Anal Bioanal Chem (2012) 403:527–535
DOI 10.1007/s00216-012-5869-y

ORIGINAL PAPER

An international assessment of the metrological equivalence of higher-order measurement services for creatinine in serum

Johanna E. Camara · Katrice A. Lippa · David L. Duewer · Hugo Gasca-Aragon · Blaza Toman

Received: 5 January 2012 / Revised: 10 February 2012 / Accepted: 12 February 2012 / Published online: 6 March 2012

Abstract The Consultative committee for amount of substance-metrology in chemistry (CCQM)-K80 Key Comparison directly assessed the equivalence of many of the world's higher-order value-assigned materials (HOVAMs) for creatinine in human serum. This 2009 international study compared the certified values and uncertainties of the materials using measurements made under repeatability conditions. The study evaluated 17 materials submitted by 6 national metrology institutes (NMIs). The creatinine quantity in these materials ranged from 3 mg/kg to 57 mg/kg (about 0.3 mg/dL to 6 mg/dL or 30 nmol/L to 500 nmol/L). All materials were stored and prepared according the specifications provided by the participating NMIs. Samples were processed and analyzed under repeatability conditions by one analyst using isotope-dilution liquid chromatography–mass spectrometry in two measurement campaigns. The certified values and repeatability measurements were compared using uncertainty-weighted generalized distance regression. The instrumental repeatability relative standard deviation was 1.2%. The measurement design required assessment of within-unit and between-campaign variability in addition to measurement repeatability. At a 95% level of confidence, the certified values for all 17 materials agreed to within their assigned uncertainties. CCQM-K80 demonstrated the metrological equivalence of the currently available HOVAMs for creatinine in human serum and of the creatinine measurement services provided by the participating NMIs.

Keywords Creatinine · Certified reference material · Degree of equivalence · Key comparison · Metrology · National metrology institute

Abbreviations

CCQM	Consultative Committee for Amount of Substance–Metrology in Chemistry
CENAM	Centro Nacional de Metrología (Mexico)
CRM	Certified reference material
GDR	Generalized distance regression
GUM	Guide to the expression of uncertainty in measurement
HOVAM	Higher-order value-assigned material
KC	Key comparison
KRISS	Korea Research Institute of Standards and Science (Korea)
LC	Liquid chromatography
LC-IDMS	Liquid chromatography-isotope dilution mass spectrometry
LGC	Limited (UK)
LOO	Leave one out cross-validation
NIM	National Institute of Metrology (China)
NIST	National Institute of Standards and Technology (USA)
NKDEP	National Kidney Disease Education Program

Electronic supplementary material The online version of this article (doi:10.1007/s00216-012-5869-y) contains supplementary material, which is available to authorized users.

J. E. Camara (✉) · K. A. Lippa · D. L. Duewer · H. Gasca-Aragon
Analytical Chemistry Division, National Institute of Standards and Technology (NIST),
100 Bureau Drive, MS 8392,
Gaithersburg, MD 20899-6392, USA
e-mail: johanna.camara@nist.gov

H. Gasca-Aragon
Department of Mathematics & Statistics, University of Massachusetts,
Amherst, MA 01003-9305, USA

B. Toman
Statistical Engineering Division, National Institute of Standards and Technology (NIST),
Gaithersburg, MD 20899-8980, USA

NMI	National metrology institute
OAWG	Organic analysis working group
PBMC	Parametric bootstrap Monte Carlo
PT	Proficiency test
PTB	Physikalisch–Technische Bundesanstalt (Germany)
SIM	Selected ion monitoring

Introduction

Creatinine is a very polar analyte and 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. Among these interferences is creatine which presents both separation and inter-conversion challenges [1]. Use of different methods of analysis, different reagents, etc., may lead to significantly different results [2]. Therefore, reference methods and higher-order value-assigned materials (HOVAMs), both certified reference materials (CRMs) and value-assigned proficiency test (PT) materials, are needed to maintain adequate accuracy in routine measurements for creatinine in blood or serum. The availability of HOVAMs for serum creatinine is of particular interest to the National Kidney Disease Education Program (NKDEP), which in collaboration with the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) and the European Communities Confederation of Clinical Chemistry has created the Creatinine Standardization Program. Reliable serum creatinine measurements are crucial for global public health efforts to improve the diagnosis and treatment of chronic kidney disease [3].

