Creatinine Measurement

State of the Art in Accuracy and Interlaboratory Harmonization

W. Greg Miller, PhD; Gary L. Myers, PhD; Edward R. Ashwood, MD; Anthony A. Killeen, MD, PhD; Edward Wang, PhD; Linda M. Thienpont, PhD; Lothar Siekmann, PhD

• ***Context.*—The National Kidney Disease Education Program recommends calculating glomerular filtration rate from serum creatinine concentration. Accurate creatinine measurements are necessary for this calculation.**

***Objective.*—To evaluate the state of the art in measuring serum creatinine, as well as the ability of a proficiency testing program to measure bias for individual laboratories and method peer groups.**

***Design.*—A fresh-frozen, off-the-clot pooled serum specimen plus 4 conventional specimens were sent to participants in the College of American Pathologists Chemistry Survey for assay of creatinine. Creatinine concentrations were assigned by isotope dilution mass spectrometry reference measurement procedures.**

***Participants.*—Clinical laboratories with an acceptable result for all 5 survey specimens (n = 5624).**

***Results.*—The fresh frozen serum (FFS) specimen had a creatinine concentration of 0.902 mg/dL (79.7 μmol/L). Mean bias for 50 instrument-method peer groups varied from −0.06 to 0.31 mg/dL (−5.3 to 27.4 μmol/L), with 30 (60%) of 50 peer groups having significant bias ($P < .001$). The bias variability was related to instrument manufacturer ($P \leq .001$) rather than method type ($P = .02$) with 24 (63%) of 38 alkaline picric acid methods and with 6 (50%) of 12 enzymatic methods having significant biases. Two conventional specimens had creatinine concentrations of 0.795 and 2.205 mg/dL (70.3 and 194.9 μmol/L) and had apparent survey biases significantly different ($P < .001$) from that of the FFS specimen for 34 (68%) and 35 (70%) of 50 peer groups, respectively.**

***Conclusions.*—Thirty of 50 peer groups had significant bias for creatinine. Bias was primarily associated with instrument manufacturer, not with type of method used. Proficiency testing using a commutable specimen measured participant bias versus a reference measurement procedure and provided trueness surveillance of instrument-method peer groups.**

(*Arch Pathol Lab Med.* 2005;129:297–304)

A long-standing objective of clinical laboratory medicine has been to produce harmonized test results among laboratories and methods.[1] This goal has received renewed emphasis with the development of clinical guidelines for interpretation of laboratory tests that require standardized test results regardless of methodology used for the test procedures. Recent recommendations from the International Federation for Clinical Chemistry[2,3] and the European Union Directive for In Vitro Diagnostic Medical Devices[4,5] have recognized the objective of harmonized results and the importance of commutable reference materials for method calibration traceability and for proficiency testing (PT) programs. Commutable means the result for a PT specimen will have a numeric value that is equivalent to that expected for a clinical specimen containing the same quantity of analyte measured by the same method.

The National Kidney Disease Education Program of the National Institutes of Health (Bethesda, Md) recommends that an estimated glomerular filtration rate (GFR) be calculated from serum creatinine and reported to physicians to assist with early identification of patients with renal disease. The National Kidney Disease Education Program recognizes the importance of standardized creatinine measurements to achieve this goal.[6] Creatinine measurements need to be standardized among all laboratories to enable a common equation to be used for estimating GFR and to allow consistent interpretive guidelines. This report addresses the state of the art for creatinine field method calibration biases and their impact on calculated GFR.

Matrix interferences are generally present in the materials used by PT providers at the present time. Matrix interferences are caused by alterations to the serum composition during the processing steps in material manufacturing. As a result, a PT survey specimen may have a serum matrix with artificial differences not present in native clinical specimens. Consequently, a PT specimen may not be commutable among all the methods used for routine analysis. Literature reports have documented that processed PT materials are frequently noncommutable and the occurrence of noncommutability has been unpredictable for any particular material-method combination.[7]

Accepted for publication September 17, 2004.

From the Virginia Commonwealth University, Richmond (Dr Miller); Centers for Disease Control and Prevention, Atlanta, Ga (Dr Myers); ARUP Laboratories, University of Utah, Salt Lake City (Dr Ashwood); University of Minnesota, Minneapolis (Dr Killeen); College of American Pathologists, Northfield, Ill (Dr Wang); Ghent University, Gent, Belgium (Dr Thienpont); and the Institute of Clinical Biochemistry, University of Bonn, Bonn, Germany (Dr Siekmann).

The authors have no relevant financial interest in the products or companies described in this article.

Reprints: Greg Miller, PhD, Virginia Commonwealth University, PO Box 980286, Richmond, VA 23298-0286 (e-mail gmiller@vcu.edu).

Noncommutability prevents direct comparison of results between methods. When PT results using noncommutable materials are evaluated, the observed difference between an individual laboratory's result and a reference measurement procedure (RMP) target value, between the peer group means for 2 different methods, or between a peer group mean and an RMP target value, has contributions from calibration bias (trueness or accuracy), random bias, and matrix bias. The calibration bias and random bias are test procedure attributes that reflect performance for patient specimens. Matrix bias is the component of the observed difference due to noncommutability between a method–PT material combination. The presence and magnitude of a matrix bias is typically unknown but adds to the sum of calibration and random biases.[8] Consequently, the total observed difference for a PT material between 2 methods can produce an incorrect inference of test procedure performance for native clinical specimens.[7–9]

Currently, most PT specimen materials are not designed to be commutable. The volumes needed and costs associated with manufacturing of the materials have limited the preparation of large quantities of fresh off-the-clot, serum-based materials. However, there has been increasing interest in use of pooled native clinical specimens for commutable materials in PT programs. For example, the College of American Pathologists (CAP) has used native pooled whole blood in its Glycohemoglobin Survey for several years. Some specialized European PT programs have reported successful use of native serum specimens.[9–12]

We report results for creatinine measurement from a PT survey of 5624 laboratories using 50 different instrument-method combinations conducted by the CAP in October 2003. This survey included 1 specimen that was a specially prepared fresh frozen serum (FFS) pool intended to be commutable among all methods and thus able to evaluate the state of the art in harmonization of results. This specimen had creatinine values assigned by higher-order isotope dilution mass spectrometry RMPs to allow evaluation of accuracy for individual laboratories and trueness for method groups.

