Imprecision of Urinary Iothalamate Clearance as a Gold-Standard Measure of GFR Decreases the Diagnostic Accuracy of Kidney Function Estimating Equations

Yuen-Ting (Diana) Kwong, BA,[1] Lesley A. Stevens, MD, MS,[2] Elizabeth Selvin, PhD, MPH,[1] Yaping (Lucy) Zhang, MS,[2] Tom Greene, PhD,[3] Frederick Van Lente, PhD,[4] Andrew S. Levey, MD,[2] and Josef Coresh, MD, PhD, MHS[1]

Background: Evaluating the accuracy of estimated glomerular filtration rate (eGFR) derived from serum creatinine (SCr) and serum cystatin C (SCysC) equations requires gold-standard measures of GFR. However, the influence of imprecise measured GFRs (mGFRs) on estimates of equation error is unknown.

Study Design: Diagnostic test study.

Setting & Participants: 1,995 participants from the Modification of Diet in Renal Disease (MDRD) Study and African American Study of Kidney Disease and Hypertension (AASK) with at least 2 baseline mGFRs from iodine 125–iothalamate urinary clearances, 1 standardized SCr value, and 1 SCysC value.

Index Tests: eGFRs calculated using the 4-variable isotope-dilution mass spectrometry (IDMS)-traceable MDRD Study equation, the Chronic Kidney Disease (CKD) Epidemiology Collaboration (CKD-EPI) SCysC equation, the CKD-EPI SCr-SCysC equation, and mGFRs collected from another prerandomization visit.

Reference Tests: A single reference mGFR, average of 2, and average of 3 mGFRs; additional analysis limited to consistent mGFRs (difference ≤25% from reference mGFR).

Results: We found that mGFRs had stable mean values, but substantial variability across visits. Of all mGFRs collected a mean of 62 days apart from the reference visit, 8.0% were outside 30% of the single reference mGFR ($1 - P_{30}$). Estimation equations were less accurate because 12.1%, 17.1%, and 8.3% of eGFRs from the MDRD Study, CKD-EPI SCysC, and CKD-EPI SCr-SCysC equations were outside 30% of the same gold standard ($1 - P_{30}$). However, improving the precision of the reference test from a single mGFR to the average of 3 consistent mGFRs decreased these error estimates ($1 - P_{30}$) to 8.0%, 12.5%, and 3.9%, respectively.

Limitations: Study population limited to those with CKD.

Conclusions: Imprecision in gold-standard measures of GFR contribute to an appreciable proportion of the cases in which eGFR and mGFR differ by >30%. Reducing and quantifying errors in gold-standard measurements of GFR is critical to fully estimating the accuracy of GFR estimates.

Am J Kidney Dis 56:39-49.

INDEX WORDS: Gold standard; measured glomerular filtration rate; kidney function–estimation equations; cystatin C; creatinine.

Accurate quantification of kidney function is central to the diagnosis, management, and research of chronic kidney disease (CKD). Glomerular filtration rate (GFR), which can be measured (mGFR) or estimated (eGFR), assesses one of the most important aspects of kidney function.[1]

mGFR computed from the clearance of injected exogenous markers (eg, iodine 125 [^{125}I]-iothalamate) is associated with little bias and usually is used as the gold standard for GFR.[2]

However, often it is not appreciated that mGFR varies substantially because of measurement error and physiologic day-to-day variations in kid-

From the [1]Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, MD; [2]Division of Nephrology, Tufts Medical Center, Boston, MA; [3]Division of Epidemiology, University of Utah, Salt Lake City, UT; and [4]Department of Clinical Pathology, Cleveland Clinic Foundation, Cleveland, OH.

Received October 18, 2009. Accepted in revised form February 10, 2010. Originally published online as doi:10.1053/j.ajkd.2010.02.347 on May 31, 2010.

Because an author of this manuscript is an editor for AJKD, the peer-review and decision-making processes were handled entirely by an Associate Editor (Kunitoshi Iseki, MD, University Hospital of The Ryukyus) who served as Acting Editor-in-Chief. Details of the journal's procedures for potential editor conflicts are given in the Editorial Policies section of the AJKD website.

Address correspondence to Josef Coresh, MD, PhD, MHS, Johns Hopkins University, 2024 E Monument St, Ste 630, Baltimore, MD 21278. E-mail: coresh@jhu.edu

0272-6386/10/5601-0009$36.00/0
doi:10.1053/j.ajkd.2010.02.347

ney function. Furthermore, the impact of this variability is largely unstudied.

In clinical practice, GFR usually is not measured directly because of the cost, invasiveness, and possible radioactive exposure associated with mGFR procedures. Alternatively, eGFR is computed from serum concentrations of endogenous markers, such as serum creatinine (SCr) and serum cystatin C (SCysC).[3,4] Although estimating GFR is less expensive, less cumbersome, and more practical for clinical use than measuring GFR, endogenous filtration markers are affected by factors other than GFR (non-GFR determinants).[5] Equations have been developed to calculate eGFR by regressing mGFR on SCr level and demographic information, which acts as a surrogate for some non-GFR determinants. For example, the Modification of Diet in Renal Disease (MDRD) Study equation includes age, sex, and race as indicators of muscle mass to provide relatively unbiased estimates, especially in populations with eGFR <60 mL/min/1.73 m^2 using SCr level alone.[6] A large pooling project recently reported that 75%-88% of eGFRs derived using the MDRD Study equation were within 30% of mGFR values in various clinical populations.[7] Later studies have indicated that GFR equations combining SCr and SCysC levels further improve the precision of GFR estimates.[8-10]

As the accuracy of GFR estimation equations increases, characterizing the differences between eGFR and mGFR becomes important. Imprecision in the gold standard, mGFR, may occur because of limitations in the accurate measurement of urine volumes and times, iothalamate concentration, and incomplete bladder emptying. In addition, mGFR reflects a short-term clearance period, 2 hours for urinary iothalamate, which differs from the person's average GFR because of physiologic day-to-day and diurnal fluctuation. Although eliminating the imprecision in mGFRs must improve the apparent accuracy of eGFR equations, the magnitude of this effect is unknown. We hypothesized that imprecision in mGFRs contributes to a significant proportion of the cases in which mGFR and eGFR differ substantially (by $>30\%$). This study extends our previous work,[8] which developed equations for estimating a single mGFR using SCr and SCysC levels, by quantifying the effect of more precise mGFR values on equation accuracy.

