Published in final edited form as:

Am J Kidney Dis. 2010 July ; 56(1): 39–49. doi:10.1053/j.ajkd.2010.02.347.

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, BA¹, Lesley A. Stevens, MD, MS², Elizabeth Selvin, PhD, MPH¹, Yaping (Lucy) Zhang, MS², Tom Greene, PhD³, Frederick Van Lente, PhD⁴, Andrew S. Levey, MD², and Josef Coresh, MD, PhD, MHS¹

¹Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, MD

²Division of Nephrology, Tufts Medical Center, Boston, MA

³Division of Epidemiology, University of Utah, Salt Lake City, Utah

⁴Department of Clinical Pathology, Cleveland Clinic Foundation, Cleveland, OH

Abstract

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—1995 participants from the Modification of Diet in Renal Disease (MDRD) Study and African American Study of Kidney Disease (AASK) with at least two baseline mGFRs from ¹²⁵I-iothalamate urinary clearances, one standardized Scr value, and one SCysC value.

Index Tests—eGFRs calculated from the 4-variable IDMS-traceable MDRD Study equation, the CKD-EPI SCysC equation, the CKD-EPI Scr-SCysC equation, and mGFRs collected from another pre-randomization visit

Reference Tests—A single reference mGFR, average of two, and average of three mGFRs; additional analysis limited to consistent mGFRs (difference $\leq$ 25% from the reference mGFR)

Results—We found that mGFRs had stable means but substantial variability across visits. Of all the mGFRs collected a mean of 62 days apart from the reference visit, 8.0% fell outside 30% of the single reference mGFR (1-P₃₀). The estimation equations were less accurate as 12.1%, 17.1%

© 2010 The National Kidney Foundation, Inc. Published by Elsevier Inc. All rights reserved.

Corresponding author and requests for single reprints: Josef Coresh, Johns Hopkins University, 2024 E. Monument St, Suite 630, Baltimore, MD 21278; Tel: 410-955-0495; Fax: 410-955-0476; coresh@jhu.edu.

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

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

N Section: 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.

and 8.3% of the eGFR from MDRD Study, CKD-EPI SCysC, and CKD-EPI SCr-SCysC equations fell outside 30% of the same gold standard (1-P₃₀). However, improving the precision of the reference test from a single mGFR to the average of three consistent mGFRs reduced these error estimates (1-P₃₀) 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 where estimated and measured GFR differs by more than 30%. Reducing and quantifying errors in gold standard measurements of GFR is critical to fully estimating the accuracy of GFR estimates.

Keywords

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 or estimated, assesses one of the most important aspects of kidney function.¹ Measured GFR (mGFR) computed from the clearance of injected exogenous markers (e.g. ¹²⁵I-iothalamate) is associated with little bias and is usually used as the gold standard for GFR.² However, it is often not appreciated that mGFR varies substantially due to measurement error and physiologic day-to-day variations in kidney function. Furthermore, the impact of this variability is largely unstudied.

In clinical practice, GFR is usually not measured directly due to the cost, invasiveness, and possible radioactive exposure associated with mGFR procedures. Alternatively, estimated GFR (eGFR) is computed from serum concentration of endogenous markers such as serum creatinine (SCr) and serum cystatin C (SCysC).^{3, 4} While estimating GFR is cheaper, 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 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/1.73m² using Scr alone.⁶ A large pooling project recently reported that 75% to 88% of eGFRs derived from the MDRD equation fell within 30% of mGFRs values in various clinical populations.⁷ Later studies have indicated that GFR equations combining Scr and SCysC further improve the precision of GFR estimates.⁸⁻¹⁰

As the accuracy of GFR estimation equations increases, characterizing whether some of the differences between estimated and measured GFR becomes important. Imprecision in the gold standard, mGFR, may occur due to 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 due to 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 more than 30%). This study extends our previous work,⁸ which developed equations for estimating a single mGFR with Scr and SCysC, by quantifying the effect of more precise mGFR values on equation accuracy.