MATERIALS AND METHODS

Preparation of FFS Pool Used for Specimen C-02

Specimen C-02 was prepared by Aalto Scientific (Carlsbad, Calif) to CAP specifications using a modification of the NCCLS C37-A Guideline.[13] Donor blood was collected into plastic donor bags with no additive, and bags were immersed in an ice water bath during collection. Within 15 minutes of collection, the donor units were centrifuged at 1500*g* for 8 minutes at 4°C to obtain a platelet-rich plasma. Plasma was transferred to a sterile plastic centrifuge bottle and allowed to clot for 4 hours at room temperature. The clotted bottles were centrifuged 18 minutes at 2100*g* at room temperature, and the serum was transferred to another sterile centrifuge bottle and flash frozen in an alcohol–frozen carbon dioxide bath. Serum units were shipped from the donor center to the processing center packed with frozen carbon dioxide and were stored at −70°C up to 2 months prior to pooling.

Frozen serum units that met the following criteria were used: bilirubin, <1.5 mg/dL (25.7 μmol/L); triglyceride, <250 mg/dL (<2.8 mmol/L); fibrin degradation products, <10 μg/mL; and negative assays for syphilis, hepatitis B surface antigen, human immunodeficiency virus antigen(s) (HIV-1 Ag), antibodies to human immunodeficiency virus (anti-HIV-1/2), hepatitis C virus (anti-HCV), and human T-lymphotropic virus types I and II (anti-HTLV I/II). Serum units (n = 670) were thawed 5 hours at room temperature with forced air circulation, incubated 4 hours at 4°C, pooled into 1-L plastic bottles and centrifuged 15 minutes at 2700*g*, pooled and mixed 3.5 hours under argon in a stainless steel container, filtered at 0.22 μm, dispensed into vials, and frozen at −50°C. The serum was maintained at 4°C during pooling and dispensing. The vials were refrozen 27 hours after individual units were thawed. The FFS vials were stored 6 months at −40°C and then bulk shipped with frozen carbon dioxide to the survey packager.

Preparation of Conventional Survey Specimens C-03 and C-04

The General Chemistry Survey specimens were prepared to CAP specifications by Bio-Rad Laboratories (Irving, Calif). They were prepared as frozen liquid serum from defibrinated human plasma. Briefly, frozen plasma collected at donor centers was converted to serum in large batches, which were dialyzed to remove anticoagulants. Various analyte concentrates, including nonhuman components, were added back to the base serum protein material to prepare 2 master pools containing minimum and maximum desired quantities of each analyte. The intermediate concentration survey specimens were prepared by admixture of the 2 master pools to achieve the range of values needed to challenge methods at different concentrations. The General Chemistry Survey specimens included specifications for 58 analytes.

Survey Logistics

For shipment to participants, specimen C-02 (FFS) was included as a regular specimen in the set of C Survey frozen vials. The survey vials were packaged frozen in polystyrene cartons containing a frozen pack intended to allow thawing but maintain cool conditions during transit.

Participants were instructed not to refreeze the specimens, to store them at 2°C to 8°C, to mix the vials by inversion 4 to 5 times, and to perform creatinine assays within 10 days of receipt. Specimen C-02 was not identified as a different preparation from other survey specimens and was handled according to usual practices by participants. Assays were performed in singlicate by participants, and results were reported as mg/dL, typically to 1 decimal place. However, 467 participants, representing all peer groups, reported results with 2 decimals. Participants also provided information on the instruments and methods used. Method stratification options were alkaline picrate with Lloyd reagent or blank for removal of interfering substances, alkaline picrate without Lloyd reagent, kinetic alkaline picrate, rate-blanked compensated kinetic alkaline picrate, enzymatic, enzymatic-amperometric, and enzymatic with ammonium blank.

RMP Value Assignment

Target values were assigned to survey specimens by isotope dilution gas chromatography/mass spectrometry (IDMS) methods. Specimens were shipped to the reference laboratories on solid carbon dioxide and stored at −70°C until assay. Each of 2 laboratories performed duplicate assays on 3 vials of each material. Both IDMS procedures are listed as RMPs in the Joint Committee on Traceability of Laboratory Medicine database provided by the International Bureau of Weights and Measures.

The IDMS RMP was performed at Ghent University (Gent, Belgium), as previously described.[14–16] For internal accuracy and precision control, certified materials were used from the National Institute of Standards and Technology, SRM 909b, and from the German National Proficiency Testing System. The relative expanded uncertainty (coverage factor k = 2) of the RMP was estimated to be 1.5%.

The IDMS RMP was also performed at the Reference Institute of Bioanalysis (Bonn, Germany), as previously described.[17] Matrix control materials, for example, National Institute of Standards and Technology SRM 909 and/or materials from the German National Proficiency Testing System, previously certified by this laboratory, are included in each analytical series. The relative expanded uncertainty (coverage factor k = 2.8) for the measurement of creatinine in specimen C-02 (FFS) was 1.5%. The Reference Institute of Bioanalysis is accredited for creatinine analysis

Table 1. Creatinine Concentrations in Survey Specimens Measured by Isotope Dilution Mass Spectrometry

Specimen*	Mean,† mg/dL‡	SEM, mg/dL	Range, mg/dL
C-02 (FFS)	0.902	0.0030	0.886–0.917
C-03	0.795	0.0018	0.787–0.810
C-04	2.205	0.0049	2.185–2.245

* FFS indicates fresh frozen serum.

† n = 12 for C-02 and C-03, and n = 11 for C-04; data derived from 2 reference measurement laboratories.