METHODS

Data Sources

The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) is a National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)-supported research team that develops and validates improved estimating equations for GFR by pooling data from research studies and clinical populations.[7,11,12] Our study focused on individual patient data from 2 multicenter randomized clinical trials with data for repeated GFR measurements. The MDRD Study included men and women aged 18-70 years with CKD who were not on dialysis therapy and had not had a kidney transplant, whereas the AASK (African American Study of Kidney Disease and Hypertension) included African American men and women of the same age who had hypertensive kidney disease and GFR of 20-65 mL/min/1.73 m^2.[13,14] The institutional review boards of all participating institutions approved the study.

To be eligible for inclusion in these analyses, participants must have had 2 prerandomization mGFRs and measurements of SCysC and SCr. The 2 mGFRs include: (1) a single reference mGFR collected from the prerandomization reference visit when SCr and SCysC also were assayed, and (2) another prerandomization mGFR collected from the visit closest to the reference mGFR visit (Fig 1). This yields 1,046 participants from the MDRD Study and 979 participants from AASK. For MDRD Study participants, the single reference mGFR was collected during the third visit before randomization (visit B3) to determine study eligibility (13-55 mL/min/1.73 m^2). The other prerandomization mGFR was

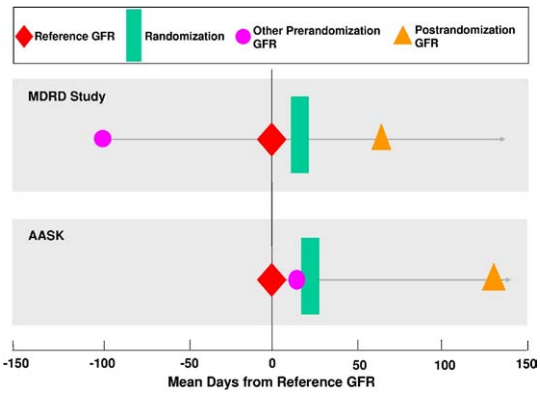

Figure 1. Timeline of glomerular filtration rate (GFR) measurements in the African American Study of Kidney Disease and Hypertension (AASK) and Modification of Diet in Renal Disease (MDRD) Study. Symbols indicate mean period of randomization (rectangle), collection of reference prerandomization mGFR visit (diamond) when serum creatinine and serum cystatin C also were measured, collection of another prerandomization GFR (circle), and collection of the first postrandomization mGFR (triangle).

collected at the first visit (B0), a mean of 104 ± 19 (standard deviation [SD] days; range, 36-153 days) before visit B3. For AASK participants, the single reference mGFR was collected during visit G1 to determine study eligibility (20-65 mL/min/1.73 m^2). Participants were invited for another prerandomization mGFR 2 weeks later (visit G2) regardless of SCr level. After excluding 30 AASK participants because their G1 and G2 mGFRs were more than 45 days apart, we had 949 AASK participants with 2 prerandomization mGFRs that were separated by a mean of 15 ± 8 days. A third mGFR collected from the closest visit after randomization in both studies also was included in the analysis to determine the effect of 3 stable mGFRs on estimates of equation accuracy.

Measurements

In the MDRD Study and AASK, GFR was measured as the weighted mean of 4 timed voluntary ^{125}I-iothalamate urinary clearances of 25-35 minutes' duration.[8,15] Comparisons of ^{125}I-iothalamate clearances to urinary clearance of inulin, the reference standard for GFR measurements, showed high correlations.[16,17] SCr was assayed using the Beckman rate–Jaffé method based on the alkaline picrate reaction (reference range, 0.8-1.4 mg/dL) and calibrated to standardized SCr values measured at the Cleveland Clinic Research Laboratory.[18] Results of the calibration procedure have been described previously.[11,19] To measure SCysC, stored serum specimens were thawed in 2005-2006 after being frozen at −70°C since collection. Samples were assayed at the Cleveland Clinic Research Laboratory using a particle-enhanced immunonephelometric assay (N Latex Cystatin C; Dade Behring, www.dadebehring.com). Interassay coefficients of variation (CVs) were 5.05% and 4.87% at mean concentrations of 0.97 and 1.90 mg/L (72.7 and 142.3 μmol/L), respectively.[8] SCysC has been shown to be robust to multiple freeze-thaw cycles.[20]

Model Evaluation

We evaluated the accuracy of 3 existing equations: the 4-variable isotope-dilution mass spectroscopy–traceable MDRD Study equation ($eGFR_{MDRD} = 175 \times$ standardized $SCr^{-1.154} \times age^{-0.203} \times 1.212$ [if black] $\times 0.742$ [if female]),[18] the CKD-EPI SCysC equation ($eGFR_{CKD\text{-}EPI\ SCysC} = 127.7 \times SCysC^{-1.17} \times Age^{-0.13} \times 0.91$ [if female] $\times 1.06$ [if black]),[8] and the CKD-EPI SCr-SCysC equation $eGFR_{CKD\text{-}EPI\ SCr+SCysC} = 177.6 \times SCr^{-0.65} \times SCysC^{-0.57} \times Age^{-0.20} \times 0.82$ [if female] $\times 1.11$ [if black].[8] In addition, we included as an additional test mGFRs collected at the other prerandomization visit.