Methods

Sources of Data

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 two multicenter randomized clinical trials with data on repeated GFR measurements. The MDRD Study included men and women aged 18-70 with chronic kidney disease who were not on dialysis and had not had a kidney transplant while the AASK (African American Study of Kidney Disease and Hypertension) study included African-American men and women of the same age who have had hypertensive kidney disease and GFR between 20 and 65 mL/min/1.73m².^{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 two pre-randomization mGFRs as well as measurements of SCysC and Scr. The two mGFRs include (1) a single reference mGFR collected from the pre-randomization reference visit when Scr and SCysC levels were also assayed and (2) another pre-randomization mGFR collected from the visit closest to the reference mGFR visit (Figure 1). This yields 1046 participants from the MDRD Study and 979 participants from AASK. For the 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/1.73m²). The other pre-randomization mGFR was collected at the first visit (B0), a mean of 104 days (SD 19, range 36-153 days) before the visit B3. For the AASK participants, the single reference mGFR was collected during the visit G1 to determine study eligibility (20-65 mL/min/1.73m²). Participants were invited for another pre-randomization mGFR two weeks later (visit G2), regardless of their serum creatinine level. After excluding 30 AASK participants where their G1 and G2 mGFRs were more than 45 days apart, we had 949 AASK participants with two pre-randomization mGFRs, which were separated by a mean of 15 days (SD 8 days). A third mGFR collected from the closest visit after randomization in both studies was also included in the analysis to determine the effect of three stable mGFRs on estimates of equation accuracy.

Measurements

In the MDRD Study and AASK, GFR was measured as the weighted mean of four timed, voluntary ¹²⁵I-iothalamate urinary clearances of 25-35 minutes duration.^{8, 15} Comparisons of ¹²⁵I-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-Jaffe method based on the alkaline picrate reaction (normal 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 previously described.^{11, 19} To measure SCysC levels, stored serum specimens were thawed in 2005-6 after being frozen at -70°C since collection. The samples were then assayed at Cleveland Clinic Research Laboratory with a particle-enhanced immuno-nephelometric assay (N Latex Cystatin C, Dade Behring, www.dadebehring.com). The interassay coefficients of variation (CV) were 5.05% and 4.87% at the mean concentrations of 0.97 and 1.90 mg/L (72.7 and 142.3 μ mol/L), respectively.⁸ SCysC has been shown to be robust to multiple freeze-thaw cycles.²⁰

Model Evaluation

We evaluated the accuracy of three existing equations: the 4-variable isotope dilution mass spectroscopy (IDMS)-traceable MDRD Study equation¹⁸, the CKD-EPI SCysC equation,⁸

and the CKD-EPI SCr-SCysC equation⁸. In addition, we included as an additional test mGFRs collected at the other pre-randomization visit.

These four 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 when Scr and SCysC in the equations were obtained, the average of two mGFRs, and the average of three mGFRs. We also varied the reference test by limiting the analysis to study participants with consistent mGFRs ($\leq 25\%$ percent difference between the reference and other mGFRs). We report the error of the index tests by considering the percentage of eGFRs from the four models that falls 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 estimated GFR ($y-\hat{y}$). Since both the eGFR and mGFRs were modeled on the log scale, $100 \times \text{RMSE}$ approximates the standard deviation of the percent 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 magnitudes of errors increase proportionately with the level of GFR. To define the significance of the difference in $1-P_{30}$ values, we used McNemar's chi-square tests for matched pairs to compare models using one 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's exact test for proportions and unpaired t-tests, respectively. When comparing models that used a single reference mGFR to the average of three mGFRs, we limited the sample to those with three mGFRs. All analyses were conducted using STATA (version 10.1, www.stata.com).