‡ The conversion factor for creatinine concentration in μmol/L is mg/dL × 88.40.

as a calibration (reference) laboratory according to ISO 17025. This includes regular comparative measurements (frequency 18 months) with the German National Metrology Institute. The laboratory is entitled to issue calibration certificates according to European co-operation for Accreditation and International Laboratory Accreditation Cooperation.

Statistical Analysis

The survey participant results were screened for error prior to statistical analysis. First, the histograms of the data were visually inspected, and errors that occurred because participants incorrectly completed the reporting form were removed. The data were then subjected to a 2-pass, 3-SD test for outliers. Laboratory data that were greater than 3 SD from their peer group mean on the first and second pass were eliminated. After outlier exclusion, participant results from peer groups with fewer than 10 laboratories were removed.

Bias between survey results and the RMP value was calculated for each participating laboratory. A 1-sample, 2-sided *t* test was applied to test the significance of absolute bias between the RMP and each peer group. To test the bias differences between C-02 (FFS) and conventional PT materials with similar concentration values (C-03 and C-04), a paired *t* test was performed within each peer group to compare the percent bias versus the RMP value for C-02 (FFS) with that of conventional PT specimens. Analysis of variance using a linear mixed model was applied to examine the fixed effects of method and instrument on participant bias versus RMP for specimen C-02 (FFS). All data analyses were performed using SAS for Windows version 8.2 software (SAS Inc, Cary, NC).

RESULTS

Acceptance criteria for manufacturing specimen C-02 (FFS) were within specifications. The results were as follows: ammonia, 235 μg/dL (138 μmol/L); glucose, 97 mg/dL (5.4 mmol/L); albumin, 4.0 g/L; total protein, 7.0 g/L; bilirubin, 0.6 mg/dL (10.3 μmol/L); triglycerides, 78 mg/dL (0.88 mmol/L); free glycerol, less than 0.4 mg/L (4.3 μmol/L); absorbance, 0.05 at 700 nm versus water; microbial colony count, fewer than 10 colony-forming units per milliliter on standard nutrient agar, blood agar, MacConkey agar, and TSI tube; and negative for *Escherichia coli*, *Salmonella*, *Staphylococcus aureus*, and *Pseudomonas aeruginosa*. Homogeneity among 15 vials was 0.6% CV using NCCLS C-37A protocol with sodium as the analyte.

The RMP target values for specimen C-02 (FFS) and for specimens C-03 and C-04 (conventional processed serum pools) are listed in Table 1. Table 2 presents summary statistics for the C-02 (FFS) specimen results. Specimen C-02 was assumed to be commutable among all methods and provided an assessment of instrument-method peer group bias versus true value from the IDMS RMP. Method bias varied from −0.06 to 0.31 mg/dL (−5.3 to 27.4 μmol/L) with 30 (60%) of 50 significantly different from zero ($P < .001$). Alkaline picric acid methods were used by 73% of participants, and 93% of these were kinetic or rate-blanked kinetic methods. The remaining 27% of participants used enzymatic methods. The data include results from 5624 laboratories that had an acceptable result for all 5 Survey specimens and belonged to a peer group with 10 or more laboratories. Results from 443 laboratories were excluded because 1 or more specimens had a result identified as an outlier.

Figure 1 presents the bias and distribution of participant peer group mean values versus the IDMS RMP value for specimen C-02 (FFS). The figure suggests peer group bias was not related to type of reaction chemistry. Endpoint alkaline picrate, kinetic alkaline picrate, rate-blanked kinetic alkaline picrate, and enzymatic procedures each had biases that ranged from 0.00 to 0.31 mg/dL (0.0 to 27.4 μmol/L), −0.06 to 0.20 mg/dL (−5.3 to 17.7 μmol/L), 0.00 to 0.11 (0.0 to 9.7 μmol/L), and −0.02 to 0.20 mg/dL (−1.8 to 17.7 μmol/L), respectively.

The peer group bias appears to be related to the instrument manufacturer. Five manufacturers accounted for 96.5% of individual laboratory results in this survey: Dade, 26%; Beckman, 24%; Ortho, 22%; Roche, 19%; and Olympus, 5%. Four Dade peer groups had biases of 0.04 to 0.07 mg/dL (3.5–6.2 μmol/L; 3 of 4, $P < .001$); 8 Beckman CX series peer groups had biases between 0.10 and 0.16 mg/dL (8.8 and 14.1 μmol/L; all $P < .001$), with 3 LX-20 groups having biases of 0.05 to 0.06 mg/dL (4.4–5.3 μmol/L; 2 of 3, $P < .001$); 4 Ortho peer groups had biases of 0.10 to 0.12 mg/dL (8.8–10.6 μmol/L), with a 0.20-mg/dL (17.7-μmol/L) bias for the DT-60 (all $P < .001$); 16 Roche peer groups had biases between −0.03 and 0.02 mg/dL (−2.7 and 1.8 μmol/L; 1 of 16, $P < .001$), with 4 groups having biases of −0.06 mg/dL (5.3 μmol/L), 0.07 mg/dL (6.2 μmol/L), 0.09 mg/dL (8.0 μmol/L), and 0.18 mg/dL (15.9 μmol/L) (all $P < .001$); 3 Olympus peer groups had biases between 0.09 and 0.12 mg/dL (8.0 and 10.6 μmol/L; 2 of 3, $P < .001$).

Table 3 presents an analysis of variance of the fixed effects of method and instrument on participant bias versus RMP for specimen C-02 (FFS). The *P* value for instrument effect is much more significant than reagent effect, suggesting that instrument manufacturer accounted for most of the bias variations.

