Evaluating the Performance of GFR Estimating Equations

Lesley A. Stevens, MD, MS, **Yaping (Lucy) Zhang, MS**, and **Christopher H. Schmid, PhD**
Tufts Medical Center, Boston, MA

Abstract

GFR is an important indicator of kidney function, critical for detection, evaluation, and management of CKD. GFR cannot be practically measured in most clinical or research settings, and therefore estimating equations are used as a primary measure of kidney function. A considerable body of literature now evaluates the performance of the GFR estimating equations. The results of these studies are often not comparable due to variation in GFR measurement methods, endogenous filtration marker assays and tools by which the equations were evaluated. In this article, methods for evaluation of GFR estimating equations is discussed. Topics addressed include: statistical methods used in development and validation of equations; explanation of measures of performance used for evaluation with focus on distinction between bias, precision and accuracy, and with reference to examples of published evaluations of creatinine and cystatin C based equations; explanation for errors in GFR estimates; and challenges and questions in reporting performance of GFR estimating equations.

Introduction

Chronic kidney disease (CKD) is now recognized as a public health problem. CKD is defined as kidney damage or glomerular filtration rate (GFR) less than 60 ml/min/1.73 m^2 for three months or more, irrespective of cause [1–3]. CKD is classified into stages according to the level of GFR, and stage specific action plans facilitate evaluation and management of CKD. GFR is therefore an important indicator of kidney function, critical for detection, evaluation, and management of CKD. National and international organizations recommend that clinical laboratories report estimated GFR and that clinicians use estimated GFR to evaluate kidney function for all patients [1,3,2,4].

GFR cannot be practically measured for routine clinical or research purposes. Creatinine clearance measured from 24 hour urine samples has been traditionally used to estimate GFR, but timed urine collections are cumbersome and susceptible to error. The serum creatinine level, an endogenous filtration marker, is commonly used as an index of GFR, but it is also affected by factors other than GFR, especially variation in muscle mass and diet [5–7]. Thus, use of a single reference range for serum creatinine to distinguish between normal and abnormal GFR can be misleading [8–10,5,11]. GFR estimating equations such as the Modification of Diet in Renal Disease (MDRD) Study equation include age, sex, and race to account for the average differences in muscle mass among subgroups and have been shown

Correspondence to: Lesley Stevens, 800 Washington St Box 391 Boston MA 02111 tel: 617 636 2569; fax 617 636 5740; lstevens1@tuftsmedicalcenter.org.

to provide a more accurate assessment of the level of kidney function than serum creatinine alone. More recently, cystatin C has been suggested as an alternative endogenous filtration marker as the serum level is thought to have less dependence on muscle mass and diet than creatinine; however, increasing data also suggest that factors other than GFR may also influence its level. Equations have been developed to account for these determinants.

A considerable body of literature now evaluates the performance of the GFR estimating equations. In this article, we will discuss methods for evaluating GFR estimating equations. With references to our published evaluations of these equations, we will first focus on the metrics with which to evaluate GFR estimating equations and then discuss factors that affect the observed equation performance.

GFR estimating equations

Equation development and validation

Estimating equations incorporate demographic and clinical variables as surrogates for unmeasured physiologic processes other than GFR that affect the serum level of the endogenous filtration marker. Measured GFR and the filtration marker are generally transformed to the logarithmic scale, as the logarithmic scale better captures the multiplicative relationship between GFR and the inverse of the filtration marker. A multiplicative relationship implies that a change in the rate of increase or decrease in the marker leads to a proportional or relative change in the level of GFR. Statistically, the transformation linearizes the relationship and stabilizes the variability around the regression line across the range of GFR. Reverse transformation (exponentiation) of the equation coefficients returns the predicted GFR to the original scale in units of ml/min/1.73 m^2.

An accurate equation provides an estimate of the measured GFR that is unbiased (i.e. on average, the estimated GFR is equal to the measured GFR) and precise (i.e. the measured GFR is close to the estimated GFR for an individual). By construction, least squares regression produces a zero average difference, so overall bias will be zero in the development dataset; therefore, in a development dataset, the fit of the equation is described by its precision. Equation performance in a development dataset may also be observed by the presence or absence of bias within subgroups. Bias within subgroups indicates that an equation may be missing an important factor or that the data are insufficient to find the correct relationship. One concern about equation development is over-fitting of the variables. This can be tested by evaluation of models in internal validation datasets, which can be created by random splits of the dataset, or other techniques such as bootstrapping or jackknife.

Demonstrating generalizability requires evaluating the performance of the estimating equations in separate populations from those in which they was developed. A valid equation is one where the bias is small overall and in subgroups, and where precision is high. The presence of bias versus imprecision may indicate the sources of error (Table 1).

Creatinine based equations

The most commonly used GFR estimating equation is the MDRD Study equation. A considerable body of literature now documents the performance of this equation in many populations (i.e. external validation) [12–32]. After accounting for differences in the creatinine assay, most studies have demonstrated that the equation has reasonably good performance in patients with CKD (estimated GFR < 60 ml/min/1.73 m^2) but poorer performance in people with higher levels of GFR, such as younger patients with type 1 diabetes, as well as transplant recipients [32,33]. Figures 1 to 3 show the performance of the MDRD Study equation in a pooled dataset of 10 studies with 5504 people, including people with and without CKD (herein called pooled creatinine dataset).

Cystatin C based equations

Serum levels of cystatin C estimate GFR better than serum creatinine alone in most studies; however, cystatin C by itself does not improve upon creatinine based estimating equations in studies to date [34]. Some recent equations based on cystatin C have included demographic factors such as age or sex [35,36]. We developed an estimating equation based on cystatin C in a pooled dataset of 4 studies of 3134 individuals with CKD (herein called pooled cystatin C dataset) [36]. Age, sex, and race coefficients were significant but were substantially smaller than in the MDRD Study equation. Although the cystatin C based estimating equation was not more accurate than the MDRD equation, combining the two markers in a single equation yielded the most accurate estimates, suggesting smaller errors due to non-GFR determinants when using multiple markers (Table 2).

Measures of Performance

Characteristics of performance of GFR estimating equations depend on the relationship between measured and estimated GFR. The most common metrics of performance may be classified as describing bias, precision or accuracy.

Bias

Bias technically describes the mean difference between the estimated and measured values, although the median sometimes replaces the mean difference. In our analyses, we have described median difference between measured GFR and estimated GFR (measured GFR – estimated GFR). It is important to specify the order of the calculation to clearly interpret negative and positive values. Because positive errors and negative errors cancel each other out, a biased estimator reflects systematic errors among populations. Units of bias are generally presented as units of the estimate; in the case of GFR, these are units of ml/min/1.73 m^2.

