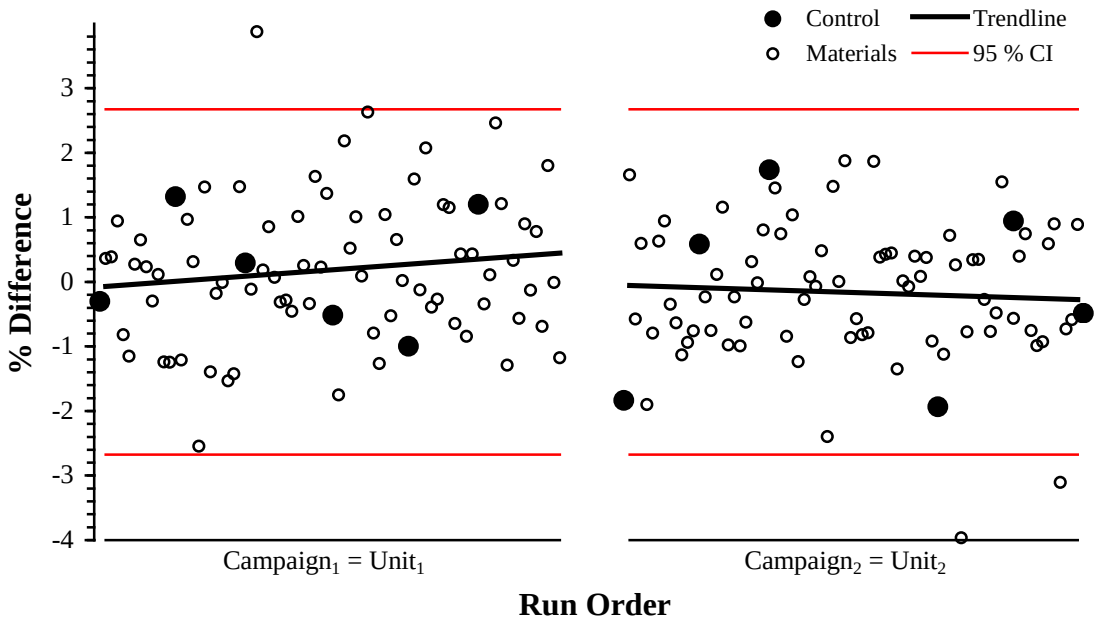
In addition to the measurements on the 17 CCQM-K80 materials, 6 measurements were made on individual aliquots of the control solution at regularly spaced intervals within each campaign. There was no significant change in the response measurement for the control solution over the course of the two measurement campaigns.

Figure D1 documents the stability of the measurements for both the control solution and the CCQM-K80 materials, plotting the percent differences of the replicate measurements from their mean value,

$$\%\text{Difference}_{ij} = 100 \times \left(\frac{R_{ij} - R_i}{R_i} \right) = 100 \left(\frac{R_{ij}}{\overline{R}_i} - 1 \right) \qquad [D1]$$

where i indexes over the 17 materials, j indexes over the 8 or 12 measurements made on the i^{th} material. Any systematic change in the measurement process would manifest as a trend and/or change in scatter; within experimental error there is little evidence of within-campaign run-order trends and no evidence for any between-campaign difference in offset, trend, or scatter.