The Organic Analysis Working Group (OAWG) of the Consultative Committee for Amount of Substance–Metrology in Chemistry (CCQM) has conducted three studies documenting the measurement capabilities of National Metrology Institutes (NMIs) for serum creatinine. CCQM Working Groups are responsible for designing and overseeing "Key Comparison (KC)" studies that address chemical measurement-related issues important for international trade, environmental, health, and safety-related decision making. Based upon the excellent agreement among participants in an exploratory or "pilot" study in 2000, CCQM–K12 "Determination of creatinine in human serum" [4] was conducted in 2002 with five NMI participants. Two additional NMIs demonstrated their creatinine measurement capability in the subsequent study CCQM–K12.1 in 2004. These studies used a traditional design whereby all participants analyzed nominally identical, limited-quantity samples of test materials. While enabling comparison of the higher-order measurement capabilities of the NMIs, this design does not address the services they deliver to customers through their HOVAMs.

At the April 2009 CCQM meeting, a KC design was presented for directly assessing the measurement services provided by NMIs through their materials. Delivery of services requires assessment of material homogeneity, packaging, storage, and transport in addition to characterizing the measurement processes themselves. The design's premise is that when a group of value-assigned materials delivers the same nominal measurand (same analyte in a "similar enough" matrix), then a comparison of the assigned values with measurements made under repeatability conditions can be used to evaluate the measurement capabilities of the institutions that value-assigned the materials.

Serum creatinine was selected as an appropriate material for this KC design based on the number of NMIs providing suitable materials, the number of available materials, and the recognized clinical relevance of the measurand. Given the potential utility of the KC, all OAWG members that deliver measurement services via value-assigned serum creatinine materials were asked to participate. Table 1 lists and describes the 17 materials submitted to the CCQM-K80 KC by six NMIs: CENAM, KRISS, LGC, NIM, NIST and PTB.

This report describes the CCQM–K80 "Comparison of value-assigned CRMs and PT materials: Creatinine in human serum" and summarizes its results. NIST coordinated this study and made the required repeatability–condition measurements. Since the quality of the repeatability measurements is critical to the success of the comparison, the experimental design and the measurement processes are described in detail.

Materials and methods

Materials

To limit the number of materials 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 it considered most representative (primarily with respect to matrix and analyte level) of the entire suite. Each participant was asked to provide three units of each value-assigned material. All serum materials were stored at the temperature specified in the provided instructions (Table 1) from the time of receipt to the time when the material was prepared for analysis.

Table 1 summarizes the assigned values, V_i, and their associated 95% level of confidence uncertainty intervals, $U_{95}(V_i)$, for the 17 materials submitted to CCQM-K80. This table also lists information provided by the participants

Table 1 Summary of all 17 certified value-assigned creatinine in human serum materials used in this comparison

Material			Certified value			Auxiliary information [a]						Condition [b]	Certification method [c]
NMI	Label	Code	V_i	$U_{95}(V_i)$	Units	Matrix	mL	MinSam	°C	Year	Expires		
CENAM	DMR 263a	A-1	7.35	0.35	mg/kg	Frozen	1		−80	2004	3–Nov–09	well frozen	ID–LC/MS
KRISS	111–01–01A	B-1	5.96	0.09	mg/kg	Frozen	3		−75	2007	31–Dec–12	thawed	ID–LC/MS
KRISS	111–01–03A	B-2	7.08	0.08	mg/kg	Lyoph	10		4	2007	31–Dec–17	ambient	ID–LC/MS
KRISS	111–01–04A	B-3	24.87	0.29	mg/kg	Lyoph	10		4	2007	31–Dec–17	ambient	ID–LC/MS
KRISS	111–01–02A	B-4	27.49	0.33	mg/kg	Frozen	3		−75	2007	31–Dec–12	thawed	ID–LC/MS
LGC	ERM–DA252a	C-1	3.1	0.2	mg/kg	Frozen	1	0.4 g	−20	2006	Shipment + 3 month	well frozen	ID–LC/MSMS
LGC	ERM–DA251a	C-2	22	2	mg/kg	Frozen	1	0.4 g	−20	2006	Shipment + 3 month	well frozen	ID–LC/MSMS
LGC	ERM–DA250a	C-3	39	2	mg/kg	Frozen	1	0.4 g	−20	2006	Shipment + 3 month	well frozen	ID–LC/MSMS
LGC	ERM–DA253a	C-4	50	2	mg/kg	Frozen	1	0.4 g	−20	2006	Shipment + 3 month	well frozen	ID–LC/MSMS
NIM	Creatinine–1	D-1	8.1	0.1	mg/kg	Frozen	1	0.011 mL	−70	2009	2013	thawing	ID–LC/MS
NIM	Creatinine–2	D-2	34.1	0.4	mg/kg	Frozen	1	0.011 mL	−70	2009	2013	thawing	ID–LC/MS
NIST	SRM 909b I	E-1	7.08	0.03	mg/kg/g	Lyoph	10		4	1996	Shipment + 5 year	ambient	ID–GC/MS
NIST	SRM 967a I	E-2	8.28	0.18	mg/kg	Frozen	1		<−60	2009	31–Dec–14	well frozen	ID–LC/MS
NIST	SRM 909b II	E-3	33.93	0.16	mg/kg/g	Lyoph	10		4	1996	Shipment + 5 year	ambient	ID–GC/MS
NIST	SRM 967a II	E-4	37.9	0.8	mg/kg	Frozen	1		<−60	2009	31–Dec–14	well frozen	ID–LC/MS
PTB	RELA 1/05 KS–A	F-1	44.89	0.92	mg/kg	Lyoph	5		4	2005		ambient	ID–GC/MS
PTB	RELA 1/05 KS–B	F-2	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 (**Lyoph**); **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 the material upon arrival