Two additional conventionally prepared survey specimens, C-03 and C-04, were assayed by IDMS, which allowed determination of the impact of matrix bias on the apparent method bias. Figure 2 shows the observed survey bias (peer group mean bias vs IDMS) for specimens C-03 and C-04 compared to the trueness bias determined from specimen C-02 (FFS). Specimen C-03 had a creatinine concentration of 0.795 mg/dL (70.3 μmol/L), which was close to that of C-02 (FFS), and it was assumed each peer group had essentially the same calibration status as determined by the C-02 (FFS) specimen. Specimen C-04 had a creatinine concentration of 2.205 mg/dL (194.9 μmol/L), which was approximately 2.5 times the value for C-02 (FFS), and the calibration bias determined for a peer group at 0.902 mg/dL (79.7 μmol/L) may not be the same at this higher concentration. The mean of percent bias for individual participants was used to compare the survey bias for specimens C-03 or C-04 to the trueness bias from specimen C-02 (FFS) to adjust for the different concentrations in the specimens. The observed survey biases versus the

Table 2. Summary Statistics for Specimen C-02 (Fresh Frozen Serum) by Survey Peer Group

Instrument*	Method†	n	Mean, mg/dL‡	SD, mg/dL‡	CV, %‡	Bias, mg/dL§	SEM Bias, mg/dL
Abbott Aeroset	AP-EP	14	1.04	0.056	5.40	0.14‖	0.015
	AP-K	89	1.04	0.059	5.69	0.14‖	0.006
Bayer Advia 1650	AP-K	47	1.10	0.035	3.14	0.20‖	0.005
Beckman Synchron CX3-7D, CX9ALX	AP-EP	59	1.01	0.048	4.76	0.11‖	0.006
	AP-K	373	1.01	0.043	4.28	0.10‖	0.002
	AP-RB	24	1.02	0.048	4.74	0.11‖	0.010
Beckman Synchron CX4/5CE, 7/RTS	AP-EP	30	1.03	0.048	4.68	0.12‖	0.009
	AP-K	156	1.04	0.067	6.45	0.13‖	0.005
Beckman Synchron CX3	AP-EP	14	1.01	0.053	5.27	0.11‖	0.014
	AP-K	29	1.04	0.063	6.05	0.14‖	0.012
Beckman Synchron CX4/5	AP-K	18	1.07	0.059	5.57	0.16‖	0.014
Beckman Synchron LX20	AP-EP	48	0.96	0.049	5.08	0.06‖	0.007
	AP-K	589	0.96	0.049	5.10	0.06‖	0.002
	AP-RB	13	0.95	0.052	5.44	0.05	0.014
Dade Behring Dimension AR	AP-EP	24	0.97	0.070	7.26	0.06‖	0.014
	AP-EP	43	0.97	0.075	7.72	0.07‖	0.011
	AP-K	1396	0.96	0.075	7.79	0.06‖	0.002
	AP-RB	18	0.94	0.070	7.46	0.04	0.017
Nova, CRT Series	E	14	0.91	0.114	12.58	0.01	0.030
Olympus 400-640/2700/5400	AP-EP	12	1.01	0.087	8.63	0.11	0.025
	AP-K	228	0.99	0.029	2.88	0.09‖	0.002
Olympus AU 5200	AP-K	26	1.02	0.065	6.37	0.12‖	0.013
Roche Cobas FARA/MIRA	AP-K	50	1.09	0.099	9.08	0.18‖	0.014
Roche Cobas Integra	AP-EP	15	0.91	0.046	5.05	0.00	0.012
	AP-K	88	0.89	0.039	4.38	−0.01	−0.004
	AP-RB	268	0.91	0.098	10.76	0.01	0.006
	E	18	0.91	0.103	11.29	0.01	0.024
Roche Modular	AP-K	78	0.88	0.034	3.87	−0.02‖	−0.004
	AP-RB	64	0.92	0.084	9.13	0.01	0.010
	E	77	0.90	0.055	6.09	0.00	0.006
Roche/Hitachi 717	AP-K	23	0.99	0.097	9.80	0.09‖	0.020
Roche/Hitachi 747	AP-K	23	0.88	0.028	3.14	−0.02	−0.006
	AP-RB	43	0.98	0.097	9.98	0.07‖	0.015
	E	19	0.90	0.048	5.36	0.00	0.011
Roche/Hitachi 911	AP-K	12	0.90	0.035	3.84	0.00	−0.010
	AP-RB	60	0.87	0.087	10.05	−0.03	−0.011
	E	15	0.91	0.131	14.32	0.01	0.034
Roche/Hitachi 912	AP-K	45	0.84	0.074	8.77	−0.06‖	−0.011
	AP-RB	16	0.93	0.105	11.32	0.02	0.026
Roche/Hitachi 917	AP-K	36	0.89	0.028	3.14	−0.01	−0.005
	AP-RB	57	0.91	0.096	10.60	0.00	0.013
	E	75	0.92	0.061	6.68	0.02	0.007
Schiaparelli Ace	AP-EP	14	1.01	0.073	7.25	0.11‖	0.020
Toshiba TBA-FR Series	AP-EP	11	1.21	0.081	6.65	0.31‖	0.024
	E	10	0.89	0.045	5.02	−0.01	−0.014
Vitros 250	E	554	1.00	0.002	0.24	0.10‖	0.000
Vitros 400, 700	E	71	1.02	0.037	3.61	0.12‖	0.004
Vitros 500, 550	E	33	1.01	0.034	3.33	0.11‖	0.006
Vitros 950	E	572	1.00	0.001	0.10	0.10‖	0.000
Vitros DT60II	E	13	1.10	0.041	3.71	0.20‖	0.011

* Abbott Laboratories, Chicago, Ill; Bayer Healthcare, Diagnostics Division, Norwood, Mass; Beckman Coulter, Fullerton, Calif; Dade Behring, Deerfield, Ill; Nova Biomedical, Waltham, Mass; Olympus America, Melville, NY; Roche Diagnostics, Indianapolis, Ind; Schiaparelli Biosystems, Fairfield, NJ; Toshiba, Tustin, Calif; and Ortho-Clinical Diagnostics, Inc, Rochester, NY.

† AP-EP indicates alkaline picrate end point; AP-K, alkaline picrate kinetic; AP-RB, alkaline picrate rate-blanked compensated kinetic; and E, enzymatic.

‡ Mean, SD, and CV of peer group participants' individual results. Conversion factor for creatinine in μmol/L is 88.4.

§ Mean of individual participant's bias versus IDMS value (0.902 mg/dL); ‖ indicates mean bias was significantly different ($P < .001$) from zero.