Results

The study population included 1995 participants with two mGFRs conducted during the baseline visits of AASK and the MDRD Study. Of these participants, 1746 (88%) individuals had consistent mGFRs (Figure 2). Three mGFRs were available for 1268 individuals, with 973 (77%) having both other mGFRs consistent with ($\leq 25\%$ difference from) the reference mGFR. Table 1 summarizes the characteristics of the MDRD Study and AASK participants as well as the combined study population. The mean reference mGFRs (5th-95th percentile) from the MDRD Study ($n=1046$), AASK ($n=949$), and combined study participants ($n=1995$) were 33 (14-57), 46 (24-64), and 39 (15-63) ml/min/1.73m² respectively. The mean (SD) Scr from MDRD Study, AASK and combined participants were 2.5 (1.1), 2.0 (0.7), 2.3 (1.0) mg/dL. The mean (SD) SCysC concentrations from 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 the groups, but given the large sample size even small differences were statistically significant. mGFRs at the other pre-randomization visit were approximately 1 ml/min/1.73m² higher than in the reference visit, possibly due to regression to the mean because the reference mGFR was used for study eligibility. Participants with a third mGFR measured after randomization (Figure 1) had a slightly lower (1-3 mL/min/1.73m²) mean mGFR compared to pre-randomization mGFRs. These participants also had a slightly lower mGFR than participants who only had two mGFRs. The subgroup of participants with consistent mGFRs was generally similar to the overall group. However, given the large sample size, the mean of several characteristics were significantly different between those with consistent and inconsistent mGFRs (Table 1).

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

Before excluding participants with discrepant pre-randomization mGFRs (in the lefthand 1-P₃₀ 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 two mGFRs, and (3) the average of three mGFRs. The model using the other pre-randomization mGFR was excluded from analysis in the last two reference tests because this pre-randomization GFR measurement was part of the averaged mGFRs, thus biasing its evaluation. In all cases, we found that the MDRD Study equation (1-P₃₀=12.1%, RMSE=0.196) performed better than the CKD-EPI SCysC equation (1-P₃₀=17.1%, RMSE=0.223) in predicting a single reference mGFR. Consistent to our previous publication,⁸ we also found that the CKD-EPI SCr-SCysC equation, which uses two markers, had greater accuracy in predicting mGFR (1-P₃₀=8.3%, RMSE=0.180) than the MDRD Study equation as well as the 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 away from the reference mGFR outperforms all eGFRs in predicting a single mGFR with a 1-P₃₀ value of 8.0% and RMSE value of 0.167. This 1-P₃₀ value is 6.4% if one develops a GFR estimate using regression of the reference mGFR on the other pre-randomization mGFR (RMSE stays the same). Figure 3 displays the general pattern of mGFR variation. This graph of the difference between the two pre-randomization 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 two or three mGFRs, the calculated accuracy of all eGFR equations improved. In fact, as the reference test was improved from a single mGFR to the average of two mGFRs, the improvement in 1-P₃₀ was significant for the CKD-EPI SCysC equation ($p < 0.001$) and CKD-EPI SCr-SCysC equation ($p < 0.01$). When the average of three mGFRs was used as the reference, accuracy improved and estimates of 1-P₃₀ and RMSE decreased but in this smaller sample the improvement was only statistically significant for the continuous RMSE analysis. Among the 1268 individuals with 3 mGFRs, the accuracy with which the CKD-EPI SCr-SCysC equation predicted the average of three mGFRs was similar to the performance of the other pre-randomization mGFR in predicting the reference mGFR (1-P₃₀=5.7%, RMSE=0.159 vs. 1-P₃₀=7.1%, RMSE=0.156).

When the analysis was repeated limited to only participants with consistent mGFRs (righthand 1-P₃₀ and RMSE columns) the accuracy of all equations was better. The percentage of eGFRs from the three estimation equations (MDRD Study, CKD-EPI SCysC, and CKD-EPI SCr-SCysC) that differed from a single reference mGFR by 30% or more decreased by 8%, 12%, and 18%, while the RMSE decreased by 4%, 5%, and 6%, respectively ($p < 0.05$ comparing the 1746 individuals with consistent mGFRs to the other 248 participants whose pre-randomization mGFRs that differed from the reference mGFR by more than 25%). When 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-P₃₀ of 5.6% and RMSE of 0.166 comparable to the similarity of two pre-randomization mGFRs (1-P₃₀ 8.0%, RMSE 0.167). Using the average of three consistent mGFRs as the reference, we eliminated 34%, 27%, and 53% of the large errors, defined by 1-P₃₀, from the MDRD, CKD-EPI SCysC and CKD-EPI

SCr-SCysC equations respectively. In fact, the CKD-EPI SCr-SCysC model had a 1-P₃₀ of only 3.9% and RMSE of 0.146. Many but not all comparisons were statistically significant.