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

required for the storage and use of their materials. The repeatability measurements were not begun until all information was supplied and compiled and the accuracy of the compilation confirmed by the participating institutions.

Preparation of controls

SRM® 914a Creatinine, (99.7±0.3)% mass fraction, was obtained from NIST. The stable isotope labeled internal standard, d_3–creatinine, was obtained from Isotec.[1] All solutions, liquid chromatography (LC) mobile phase, and reconstituted sera (with the exception of the lyophilized E-1 and E-3 materials, see below) were prepared using LC-grade water. Fluka ammonium acetate (>99%) from Sigma–Aldrich was used to prepare the mobile phase.

Stocks of creatinine and d_3–creatinine were prepared gravimetrically in water, yielding solutions with nominal mass fractions of 10.0465 μg/g and 17.3287 μg/g. These solutions were combined to yield a control solution having an approximately equal mass ratio (≈3.5 μg of each component) and were stored at −20 °C prior to use.

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 °C and allowed to equilibrate at room temperature. For KRISS materials B-2 and B-3, (10.00±0.02) mL water was added to each vial using a Type I Class A volumetric pipette. Stoppers were replaced and the vials were swirled to mix contents. PTB materials F-1 and F-2 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 materials E-1 and E-3, the diluent water provided with the material was removed from storage at 4 °C with the sample and allowed to equilibrate to room temperature. Instructions for fill–weight correction were followed. The bottle containing the lyophilized serum was weighed and the sample was reconstituted with (10.00± 0.02) mL of the provided diluent water. This serum was swirled 2 to 3 times and allowed to stand for 10 min. The contents were again swirled, and samples were 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 E-1 and E-3 bottles were cleaned, dried, and weighed to determine accurate fill–weights. All materials remained at room temperature until all samples were reconstituted, and then were stored at 4 °C until required for further processing (≈2 h).

Analytical method

All materials were analyzed under repeatability conditions using the Joint Committee for Traceability in Laboratory Medicine–recognized definitive method utilizing exact matching isotope dilution reversed–phase liquid chromatography with electrospray ionization mass spectrometry (LC-IDMS) [1].

Sampling and measurement protocol

The participating institutions provided the measurement laboratory with three units of each of their submitted materials, two to be analyzed and one spare in case of technical failure and/or to facilitate investigation of disputed results. Given the number of materials and the time required for each analysis, the repeatability measurements were made in two measurement sessions ("campaigns"). Each campaign involved replicate analyses (two injections of one preparation separated in time) on each of two (or three for LGC materials C-2, C-3, and C-4; see below) independently prepared aliquots from one randomly selected unit of each of the 17 materials. This three-level nested design is summarized in Fig. 1. Sample run order was randomized and control measurements were interspersed at regular intervals throughout each campaign. The two campaigns were separated in time by 5 days, but no changes were made to the equipment, reagents, instrumentation, or controls between campaigns. The same analyst performed all sample preparations and measurements.