IDMS values for the conventionally prepared materials C-03 and C-04 were statistically different ($P < .001$) from the trueness bias for the commutable C-02 (FFS) specimen in 34 (68%) of 50 and 35 (70%) of 50 peer groups, respectively.

The dispersion of participant results within a peer group was measured as the SD for the peer group mean value. Specimens C-02 (FFS) and C-03 had creatinine concentrations close to each other (0.902 and 0.795 mg/dL [79.7 and 70.3 μmol/L], respectively), which permitted the variance for each specimen to be compared by an F test. The ranges of SDs for C-02 and C-03 were 0.004 to 0.131 mg/dL (0.4–11.6 μmol/L) and 0.003 to 0.210 mg/dL (0.3–18.6 μmol/L), respectively. There was no significant difference ($P > .001$) in the variance between C-02 (FFS) and C-03 for 41 (82%) of 50 peer groups. Where differences were significant, 4 Vitros peer groups had smaller SDs for C-02 (range, 0.001–0.036 mg/dL [0.09–3.18 μmol/L]) than for C-03 (range, 0.040–0.073 mg/dL [3.5–6.5 μmol/L]); and 5 peer groups (Bayer Advia 1650, Beckman Synchron

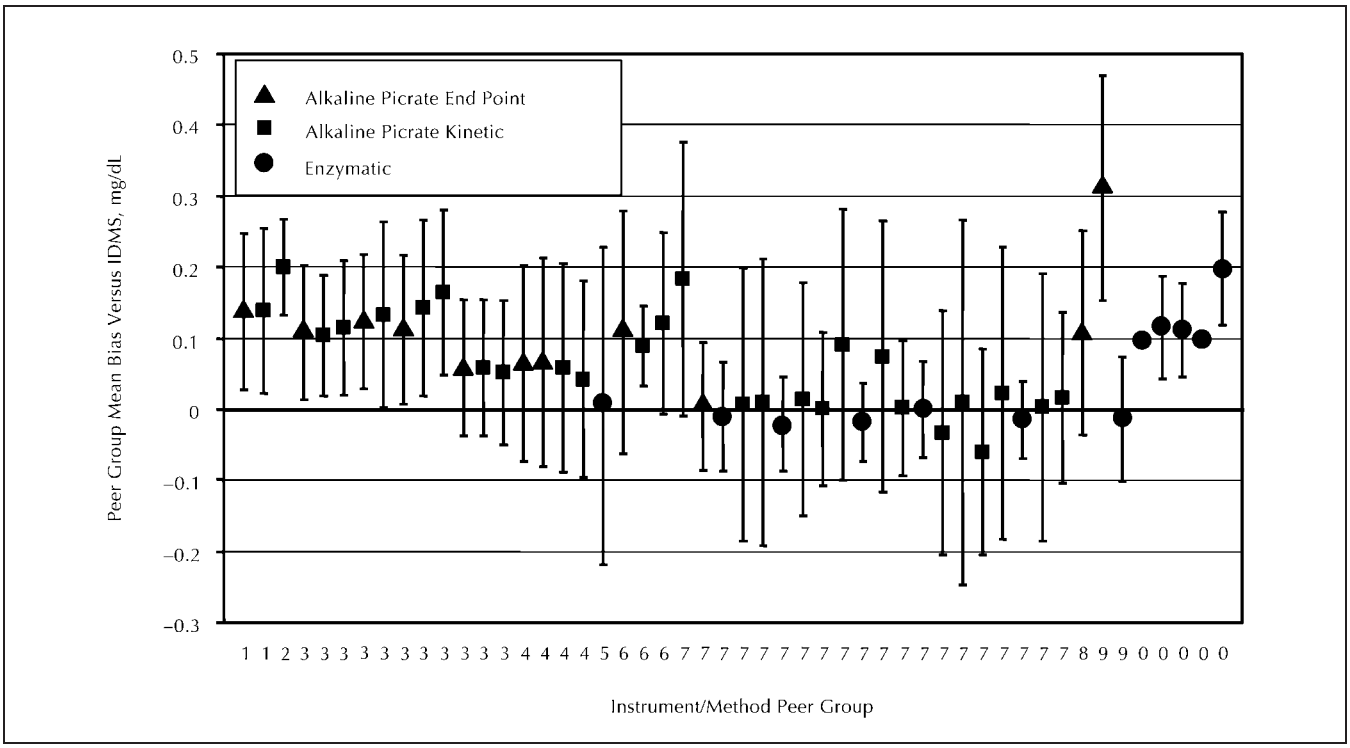

Figure 1. *Mean bias by peer group versus isotope dilution gas chromatography mass spectrometry (IDMS) reference measurement procedure for specimen C-02 (fresh frozen serum). Error bars indicate 1.96 × SD for distribution of participant results. The error bars that appear missing are smaller than the plot symbol. The numbers on the horizontal axis identify the instrument manufacturer and are in the same sequence as Table 2: 1, is Abbott; 2, Bayer; 3, Beckman Coulter; 4, Dade Behring; 5, Nova; 6, Olympus; 7, Roche; 8, Schiapparelli; 9, Toshiba; and 0, Vitros.*

Table 3. Analysis of Variance

Source of Variation	*df*	Sum of Squares	Mean Square	F	*P*
Method	3	0.0401	0.0137	3.72	.02
Instrument	10	0.1047	0.0105	5.36	<.001
Error	36	0.0703	0.0019		
Total	**49**	**0.2151**			

LX20 with 2 reagent types, Olympus 400-640/2700/5400 with kinetic alkaline picrate reagent, and Roche Cobas Fara/Mira) had larger SDs for C-02 (range, 0.029–0.099 mg/dL [2.6–8.6 μmol/L]) than for C-03 (range, 0.003–0.062 mg/dL [0.3–5.5 μmol/L]). These observations suggest the FFS and conventional materials gave generally comparable measures of imprecision in each peer group.

