

Research Article

Cite this article: Zhu X, Shafi T, Norris KC, Simino J, Shrestha S, Mosley TH, Griswold ME, and Lirette ST. Comparing methods for glomerular filtration rate estimation. *Journal of Clinical and Translational Science* 9: e148, 1–8. doi: [10.1017/cts.2025.10057](https://doi.org/10.1017/cts.2025.10057)

Received: 19 December 2024

Revised: 18 April 2025

Accepted: 13 May 2025

Keywords:

GFR; estimation; equation; serum creatinine; performance

Corresponding author:

X. Zhu; Email: xzhu3@umc.edu

Comparing methods for glomerular filtration rate estimation

Xiaoqian Zhu¹ , Tariq Shafi², Keith C. Norris³ , Jeannette Simino¹, Srishti Shrestha⁴, Thomas H. Mosley⁴, Michael E. Griswold⁴ and Seth T. Lirette¹

¹Department of Data Science, University of Mississippi Medical Center, Jackson, MS, USA; ²Baylor Scott and White Health, Temple, TX, USA; ³Division of General Internal Medicine and Health Services Research, University of California Los Angeles, Los Angeles, CA, USA and ⁴The Memory Impairment and Neurodegenerative Dementia (MIND) Center, University of Mississippi Medical Center, Jackson, MS, USA

Abstract

Background: The glomerular filtration rate (GFR), estimated from serum creatinine (SCr), is widely used in clinical practice for kidney function assessment, but SCr-based equations are limited by non-GFR determinants and may introduce inaccuracies across racial groups. Few studies have evaluated whether advanced modeling techniques enhance their performance. **Methods:** Using multivariable fractional polynomials (MFP), generalized additive models (GAM), random forests (RF), and gradient boosted machines (GBM), we developed four SCr-based GFR-estimating equations in a pooled data set from four cohorts ($n = 4665$). Their performance was compared to that of the refitted linear regression-based 2021 CKD-EPI SCr equation using bias (median difference between measured GFR [mGFR] and estimated GFR [eGFR]), precision, and accuracy metrics (e.g., P10 and P30, percentage of eGFR within 10% and 30% of mGFR, respectively) in a pooled validation data set from three additional cohorts ($n = 2215$). **Results:** In the validation data set, the greatest bias and lowest accuracy, were observed in Black individuals for all equations across subgroups defined by race, sex, age, and eGFR. The MFP and GAM equations performed similarly to the refitted CKD-EPI SCr equation, with slight improvements in P10 and P30 in subgroups including Black individuals and females. The GBM and RF equations demonstrated smaller biases, but lower accuracy compared to other equations. Generally, differences among equations were modest overall and across subgroups. **Conclusions:** Our findings suggest that advanced methods provide limited improvement in SCr-based GFR estimation. Future research should focus on integrating novel biomarkers for GFR estimation and improving the feasibility of GFR measurement.

Introduction

The glomerular filtration rate (GFR), estimated from serum creatinine (SCr), is a standard tool in diagnosing, staging, and managing chronic kidney disease (CKD) in clinical practice. Over the years, many SCr-based equations have been developed, with current clinical guidelines recommending the race-free 2021 CKD-EPI SCr equation [1]. However, the performance of these equations is limited by non-GFR determinants that influence SCr, such as muscle mass and nutritional intake [2,3]. In the study that developed the 2021 CKD-EPI equations, external validation has shown that the CKD-EPI SCr equation may introduce inaccuracies in GFR estimation for both Black and non-Black populations, and lead to differential bias between race groups (namely, overestimation in non-Black individuals and underestimation in Black individuals). Furthermore, the 2021 CKD-EPI SCr equation may be less precise than the 2009 race-based CKD-EPI SCr equation [4].