These 4 index tests were analyzed for their ability to predict increasingly precise reference tests to reflect true kidney function. These reference tests include the single reference mGFR collected at the same visit that SCr and SCysC in the equations were obtained, the average of 2 mGFRs, and the average of 3 mGFRs. We also varied the reference test by limiting the analysis to study participants with consistent mGFRs (≤25% difference between the reference and other mGFRs). We report the error of the index tests by considering the percentage of eGFRs from the 4 models that is outside 30% of the gold standard ($1 - P_{30}$) and by the regression root mean squared error (RMSE). Derived from the regression of logarithmically transformed GFR, RMSE is the standard deviation of the distance between the gold standard and eGFR (y-ŷ). Because both the eGFR and mGFRs were modeled on the log scale, 100 × RMSE approximates the standard deviation of the percentage of difference between the estimate and the gold standard, approximating the CV for mGFR. By using $1 - P_{30}$ and the logarithmic transformation of GFR, we account for the expectation that the absolute magnitude of errors increases proportionately with GFR level. To define the significance of the difference in $1 - P_{30}$ values, we used McNemar χ^2 tests for matched pairs to compare models using 1 reference mGFR and those with multiple and/or consistent mGFRs. To define the significance of RMSE differences, we used paired *t* tests to compare squared residuals derived from the regression of logarithmically transformed GFR. We similarly defined the significance of $1 - P_{30}$ and RMSE differences between those with consistent and inconsistent mGFRs using Fisher exact test for proportions and unpaired *t* tests, respectively. Comparing models that used a single reference mGFR with the average of 3 mGFRs, we limited the sample to those with 3 mGFRs. All analyses were conducted using STATA (version 10.1; www.stata.com).

RESULTS

The study population included 1,995 participants with 2 mGFRs obtained during the baseline visits of the AASK and MDRD Study. Of these participants, 1,746 (88%) individuals had consistent mGFRs (Fig 2). Three mGFRs were available for 1,268 individuals, with 973 (77%) having both other mGFRs consistent with (≤25% difference) the reference mGFR. Table 1 lists characteristics of the MDRD Study and AASK participants and the combined study population. Mean reference mGFRs from the MDRD Study (n = 1,046), AASK (n = 949), and combined study participants (n = 1,995) were 33 (5th-95th percentile, 14-57), 46 (5th-95th percentile, 24-64), and 39 mL/min/1.73 m^2 (5th-95th percentile, 15-63), respectively. Mean SCr levels from MDRD Study, AASK, and combined participants were 2.5 ± 1.1, 2.0 ± 0.7, 2.3 ± 1.0 mg/dL. Mean SCysC concentrations from the MDRD Study, AASK, and combined participants were 2.2 ± 0.8, 1.7 ± 0.6, and 2.0 ± 0.7 mg/dL, respectively.

Mean mGFRs tended to be similar across groups, but given the large sample size, even small differences were statistically significant. mGFRs at the other prerandomization visit were approximately 1 mL/min/1.73 m^2 higher than in the reference visit, possibly caused by regression to the mean because the reference mGFR was

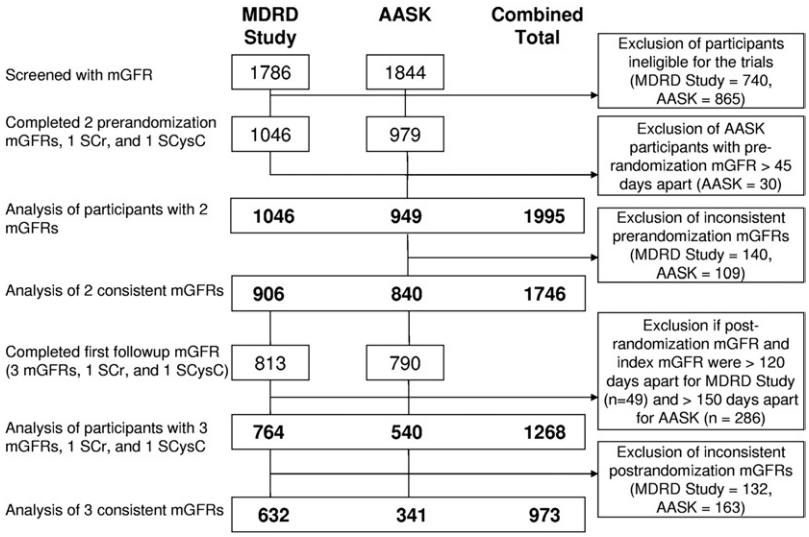

Figure 2. Flow chart with exclusion and inclusion criteria. Abbreviations: AASK, African American Study of Kidney Disease and Hypertension; MDRD, Modification of Diet in Renal Disease; mGFRs, measured glomerular filtration rates; SCr, serum creatinine; SCysC, serum cystatin C.

used for study eligibility. Participants with a third mGFR measured after randomization (Fig 1) had a slightly lower (1-3 mL/min/1.73 m^2) mean mGFR compared with prerandomization mGFRs. These participants also had a slightly lower mGFR than participants who only had 2 mGFRs. The subgroup of participants with consistent mGFRs generally was similar to the overall group. However, given the large sample size, mean values for several characteristics were significantly different between those with consistent and inconsistent mGFRs (Table 1).

Table 2 lists the performance of the MDRD Study, CKD-EPI SCysC, and CKD-EPI SCr-SCysC equations and the other prerandomization mGFR in predicting progressively more precise gold standards within the MDRD Study, AASK, and combined populations. We focus on results from the combined study population (listed in the last 4 columns); within-study analyses yielded similar trends, although findings often were less statistically significant in the AASK.

Before excluding participants with discrepant prerandomization mGFRs (Table 2; left-hand $1 - P_{30}$ and RMSE columns), we observe that the 4 models (considered in order of listing) performed progressively better in predicting: (1) a single mGFR, (2) the average of 2 mGFRs, and (3) the average of 3 mGFRs. The model using the other prerandomization mGFR was excluded from analysis in the last 2 reference tests because this prerandomization GFR measurement was part of the averaged mGFRs, thus biasing its evaluation. In all cases, we found that the MDRD Study equation ($1 - P_{30} = 12.1\%$; RMSE = 0.196) performed better than the CKD-EPI SCysC equation ($1 - P_{30} = 17.1\%$; RMSE = 0.223) in predicting a single reference mGFR. Consistent with our previous publication,[8] we also found that the CKD-EPI SCr-SCysC equation, which uses 2 markers, had greater accuracy in predicting mGFR ($1 - P_{30} = 8.3\%$; RMSE = 0.180) than the MDRD Study equation and CKD-EPI SCysC equation, which use single markers. To provide a context for understanding the accuracy of eGFR, we have found that the gold-standard mGFR collected 62 days from the reference mGFR outperforms all eGFRs in predicting a single mGFR with a $1 - P_{30}$ value of 8.0% and RMSE value of 0.167. This $1 - P_{30}$ value is 6.4% if one develops a GFR estimate using regression of the reference mGFR on the other prerandomization mGFR (RMSE stays the same). Figure 3 shows the general pattern of mGFR variations. This graph of the difference between the 2 prerandomization mGFRs versus their average shows greater differences at higher levels of kidney function, supporting the use of percentage differences and log transformation of mGFRs.