Serum sample preparation

Serum materials to be analyzed were removed from −80 °C (frozen) or 4 °C (reconstituted lyophilized) storage and

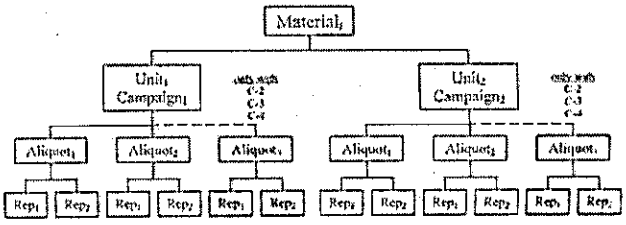

Fig. 1 Repeatability measurement design for CCQM-K80. Three LGC materials (C-2, C-3 and C-4) required a minimum sample mass of 0.4 g for sample preparation; thus, Aliquot$_1$ and Aliquot$_2$ were prepared using this certificate-specified sample mass while Aliquot$_3$ was prepared following the routine protocol used for the all other study materials

[1] Certain commercial materials are identified in this report to specify the experimental procedure as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials are necessarily the best available for the purpose.

allowed to equilibrate to room temperature. Based on provided assigned creatinine values (Table 1), these samples were prepared gravimetrically using the exact matching technique by dilution with d_3–creatinine internal standard solution, resulting in an approximately equal mass ratio of creatinine:d_3–creatinine (≈2.6 μg each). In each campaign, duplicate aliquots of each material were prepared and processed independently. To prepare 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 added and weighed accurately. All samples were vortex-mixed and allowed to equilibrate overnight at 4 °C.

As an exception, all LGC materials came with the stipulation that a minimum serum amount of 0.4 g be processed per analysis. Therefore, C-2, C-3, and C-4 were prepared by diluting ≈0.4 g aliquots with matching levels of d_3–creatinine (8.8 μg/g to 19.9 μg/g). To determine if this increased sample size introduced bias, normal-scale samples were also prepared from the same materials according to the previously described protocol. As an additional exception, C-1 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 (≈1.5 μg).

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 centrifuged at 1,500 g_n for 20 min at room temperature. The supernatant from each sample was transferred via plastic pipette to a 5–mL amber glass vial. The supernatants from LGC materials with larger volumes (C-2, C-3 and C-4) were transferred to 200 mL TurboVap tubes. Due to capacity limitations of the centrifuge and evaporation apparatus, samples were processed in two sets during each measurement campaign. All samples were evaporated to dryness under nitrogen at 40 °C. Dried residues were reconstituted in 500 μL water and vortex-mixed. Samples were then filtered through 0.45 μm polyvinylidene fluoride syringe filters into 2–mL screw–cap plastic tubes. The high-level LGC samples were diluted 1→10 by volume with water after filtration. Once the control solution was removed from –20 °C and thawed at room temperature, 200 μL aliquots of the control and samples were transferred to amber glass LC vials with conical inserts for analysis.

Instrumentation

An Agilent Technologies LC-MS was used to analyze all samples for creatinine response. The column utilized was a Phenomenex 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; isocratic; column temperature, 22 °C; injection volume, 5 μL. The MS detection parameters were: positive–mode electrospray ionization; gas temperature, 350 °C; vaporizer temperature, 150 °C; drying gas, 12.0 L/min; nebulizer pressure, 345 kPa (50 psig); capillary, 1,500 V; charge, 2,000 V. Selected ion monitoring (SIM) was used to detect creatinine at *m/z* 114 and d_3–creatinine at *m/z* 117.

Results

Determination of creatinine response

Figure 2 displays representative chromatograms for each material. None of the target *m/z* 114 creatinine or *m/z* 117 d_3–creatinine peaks appeared irregular upon 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.

The SIM peak areas and gravimetric measurements were transformed into responses having nominal units of mass fraction expressed as mg/kg as follows:

$$R = \frac{Area_{114} \times Mass_{IS}}{Area_{117} \times Mass_{material}} \quad (1)$$

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 the KC material in the solution, and $Mass_{IS}$ is the mass of the d_3–creatinine internal standard in the solution. Electronic Supplementary Material Table S2 lists the estimated mean responses, R, for each of the materials.

Measurement quality assurance

In addition to the measurements made on the study materials, a control 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 material C-1 and of the lyophilized materials (B-2, B-3, E-1, E-3, F-1 and F-2) suggests that the modified preparation procedures required for these materials did not affect measurement precision.

Since a significantly modified preparation procedure was required to meet the minimum sample mass requirements of 0.4 g for three of the LGC materials (C-2, C-3 and C-4), 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 nor affect precision.