To address these limitations, guidelines from the National Kidney Foundation (NKF) and the American Society of Nephrology (ASN) recommend using the CKD-EPI creatinine-cystatin C combined equation for further confirmation and advocate for further investigation into other endogenous filtration markers that may enhance GFR estimation [1]. Several studies have explored the potential of novel biomarkers like beta-trace protein, β_2 -microglobulin, and citrulline for estimating GFR [5,6], but SCr remains a valuable marker for GFR estimation due to its low cost, widespread availability, and routine inclusion in the basic metabolic panel. While it is important to continue exploring alternatives to SCr, until guidelines update their recommendations for GFR estimation, the SCr-based 2021 CKD-EPI equation will likely continue to be the standard in clinical settings. Given its clinical relevance, we sought to investigate whether SCr-based equations can be improved using advanced modeling techniques without incorporating additional biomarkers, potentially offering a cost-effective approach to enhance clinical practice.

It is well-known that SCr has a nonlinear relationship with measured GFR (mGFR) [7]. Existing SCr-based equations typically employ linear regression methods with piecewise spline

© The Author(s), 2025. Published by Cambridge University Press on behalf of Association for Clinical and Translational Science. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.



Clinical Research
FORUM
Analysis. Advocacy. Action.

terms for SCr to capture the nonlinearity, using log-transformed values for both mGFR and SCr [4]. However, several advanced smoothing and machine learning techniques can model complex nonlinear associations and identify the most appropriate relationships between the outcome and predictors based on the data itself, without assuming a predefined functional form. We, therefore, assessed whether these advanced methods could improve equation performance by more accurately capturing the nonlinear relationship between SCr and mGFR compared to the traditional linear regression-based approach. Specifically, we developed four new SCr-based equations using advanced approaches and compared their performance to the refitted linear regression-based 2021 CKD-EPI SCr equation.

Materials and methods

Data sources

This investigation utilized existing data from seven study cohorts that had mGFR data. For the development of new equations, we used the Genetic Epidemiology Network of Arteriopathy Study (GENOA) ($N = 1010$) [8], African American Study of Kidney Disease and Hypertension Study (AASK) ($N = 1807$) [9], Modification of Diet in Renal Disease (MDRD) Study ($N = 1628$) [10], and Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) ($N = 168$) [11]. For external validation, we utilized three studies: the Epidemiology of Coronary Artery Calcification (ECAC) cohort study ($N = 406$) [12], the Assessing Long Term Outcomes in Living Kidney Donors (ALTOLD) ($N = 386$) [13], and the Chronic Renal Insufficiency Cohort (CRIC) Study ($N = 1423$) [14]. Additional information about these cohorts can be found in the Supplemental Materials (Section: Details of Study Cohorts).

Measured GFR, serum creatinine, and covariates

Details of mGFR protocols and laboratory measurements were published elsewhere [8,10,11,13–16]. GFR was measured using urinary clearance of non-radiolabeled iothalamate in GENOA, ECAC, and CRISP; radiolabeled ^{125}I -iothalamate urinary clearances in MDRD, AASK, and CRIC; and plasma clearance of iothexol in ALTOLD. mGFRs from all studies were standardized to 1.73 m^2 of body surface area. SCr measurement was standardized in all studies. Age, sex, and race were self-reported.

Similar to the guideline-recommended 2021 CKD-EPI SCr equation, all models included age, sex, and SCr as predictors, with mGFR as the outcome. mGFR was expressed in $\text{ml/min}/1.73\text{m}^2$, SCr in mg/dL , and age in years for equation development. mGFR was log-transformed due to its high variability and to ensure positive predictions. SCr was kept on its original scale in all models except for the linear regression model used to refit the CKD-EPI SCr equation, where it was log-transformed.

Statistical analysis

We selected four distinct methods to estimate GFR – each with unique strengths in capturing nonlinear relationships between predictors and the outcome – to estimate GFR: multivariable fractional polynomials (MFP), generalized additive models (GAM), random forests (RF), and gradient boosted machines (GBM). MFP employs built-in variable selection and applies appropriate power transformations to continuous predictor variables when nonlinearity is identified [17]. GAM utilizes smooth functions, represented

by penalized regression splines, to flexibly model nonlinear relationships when present and can effectively handle complex interactions between variables [18]. RF and GBM are ensemble machine learning techniques that combine predictions from multiple models to improve the overall accuracy and robustness of predictions. RF aggregates predictions from multiple deep decision trees, each trained on a random subset of the data, ensuring diversity among the trees and reducing overfitting [19]. It captures nonlinearity by allowing individual decision trees to model different aspects of the relationship between dependent and independent variables. GBM uses a stochastic gradient boosting strategy, sequentially building a series of shallow decision trees, where each tree corrects the residuals of the previous ones [20,21]. The technical details of these methods are provided in the Supplemental Materials (Technical Notes of Statistical Methods).