As the precision of the reference test increased by averaging 2 or 3 mGFRs, the calculated accuracy of all eGFR equations improved. As the reference test was improved from a single mGFR to the average of 2 mGFRs, the improvement in $1 - P_{30}$ was significant for the CKD-EPI SCysC equation ($P < 0.001$) and CKD-EPI SCr-SCysC

Table 1. Participant Characteristics Before and After Exclusion of Participants With Discrepant mGFR Values

	MDRD Study		AASK		MDRD Study + AASK	
	All mGFRs	Consistent[a] mGFRs	All mGFRs	Consistent[a] mGFRs	All mGFRs	Consistent[a] mGFRs
Sample size (no. of participants)	1,046	906	949	840	1,995	1,746
Women (%)	39	39	39	39	39	39
Age (y)	52 ± 13	52 ± 12	55 ± 11	54 ± 11[b]	53 ± 12	53 ± 12
Black (%)	10	10	100	100	53	53
Diabetes (%)	6	6	0	0	3	3
Height (cm)	171 ± 10	171 ± 10	171 ± 10	171 ± 10	171 ± 10	171 ± 10
Weight (kg)	79 ± 16	79 ± 16	90 ± 21	90 ± 21	84 ± 19	84 ± 19
Body mass index (kg/m^2)	27 ± 4	27 ± 4	31 ± 7	31 ± 7	29 ± 6	29 ± 6
Body surface area (m^2)	1.9 ± 0.2	1.9 ± 0.2	2.0 ± 0.2	2.0 ± 0.2	2.0 ± 0.2	2.0 ± 0.2
Serum cystatin (mg/L)	2.2 ± 0.8	2.2 ± 0.7[b]	1.7 ± 0.6	1.7 ± 0.6	2.0 ± 0.7	2.0 ± 0.7[b]
Serum creatinine (mg/dL)	2.5 ± 1.1	2.4 ± 1.0[b]	2.0 ± 0.7	2.0 ± 0.7	2.3 ± 1.00	2.2 ± 0.9[b]
mGFR (mL/min/1.73 m^2)						
Reference (5th-95th percentile)	33 ± 14 (13-57)	34 ± 14[b] (14-57)	46 ± 13 (24-64)	46 ± 13[b] (24-64)	39 ± 15 (15-63)	40 ± 15[b] (16-63)
Other prerandomization visit (5th-95th percentile)	34 ± 13[c] (14-56)	34 ± 13[b] (14-56)	47 ± 15[c] (23-72)	46 ± 14[b] (23-67)	40 ± 16[c] (16-66)	40 ± 15[c] (17-65)
Difference from reference mGFR	0.7 ± 5.4	0.1 ± 3.9[b]	1.4 ± 7.8	0.3 ± 5.1[b]	1.0 ± 6.7	0.2 ± 4.5[b]
% Difference from reference mGFR	5.6 ± 19.8	1.8 ± 10.9[b]	3.3 ± 17.8	0.5 ± 11.0[b]	4.5 ± 18.9	1.2 ± 11.0[b]
Time between mGFR visits (d)	104 ± 19	104 ± 19	15 ± 8	15 ± 8	62 ± 19	61 ± 47
Subgroup with 3rd mGFR (postrandomization)						
No. of participants	764	632	504	341	1268	973
Third mGFR (5th-95th percentile)	31 ± 12[c] (13-51)	32 ± 12[b,c] (14-52)	46 ± 17 (20-74)	45 ± 14[c] (22-68)	37 ± 16[c] (14-66)	37 ± 14[c] (15-61)
Difference from reference mGFR	−1.3 ± 4.5	−1.1 ± 10.2[b]	0.2 ± 11.1	−0.7 ± 5.9[b]	−0.7 ± 7.9	−1.0 ± 4.5[b]
% Difference from reference mGFR	−4.2 ± 13.9	−3.7 ± 10.2[b]	0.7 ± 26.0	−1.9 ± 12.4[b]	−2.2 ± 20	−3.0 ± 11.0[b]
Time between reference and 3rd mGFR (d)	65 ± 9	65 ± 9	134 ± 9	134 ± 9	93 ± 35	89 ± 34[b]

Note: Mean ± standard deviation reported for continuous data. Conversion factors for units: serum creatinine in mg/dL to μmol/L, ×88.4; serum cystatin C in mg/L to nmol/L, ×74.9; GFR in mL/min/1.73 m^2 to mL/s/1.73 m^2, × 0.01667.

Abbreviations: AASK, African American Study of Kidney Disease; MDRD, Modification of Diet in Renal Disease; mGFR, measured glomerular filtration rate.

[a]mGFR ≤25% different from the reference mGFR.

[b]*t* Test comparing consistent with inconsistent mGFRs; $P < 0.05$.

[c]$P < 0.05$ in a paired *t* test compared with the reference mGFR.