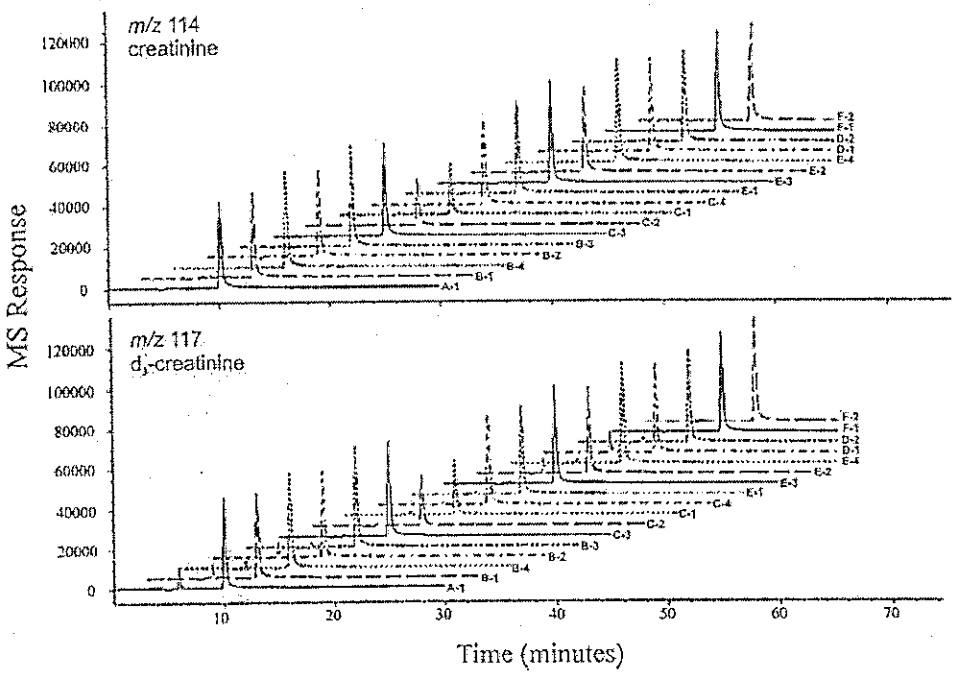

Fig. 2 Example chromatograms from all tested CRMs for mass spectral analytes creatinine at *m/z* 114 and d_3-creatinine at *m/z* 117

Variance components and uncertainty estimation

The practical need for splitting the workload into two campaigns led to the use of the three-level nested measurement design shown in Fig. 1. Two replicate measurements were made on at least two independent aliquots of two different units of each material. To minimize the impact of any between-campaign measurement bias, one unit of each material was analyzed per campaign. This design confounds between-unit and between-campaign variability. Evaluation of the uncertainty of the mean value of these measurements requires consideration of within-unit and between-campaign/unit variability in addition to measurement repeatability. Traditional long-term frequency ("frequentist") estimates for these components and for the standard uncertainty of the measurement mean, $u(R_i)$, are provided in Electronic Supplementary Material Table S2.

However, estimating uncertainties for the R_i that can be uniformly interpreted as similar to the $U_{95}(V_i)$ of the assigned values is more complicated than multiplying $u(R_i)$ by a factor of 2. We investigated two approaches to estimating expanded uncertainties at the 95% level of confidence, $U_{95}(R_i)$: the frequentist method recommended in the JCGM 100:2008 "Guide to the expression of uncertainty in measurement" (GUM) [5] and a Bayesian approach that yields a probability interval interpretable as an uncertainty interval as defined in the JCGM 101:2008 "Evaluation of measurement data—Supplement 1 to the GUM—Propagation of distributions using a Monte Carlo method" [6]. The Bayesian estimates were obtained using a purpose-designed model and the WinBUGS system [7]. This Bayesian approach facilitated incorporating constraints based on prior evidence into the estimates (see Electronic Supplementary Material).