To compare these four methods with the linear regression-based CKD-EPI SCr equation, we refitted the 2021 CKD-EPI SCr equation using a linear regression model (hereafter referred to as LM). We preserved the same expression for the linear predictors as in the original 2021 CKD-EPI SCr equation, which included a linear effect for age and sex-specific linear splines for SCr, with cut-off values of 0.9 for females and 0.7 for males.

The MFP, GAM, RF, and GBM methods were implemented using R packages *mfp*, *mgcv*, *RandomForest*, and *gbm*, respectively, while the LM was implemented using base R. R packages *pdp* and *visreg* were used to visualize the function form of equations.

External validation

We evaluated the performance of the new GFR-estimating equations in the pooled external validation data set using several population-level and individual-level metrics. Bias, the population-level systematic difference, was evaluated as the median difference between mGFR and estimated GFR (eGFR) (i.e., mGFR minus eGFR). Precision was assessed as the interquartile range of the difference. Population-level accuracy was assessed as root mean square error (RMSE) of log mGFR and log eGFR, mean absolute error (MAE) of the difference between mGFR and eGFR, percentage of eGFR within 10% of mGFR (P10), and percentage of eGFR within 30% of mGFR (P30) [22,23].

We examined individual-level accuracy by calculating 95% prediction intervals (PIs) of mGFR at different eGFR values. The lower and upper bounds of these intervals were determined using the 2.5th and 97.5th percentiles of mGFR, predicted from quantile regression models [24,25]. Similarly, we constructed the 50% prediction interval (PI) using the 25th and 75th percentiles of mGFR obtained from quantile regression models.

We obtained 95% confidence intervals (CIs) (2.5th percentile, 97.5th percentile) for all population-level metrics using 2000 bootstrap samples (see details in Supplemental Materials: Bootstrapping Methods to Calculate Obtain Estimates (95% CI) of Metrics for Equation Performance) [26]. All analyses were performed in R (R Foundation for Statistical Computing, Vienna, Austria) and STATA/SE version 18.0 (StataCorp LLC, College Station, TX). Exemplary R code is provided in Supplementary Materials.

Results

Characteristics of study participants

Table 1 shows the characteristics of the participants in both the development and external validation data sets. In the development data set, the mean $\pm$ SD age of participants was