Bias may be computed as either an arithmetic or relative difference. For example, an underestimate of 5 ml/min/1.73 m^2 at an estimated GFR of 15 ml/min/1.73 m^2 indicates a measured GFR of 20 ml/min/1.73 m^2, a clinically relevant arithmetic difference. In contrast, an underestimate of 5 ml/min/1.73 m^2 for measured GFR of 100 ml/min/1.73 m^2 would not be considered clinically relevant. Considering bias on a relative scale as a percentage of measured GFR may provide a more relevant metric. In the two examples above, a bias of 5

ml/min/1.73 m^2 leads to a relative bias of 25% for measured GFR of 20 ml/min/1.73 m^2 but a relative bias of 5% when measured GFR is 100 ml/min/1.73 m^2. The relative bias may agree more closely with the clinical implications of the difference. Expression of the bias in log units of the original regression equation is an alternative way of expressing this percent change. Conversely, large relative differences at very low GFR levels may equate to insignificant clinical changes. A 30% decrease from a GFR of 10 ml/min/1.73 m^2 gives a GFR of 7 ml/min/1.73 m^2. Use of both arithmetic and relative scales helps to compare the equation errors at different ranges of GFR.

The arithmetic and relative differences for the MDRD Study equation in the pooled creatinine dataset are 2.7 ml/min/1.73 m^2 and 5.8%, respectively. Both the arithmetic and relative differences are greater at higher levels of estimated GFR. Figures 1 and 2 demonstrate the change in the arithmetic difference across the range of estimated GFR for the MDRD Study equation when tested in the pooled creatinine dataset, and Figure 3 shows the relative difference. For all three figures, each point represents the estimated and measured GFR for one of the 5504 people in the dataset. The thin black line in Figure 1 and the horizontal lines at 0 in Figures 2 and 3 are lines of identity; that is, where estimated GFR equals measured GFR. If median difference were zero, all the points would fall along these lines. The thick black curves on the figures display bias as a function of estimated GFR where the average difference is estimated by a nonparametric smooth regression function (lowess function). The figures demonstrate that the MDRD study equation has little bias at lower levels of estimated GFR but underestimates measured GFR at higher levels. At estimated GFR < 60 ml/min/1.73 m^2, the bias is 0.8 ml/min/1.73 m^2, whereas at levels of estimated GFR > 60 ml/min/1.73 m^2, the bias is 8.3 ml/min/1.73 m^2.

Table 2 shows that equations based on cystatin C level overestimated measured GFR, whereas equations based on serum creatinine level underestimated measured GFR. The equation incorporating both cystatin C and creatinine levels has almost no bias.

Precision

Precision describes the variability of the differences about the average difference. Estimates that are unbiased, but imprecise, may arise for two reasons. First, the measurements themselves may be imprecise. Second, key elements may be missing. These missing elements may not lead to overall bias but may be relevant for a subgroup. For example, a missing interaction between sex and race may lead to imprecise estimates for Black females, although females on average are correctly estimated.

Several different metrics, including standard deviation, variance, and interquartile range (IQR) of the differences between measured and estimated can summarize precision. The units of the precision metrics depend on the metric. For the standard deviation and IQR, units are ml/min/1.73 m^2. For variance, the units are squared. Precision may also be expressed on the relative scale. Relative precision, like relative bias, may help to standardize differences with respect to differing levels of GFR. As with bias, relative precision is equivalent to arithmetic precision on the logarithmic scale.

The IQR of differences for the cystatin C equation is 8 ml/min/1.73 m^2 (Table 2). This indicates that the middle 50% of the distribution of differences covers a range of 8 ml/min/1.73 m^2. In other words, typical differences are fairly tightly concentrated around the median value. Table 2 also compares arithmetic and relative precision, where relative precision is expressed as IQR of percentage change (or IQR of difference on the logarithmic scale). In the comparison of equations that use cystatin C or creatinine, relative precision was better in models that used creatinine; whereas arithmetic precision was similar across models regardless of the inclusion of creatinine. Figure 4 shows that the relative measures of precision differ between equations at both the high and low levels of estimated GFR.

The width of confidence intervals around the difference also reflects the standard deviation of the differences and thus indicates precision. In Figures 1 to 3, the dashed curves around the solid curve indicate the 95% confidence intervals for the difference between the measured and estimated GFR for the MDRD Study equation (relative difference for Figure 3). In Figure 2, the increase in the width of the intervals widens as estimated GFR increases shows that precision decreases with increasing GFR; however, the relative precision slightly improves at higher levels of GFR (Figure 3).

Accuracy

Accuracy incorporates both bias and precision and, as such, reflects systematic as well as random errors. Accurate estimates have both low bias and high precision. Metrics include arithmetic difference or absolute percent difference, mean squared error (MSE) or its square root (RMSE), or percentage of estimates within k% of the measured value (P_k).

Mean squared error is the average of the squared differences between measured and estimated GFR. MSE equals the sum of the variance of the differences plus the square of their average difference. In other words, this measure of accuracy is a sum of a measure of precision plus the square of a measure of bias. RMSE has the advantage of being measured on the same scale as the difference, rather than in squared units. Generally, MSE and RMSE are expressed on the scale on which the regression is estimated; therefore, because we usually model GFR on the logarithmic scale, MSE or RMSE expresses a relative change on the GFR scale. For example, a RMSE of 0.2 means that on average, estimated GFR is within 20% of measured GFR. If bias is zero, RMSE is equivalent to the standard deviation of the errors; however, when bias is non-zero, RMSE will be greater than the standard deviation because it will include the bias as well.

Another common measure of goodness of fit in linear regression is the proportion of total variance in the outcome explained by the model. This is usually labeled R-squared (R^2) because it is also the squared correlation between the observed and predicted outcomes (here the measured and estimated GFR). Because R^2 is measured relative to the total variance of the outcomes, it may vary for the same model when applied to different data sources with varying ranges of observed and predicted values. Neither RMSE nor MSE are affected by the ranges of the predictor and outcome variables.

As a quantile-based measure, P_k is robust to large differences (outliers) that may inflate both MSE and RMSE. If the number of outliers is large relative to the sample size P_k naturally

increases with k (Figure 5). P_k is a relative measures; therefore, the accuracy varies according to the level of GFR. For example, a 30% error at a GFR of 100 ml/min/1.73 m^2 is 30 ml/min/1.73 m^2 whereas at a GFR of 20 ml/min/1.73 m^2, it is only 6 ml/min/1.73 m^2; therefore, P_k does not have a consistent meaning across the whole range of kidney function. Because MSE and RMSE are measured on the log scale (and, therefore, are relative metrics), they too have the same drawback.