Figure 4 summarizes the improvement in accuracy with progressively more precise gold standards. Analyses testing GFR estimation in the MDRD Study and AASK study populations separately had similar results. Unlike measures of precision (RMSE) and accuracy (1-P₃₀), 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).

While measures of precision and accuracy were quite different when the precision of the gold standard increased, the GFR estimation equations were quite similar. The difference between eGFR calculated from an equation developed with 1 mGFR and the average of 2 mGFRs as the gold standard was 0.5 with a standard deviation of 0.4, indicating most differences were < 1 ml/min/1.73m².

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/1.73m²). In this study, we tested the hypothesis that imprecision in the 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 two longitudinal randomized clinical trials (the MDRD Study and AASK), we determined the accuracy of estimating a reference ¹²⁵I-iothalamate mGFR by another mGFR two months away. We found that mGFRs an average of 62 days apart had substantial variability across visits with 8.0% of mGFRs more than 30% away from the reference mGFR (1-P₃₀; RMSE=0.167; SD of % difference = 18.9). Nonetheless, this repeat gold standard mGFRs performed better than eGFRs from the MDRD Study, CKD-EPI SCysC, and CKD-EPI SCr-SCysC equations. However, the use of averaged and consistent mGFRs as “better” gold standard in model development reduced up to half of the large (>30%) inaccuracies observed in GFR estimating equations. In fact, using the average of two mGFRs as the gold standard, the CKD-EPI SCr-SCysC equation (1-P₃₀ = 6.7% and RMSE=0.166) achieves similar accuracy to a single mGFR estimating a reference mGFR (1-P₃₀ of 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 the observed accuracy of estimated GFR. This holds regardless of the endogenous marker or estimating equation used. Thus, we demonstrated 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, day-to-day variation, and deviations from measurement protocol. Furthermore, urinary clearance methods require both urine and blood samples, which may introduce imprecision due to random error in collection and measurement of samples. A previous study conducted by Perrone *et al.* found that between-day CV of ¹²⁵I-iothalamate clearance fell within the ranges of 11.6% and 16.6%.¹⁷ Another study found a median 6.3% inter-test CV for ¹²⁵I-iothalamate clearance in the MDRD Study.²¹ Using data from the MDRD Study and AASK combined, the CV of two mGFR 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 on completeness of voiding.

Imprecise mGFRs do not account for all of the errors in the estimation equations. As we observed, even with the average of three consistent mGFRs as the reference standard, 3.9% of estimated GFR differed more than 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 inter-individual variation in these determinants may contribute to the remaining inaccuracies found within GFR estimates using the MDRD Study equation.^{7, 8} Furthermore, Scr, like mGFR, has been shown to fluctuate throughout the day.²² These physiological fluctuations as well as measurement errors in Scr and SCysC may account for a substantial portion of the remaining inaccuracies in eGFR 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 CKD patients, the number of people with eGFR > 90mL/min/1.73m² was small. Thus, we could not determine the influence of imprecise mGFR on eGFR among healthy individuals where both serum creatinine and measured GFR 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 the mGFRs were ¹²⁵I-iothalamate urinary clearances. It is possible that increasing precision in ^{99m}Tc-diethylenetriamine-pentaacetic acid (DTPA), ¹⁶⁹Yb-DTPA, and inulin could diminish or improve the accuracy of the chosen estimating equations.