Generalized distance regression (GDR)

A generalized form of the Deming regression method often termed generalized distance regression (GDR) was used to define the linear relationship between the certified value (V_i) and repeatability measurements (R_i) for each material:

$$R_i = \alpha + \beta \times V_i + \varepsilon_i \tag{2}$$

where α is the intercept, β is the slope, and ε_i is the residual random error. GDR has been previously used in metrological applications [8–13] and provides the appropriate parameters by iteratively minimizing:

$$E = \sum_i^N \varepsilon_i^2; \quad \varepsilon_i^2 = \left(\frac{R_i - \widehat{R}_i}{U_{95}(R_i)/2}\right)^2 + \left(\frac{V_i - \widehat{V}_i}{U_{95}(V_i)/2}\right)^2; \quad \widehat{R}_i = \widehat{\alpha} + \widehat{\beta} \times \widehat{V}_i \tag{3}$$

where i indexes the materials, N is the number of materials, and $\widehat{V}_i, \widehat{R}_i, \widehat{\alpha}$, and $\widehat{\beta}$ are the GDR estimates of the assigned values, measured values, intercept, and slope. The parametric bootstrap Monte Carlo (PBMC) technique [14] facilitates computing approximate 95% level of confidence intervals, $\pm U_{95}$, for all of the GDR estimates. A graphically oriented GDR tool developed at NIST was used in this study [13].

"Leave one out" (LOO) cross-validation is routinely used to explore the predictive utility of a modeled relationship [15].

LOO analysis was used to determine if the GDR function was strongly influenced by materials that have relatively small $U_{95}(V_i)$ and/or $U_{95}(R_i)$ or very low or very high $\{V_i, R_i\}$ values. The analysis was accomplished by excluding each individual material in turn from its own evaluation.

Figure 3 displays the overall GDR results with four inset plots providing high-resolution views for selected materials that represent the range of results observed. Materials B-1 and F-2 illustrate U_{95} ellipses typical for this study that are well centered within the 95% level of confidence region of the GDR function. The ellipse for material C-1 is relatively large, reflecting its relatively large $U_{95}(V_i)$. The ellipse for material E-1 is not centered within the 95% level of confidence region of the GDR function and the $U_{95}(V_i)$ is quite small, suggesting that uncertainty assigned to this material in 1996 may be somewhat too small. This material was therefore excluded from the GDR determination. Nonetheless, all 17 materials are compatible with the GDR relationship at the 95% level of confidence.

Degrees of equivalence determination

The degree of equivalence for each material, d_i, is an estimate of the extent of agreement between a reported and a reference value taking into account the uncertainties associated with both. If the interval $d_i \pm U_{95}(d_i)$ contains zero, the reported value is consistent with the reference value at about a 95% level of confidence. For GDR comparisons, the d_i can be estimated as the signed distance between the observed, $\{V_i, R_i\}$, and estimated, $\{\widehat{V}_i, \widehat{R}_i\}$ values. PBMC analysis was used to estimate the $U_{95}(d_i)$.

The resulting degrees of equivalence for each material are displayed in Fig. 4a. All materials are deemed equivalent as their $d_i \pm U_{95}(d_i)$ intervals include zero bias. The median absolute value of the $\%d_i$ is 0.3; the median $U_{95}(d_i)$ is 1.9. For most of the materials, the $U_{95}(d_i)$ is dominated by the certified uncertainty, $U_{95}(V_i)$.

The degree of equivalence for a participating NMI, D, combines the d_i of all materials submitted by that NMI. Each D estimates the expected extent of agreement between the serum creatinine measurement services provided by that NMI relative to those from the other NMIs in the KC. The $U_{95}(D)$ were estimated from the d_i and $U_{95}(d_i)$ estimates. See the Electronic Supplementary Material for further information.

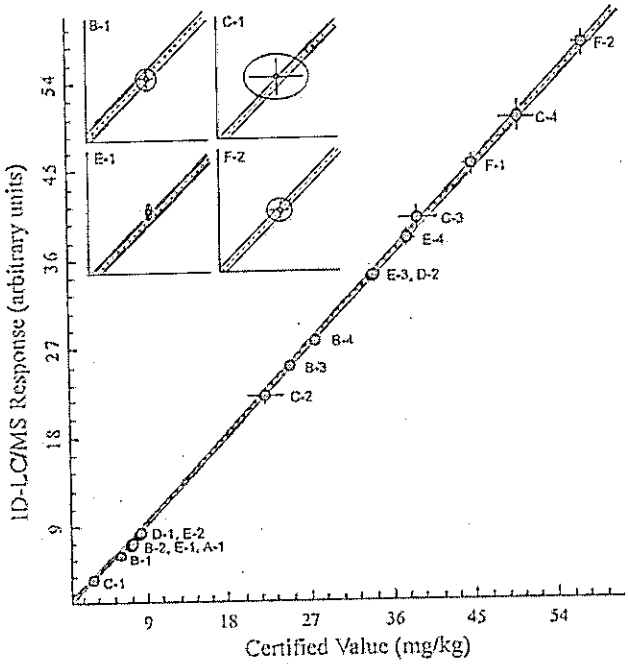

Fig. 3 Generalized distance regression (GDR) relationship (- - - - -) and its 95% level of confidence interval (——). The *black crosses* represent the $\{V_i \pm U_{95}(V_i),\ R_i \pm U_{95}(R_i)\}$ pairs. The four inset plots represent a high-resolution display for selected materials with the ellipses indicating the approximate 95% level of joint confidence region for the materials