P_{30} has traditionally been used in description of performance of GFR estimating equations. The thin dashed lines in Figure 1 delineate the boundaries defined by P_{30}. Dots that fall inside these lines indicate individuals with estimated GFR within 30% of measured GFR; dots outside represent errors of more than 30%. The proportion of dots within the lines is P_{30}. For the MDRD Study equation in the pooled creatinine dataset, P_{30} was 83% and for the equation that combined creatinine and cystatin C in the pooled cystatin C dataset, P_{30} was 90% [33,36]. Of note, the P_{30} in the MDRD Study population was 90%, demonstrating the importance of testing for generalizability in datasets separate from those in which the equation was developed [11]. Below, we discuss why P_{30} vs P_{10} or P_{20} have been routinely used.

Explanation for errors in GFR estimates

Several factors may lead to observed inaccuracy in estimated GFR. Table 1 lists these factors according to whether they affect bias and precision. These factors are described in detail below.

Measurement error in GFR

It is difficult to measure GFR; thus, values for measured GFR often contain an element of error, which differentiates it from the "true" GFR. The measurement error itself varies among filtration markers, clearance methods used and even among centers who use the same marker and method. In evaluating an estimating equation, we are interested in the comparison to true GFR, yet only measured GFR can be observed. As such, the observed differences between estimated GFR and measured GFR may overstate the difference between estimated GFR and true GFR, as the observed difference includes measurement error in GFR itself.

Variation in measurement of endogenous filtration markers

The variability in the creatinine assay even among laboratories that use the same instruments and assays is well known [37]. In the College of American Pathologists survey in 2004, most of the clinical laboratories in the United States had a positive bias compared to the standardized serum creatinine sample [37]. Recently, variability among cystatin C assays have also been demonstrated [38]. We previously demonstrated that differences among creatinine assays can have a substantial impact on accuracy of GFR estimates, particularly at higher levels of GFR [39–43]. The extent and the direction of the effect will depend upon the bias of the assay compared to the assay in the laboratory where the equation was developed. Similar effects would be expected for estimating equations that use cystatin C if the assays are not standardized.

Modeling errors

There may be deviations from the relationship between the surrogates and the non-GFR determinants captured in an equation's regression coefficients when this equation is applied to a specific person or population. Deviations reflect systematic differences among populations, variability among individuals within a population, or within an individual over time and are captured as bias for differences among populations and imprecision for differences among individuals.

We, and others, have observed a differential bias in the MDRD Study equation in the GFR range of 60–90 ml/min/1.73 m^2 between people known to have CKD compared to those without CKD [44,14]. This differential bias (or deviation) among populations of GFR estimating equations may vary among populations due to several reasons [45]:

1.   Differences in GFR measurement error that may be greater at the higher range of GFR when evaluated on the natural scale (as described above).

2.   Residual differences in calibration of creatinine or differences in methods to measure GFR among studies (as described above).

3.   Regression to the mean, a phenomenon in which the estimated GFR is "shrunken" towards the mean GFR of the study used to fit the regression, which may differ from the mean GFR in other studies.

4.   Differences in the presence of non-GFR determinants of creatinine. For example, healthy people are more likely to have larger muscle mass and increased dietary protein intake than people with kidney disease.

5.   Variation among populations in the proportional distribution of GFR, thereby altering the implications of a particular serum creatinine value. Populations with CKD have a larger proportional variation in GFR (approximately 10-fold, from 6 to 60 ml/min/1.73 m^2) than in populations without CKD (approximately 3 fold, from 60 to 180 ml/min/1.73 m^2). As such, a larger proportion of the variation in serum creatinine in CKD populations is likely due to variation in GFR than to variation in the other determinants. This is analogous to the dependence of the interpretation of any diagnostic test's results on the pre-test likelihood of disease

5.   Effects of selection of participants within a study or clinical population. Signs of CKD, or its absence, are often included in the screening process leading to distortions in the observed relationship of measured GFR with serum creatinine. Non-CKD studies likely disproportionately excluded individuals with prior signs of CKD, while CKD studies disproportionately retained patients with prior evidence of CKD.

Challenges and questions in reporting performance of GFR estimating equations

Classification with respect to estimated GFR

Our recommendation is to use estimated GFR, not measured GFR, as a category for evaluation. We acknowledge that this may appear to be counterintuitive on first glance. This decision has relevance for creation of plots (measured GFR should not be on the x-axis) and for classifying participants (measured GFR should not be categorized to determine stage of CKD and for determination of sensitivity and specificity). For example, we have used estimated GFR on the x axis for Figures 1–3 rather than the average of measured and estimated GFR, as is the usual format for Bland Altman plots [46]. There are two primary reasons for this recommendation:

1. As described above, measured GFR is not identical to true GFR. The extremes of the measured values will therefore over- or underestimate the true level of GFR. As such the observed bias will always be positive at the high end and negative at the low end. In contrast, the estimated GFR is the predicted average so should have as many measured above as below it and thus should on average have zero bias in the development dataset.

2. The objective of the regression methodology utilized in development of GFR estimating equations is to determine an estimate of GFR that is unbiased for the measured GFR for populations with a given level of estimated GFR. This regression methodology is not designed to obtain an estimate of measured GFR that is unbiased for populations with a given level of measured GFR; hence, the standard statistical approach for evaluating the performance of regression equations is to use the predictor (i.e. the GFR estimate) on the x-axis when creating residual plots or when creating subgroups for evaluating performance.

Use of sensitivity and specificity as an indicator of performance

CKD status, as defined by the threshold value of 60 ml/min per 1.73 m^2, is clearly an outcome of interest; however, methodologically, it is not as straightforward to calculate as would appear. First, sensitivity and specificity are defined in the context of the gold standard measure or the "truth". As described above, since GFR is measured with error, the classification of CKD will be made with error. In addition, the accuracy of the classification of CKD based on measured GFR will be very strongly dependent on the distribution of GFR in the population studied. For example, if many people have a true GFR between 55 and 65 ml/min/1.73 m^2, then no GFR estimate, or current measure of GFR, will be likely to classify individuals correctly. Conversely, the classification of CKD will appear very accurate, but is inflated, all of the GFR values are below 60 ml/min/1.73 m^2.

Use of P_{10} or P_{20} instead of P_{30} as a metric for accuracy

The use of P_{30} was initially suggested in the K/DOQI Guidelines for Classification Stratification and Evaluation of Chronic Kidney Disease in the section on clinical interpretation to identify individuals with CKD [3]. The rationale for this was that the 30%

was thought to be reasonable given that it reflects not only errors in the equation itself but also all errors describe above. In addition, P_{10} or P_{20} reflects the central tendency of the percent differences, which is already captured in the median percent bias, whereas the P_{30} also captures information on outliers. At present, current estimating equations are unbiased, yet they remain imprecise; therefore, improvement in P_{30} from one equation to the other should capture improvement in precision. Finally, since P_{30} values remain less than 95% for GFR estimating equations, it seems premature to move to P_{20} or P_{10}.