A second limitation of our study is that the mGFRs from different clinical visits were assumed to be equivalent. We could not quantify true variability in mGFR and we recognize that the third post-randomization mGFR included as the gold standard in the average of three mGFRs was collected after both eligibility criteria and specific interventions such as the intake of low protein diet or the administration of ACE inhibitors may have influenced measurements. Thus, this mGFR may have introduced more “noise” in the evaluation of the true GFR at the reference visit. Similar means for post-randomization mGFRs to pre-randomization mGFRs suggests that short-term intervention effects are small compared to intra-individual variation. This is probably because kidney disease progresses slowly in general. Results of the MDRD and AASK clinical trials indicate that average rate of progression in all treatment groups post randomization was 3.3 mL/min/year and 2.1 mL/min/1.73m²/year, respectively.^{23, 24} With >25% difference as the exclusion criteria of discrepant mGFRs a few months apart, progression of kidney disease (i.e. true GFR decline) is unlikely to be a primary cause of discrepancies. The MDRD Study and AASK had a number of design differences (AASK mGFRs pre-randomization were closer, mean GFR was higher, and sample sizes were smaller, particularly for those with three mGFRs). Improved precision tended to be smaller in AASK but the overall conclusions tended to be similar.

Selection bias may have occurred since patients with more variable GFRs may differ systematically from those with more stable GFRs. However, the clinical characteristics of the overall group were generally similar to the subgroup restricted to participants with consistent GFRs. The characteristics that were significantly different between participants with consistent and inconsistent mGFRs were generally related to kidney function rather than other demographic characteristics (Table 1). Selection bias may have also occurred because only a subset of individuals had three mGFRs. Because patients were excluded from post-randomization visits in part based on GFRs from earlier visits, data may not be missing

completely at random. However, the results did not change substantially when all the models were limited to those with three mGFRs.

Lastly, the performance of the MDRD Study, CKD-EPI SCysC, and CKD-EPI SCr-CysC equations may be better in our study as these equations were developed in this study population. However, the intra-individual 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 among other CKD populations. In this study, we estimated equation accuracy using the percentage estimates that differed >30% of the reference mGFR (1-P₃₀) and RMSE, but other measures such as P₂₀ and P₁₀ 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 eGFR estimation equations is due to variability of the gold standard measurements like ¹²⁵I-iothalamate mGFR used to evaluate these equations. By excluding discrepant mGFR measures and averaging mGFRs from multiple visits, we eliminated 53% of instances where the reference test differed by 30% or more from the CKD-EPI SCr-SCysC eGFR equation. Reducing such sources of error in mGFR is important as equations that use multiple kidney markers like the CKD-EPI SCr-SCysC equation are close to achieving similar accuracy as 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.

Acknowledgments

This work was presented in abstract form at the 48th Cardiovascular Disease Epidemiology and Prevention Conference 2008 in Colorado Springs on March 13th, 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 and Aghogho Okparavero (Tufts Medical Center); Liang Li, PhD (Cleveland Clinic); Jane Manzi, PhD, Brad Astor, PhD, MPH (John Hopkins University); Harold I. Feldman, MD, MSCE, J. Richard Landis, PhD; Marshall Joffe MD MPH PhD (University of Pennsylvania); John W. Kusek, PhD, Paul W. Eggers, PhD, Robert Starr MD (National Institute of Diabetes and Digestive and Kidney Diseases).

Collaborators contributing data for this study are Gerald Beck, PhD (MDRD Study) and Gabriel Contreras, MD and Julie Lewis, MD (African American Study of Kidney Disease and Hypertension).

Support: CKD-EPI is funded by grants (U01 DK 053869, U01 DK 067651 and U01 DK 35073) from the National Institute of Diabetes, Digestive and Kidney Disease (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.

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. [PubMed: 16743680]
2. Malamos B, Dontas AS, Koutras DA, Marketos S, Sfontouris J, Papanicolaou N. ¹²⁵I-sodium iothalamate in the determination of the glomerular filtration rate. *Nucl Med (Stuttg).* 1967; 6(3): 304–310. [PubMed: 6059878]
3. Rapoport A, Husdan H. Endogenous creatinine clearance and serum creatinine in the clinical assessment of kidney function. *Can Med Assoc J.* 1968; 99(4):149–156. [PubMed: 5673222]