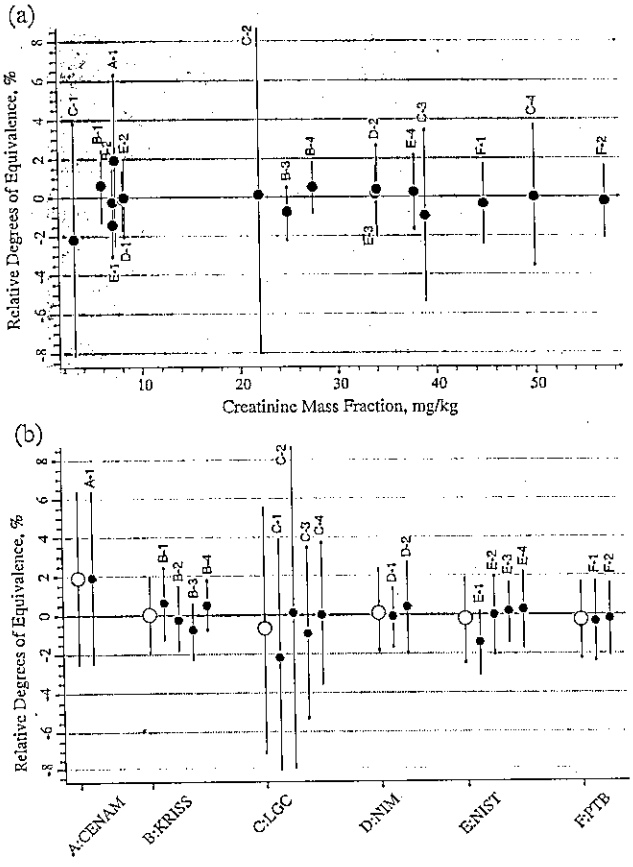

Fig. 4 The relative degree of equivalence and its 95% level of confidence interval for (a) each material and (b) the individual NMI participants and the materials. Each $\%d_i \pm U_{95}(\%d_i)$ is indicated as a *blue circle dot&bar*. In panel (b), the $\%D \pm U_{95}(\%D)$ for each NMI is indicated as *open black circle dot&bar*. The *red line* denotes zero bias relative to the GDR function

The individual participants' $\%D \pm U_{95}(\%D)$ results are displayed in Fig. 4b. The median absolute value of D is 0.3 with a median $U_{95}(D)$ of 2.1. This small increase in the median magnitude of the $U_{95}(D)$ above the $U_{95}(d_i)$ reflects within-NMI variability. Since all of the $\%D \pm U_{95}(\%D)$ intervals include zero, all of the NMIs in the study have demonstrated their ability to appropriately value-assign higher-order creatinine materials.

Conclusion

The CCQM-K80 study was the first OAWG KC to directly assess NMI capabilities as they are delivered to customers. Creatinine in human serum was an attractive candidate for this study due to its global clinical importance and the relatively large number of serum creatinine CRMs and value-assigned PT materials that are currently available. The results of this study establish that the serum creatinine HOVAMs provided by the participating NMIs are equivalent at the 95% level of confidence and demonstrated the metrological equivalence of these materials over a broad range of clinical values (3 mg/kg to 57 mg/kg, ≈0.3 mg/dL to 6 mg/dL), which covers pediatric as well as low and high adult levels [16, 17]. The availability of HOVAMs with IDMS-assigned creatinine values is in line with the NKDEP Creatinine Standardization Program guidelines for providing tools to in vitro diagnostics manufacturers in an effort to reduce analytical bias [3]. Further, the results verified the capabilities of the participating NMIs to deliver services through their HOVAMs, as the certified value for each material corresponded to the measurement response after shipping and storage according to NMI instructions.

Participation in CCQM-K80 required all materials, with the exception of those from NIST, to be shipped internationally. Creatinine in serum demonstrated stability under shipping, storage, and measurement conditions, even in situations where frozen samples had thawed before arrival at NIST. Conditions of materials upon arrival should be well documented as less stable analytes may be significantly affected by shipping conditions.