Implications of Inaccurate GFR estimates and Next Steps

The major limitation of GFR estimating equations are bias in some groups and lack of precision overall

These limitations are barriers to effective clinical practice, research studies and public health initiatives in CKD. For example, the high prevalence muscle wasting and inflammation in the elderly suggests that creatinine- or cystatin C-based estimating equations might also be inaccurate in elderly people with frailty or comorbid illnesses. Yet, precise estimates of kidney function are critically important for such patients because of their large implications for calculating the prevalence of CKD, for appropriate management of comorbid conditions and medication dosing. Furthermore, clinical trials for treatments of kidney disease require long duration to assess kidney disease progression. More precise GFR estimates would be more sensitive to changes in kidney function, thereby allowing for shorter (and less expensive) trials [47].

Improving both bias and precision requires research on several fronts. We will highlight two key areas: First, incorporation of estimates of the measurement error into measures of performance of the equation may allow us to compare the GFR estimates to true GFR itself. However, this will involve complex mathematical modeling as well as, and more importantly, knowledge of the magnitude of the error which will differ among GFR measurement techniques. Research efforts should be devoted to understanding the differences among methods. Second, to a great extent, inaccuracy in the GFR estimate is related to the filtration marker itself. The non-GFR determinants of serum creatinine are well understood and this knowledge may be used to improve equations. For example, one study reported improved performance of a creatinine based estimating equation by inclusion of lean body mass [48]. Lean body mass itself is derived from an equation which is likely to be population specific, and therefore this requires further investigation. Research into the non-GFR determinants of cystatin C (or other endogenous filtration markers) should be performed. Finally, best methods to combine engenous filtration markers needs to be determined. We demonstrated that the combination of creatinine and cystatin C in a single population yielded the most accurate equation; however whether this is transportable to other populations requires further testing.

Summary

GFR estimating equations have proven to be valuable, important tools for clinical research and healthy policy related to CKD. At present, there is no GFR estimating equation that provides GFR estimates that are accurate for all people. Understanding the strength and

limitation of equations facilitates their use. Part of this process is understanding the statistical issues that underlie their development and validation, information that can be learned from the metrics used to describe bias, precision and accuracy, as well as how limitations in the gold standard itself affects the observed error. In this article, we have presented these issues in the context of an in-depth discussion of published reports of currently available estimating equations.

References

1. Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, Hogg RJ, Perrone RD, Lau J, Eknoyan G. National Kidney Foundation practice guidelines for chronic kidney disease: Evaluation, classification, and stratification. Ann Intern Med. 2003; 139 (2):137–147. [PubMed: 12859163]

2. Levey AS, Eckardt KU, Tsukamoto Y, Levin A, Coresh J, Rossert J, Zeeuw Dd, Hostetter TH, Lameire N, Eknoyan G. Definition and classification of chronic kidney disease: A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). Kidney Int. 2005; 67 (6): 2089–2100. [PubMed: 15882252]

3. National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: Evaluation, classification, and stratification. Am J Kidney Dis. 2002; 39(2):S1–S266. [PubMed: 11904577]

4. National Kidney Disease Education Program . [Accessed February 9, 2006] Information of Health Professionals. National Institute for Diabetes and Digestives and Kidney Disease. 2004. http:// www.nkdep.nih.gov/labprofessionals/index.htm

5. Levey AS. Measurement of renal function in chronic renal disease. Kidney Int. 1990; 38:167–184. [PubMed: 2200925]

6. Jones C, McQuillan G, Kusek J, Eberhardt M, Herman W, Coresh J, Salive M, Jones C, Agodoa L. Serum creatinine levels in the US population: Third National Health and Nutrition Examination Survey. Am J Kidney Dis. 1998; 32:992–999. (erratum: Am J Kidney Dis 2000;2035:2178). [PubMed: 9856515]

7. Jafar TH, Chaturvedi N, Gul A, Khan AQ, Schmid CH, Levey AS. Ethnic differences and determinants of proteinuria among South Asian subgroups in Pakistan. Kidney Int. 2003; 64 (4): 1437–1444. [PubMed: 12969163]

8. Shemesh O, Golbetz H, Kriss J, Myers B. Limitations of creatinine as a filtration marker in glomerulopathic patients. Kidney Int. 1985; 28:830–838. [PubMed: 2418254]

9. Myers GL, Miller WG, Coresh J, Fleming J, Greenberg N, Greene T, Hostetter T, Levey AS, Panteghini M, Welch M, Eckfeldt JH. Recommendations for improving serum creatinine measurement: A report from the laboratory working group of the National Kidney Disease Education Program. Clin Chem. 2006; 52 (1):5–18. [PubMed: 16332993]

10. Stevens, LA.; Levey, AS. Measurement of kidney function. In: Singh, AK., editor. Medical Clinics of North America. Vol. 89. W.B. Saunders; Philadelphia: 2005. p. 457-473.

11. 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]

12. Poggio ED, Wang X, Greene T, Van Lente F, Hall P. Performance of the MDRD and Cockcroft-Gault equations in the estimation of glomerular filtration rate in health and in chronic kidney disease. J Am Soc Nephrol. 2005; 16 (2):459–466. [PubMed: 15615823]

13. Lewis JB, Agodoa L, Cheek D, Greene T, Middleton J, O'Connor D, Akinlou O, Philips R, Sika M, Wright J Jr. for the African American Study of Hypertension and Kidney Disease . Comparison of cross-sectional renal function measurements in African-Americans with hypertensive nephrosclerosis and of primary formulas to estimate glomerular filtration rate. Am J Kidney Dis. 2001; 38 (4):744–753. [PubMed: 11576877]

14. Rule AD, Larson TS, Bergstralh EJ, Slezak JM, Jacobsen SJ, Cosio FG. Using serum creatinine to estimate glomerular filtration rate: accuracy in good health and in chronic kidney disease. Ann Intern Med. 2004; 141 (12):929–937. [PubMed: 15611490]

15. Gonwa TA, Jennings L, Mai ML, Stark PC, Levey AS, Klintmalm GB. Estimation of glomerular filtration rates before and after orthotopic liver transplantation: evaluation of current equations. Liver Transpl. 2004; 10 (2):301–309. [PubMed: 14762871]