COMMENT

Trueness for a measurement is the agreement between replicate measurements of a sample and the numeric value assigned to that sample by an RMP. Trueness is typically expressed as mean systematic bias because the replication reduces random bias (imprecision) to a low level. The related term *accuracy* is the agreement between an individual measurement on a sample and a value assigned by an RMP with known uncertainty. Accuracy for individual patient specimens includes contributions from both systematic and random bias.

Trueness is an attribute that can be evaluated from a large PT survey when the PT material is commutable among all the methods used by participants in the Survey and the RMP. The C-02 specimen was prepared by a protocol that produced a nonadulterated FFS pool, which was expected to be free of matrix interferences and commutable among the RMP and all routine methods for creatinine reported in the PT survey.

Trueness of Routine Methods

There are several types of chemical methods used to measure creatinine in routine clinical laboratories. Enzymatic methods were reported by 27% of participants. The alkaline picric acid (Jaffe) method was used by 73% of participants and represented 76% of the method peer groups reported. The survey allowed participants to categorize the alkaline picrate reaction into methods with or without use of Lloyd pretreatment, kinetic methods, and rate-blanked kinetic methods. As reported by participants, 93% of laboratories using alkaline picrate used a kinetic method (with or without rate blanking). Review of Table 2 suggests there may be misclassification of reaction type by participants who are not clear what type of alkaline picrate reaction was supplied by manufacturers. For example, users of the Dade Dimension series reported all 4 reaction types, although only the kinetic alkaline picrate reaction is used by the manufacturer in this ''closed'' system. This type of misclassification causes a greater number of peer groups than is appropriate. Correct method classification is necessary if the survey is to provide useful feedback to manufacturers regarding method trueness.

Trueness bias was statistically significant for 60% of the method peer groups listed in Table 2. Anomalies in statistical significance in a *t* test can be caused by the magnitude of variability in individual results, indicated by the SD, and by the number of results in the group. For ex-

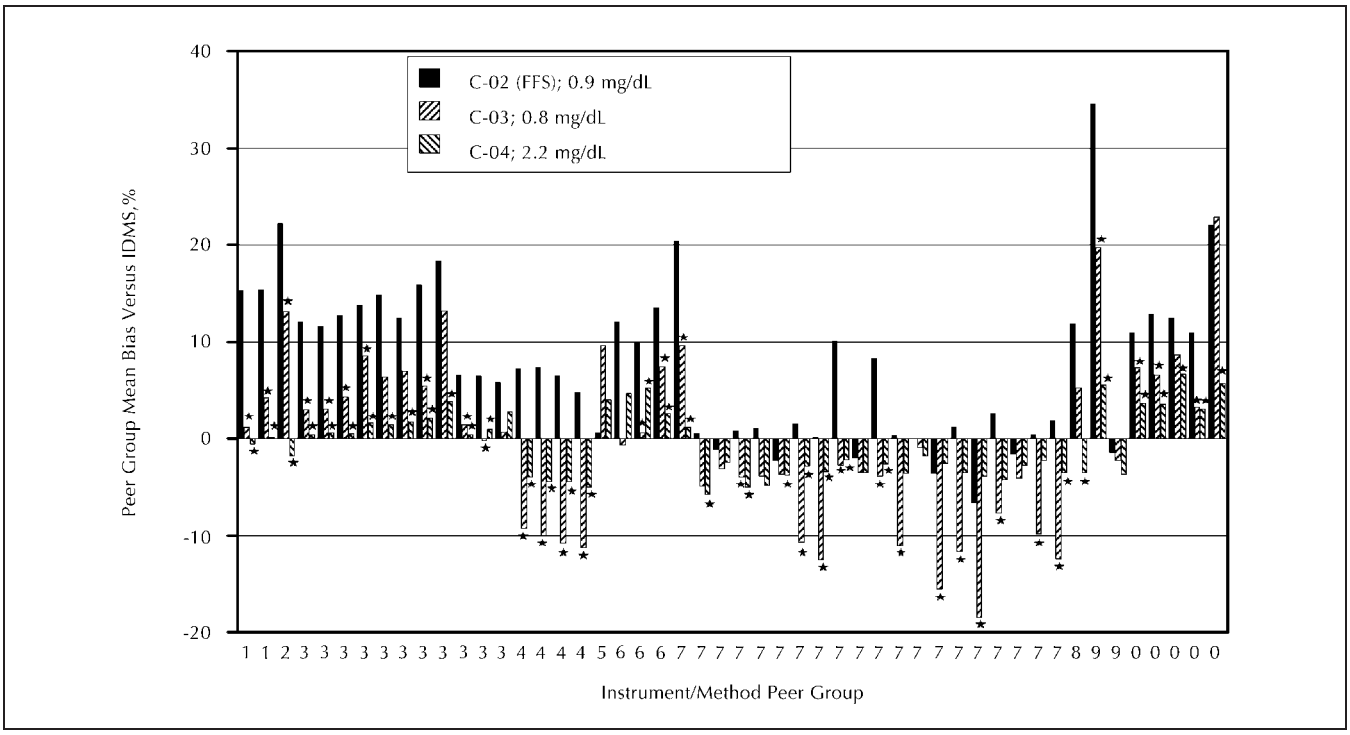

Figure 2. *Commutability of conventional specimens, C-03 (creatinine, 0.8 mg/dL [70.7 μmol/L]) and C-04 (creatinine, 2.2 mg/dL [194.4 μmol/L]) with fresh frozen serum specimen C-02 (creatinine, 0.9 mg/dL [79.6 μmol/L]). Asterisk indicates the bias for C-03 or C-04 was significantly different (*$P < .001$*) from the bias for C-02. The numbers on the horizontal axis are as in Figure 1. IDMS indicates isotope dilution gas chromatography mass spectrometry.*

ample, the Roche Modular alkaline picrate kinetic group had −0.02 mg/dL (1.8 μmol/L) bias, which had a *P* value less than .001, because variability of the individual results was small (SD = 0.034 mg/dL [3.0 μmol/L]) and the number of results was large (n = 78), giving a small SEM for the bias (−0.004 mg/dL [−0.4 μmol/L]). In this example, while the statistical significance was correct, the method bias would not be an issue for clinical interpretation. The opposite can also occur; for example, the Olympus 400-640/2700/5400 alkaline picrate end point group had a 0.11 mg/dL (9.7 μmol/L) bias, which was not statistically significant because variability of the individual results was large (SD = 0.087 mg/dL [7.7 μmol/L]) and the number of results was small (n = 12), giving a relatively large SEM for the bias (0.025 mg/dL [2.2 μmol/L]). In the Olympus example, the 0.11 mg/dL (9.7 μmol/L) bias in creatinine would produce a clinically important bias in GFR calculated from serum creatinine.