The study was designed to provide repeatability condition measurements given the practical constraints of the measurement process and the usage conditions specified for the materials. The sample processing parameters accommodated the majority of samples. Deviations from this protocol were performed only when necessary due to specifications provided by the NMI or when the available amount of sample was limited. The consequences of all such deviations were evaluated and found immaterial.

Prepared serum creatinine samples were required to sit in a temperature-controlled autosampler tray for an extended period of time (≈24 h) in order to analyze all the samples in a campaign under repeatability conditions. Less stable analytes may be affected under such conditions. Run order controls and replicate injections were used to enable investigation of sample stability and/or instrumental fluctuations during the extended run time. For future KCs involving value-assigned materials, it will be important to take the measurement protocol and instrumentation constraints into consideration prior to determining the number of materials that can be evaluated.

Acknowledgments The following individuals are acknowledged for their contributions to CCQM-K80: William F. Guthrie, Reenie Parris, Karen W. Phinney, Antonio Possolo, Katherine E. Sharpless, Jolene D. Splett, Melina Pérez Urquiza, Sang-Ryoul Park, Gavin O'Connor, Xinhua Dai, Ruediger Ohlendorf, and André Henrion.

References

1. Stokes P, O'Connor G (2003) Development of a liquid chromatography-mass spectrometry method for the high-accuracy determination of creatinine in serum. J Chrom B 794:125–136
2. Van Nevel L, Ornemark U, Smeyers P, Harper C, Taylor PDP (2002) IMEP-17 report to participants. Part 1 International comparability
3. Myers GL, Miller WG, Coresh J, Fleming J, Greenberg N, Greene T et al (2006) Recommendations for improving serum creatinine measurement: a report from the laboratory working group of the National Kidney Disease Education Program. Clin Chem 52:5–18
4. Welch MJ, Phinney CP, Parris RM, May WE, Heo GS, Henrion A et al (2003) CCQM-K12: the determination of creatinine in serum. Metrologia 40:08005
5. JCGM 100:2008 (2008) Guide to the expression of uncertainty in measurement (GUM). Sèvres, France: BIPM http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf
6. JCGM 101:2008 (2008) Evaluation of measurement data–Supplement 1 to the GUM—Propagation of distributions using a Monte Carlo method. Sèvres, France: BIPM http://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf
7. Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 10:325–337
8. ISO 6143:2001 (2001) Gas analysis–comparison methods for determining and checking the composition of calibration gas mixtures. International Organization for Standardization, Geneva
9. van der Veen AMH, Brinkmann FNC, Arnautovic M, Besley L, Heine HJ, Estban TL et al (2007) International comparison CCQM-P41 Greenhouse gases. 2. Direct comparison of primary standard gas mixtures. Metrologia 44:08003
10. Wielgosz RI, Esler M, Viallon J, Moussay P, Oh SH, Kim BM et al (2008) Final report on CCQM-P73: international comparison of nitrogen monoxide in nitrogen gas standards (30–70) μmol/mol. Metrologia 45:08002
11. van der Veen AMH, Chander H, Ziel PR, Wessel RM, de Leer EWB, Smeulders D et al (2010) International comparison CCQM-K54: primary standard gas mixtures of hexane in methane. Metrologia 47:08019
12. Lee J, Lee JB, Moon DM, Kim JS, van der Veen AMH, Besley L et al (2010) Final report on international key comparison CCQM-K53: oxygen in nitrogen. Metrologia 47:08005

13. Duewer DL, Lippa K, Long S, Murphy K, Sharpless K, Sniegoski L et al (2009) Demonstrating the comparability of certified reference materials. Anal Bioanal Chem 395:155–169
14. Duewer DL, Kowalski BR, Fasching JL (1976) Improving reliability of factor-analysis of chemical data by utilizing measured analytical uncertainty. Anal Chem 48:2002–2010
15. Picard RR, Cook RD (1984) Cross-validation of regression-models. J Am Stat Assoc 79:575–583
16. Finney H, Newman DJ, Thakkar H, Fell JME, Price CP (2000) Reference ranges for plasma cystatin C and creatinine measurements in premature infants, neonates, and older children. Arch Dis Child 82:71–75
17. Ceriotti F, Boyd JC, Klein G, Henny J, Queralto J, Kairisto V et al (2008) Reference intervals for serum creatinine concentrations: assessment of available data for global application. Clin Chem 54:559–566