16. Froissart M, Rossert J, Jacquot C, Paillard M, Houillier P. Predictive performance of the modification of diet in renal disease and Cockcroft-Gault equations for estimating renal function. J Am Soc Nephrol. 2005; 16 (3):763–773. [PubMed: 15659562]

17. Hallan S, Assberg A, Lindberg M, Johnsen H. Validation of the Modification of Diet in Renal Disease Formula for estimating GFR with special emphasis on calibration of the serum creatinine assay. Am J Kidney Dis. 2004; 44 (1):84–93. [PubMed: 15211442]

18. Lin J, Knight E, Hogan ML, Singh A. A comparison of prediction equations for estimating glomerular filtration rate in adults without kidney disease. J Am Soc Nephrol. 2003; 14:2573–2580. [PubMed: 14514734]

19. Bostom A, Kronenberg F, Ritz E. Predictive performance of renal function equations for patients with chronic kidney disease and normal serum creatinine levels. J Am Soc Nephrol. 2002; 13 (8):2140–2144. [PubMed: 12138147]

20. Zuo L, Ma YC, Wang M, Zhou Y, Xu GB, Wang HY. Application of glomerular filtration rate estimating equations in Chinese patients with chronic kidney disease. Am J Kidney Dis. 2005; 45 (3):463–472. [PubMed: 15754268]

21. Gaspari F, Ferrari S, Stucchi N, Centemeri E, Carrara F, Pellegrino M, Gherardi G, Gotti E, Segoloni G, Salvadori M, Rigotti P, Valente U, Donati D, Sandrini S, Sparacino V, Remuzzi G, Perico N. Performance of Different Prediction Equations for Estimating Renal Function in Kidney Transplantation. Am J Trans. 2004; 4 (11):1826–1835.

22. Lamb E, Webb M, Simpson D, Coakley A, Newman D, O'Riordan S. Estimation of glomerular filtration rate in older patients with chronic renal insufficiency: is the modification of diet in renal disease formula an improvement? J Am Geriatr Soc. 2003; 51 (7):1012–1017. [PubMed: 12834524]

23. Vervoort G, Willems H, Wetzels J. Assessment of glomerular filtration rate in healthy subjects and normoalbuminuric diabetic patients: validity of a new (MDRD) prediction equation. Nephrol Dial Transplant. 2002; 17 (11):1909–1913. [PubMed: 12401845]

24. Skluzacek P, Szewc R, Nolan C, Riley D, Lee S, Pergola P. Prediction of GFR in liver transplant candidates. Am J Kidney Dis. 2003; 42 (6):1169–1176. [PubMed: 14655188]

25. Ibrahim H, Mondress M, Tello A, Fan Y, Koopmeiners J, Thomas W. An alternative formula to the Cockcroft-Gault and the Modification of Diet in Renal Diseases formulas in predicting GFR in individuals with type 1 diabetes. J Am Soc Nephrol. 2005; 16 (4):1051–1060. [PubMed: 15716336]

26. Rigalleau V, Lasseur C, Perlemoine C, Barthe N, Raffaitin C, Liu C, Chauveau P, Baillet-Blanco L, Beauvieux MC, Combe C, Gin H. Estimation of glomerular filtration rate in diabetic subjects: Cockcroft formula or modification of Diet in Renal Disease study equation? Diabetes Care. 2005; 28 (4):838–843. [PubMed: 15793182]

27. Poge U, Gerhardt T, Palmedo H, Klehr H-U, Sauerbruch T, Woitas RP. MDRD Equations for Estimation of GFR in Renal Transplant Recipients. Am J Trans. 2005; 5 (6):1306–1311.

28. Grubb A, Bjork J, Lindstrom V, Sterner G, Bondesson P, Nyman U. A cystatin C-based formula without anthropometric variables estimates glomerular filtration rate better than creatinine clearance using the Cockcroft-Gault formula. Scand J Clin Lab Invest. 2005; 65 (2):153–162. [PubMed: 16025838]

29. Fehrman-Ekholm I, Skeppholm L. Renal function in the elderly (≫70 years old) measured by means of iohexol clearance, serum creatinine, serum urea and estimated clearance. Scand J Clin Lab Invest. 2005; 38 (1):73–77.

30. Poggio ED, Nef PC, Wang X, Greene T, Van Lente F, Dennis VW, Hall PM. Performance of the Cockcroft-Gault and modification of diet in renal disease equations in estimating GFR in ill hospitalized patients. Am J Kidney Dis. 2005; 46 (2):242–252. [PubMed: 16112042]

31. Verhave JC, Fesler P, Ribstein J, du Cailar G, Mimran A. Estimation of renal function in subjects with normal serum creatinine levels: Influence of age and body mass index. Am J Kidney Dis. 2005; 46 (2):233–241. [PubMed: 16112041]

32. Coresh J, Stevens LA. Kidney function estimating equations: Where do we stand? Curr Opin Neph Hyper. 2006; 15 (3):276–284.

33. Stevens LA, Coresh J, Deysher AE, Feldman HI, Lash JP, Nelson R, Rahman M, Schmid CH, Zhang Y, Greene T, Levey AS. Evaluation of the MDRD Study equation in a large diverse population. J Am Soc Nephrol. 2007; 18 (10):2749–2757. [PubMed: 17855641]

34. Madero M, Sarnak MJ, Stevens LA. Serum cystain C as a marker of glomerular filtration rate. Curr Opin Neph Hyper. 2006; 15 (6):610–616.

35. Rule AD, Bergstralh EJ, Slezak JM, Bergert J, Larson TS. Glomerular filtraton rate estimated by cystatin C among different clinical presentations. Kidney Int. 2006; 69 (2):399–405. [PubMed: 16408133]

36. Stevens LA, Coresh J, Schmid CH, Feldman HI, Froissart M, Kusek J, Rossert J, Van Lente F, Bruce RD 3rd, Zhang YL, Greene T, Levey AS. 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]

37. Miller W, Myers G, Ashwood E, Killeen A, Wang E, Thienpont L, Siekmann L. Creatinine measurement: State of the art in accuracy and interlaboratory harmonization. Arch Pathol Lab Med. 2005; 129 (3):297–304. [PubMed: 15737021]

38. Flodin M, Hansson LO, Larsson A. Variations in assay protocol for the Dako cystatin C method may change patient results by 50% without changing the results for controls. Clin Chem Lab Med. 2006; 44 (12):1481–1485. [PubMed: 17163826]

39. Clase CM, Garg AX, Kiberd BA. Reply from the authors: Estimating the prevalence of low glomerular filtration rate requires attention to the creatinine calibration assay. J Am Soc Nephrol. 2002; 13:2812–2816.