A possible contributor to the systematic bias was nonspecificity of the affected methods for 1 or more components in the fresh frozen pooled serum. The FFS was pooled from 670 donor serum units, which argues that any unit that may have contributed an interfering substance would have represented 0.15% of the pool, thus making it unlikely that an interfering concentration would remain. Previous reports have supported that creation of pooled sera from a large number of healthy donors minimizes the impact of any potentially interfering substances.[13,18] The alkaline picrate method is known to have a number of nonspecificities, which vary with the details of method implementation.[19,20] Common interfering agents include ketone bodies, glucose, protein, bilirubin, ascorbate, and cephalosporin antibiotics.[20] Young's compendium includes numerous drug interferences with alkaline picrate and, with lesser frequency, enzymatic methods.[21] Even so, the large number of serum units in the pool makes it unlikely any abnormal quantity of a physiologic substance or drug was present.

The observed trueness bias is more likely to be a miscalibration than a method nonspecificity effect based on the highly significant instrument effect compared to method effect for analysis of variance results in Table 3. Examination of Table 2 and Figure 1 shows that both small and large biases are associated with all method types, both alkaline picrate and enzymatic. However, peer group bias appears clustered with instrument manufacturer. For example, the 20 Roche peer groups represent 8 instruments, and 5 of these used versions of both enzymatic and kinetic alkaline picrate reagents. For these 5 instruments (16 peer groups), all had biases between −0.03 and 0.07 mg/dL (−2.7 and 6.2 μmol/L). The 8 Beckman CX series peer groups all used alkaline picrate reagents (kinetic, rate-blanked kinetic, and end point) and had biases from 0.10 to 0.16 mg/dL (8.8 to 14.1 μmol/L), while the 3 LX series peer groups used the same range of reagents and had biases from 0.05 to 0.06 mg/dL (4.4 to 5.3 μmol/L). Four Vitros peer groups used identical enzymatic reagents (the DT60 uses a different enzyme reaction), and all had biases from 0.10 to 0.12 mg/dL (8.8 to 10.6 μmol/L).

The data available from this survey suggest the trueness bias can be substantially corrected by recalibration by method manufacturers. Since 96.5% of the participants in this survey used instruments from 5 manufacturers, availability of a metrologically traceable reference system

would allow rapid standardization of creatinine. Standardized calibration would minimize the impact of trueness bias on the clinical usefulness of calculating GFR from serum creatinine. Many instruments are ''open,'' meaning they can use reagents and calibrators from third-party manufacturers. Consequently, less than 96.5% of participants would be impacted by the 5 most common instrument manufacturers. Adoption of a standardized calibration scheme for creatinine by both instrument and reagent manufacturers is necessary to improve the overall state of the art.

Standardized calibration is a critical component of trueness, but would not address nonspecificity issues that may be present in some methods. It is possible for a method to be calibrated to compensate for ''average'' nonspecificity bias from normally occurring substances in clinical specimens, but individual patients with various pathologic conditions could still have nonspecific interferences. Thus, attempting to compensate for average nonspecificity would not achieve trueness for all categories of patients nor for patients taking therapeutic drugs for which interfering components may contribute to a method-dependent bias.

Suitability of Survey Materials to Evaluate Trueness

The observed bias for peer groups using conventionally prepared survey materials was frequently different than the trueness bias observed from the commutable C-02 (FFS) specimen. Review of Figure 2 shows overall for C-03 and/or C-04 that 9 (38%) of 24 enzymatic and 60 (79%) of 76 alkaline picrate methods had results noncommutable with those for C-02 (FFS). This observation supports the existence of matrix-related bias in the conventional materials. It is not possible to resolve if the matrix-related differences in bias are caused by alterations in the survey materials due to manufacturing processes or due to nonspecificity of the field methods for substances present in these materials that could be present in some pathologic sera. While not conclusive, it appeared enzymatic methods were less frequently affected by matrix-related interferences than the conventional survey specimens. In any event, the overall 69% disagreement between the commutable C-02 (FFS) results and the conventional survey materials results precluded use of the conventional materials for field assessment of method trueness bias or for assessment of accuracy for individual participant results. A commutable material, such as the FFS pool used for specimen C-02, is necessary for a PT survey to be useful for surveillance of field method trueness and individual participant accuracy.

Impact of Creatinine Bias on Calculated GFR

Using the 4-parameter equation from the Modification of Diet in Renal Disease (MDRD) recommended by the National Kidney Disease Education Program,[22] a 0.1-mg/dL (8.8-μmol/L) change in creatinine from 1.0 to 1.1 mg/dL (88.4 to 97.2 μmol/L) for a 60-year-old non–African American woman causes a 10% change in calculated GFR from 60 to 54 mL/min/1.73 m^2. In actual laboratory practice, the variability among laboratories in a peer group, expressed as the SD, adds to the systematic bias and produces a greater uncertainty in the calculated GFR. The midpoint SD for all peer groups in Table 2 was 0.06 mg/dL (5.3 μmol/L). Combining a calibration bias of 0.1 mg/dL (8.8 μmol/L) with 1.96 × 0.06 mg/dL SD, gives a 95% confidence limit for the total error of 0.22 mg/dL (19.4 μmol/L). At this total error limit, the calculated GFR is 48 mL/min/1.73 m^2, which is a 20% error. For the largest interlaboratory SD observed, 0.131 mg/dL (11.6 μmol/L), and the same calibration bias, the calculated GFR would have a 30% error for the 95% confidence limit.