Table 2. Accuracy of Models for Estimating Reference mGFRs for All Observations and Consistent mGFRs Only

	MDRD Study				AASK				MDRD Study + AASK			
	$1 - P_{30}$		RMSE		$1 - P_{30}$		RMSE		$1 - P_{30}$		RMSE	
	All GFRs	Consistent mGFRs[a]	All GFRs	Consistent mGFRs[a]	All GFRs	Consistent mGFRs[a]	All GFRs	Consistent mGFRs[a]	All GFRs	Consistent mGFRs[a]	All GFRs	Consistent mGFRs[a]
(1) Single mGFR as Reference Test												
Sample size	1,046	906	1,046	906	949	840	949	840	1,995	1,746	1,995	1,746
MDRD Study eqn	8.8%	7.6%[b]	0.180	0.174[b]	15.8%	14.9%	0.193	0.186	12.1%	11.1%[b]	0.196	0.188[b]
CKD-EPI SCysC eqn	17.2%	14.7%[c]	0.219	0.209[b]	17.2%	15.4%[c]	0.196	0.190	17.1%	15.0%[c]	0.223	0.211[b]
CKD-EPI SCr & SCysC eqn	6.7%	4.7%[c]	0.165	0.157[b]	10.1%	8.8%[b]	0.173	0.163	8.3%	6.7%[c]	0.180	0.169[c]
Prerandomization mGFR	8.9%	—[d]	0.170	—[d]	6.9%	—[d]	0.146	—[d]	8.0%	—[d]	0.167	—[d]
(2) Average of 2 mGFRs as Reference Test												
Sample size	1,046	906	1,046	906	949	840	949	840	1,995	1,746	1,995	1,746
MDRD Study eqn	9.1%	7.7%[c]	0.173[e]	0.168[e]	14.6%	13.5%[b]	0.188	0.184[b]	11.7%	10.5%[c]	0.186[f]	0.181[b,f]
CKD-EPI SCysC eqn	13.7%[f]	13.1%[e]	0.201[f]	0.198[f]	15.4%[e]	14.3%[b]	0.193	0.190	14.5%[f]	13.7%[b,e]	0.206[f]	0.203[b,f]
CKD-EPI SCr & SCysC eqn	5.3%[e]	4.3%[b]	0.155[f]	0.149[b,f]	8.3%[e]	7.0%[c,e]	0.166[e]	0.160[b]	6.7%[e]	5.6%[c,e]	0.166[f]	0.160[c,f]
(3) Average of 3 mGFRs as Reference Test												
Sample size	764	632	764	632	504	341	504	341	1,268	973	1,268	973
MDRD Study eqn	6.9%	5.4%[b]	0.162[e]	0.155[b]	14.7%	12.9%	0.189	0.176	10.0%	8.0%[c]	0.179[e]	0.168[b]
CKD-EPI SCysC eqn	12.0%	11.4%	0.187[e]	0.181[e]	15.7%	14.7%	0.197	0.186	13.5%	12.5%	0.196[f]	0.189
CKD-EPI SCr & SCysC eqn	4.2%	3.3%[b]	0.144[e]	0.136[b,e]	7.9%	5.0%[b,e]	0.168	0.149	5.7%	3.9%[c]	0.159[f]	0.146[b,e]

Abbreviations and definitions: mGFR, measured glomerular filtration rate; P_{30}, percentage of estimated GFR within 30% of the gold-standard reference test; RMSE, root mean squared error; SCr, serum creatinine; SCysC, serum cystatin C.

[a]mGFRs excludes those that differed from the reference mGFR by ≥25%.

[b]$P < 0.05$ or [c]$P < 0.001$, when comparing individuals with consistent and inconsistent mGFRs. $1 - P_{30}$ tested using χ^2 (eg, when the average of 2 mGFRs is the reference test, using the CKD-EPI SCysC equation model in the combined population, $1 - P_{30} = 13.7\%$ when limited to the 1,746 with consistent mGFRs vs 20.1% in the 249 participants with inconsistent GFRs [$P = 0.02$], leading to an overall $1 - P_{30}$ of 14.5% in all GFRs). RMSE tested using unpaired *t* test of the squared residuals. Symbols are placed in column of consistent mGFRs.

[d]Biased estimate because the prerandomization mGFR (nonreference visit) also was used to define consistent GFRs.

[e]$P < 0.05$ or [f]$P < 0.001$, in a paired test compared with the same model when a single mGFR is used as the reference in the same population sample. $1 - P_{30}$ tested using McNemar χ^2 for paired proportions (eg, using the CKD-EPI SCysC equation model in the combined population, $1 - P_{30} = 14.5\%$ when the average of 2 mGFRs is the reference test vs 17.1% when single mGFR is the reference test [$P < 0.001$]; when using the CKD-EPI SCysC equation model in the combined population, $1 - P_{30} = 13.5\%$ when the average of 3 mGFRs is the reference test vs 10.6% in the CKD-EPI SCysC model limited to the 1,268 participants with 3 GFRs and for which a single mGFR is the reference test [$P = 0.07$]). RMSE tested using paired *t* test of the squared residuals.

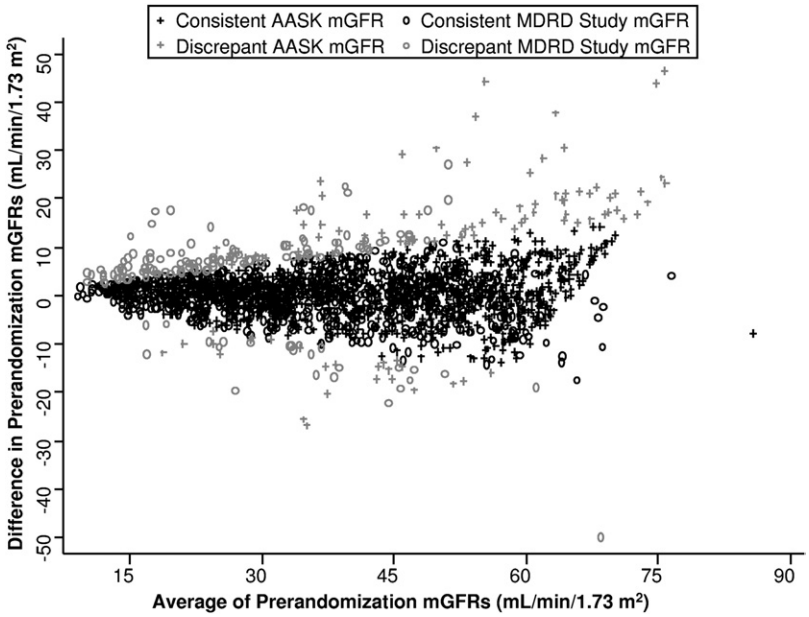

Figure 3. Bland-Altman plot of the difference of the 2 prerandomization glomerular filtration rate measurements (mGFRs) by the average of prerandomization mGFRs. Abbreviations: AASK, African American Study of Kidney Disease and Hypertension; MDRD, Modification of Diet in Renal Disease.

equation ($P < 0.01$). When the average of 3 mGFRs was used as the reference, accuracy improved and estimates of $1 - P_{30}$ and RMSE decreased; however, in this smaller sample, the improvement was statistically significant for only the continuous RMSE analysis. For the 1,268 individuals with 3 mGFRs, the accuracy with which the CKD-EPI SCr-SCysC equation predicted the average of 3 mGFRs was similar to the performance of the other prerandomization mGFR in predicting the reference mGFR ($1 - P_{30} = 5.7\%$; RMSE = 0.159 vs $1 - P_{30} = 7.1\%$; RMSE = 0.156).