Figure D1: Measurement Differences as a Function of Run-Order

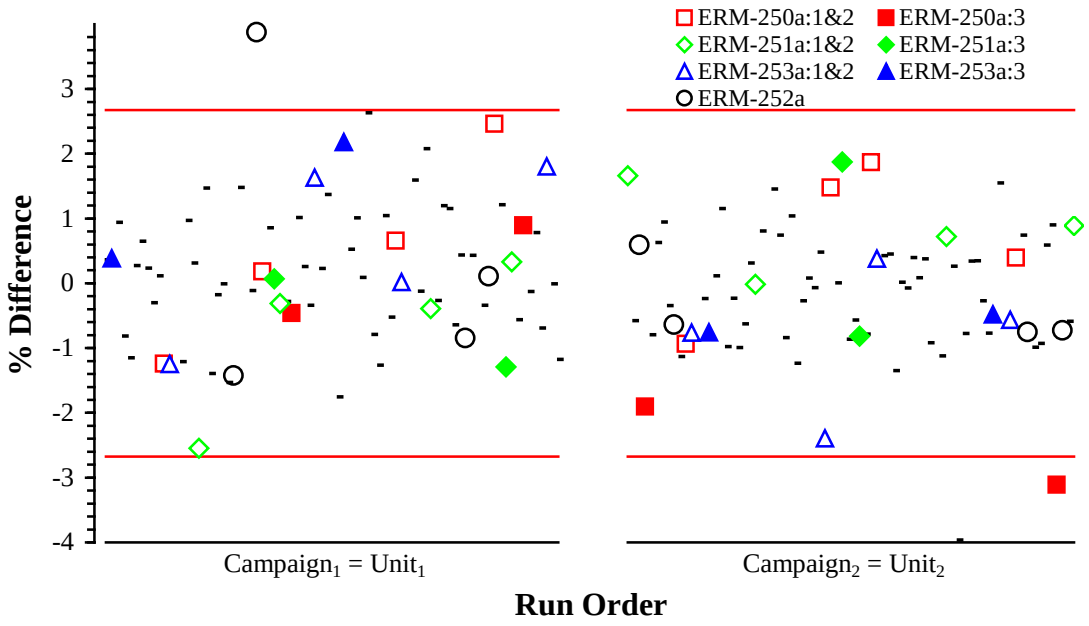


The relative standard measurement repeatability estimated from the 12 independent control solution %Differences is 1.2 %, giving a 95 % level of confidence interval of 2.7 %. The 148 individual %Differences for the CCQM-K80 materials are not independent and thus their standard deviation of 1.1 % is indicative but does not directly estimate measurement repeatability.

D.2 Sample Preparation Quality Assurance for LGC Materials

Sample preparation for the LGC materials differed from that for the other CCQM-K80 materials: ERM-252a because of its very low creatinine level and ERM-250a, -251a, and -253a because of the specified minimum sample mass. Figure D2 displays %Difference values for all of the LGC materials relative to the other materials, where the open symbols display data for the "specified preparation" ($Aliquot_1$ and $Aliquot_2$) and the large solid symbols display data for the "routine preparation" ($Aliquot_3$) measurements.

Figure D2: Measurement Differences for LGC Materials



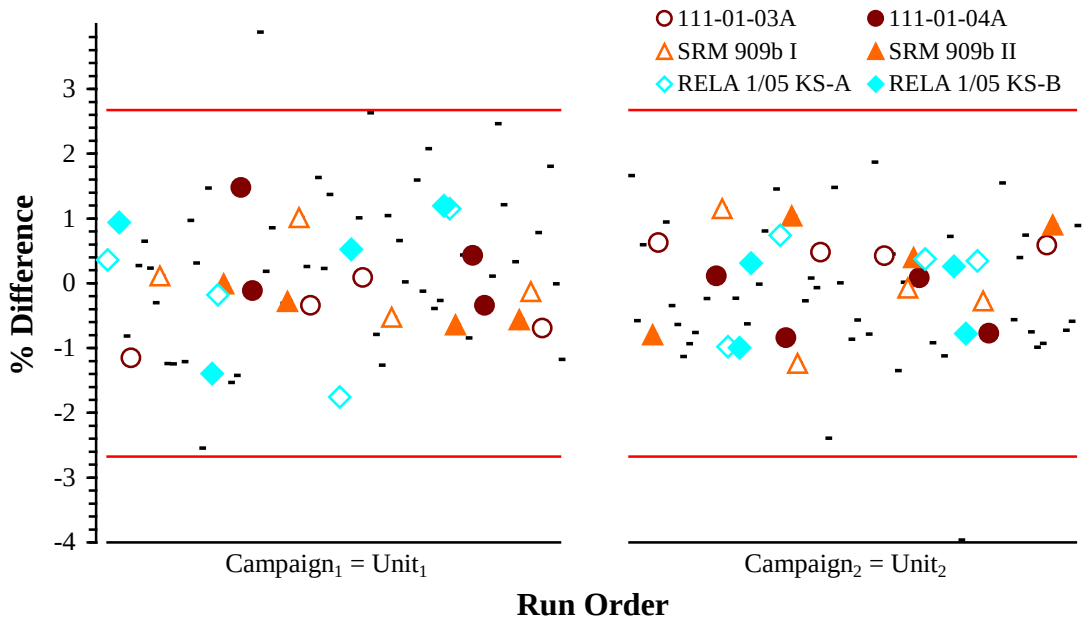
While the %Difference for one of the four ERM-252a measurements in $campaign_1$ is above the 95 % CI, the %Differences for others are unexceptional. The single large %Difference thus reflects measurement repeatability (and the low creatinine level of the material) rather than either sample preparation or within-unit heterogeneity. There is no apparent bias or excess variability in the "specified" measurements of the other LGC materials.