40. Coresh J, Astor B, McQuillan G, Kusek J, Greene T, Van Lente F, Levey A. Calibration and random variation of the serum creatinine assay as critical elements of using equations to estimate the glomerular filtration rate. Am J Kidney Dis. 2002; 39:920–929. [PubMed: 11979335]

41. Coresh J, Eknoyan G, Levey AS. Estimating the prevalence of low glomerular filtration rate requires attention to the creatinine assay calibration. J Am Soc Nephrol. 2002; 13 (11):2811–2812. author reply 2812–2816. [PubMed: 12397055]

42. Murthy K, Stevens LA, Stark PC, Levey AS. Variation in serum creatinine assay calibration: A practical application to glomerular filtration rate estimation. Kidney Int. 2005; 68 (4):1884–1887. [PubMed: 16164667]

43. Stevens LA, Manzi J, Levey AS, Chen JL, Deysher AE, Ojo A, Poggio E, Steffes M, Zhang Y, Van Lente F, Coresh J. 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]

44. Greene T. Effect of source population on the relationship of GFR estimates with "true GFR". J Am Soc Nephrol. 2006; 17:142A.

45. Stevens LA, Coresh J, Greene T, Levey AS. Assessing kidney function - measured and estimated glomerular filtration rate. N Engl J Med. 2006; 354:2473–2483. [PubMed: 16760447]

46. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1 (8476):307–310. [PubMed: 2868172]

47. Chakravarty, AG. Surrogate Markers - Their Role in Regulatory Decision Process. FDA; 2003. http://www.fda.gov/cder/Offices/Biostatistics/Chakravarty_376/index.htm [Accessed Feburary 12 2006]

48. Bjork J, Back SE, Sterner G, Carlson J, Lindstrom V, Bakoush O, Simonsson P, Grubb A, Nyman U. Prediction of relative glomerular filtration rate in adults: new improved equations based on Swedish Caucasians and standardized plasma-creatinine assays. Scand J Clin Lab Invest. 2007; 67 (7):678–695. [PubMed: 17852799]
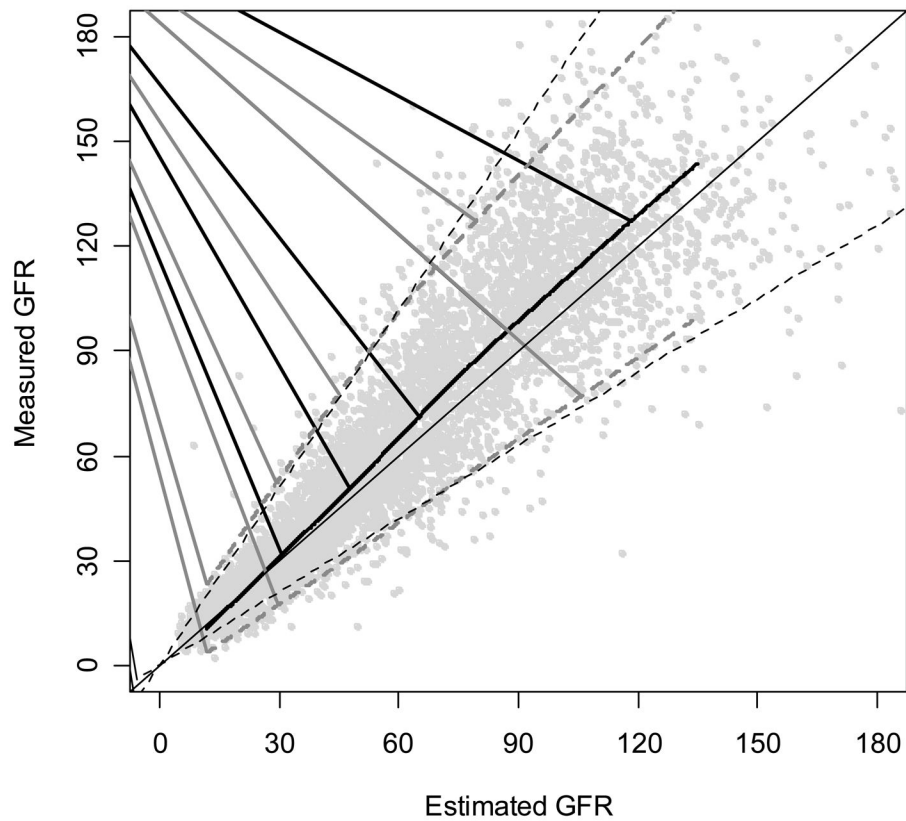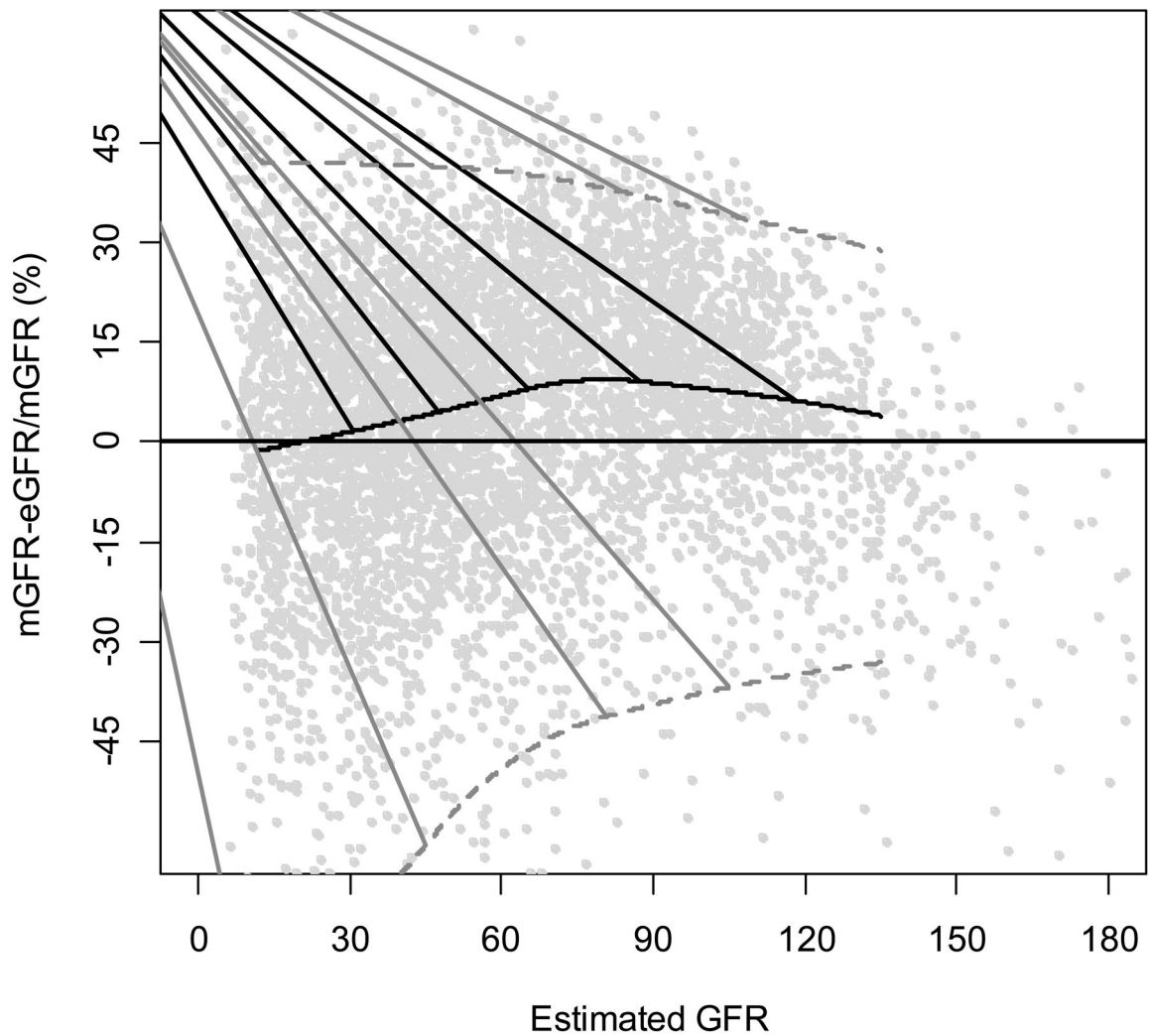
**Figure 1. Comparison of MDRD Study Equations to Measured GFR**