When the analysis was repeated limited to only participants with consistent mGFRs (Table 2; right-hand $1 - P_{30}$ and RMSE columns), the accuracy of all equations was better. The percentage of eGFRs from the 3 estimation equations (MDRD Study, CKD-EPI SCysC, and CKD-EPI SCr-SCysC) that differed from a single reference mGFR by ≥30% decreased by 8%, 12%, and 18%, whereas RMSE decreased by 4%, 5%, and 6%, respectively ($P < 0.05$ comparing the 1,746 individuals with consistent mGFRs with the other 248 participants with prerandomization mGFRs that differed from the reference mGFR by >25%). Combining both approaches by using the average of consistent mGFRs as the reference, the accuracy of GFR-estimating equations improved further. For example, the CKD-EPI SCr-SCysC equation had a $1 - P_{30}$ of 5.6% and RMSE of 0.166, comparable to the similarity of 2 prerandomization mGFRs ($1 - P_{30} = 8.0\%$; RMSE = 0.167). Using the average of 3 consistent mGFRs as the reference, we eliminated 34%, 27%, and 53% of the large errors, defined using $1 - P_{30}$, from the MDRD, CKD-EPI SCysC, and CKD-EPI SCr-SCysC equations, respectively. The CKD-EPI SCr-SCysC model had a $1 - P_{30}$ of only 3.9% and RMSE of 0.146. Many, but not all, comparisons were statistically significant.

Figure 4 shows the improvement in accuracy with progressively more precise gold standards. Analyses testing GFR estimation in the MDRD Study and AASK populations separately had similar results. Unlike measures of precision (RMSE) and accuracy ($1 - P_{30}$), GFR estimation regression equation coefficients were similar whether a single mGFR or progressively more precise gold standards were used in their development (data not shown).

Although measures of precision and accuracy were different when the precision of the gold standard increased, the GFR-estimation equations were similar. The difference between eGFR calculated using an equation developed with 1 mGFR and the average of 2 mGFRs as the gold standard was 0.5 ± 0.4, indicating that most differences were <1 mL/min/1.73 m^2.

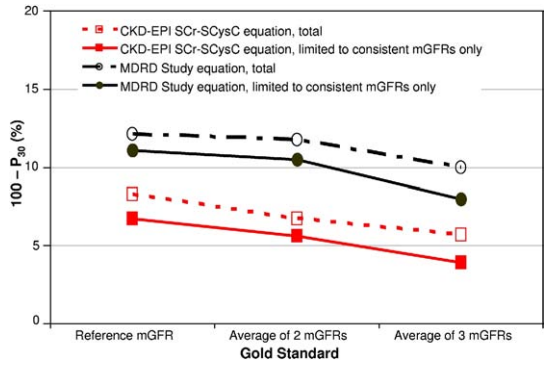

Figure 4. Estimated accuracy of Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) serum creatinine (SCr)-serum cystatin C (SCysC) equation and the Modification of Diet in Renal Disease (MDRD) Study equation in estimating 6 different gold-standard measured glomerular filtration rates (mGFRs): reference mGFR, average of 2 mGFRs, and the average of 3 mGFRs without and with limiting the analysis to consistent (within 25% of each other) mGFRs. P_{30} is calculated as the percentage of estimated GFR within 30% of the gold standard (the average of 3 mGFRs).

DISCUSSION

GFR measured as clearance of exogenous filtration markers is the gold standard for developing GFR-estimating equations. Current estimating equations are relatively unbiased, but imprecise, in CKD populations (GFR <60 mL/min/1.73 m^2). In this study, we tested the hypothesis that imprecision in gold-standard mGFRs contributes substantially to the estimated error in GFR-estimation equations. To our knowledge, no study has quantified the impact of imprecise mGFRs on kidney function–estimation equations. Using data from 2 longitudinal randomized clinical trials (the MDRD Study and AASK), we determined the accuracy of estimating a reference ^{125}I-iothalamate mGFR by another mGFR 2 months away. We found that mGFRs an average of 62 days apart had substantial variability across visits, with 8.0% of mGFRs >30% away from the reference mGFR ($1 - P_{30}$; RMSE = 0.167; SD of percentage of difference = 18.9). Nonetheless, this repeated gold-standard mGFR performed better than eGFRs from the MDRD Study, CKD-EPI SCysC, and CKD-EPI SCr-SCysC equations. However, use of averaged and consistent mGFRs as "better" gold standards in model development reduced up to half of the large (>30%) inaccuracies observed in GFR-estimating equations. Using the average of 2 mGFRs as the gold standard, the CKD-EPI SCr-SCysC equation ($1 - P_{30}$ = 6.7%; RMSE = 0.166) achieves accuracy similar to a single mGFR estimating a reference mGFR ($1 - P_{30}$ = 8.0% for mGFR and 6.4% for a regression estimate based on the mGFR). From these results, we infer that the level of variation in urinary iothalamate clearance used as the gold standard in our studies substantially impacts on the observed accuracy of eGFR. This holds regardless of the endogenous marker or estimating equation used. Thus, we showed that in CKD populations, although errors in GFR measurement do not bias regression estimates (intercept and slopes) in eGFR equations, they inflate the error estimates and decrease accuracy estimates of eGFR equations.