The %Differences for the "routine" measurements of ERM-251a and -253a are unexceptional. However, both $Aliquot_3$ measurements of ERM-250a in $campaign_2$ are low. The measurement design does not enable distinguishing whether this bias is from material handling or within-unit heterogeneity. In either case, the measurements performed on the aliquots prepared as "specified" are not intrinsically biased relative to the other materials.

D.3 Sample Preparation Quality Assurance for lyophilized Materials

Figure D3 displays the %Difference values for the six lyophilized materials in the study: 111-010-03A, 111-010-03A, SRM 909b I, SRM 909b II, RELA 1/05 KS-A, and RELA 1/05 KS-B. The standard deviation of the %Difference values for these materials is 0.8 %, somewhat less than that of the control measurements and the other samples. While we cannot rationalize this apparent *decrease*, the reconstitution step required for these materials did not *increase* the measurement variability.

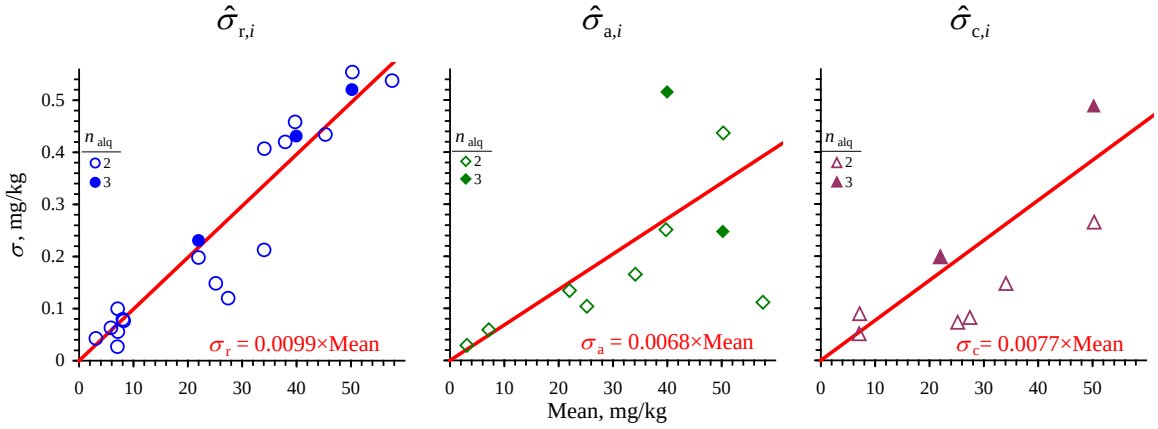
Figure D3: Measurement Differences for Lyophilized Materials



APPENDIX E
RELATIVE VARIANCE

Figure E1 displays the estimated between-replicate, $\hat{\sigma}_{r,i}$, between-aliquot, $\hat{\sigma}_{a,i}$, and between-campaign, $\hat{\sigma}_{c,i}$, estimates as functions of the nominal measured creatinine quantities, R_i. The red lines represent the expected value for each component based upon pooling of non-zero estimates.

Figure E1: Variance Components as Functions of Creatinine Level



The 0.99 % pooled relative $\hat{\sigma}_{r,i}$ estimates the instrumental sources of variance. There are no exceptionally large values. There are no significant differences in the estimates based on duplicate analyses of two aliquots per unit versus those on duplicates of three aliquots.

Since many of the $\hat{\sigma}_{a,i}$ are estimated as zero, the 0.68 % pooled relative standard deviation provides only a worst-case bound on the aliquot preparation variance. Since many of the $\hat{\sigma}_{c,i}$ are likewise estimated as zero, the 0.77 % pooled relative standard deviation also provides only an upper-bound on the sample preparation variance sources.