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. [PubMed: 3911736]
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. [PubMed: 15086483]
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. [PubMed: 10075613]
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. [PubMed: 17855641]
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. [PubMed: 18295055]
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. [PubMed: 17898698]
10. Tidman M, Sjoström 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. [PubMed: 17911090]
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. [PubMed: 17332152]
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. [PubMed: 17591522]
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. [PubMed: 1664792]
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):S154–65. [PubMed: 12819322]
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. [PubMed: 15737021]
16. Israelit AH, Long DL, White MG, Hull AR. Measurement of glomerular filtration rate utilizing a single subcutaneous injection of ¹²⁵I-iothalamate. *Kidney Int.* 1973; 4(5):346–349. [PubMed: 4762581]
17. Perrone RD, Steinman TI, Beck GJ, et al. Utility of radioisotopic filtration markers in chronic renal insufficiency: Simultaneous comparison of ¹²⁵I-iothalamate, ¹⁶⁹Yb-DTPA, ^{99m}Tc-DTPA, and inulin. The Modification of Diet in Renal Disease study. *Am J Kidney Dis.* 1990; 16(3):224–235. [PubMed: 2205098]
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. [PubMed: 16908915]
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. [PubMed: 15579521]
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. [PubMed: 7923773]
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. [PubMed: 8305642]

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. [PubMed: 5725198]
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. [PubMed: 8114857]
24. Wright J, J. T, 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. [PubMed: 12435255]

Figure 1.

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

Figure 2. Flowchart with exclusion and inclusion criteria. Abbreviations: measured glomerular filtration rates (mGFRs), serum creatinine (Scr), serum cystatin C (SCysC), Modification of Diet in Renal Disease (MDRD), and African American Study of Kidney Disease (AASK)

Figure 3.
Bland-Altman plot of the difference of the two prerandomization glomerular filtration rate measurements (mGFRs) by the average of pre-randomization mGFRs.

Figure 4. Estimated accuracy of CKD-EPI SCr-SCysC equation and the Modification of Diet in Renal Disease (MDRD) Study equation in estimating six different gold standard measured glomerular filtration rates (mGFRs): reference mGFR, average of two mGFRs, and the average of three 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 three mGFRs).

Table 1

Participant characteristics before and after the exclusion of participants with discrepant values of mGFR

	MDRD Study		AASK		MDRD Study + AASK	
	All mGFRs	Consistent* mGFRs	All mGFRs	Consistent* mGFRs	All mGFRs	Consistent* mGFRs
Sample Size (no. participants)	1046	906	949	840	1995	1746
Female (%)	39%	39%	39%	39%	39%	39%
Age, years	52 +/- 13	52 +/- 12	55 +/- 11	54 +/- 11 [‡]	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 [‡]	1.7 +/- 0.6	1.7 +/- 0.6	2.0 +/- 0.7	2.0 +/- 0.7 [‡]
Serum Creatinine, mg/dL	2.5 +/- 1.1	2.4 +/- 1.0 [‡]	2.0 +/- 0.7	2.0 +/- 0.7	2.3 +/- 1.00	2.2 +/- 0.9 [‡]
mGFR						
Reference (5 th -95 th %ile)	33 +/- 14 (13-57)	34 +/- 14 [‡] (14-57)	46 +/- 13 (24-64)	46 +/- 13 [‡] (24-64)	39 +/- 15 (15-63)	40 +/- 15 [‡] (16-63)
Other pre-randomization visit (5 th -95 th %ile)	34 +/- 13 [‡] (14-56)	34 +/- 13 [‡] (14-56)	47 +/- 15 [‡] (23-72)	46 +/- 14 [‡] (23-67)	40 +/- 16 [‡] (16-66)	40 +/- 15 [‡] (17-65)
Difference from reference mGFR	0.7 +/- 5.4	0.1 +/- 3.9 [‡]	1.4 +/- 7.8	0.3 +/- 5.1 [‡]	1.0 +/- 6.7	0.2 +/- 4.5 [‡]
% Difference from reference mGFR	5.6 +/- 19.8	1.8 +/- 10.9 [‡]	3.3 +/- 17.8	0.5 +/- 11.0 [‡]	4.5 +/- 18.9	1.2 +/- 11.0 [‡]
Time between mGFR visits, days	104 +/- 19	104 +/- 19	15 +/- 8	15 +/- 8	62 +/- 19	61 +/- 47
Subgroup with 3 rd mGFR (post randomization)						
No.	764	632	504	341	1268	973
Third mGFR (5 th -95 th %ile)	31 +/- 12 [‡] (13-51)	32 +/- 12 ^{‡‡} (14-52)	46 +/- 17 (20-74)	45 +/- 14 [‡] (22-68)	37 +/- 16 [‡] (14-66)	37 +/- 14 [‡] (15-61)
Difference from reference mGFR	-1.3 +/- 4.5	-1.1 +/- 10.2 [‡]	0.2 +/- 11.1	-0.7 +/- 5.9 [‡]	-0.7 +/- 7.9	-1.0 +/- 4.5 [‡]
% Difference from reference mGFR	-4.2 +/- 13.9	-3.7 +/- 10.2 [‡]	0.7 +/- 26.0	-1.9 +/- 12.4 [‡]	-2.2 +/- 20	-3.0 +/- 11.0 [‡]
Time between reference and 3 rd mGFR, days	65 +/- 9	65 +/- 9	134 +/- 9	134 +/- 9	93 +/- 35	89 +/- 34 [‡]