There are many sources of imprecision in mGFRs, including hour-to-hour and day-to-day variation and deviations from measurement protocol. Furthermore, urinary clearance methods require both urine and blood samples, which may introduce imprecision because of random error in collection and measurement of samples. A previous study conducted by Perrone et al[17] found that between-day CVs of ^{125}I-iothalamate clearance decreased within the range of 11.6%-16.6%. Another study found a median 6.3% intertest CV for ^{125}I-iothalamate clearance in the MDRD Study.[21] Using data from the MDRD Study and AASK combined, the CV of 2 mGFRs measured a mean of 62 days apart was 11.9% (data not shown). Studies can improve mGFR precision using accepted techniques, such as ensuring adequate hydration and high urine flow rates, as well as standardized training. Bladder ultrasound devices, which were not widely available during the MDRD Study and AASK, can help check the completeness of voiding.

Imprecise mGFRs do not account for all errors in the estimation equations. As we observed, even with the average of 3 consistent mGFRs as the reference standard, 3.9% of eGFRs differed by >30% of the gold standard. Such remaining errors in eGFR equations may reflect random variation in non-GFR determinants of SCr and SCysC. Although estimating equations adjust for the average effect of non-GFR determinants represented by age, sex, and race, the interindividual variation in these determinants may contribute to the remaining inaccuracies found within GFR

estimates using the MDRD Study equation.[7,8] Furthermore, SCr level, like mGFR, has been shown to fluctuate throughout the day.[22] These physiologic fluctuations, as well as measurement errors in SCr and SCysC, may account for a substantial portion of the remaining inaccuracies in GFR-estimating equations.

Our study has several limitations. First, our results may be of limited generalizability to other study populations and mGFR protocols. Because the MDRD Study and AASK limited their recruitment populations to patients with CKD, the number of people with eGFR >90 mL/min/1.73 m^2 was small. Thus, we could not determine the influence of imprecise mGFR on eGFR in healthy individuals, for whom both SCr level and mGFR are likely to have more variability. Similarly, the variability in mGFR may differ across other patient populations (those with end-stage renal failure, kidney transplants, etc). Furthermore, all mGFRs were ^{125}I-iothalamate urinary clearances. It is possible that increasing precision in technetium 99m–diethylenetriamine-penta-acetic acid (DTPA), ytterbium 169-DTPA, and inulin could diminish or improve the accuracy of the chosen estimating equations.

A second limitation of our study is that mGFRs from different clinical visits were assumed to be equivalent. We could not quantify true variability in mGFR and we recognize that the third postrandomization mGFR included as the gold standard in the average of 3 mGFRs was collected after both eligibility criteria and specific interventions, such as intake of a low-protein diet or administration of angiotensin-converting enzyme inhibitors, may have influenced measurements. Thus, this mGFR may have introduced more "noise" into the evaluation of the true GFR at the reference visit. Mean values for postrandomization mGFRs similar to prerandomization mGFRs suggests that short-term intervention effects are small compared with intraindividual variation. This probably occurs because kidney disease progresses slowly in general. Results of the MDRD Study and AASK clinical trials indicate that average rates of progression in all treatment groups postrandomization were 3.3 mL/min/y and 2.1 mL/min/1.73 m^2/y, respectively.[23,24] With >25% difference as the exclusion criteria of discrepant mGFRs a few months apart, progression of kidney disease (ie, true GFR decrease) is unlikely to be a primary cause of discrepancies. The MDRD Study and AASK had a number of design differences (AASK mGFRs prerandomization were closer, mean GFR was higher, and sample sizes were smaller, particularly for those with 3 mGFRs). Improved precision tended to be smaller in the AASK; however, overall conclusions tended to be similar.

Selection bias may have occurred because patients with more variable GFRs may differ systematically from those with more stable GFRs. However, clinical characteristics of the overall group generally were similar to the subgroup restricted to participants with consistent GFRs. Characteristics that were significantly different between participants with consistent and inconsistent mGFRs generally were related to kidney function rather than other demographic characteristics (Table 1). Selection bias also may have occurred because only a subset of individuals had 3 mGFRs. Because patients were excluded from postrandomization visits in part based on GFRs from earlier visits, data may not be missing completely at random. However, results did not change substantially when all models were limited to those with 3 mGFRs.

Last, performance of the MDRD Study, CKD-EPI SCysC, and CKD-EPI SCr-CysC equations may be better in our study because these equations were developed in this study population. However, the intraindividual variation in mGFRs at multiple visits has not been included in the development of these equations. Thus, substantial improvement in estimates of equation accuracy should be observed if the same equations were tested in other CKD populations. In this study, we estimated equation accuracy using percentage estimates that differed >30% from the reference mGFR ($1 - P_{30}$) and RMSE; however, other measures, such as P_{20} and P_{10}, could provide additional detail. More stringent measures of accuracy will be relevant as additional markers allow better kidney function–estimation equation accuracy.

In summary, this study suggests that a substantial proportion of the reported imprecision of GFR-estimation equations is caused by variability in the gold-standard measurements, such as the ^{125}I-iothalamate mGFR used to evaluate these equations. By excluding discrepant mGFR measures and averaging mGFRs from multiple visits,

we eliminated 53% of instances in which the reference test differed by ≥30% from the CKD-EPI SCr-SCysC eGFR equation. Reducing such sources of error in mGFR is important because equations that use multiple kidney markers, such as the CKD-EPI SCr-SCysC equation, are close to achieving accuracy similar to another gold-standard mGFR. Future studies of GFR estimation should aim to both improve GFR measurement precision and incorporate multiple GFR measurements to distinguish error attributable to GFR measurement rather than estimation. Use of these strategies in developing better GFR-estimation methods and assessing their accuracy will improve CKD diagnostic accuracy, which is central to the development and implementation of cost-effective treatments for decreasing the economic and social burden of CKD.

ACKNOWLEDGEMENTS

This work was presented in abstract form at the 48th Cardiovascular Disease Epidemiology and Prevention Conference 2008 in Colorado Springs, CO, on March 13, 2008.

In addition to Drs Stevens, Selvin, Zhang, Greene, Van Lente, Levey, and Coresh, investigators and research staff of the CKD-EPI are Christopher H. Schmid, PhD, Aghogho Okparavero (Tufts Medical Center); Liang Li, PhD (Cleveland Clinic); Jane Manzi, PhD, Brad Astor, PhD, MPH (Johns Hopkins University); Harold I. Feldman, MD, MSCE, J. Richard Landis, PhD, Marshall Joffe, MD, MPH, PhD (University of Pennsylvania); and John W. Kusek, PhD, Paul W. Eggers, PhD, Robert Starr, MD (NIDDK).