Solid horizontal line is the line of identity. Black curved line is a smooth curve through the points and was created using 95% of the data. Dashed grey lines are quantile regressions of the 5th and 90th percentiles of the differences which measures precision. The grey dotted vertical line indicates 30% errors [45].
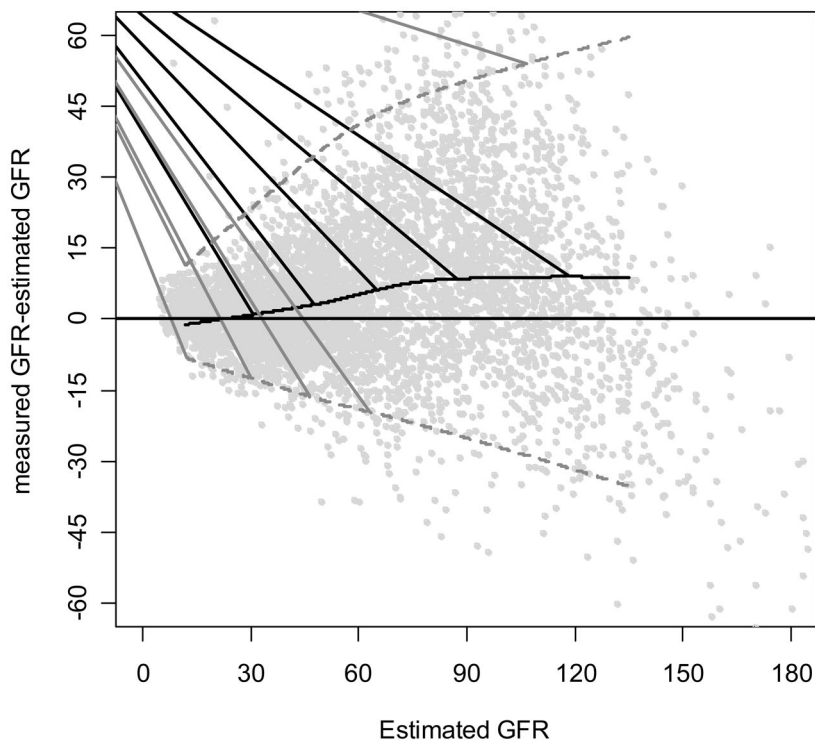
Figure 2. Difference Between the MDRD Study Equations by Level of Estimated GFR
Difference is calculated as [(measured GFR)-(estimated GFR)]. Solid horizontal line
indicates no difference. Solid black curve is a non-linear regression of the mean difference,
which measures bias. Black curved line is a smooth curve through the points and was
created using 95% of the data. Dashed grey lines are quantile regressions of the 5th and 90th
percentiles of the differences which measures precision [33].

Figure 3. Percent Difference of the MDRD Study Equations by Level of Estimated GFR
Percent difference is calculated as [(measured GFR)-(estimated GFR)]/measured GFR. Solid horizontal line indicates no difference. Solid black curve is a non-linear regression of the mean difference, which measures bias. Black curved line is a smooth curve through the points and was created using 95% of the data. Dashed grey lines are quantile regressions of the 5th and 90th percentiles of the differences, which measures precision.
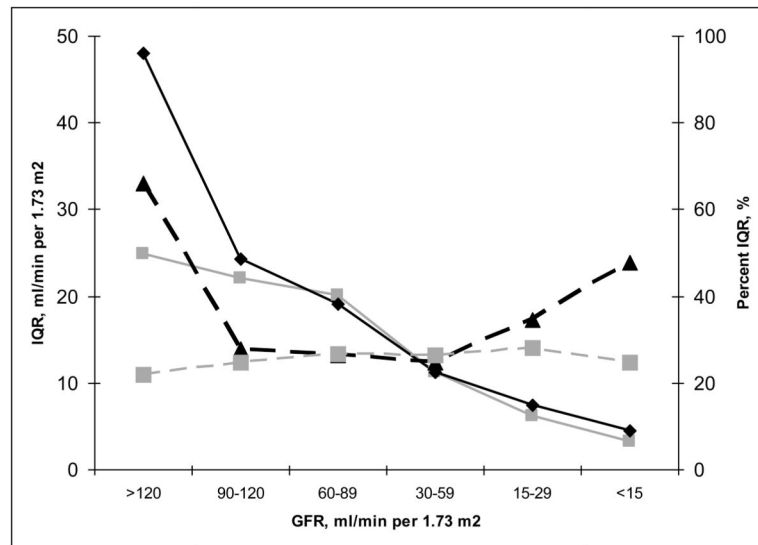
Figure 4. Difference Between Arithmetic and Relative Metrics: Precision for Cystatin vs Creatinine Estimating Equations