Note: Mean +/- SD reported for continuous data. Units of GFR are mL/min/1.73m². Conversion factors for units: serum creatinine in mg/dL to $\mu\text{mol/L}$, X88.4; serum cystatin C in mg/L to nmol/L, X74.9; GFR in mL/min/1.73m² to mL/s/1.73m² $\times$ 0.01667

* mGFR is $\geq 25\%$ different from the reference mGFR

[‡] p < 0.05 in a paired t-test compared to the reference mGFR

[†]t-test comparing consistent to inconsistent mGFRs $p < 0.05$

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

Table 2

Accuracy of models for estimating reference mGFRs for all observations and for 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 [^]	All GFRs	Consistent mGFRs [^]	All GFRs	Consistent mGFRs [^]	All GFRs	Consistent mGFRs [^]	All GFRs	Consistent mGFRs [^]	All GFRs	Consistent mGFRs [^]
Single mGFR as reference test												
Sample Size	1046	906	1046	906	949	840	949	840	1995	1746	1995	1746
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% ^{‡‡,b}	0.219	0.209 ^{‡,b}	17.2%	15.4% ^{‡‡,b}	0.196	0.190	17.1%	15.0% ^{‡‡,b}	0.223	0.211 ^{‡,b}
CKD-EPI SCR & SCysC eqn ^{***}	6.7%	4.7% ^{‡‡,b}	0.165	0.157 ^{‡,b}	10.1%	8.8% ^{‡,b}	0.173	0.163	8.3%	6.7% ^{‡‡,b}	0.180	0.169 ^{‡‡,b}
Prerandomization mGFR eqn	8.9%	--- <i>I</i>	0.170	--- <i>I</i>	6.9%	--- <i>I</i>	0.146	--- <i>I</i>	8.0%	--- <i>I</i>	0.167	--- <i>I</i>
Average of 2 mGFRs as reference test												
Sample Size	1046	906	1046	906	949	840	949	840	1995	1746	1995	1746
MDRD Study eqn	9.1%	7.7% ^{‡‡,b}	0.173 ^{‡,a}	0.168 ^{‡,a}	14.6%	13.5% ^{‡,b}	0.188	0.184 ^{‡,b}	11.7%	10.5% ^{‡‡,b}	0.186 ^{‡‡,a}	0.181 ^{‡‡,a,‡,b}
CKD-EPI SCysC eqn	13.7% ^{‡‡,a}	13.1% ^{‡,a}	0.201 ^{‡‡,a}	0.198 ^{‡‡,a}	15.4% ^{‡,a}	14.3% ^{‡,b}	0.193	0.190	14.5% ^{‡‡,a}	13.7% ^{‡,a,‡,b}	0.206 ^{‡‡,a}	0.203 ^{‡‡,a,‡,b}
CKD-EPI SCR & SCysC eqn	5.3% ^{‡,a}	4.3% ^{‡,b}	0.155 ^{‡‡,a}	0.149 ^{‡‡,a,‡,b}	8.3% ^{‡,a}	7.0% ^{‡,a,‡‡,b}	0.166 ^{‡,a}	0.160 ^{‡,b}	6.7% ^{‡,a}	5.6% ^{‡,a,‡‡,b}	0.166 ^{‡‡,a}	0.160 ^{‡‡,a,‡‡,b}
Average of 3 mGFRs as reference test												
Sample Size	764	632	764	632	504	341	504	341	1268	973	1268	973
MDRD Study eqn	6.9%	5.4% ^{‡,b}	0.162 ^{‡,a}	0.155 ^{‡,b}	14.7%	12.9%	0.189	0.176	10.0%	8.0% ^{‡‡,b}	0.179 ^{‡,a}	0.168 ^{‡,b}
CKD-EPI SCysC eqn	12.0%	11.4%	0.187 ^{‡,a}	0.181 ^{‡,a}	15.7%	14.7%	0.197	0.186	13.5%	12.5%	0.196 ^{‡‡,a}	0.189
CKD-EPI SCR & SCysC eqn	4.2%	3.3% ^{‡,b}	0.144 ^{‡,a}	0.136 ^{‡,a,‡,b}	7.9%	5.0% ^{‡,a,‡,b}	0.168	0.149	5.7%	3.9% ^{‡‡,b}	0.159 ^{‡‡,a}	0.146 ^{‡,a,‡,b}