Collaborators contributing data for this study are Gerald Beck, PhD (MDRD Study), Gabriel Contreras, MD, and Julie Lewis, MD (AASK).

Support: CKD-EPI is funded by grants (UO1 DK 053869, UO1 DK 067651, and UO1 DK 35073) from the NIDDK as part of a cooperative agreement in which the NIDDK has substantial involvement in the design of the study and the collection, analysis, and interpretation of the data. The NIDDK was not required to approve publication of the finished manuscript.

Financial Disclosure: The authors declare that they have no relevant financial interests.

REFERENCES

1. Rehberg PB. Studies on kidney function: the excretion of urea and chlorine analysed according to a modified filtration-reabsorption theory. *Biochem J*. 1926;20(3):461-482.

2. Malamos B, Dontas AS, Koutras DA, Marketos S, Sfontouris J, Papanicolaou N. ^{125}I-sodium iothalamate in the determination of the glomerular filtration rate. *Nucl Med (Stuttg)*. 1967;6(3):304-310.

3. Rapoport A, Husdan H. Endogenous creatinine clearance and serum creatinine in the clinical assessment of kidney function. *CMAJ*. 1968;99(4):149-156.

4. Grubb A, Simonsen O, Sturfelt G, Truedsson L, Thysell H. Serum concentration of cystatin C, factor D and beta 2-microglobulin as a measure of glomerular filtration rate. *Acta Med Scand*. 1985;218(5):499-503.

5. Knight EL, Verhave JC, Spiegelman D, et al. Factors influencing serum cystatin C levels other than renal function and the impact on renal function measurement. *Kidney Int*. 2004;65(4):1416-1421.

6. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med*. 1999;130(6):461-470.

7. Stevens LA, Coresh J, Feldman HI, et al. Evaluation of the Modification of Diet in Renal Disease Study equation in a large diverse population. *J Am Soc Nephrol*. 2007;18(10): 2749-2757.

8. Stevens LA, Coresh J, Schmid CH, et al. Estimating GFR using serum cystatin C alone and in combination with serum creatinine: a pooled analysis of 3,418 individuals with CKD. *Am J Kidney Dis*. 2008;51(3):395-406.

9. Ma YC, Zuo L, Chen JH, et al. Improved GFR estimation by combined creatinine and cystatin C measurements. *Kidney Int*. 2007;72(12):1535-1542.

10. Tidman M, Sjostrom P, Jones I. A comparison of GFR estimating formulae based upon S-cystatin C and S-creatinine and a combination of the two. *Nephrol Dial Transplant*. 2008;23(1):154-160.

11. Levey AS, Coresh J, Greene T, et al. Expressing the Modification of Diet in Renal Disease Study equation for estimating glomerular filtration rate with standardized serum creatinine values. *Clin Chem*. 2007;53(4):766-772.

12. Stevens LA, Manzi J, Levey AS, et al. Impact of creatinine calibration on performance of GFR estimating equations in a pooled individual patient database. *Am J Kidney Dis*. 2007;50(1):21-35.

13. Beck GJ, Berg RL, Coggins CH, et al. Design and statistical issues of the Modification of Diet in Renal Disease trial. The Modification of Diet in Renal Disease Study Group. *Control Clin Trials*. 1991;12(5):566-586.

14. Gassman JJ, Greene T, Wright JT Jr, et al. Design and statistical aspects of the African American Study of Kidney Disease and Hypertension (AASK). *J Am Soc Nephrol*. 2003;14(7 suppl 2):S154-165.

15. Miller WG, Myers GL, Ashwood ER, et al. Creatinine measurement: state of the art in accuracy and interlaboratory harmonization. *Arch Pathol Lab Med*. 2005;129(3): 297-304.

16. Israelit AH, Long DL, White MG, Hull AR. Measurement of glomerular filtration rate utilizing a single subcutaneous injection of ^{125}I-iothalamate. *Kidney Int*. 1973;4(5): 346-349.

17. Perrone RD, Steinman TI, Beck GJ, et al. Utility of radioisotopic filtration markers in chronic renal insufficiency: simultaneous comparison of ^{125}I-iothalamate, ^{169}Yb-DTPA, ^{99m}Tc-DTPA, and inulin. The Modification of Diet in Renal Disease Study. *Am J Kidney Dis*. 1990; 16(3):224-235.

18. Levey AS, Coresh J, Greene T, et al. Using standardized serum creatinine values in the Modification of Diet in Renal Disease Study equation for estimating glomerular filtration rate. *Ann Intern Med*. 2006;145(4):247-254.

19. Lewis J, Greene T, Appel L, et al. A comparison of iothalamate-GFR and serum creatinine-based outcomes: acceleration in the rate of GFR decline in the African American Study of Kidney Disease and Hypertension. *J Am Soc Nephrol*. 2004;15(12):3175-3183.

20. Kyhse-Andersen J, Schmidt C, Nordin G, et al. Serum cystatin C, determined by a rapid, automated particle-enhanced turbidimetric method, is a better marker than serum creatinine for glomerular filtration rate. *Clin Chem*. 1994;40(10):1921-1926.

21. Levey AS, Greene T, Schluchter MD, et al. Glomerular filtration rate measurements in clinical trials. Modification of Diet in Renal Disease Study Group and the Diabetes Control and Complications Trial Research Group. *J Am Soc Nephrol*. 1993;4(5):1159-1171.

22. Scott PJ, Hurley PJ. Demonstration of individual variation in constancy of 24-hour urinary creatinine excretion. *Clin Chim Acta*. 1968;21(3):411-414.

23. Klahr S, Levey AS, Beck GJ, et al. The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease. Modification of Diet in Renal Disease Study Group. *N Engl J Med*. 1994;330(13):877-884.

24. Wright JT Jr, Bakris G, Greene T, et al. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease results from the AASK trial. *JAMA*. 2002;288(19):2421-2431.