If the clinically acceptable maximum creatinine measurement contribution to the total error for calculated GFR were 15%, and a method had an interlaboratory SD of 0.06 mg/dL (5.3 μmol/L), the maximum allowable calibration bias would be 0.034 mg/dL (3.0 μmol/L) at a creatinine level of 1.0 mg/dL (88.4 μmol/L). Applying this clinical criterion for maximum allowable bias and SD to the data in Table 2 identifies 9 (18%) of 50 peer groups whose performance met that goal.

The authors appreciate the efforts of Sharon Burr of the College of American Pathologists, Northfield, Ill, who coordinated the preparation of the fresh frozen serum specimen, its inclusion in the Survey, and the reference measurement procedure value assignment.

References

1. Miller WG, Kauffman HW. Matrix effects and accuracy assessment in clinical chemistry. *Arch Pathol Lab Med.* 1993;117:343–436.

2. Libeer JC, Baadenhuijsen H, Fraser CG, et al. Characterization and classification of external quality assessment schemes (EQA) according to objectives such as evaluation of method and participant bias and standard deviation. *Eur J Clin Chem Clin Biochem.* 1996;34:665–678.

3. Thienpont L, Franzini C, Kratochvila J, et al. Analytical quality specifications for reference methods and operating specifications for networks of reference laboratories. *Eur J Clin Chem Clin Biochem.* 1995;33:949–957.

4. Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on *in vitro* diagnostic medical devices. *Official Journal of the European Communities* L 331. December 7, 1998:1–37.

5. ISO 17511:2003. In vitro diagnostic medical devices: measurement of quantities in biological samples: metrological traceability of values assigned to calibrators and control materials. International Organization for Standardization. Available at: http://www.iso.org/. Accessed August 11, 2004.

6. Coresh J, Astor BC, McQuillan G, et al. Calibration and random variation of the serum creatinine assay as critical elements of using equations to estimate glomerular filtration rate. *Am J Kidney Dis.* 2002;39:920–929.

7. Miller WG. Specimen materials, target values and commutability for external quality assessment (proficiency testing) schemes. *Clin Chim Acta.* 2003;327:25–37.

8. Ross JW, Miller WG, Myers GL, Praestgaard J. The accuracy of laboratory measurements in clinical chemistry: a study of 11 routine chemistry analytes in the College of American Pathologists Chemistry Survey with fresh frozen serum, definitive methods, and reference methods. *Arch Pathol Lab Med.* 1998;122:587–608.

9. Thienpont LM, Stockl D, Friedecky B, Kratochvila J, Budina M. Trueness verification in European external quality assessment schemes: time to care about the quality of the samples. *Scand J Clin Lab Invest.* 2003;63:195–202.

10. Cobbaert C, Weykamp C, Baadenhuijsen H, Kuypers A, Lindemans J, Jansen R. Selection, preparation, and characterization of commutable frozen human serum pools as potential secondary reference materials for lipid and apolipoprotein measurements: study within the framework of the Dutch project ''Calibration 2000.'' *Clin Chem.* 2002;48:1526–1538.

11. Stockl D, Libeer JC, Reinauer H, Thienpont LM, De Leenheer AP. Accuracy-based assessment of proficiency testing results with serum from single donation: possibilities and limitations. *Clin Chem.* 1996;42:469–470.

12. Linko S, Himberg JJ, Thienpont L, Stockl D, De Leenheer A. Assessment of the state-of-the-art trueness and precision of serum total-calcium and glucose measurements in Finnish laboratories: the QSL-Finland study. *Scand J Clin Lab Invest.* 1998;58:229–240.

13. *Preparation and Validation of Commutable Frozen Human Serum Pools as Secondary Reference Materials for Cholesterol Measurement Procedures: Approved Guideline.* NCCLS document C37-A (ISBN 1-56238-392-2). Wayne, Pa: NCCLS; 1999.

14. Stöckl D, Reinauer H. Candidate reference methods for the determination of target values for cholesterol, creatinine, uric acid, and glucose in external quality assessment and internal accuracy control, I: method setup. *Clin Chem.* 1993;39:993–1000.

15. Thienpont LM, De Leenheer AP, Stöckl D, Reinauer H. Candidate reference methods for determining target values for cholesterol, creatinine, uric acid, and glucose in external quality assessment and internal accuracy control, II: method transfer. *Clin Chem.* 1993;39:1001–1006.

16. Thienpont LM, Van Nieuwenhove B, Stöckl D, Reinauer H, De Leenheer AP. Determination of reference method values by isotope dilution–gas chroma-

tography/mass spectrometry: a five years' experience of two European reference laboratories. *Eur J Clin Chem Clin Biochem*. 1996;34:853–860.

17. Siekmann L. Measurement of creatinine in human serum by isotope dilution mass spectrometry. *J Clin Chem Clin Biochem*. 1985;23:137–144.

18. Hoelzel W, Weykamp C, Jeppsson J-O, et al. IFCC reference system for measurement of hemoglobin A1c in human blood and the national standardization schemes in the United States, Japan, and Sweden: a method-comparison study. *Clin Chem*. 2004;50:166–174.

19. Newman DJ, Price CP. Renal function and metabolites, creatinine and creatine. In: Burtis CA, Ashwood ER, eds. *Tietz Textbook of Clinical Chemistry*. 3rd ed. Philadelphia, Pa: Saunders; 1999:1241–1245.

20. Spencer K. Analytical reviews in clinical biochemistry: the estimation of creatinine. *Ann Clin Biochem*. 1986;23:1–25.

21. Young DS. *Effects of Drugs on Clinical Laboratory Tests*. 5th ed. Washington, DC: AACC Press; 2000.

22. National Kidney Disease Education Program Web site. Information for health professionals: laboratory support: suggestions for laboratories. Available at: http://www.nkdep.nih.gov/. Accessed August 11, 2004.