Interquartile range (solid lines) and percent interquartile range (dashed lines) are plotted for cystatin, age, sex and race (black lines) compared to creatinine, age, sex, and race (gray lines) (equations 2 and 3 in Table 2, respectively).
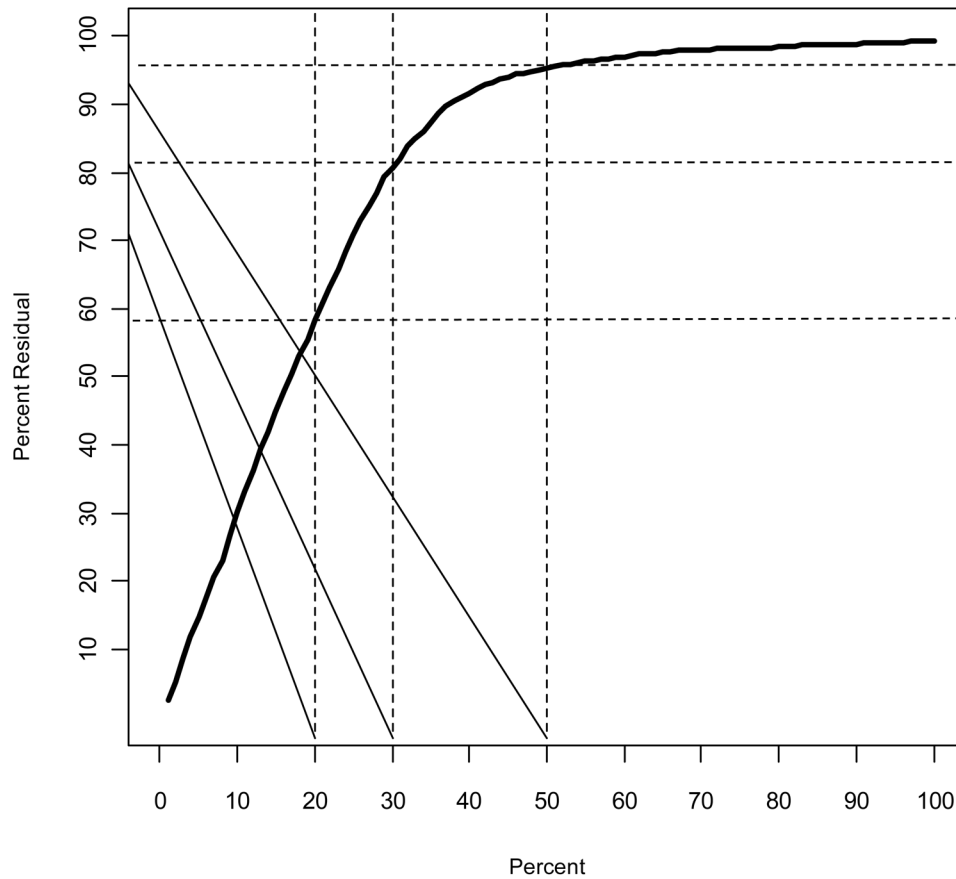
Figure 5. Distribution Function for the Relative Difference between Measured and Estimated GFR for the MDRD Study Equation in the Pooled Creatinine Dataset

The y-axis reflects the probability that a difference is less than or equal to the value on the x-axis. The intersection of the dashed horizontal and vertical lines highlight different examples. The right-most and top vertical and horizontal lines, respectively, show that 95% of individuals have a relative difference of 50% or less, compared to the left-most and bottom vertical and horizontal lines, respectively, that show that 58% of individuals have a relative difference of 20% or less.

Table 1

Factors Affecting Errors in GFR estimates

Factors	Bias	Precision
GFR		
Biological variation in GFR		X
Mean level of GFR	X	
GFR measurement error		
Systematic differences among exogenous filtration markers	X	
Random errors in urine collection		X
Edematous conditions for plasma clearances	X	X
Endogenous filtration markers		
Differences in assays	X	
Non-GFR determinants of endogenous filtration markers	X	X
Selection bias	X	

Table 2

Performance of Cystatin C and Creatinine-Based Estimating Equations Developed in 3134 People with CKD

Model	Difference		Percent Difference		P30 (95% CI)	RMSE (95% CI)
	Median (95% CI)	IQR (95% CI)	Median (95% CI)	IQR (95% CI)		
External Validation (n=438)						
1. Cystatin C	−3 (−3, −2)	8 (7, 9)	−10 (−13, −7)	31 (28, 36)	73 (72,74)	0.264 (0.239, 0.289)
2. Cystatin C, age, sex, and race	−2 (−2, −1)	8 (7, 9)	−6 (−8, −4)	30 (26, 32)	79 (78,80)	0.248 (0.223, 0.271)
3. Creatinine, age, sex, and race	2 (1,3)	8 (7, 9)	7 (4,9)	25 (22, 29)	84 (83,85)	0.229 (0.210, 0.247)
4. Cysatin C and creatinine, age, sex, and race	0 (0,1)	7 (6, 8)	1 (−1,3)	22 (20, 25)	90 (89,91)	0.193 (0.174, 0.211)

Note: The difference is calculated as measured GFR - estimated GFR. Units are in mL/min/1.73 m². Percent difference calculated as (measured GFR - estimated GFR)/measured GFR, and units are in percent. Median values measure bias and IQRs measure precision. To convert GFR from mL/min/1.73 m² to mL/s/1.73 m², multiply by 0.01666. Used with permission [36].

Abbreviations: P30, percentage of estimated GFR within 30% of mGFR; IQR, interquartile range; RMSE, root mean square error; CI, confidence interval; MDRD, Modification of Diet in Renal Disease; eGFR, estimated glomerular filtration; mGFR, measured glomerular filtration rate.

*
The 95% CI around the estimates of bias, IQR, P30, and RMSE provides the range of values likely to include parameter in 95% of circumstances. Comparison of CIs around any metric for 2 equations provides information about difference between them. If the 2 metrics were the same, the chance that the upper and lower bounds of the CIs would overlap is less than 5%; therefore, it can be concluded that they are statistically significantly different.

Table 3

Criteria for Bias, Precision and Accuracy

Criteria	Metric	Definition
Bias	Median difference	Measured GFR-Estimated GFR
	Median percent difference	(Measured GFR-Estimated GFR)/Measured GFR
Precision	SD difference	
	IQR difference	Interquartile range of (Measured GFR-Estimated GFR)
	IQR % difference	Interquartile range of [(Measured GFR-Estimated GFR)/Measured GFR]* 100
Accuracy	Median absolute	Median of the absolute value of eGFR-mGFR
	P_{30}	Percent of estimates within 30% of measured GFR
	RMSE	Square root of mean (log Measured GFR-log Estimated GFR)^2

RMSE measures precision when bias is 0 (development dataset)

[†] IQR is the width of the 25th to 75th percentile.