[/] Biased estimate because the pre-randomization mGFR (non-reference visit) was also used to define consistent GFRs.

[‡] p<0.05

^{‡‡} p<0.001

^a_p values measured in a paired test compared to the same model when a single mGFR is used as the reference in the same population sample. 1-P₃₀ tested using McNemar chisquare for paired proportions (for example using the CKD-EPI SCysC equation model in the combined population, 1-P₃₀=14.5% when the average of 2 mGFRs is the reference test versus 17.1% when single mGFR is the reference test [p<0.0001]; when using the CKD-EPI SCysC model in the combined population, 1-P₃₀=13.5% when the average of 3 mGFRs is the reference test versus 10.6% in the CKD-EPI SCysC model limited to the 1268 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

[†] p< 0.05

^{††} p< 0.001

^b Comparison of individuals with consistent and inconsistent mGFRs. 1-P₃₀ tested using chi-square (for example, when the average of 2 mGFRs is the reference test, using the CKD-EPI SCysC equation model in the combined population, 1-P₃₀=13.7% when limited to the 1746 with consistent mGFRs versus 20.1% in the 249 participants with the inconsistent GFRs [p=0.02], leading to an overall 1-P₃₀ of 14.5% in all GFRs). RMSE tested using unpaired t-test of the squared residuals. Symbols are placed in the consistent mGFRs' column.

^{*} The MDRD Study equation is $eGFR_{MDRD} = 175 \times standardized\ Ser^{-1.154} \times age^{-0.203} \times 1.212 [if\ black] \times 0.742 [if\ female]$

^{**} CKD-EPI SCysC equation is $eGFR_{CKD-EPI_{Cys}} = 127.7 \times SCysC^{-1.17} \times age^{-0.13} \times 0.91 [if\ female] \times 1.06 [if\ black]$

^{***} the CKD-EPI Scr-SCysC is $eGFR_{CKD-EPI\ Scr+SCysC} = 177.6 \times Scr^{-0.65} \times SCysC^{-0.57} \times Age^{-0.20} \times 0.82 [if\ female] \text{ and } 1.11 [if\ black]$

[^] consistent mGFRs excludes those which differed from the reference mGFR by 25% or more

Abbreviations and definitions: eqn, equation; P₃₀, percentage of estimated GFR within 30% of the gold standard reference test; root mean squared error (RMSE); measured glomerular filtration rate (mGFR); Modification of Diet in Renal Disease (MDRD) Study; Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI); AASK, African American Study of Kidney Disease