Table 1. Characteristics of study participants

	Development data set			External validation data set		
	Total <i>n</i> = 4665 (100.0%)	Non-Black <i>n</i> = 2149 (46.1%)	Black <i>n</i> = 2516 (53.9%)	Total <i>n</i> = 2215 (100.0%)	Non-Black <i>n</i> = 1680 (75.8%)	Black <i>n</i> = 535 (24.2%)
<i>N</i>						
Age, years	54.0 (12.9)	52.3 (14.1)	55.4 (11.7)	55.6 (13.5)	55.6 (13.9)	55.7 (12.3)
Categories, <i>n</i> (%)						
<40	716 (15.3)	462 (21.5)	254 (10.1)	352 (15.9)	278 (16.5)	74 (13.8)
40–65	2946 (63.2)	1255 (58.4)	1691 (67.2)	1255 (56.7)	929 (55.3)	326 (60.9)
>65	1003 (21.5)	432 (20.1)	571 (22.7)	608 (27.4)	473 (28.2)	135 (25.2)
Female, <i>n</i> (%)	2099 (45.0)	992 (46.2)	1107 (44.0)	1083 (48.9)	831 (49.5)	252 (47.1)
Body mass index, kg/m²	29.5 (6.2)	28.0 (5.4)	30.7 (6.6)	30.1 (6.4)	29.2 (6.1)	32.9 (6.7)
Categories, <i>n</i> (%)						
Underweight	37 (0.8)	18 (0.8)	19 (0.8)	10 (0.5)	9 (0.5)	1 (0.2)
Normal weight	1092 (23.5)	663 (31.0)	429 (17.1)	485 (21.9)	425 (25.3)	60 (11.2)
Overweight	3225 (69.3)	1391 (65.1)	1834 (73.0)	1541 (69.6)	1149 (68.5)	392 (73.3)
Obese	297 (6.4)	65 (3.0)	232 (9.2)	177 (8.0)	95 (5.7)	82 (15.3)
Body surface area, m²	2.0 (0.2)	1.9 (0.2)	2.0 (0.2)	2.0 (0.3)	2.0 (0.3)	2.1 (0.3)
Measured GFR, ml/min/1.73 m²	57.8 (28.3)	54.6 (31.0)	60.6 (25.4)	62.1 (27.6)	66.7 (28.1)	47.6 (20.0)
Categories, <i>n</i> (%)						
≥ 90	640 (13.7)	325 (15.1)	315 (12.5)	434 (19.6)	414 (24.6)	20 (3.7)
60–89	1454 (31.2)	513 (23.9)	941 (37.4)	644 (29.1)	529 (31.5)	115 (21.5)
45–59	871 (18.7)	345 (16.1)	526 (20.9)	432 (19.5)	301 (17.9)	131 (24.5)
30–44	815 (17.5)	401 (18.7)	414 (16.5)	432 (19.5)	272 (16.2)	160 (29.9)
<30	885 (19.0)	565 (26.3)	320 (12.7)	273 (12.3)	164 (9.8)	109 (20.4)
Serum creatinine, mg/dL	1.6 (1.0)	1.7 (1.1)	1.6 (0.9)	1.4 (0.7)	1.3 (0.6)	1.9 (0.7)

Note: GFR = glomerular filtration rate; mGFR = measured glomerular filtration rate. Data are presented as mean (standard deviation) for continuous variables and as *n* (%) for categorical variables.

54 ± 13 years; 2099 (45%) participants were female, and 2516 (54%) self-reported Black race. The mean mGFR was 58 ± 28 ml/min/1.73 m², and 2571 (55%) had mGFR < 60 ml/min/1.73 m². The external validation data set showed a slightly higher mean mGFR (62 ± 28 ml/min/1.73 m²), an older mean age (56 ± 14 years), a higher proportion of females (49%), and a lower proportion of Black race (24%). Characteristics of participants from each cohort are shown in Supplemental Table 1.

Formulation of equations

MFP equation

Nonlinear relationships of both SCr and age with log(mGFR) were selected by the MFP model based on the deviance criterion. The following equation was used to estimate GFR:

$$\exp \left[1.869 - 2.389 * \frac{1}{\text{SCr}} + 6.8489 * \text{SCr}^{-0.5} + 1.2049 * \left(\frac{\text{age}}{100} \right)^{0.5} - 1.610 * \left(\frac{\text{age}}{100} \right)^{0.5} * \log \left(\frac{\text{age}}{100} \right) + 0.294 \text{ if male} \right]$$

Supplemental Figure 1 displays the relationships of MFP-based eGFR with SCr and age.

GAM equation

Nonlinear relationships with log(mGFR) for both SCr and age were statistically supported in the GAM-based equation; Penalized cubic regression splines were used in the GAM model, as they resulted in the lowest generalized cross-validation scores compared to other available spline functions. The equation to estimate GFR is:

$\exp [3.755293 + f(\text{SCr}) + f(\text{age}) + 0.291015 \text{ if male}]$ where *f* represents the smooth functions.

However, unlike MFP, it is difficult to transcribe the mathematical forms of the smooth terms concisely. Supplemental Figure 2 shows the relationships of GAM-based eGFR with SCr and age.

RF equation

In the RF approach, a grid search was conducted across various parameter combinations, including the number of trees and the minimum size of terminal nodes, to identify the optimal parameters for the final model, based on the lowest out-of-bag RMSE. The final model used 1000 trees with a node size of nine. Two randomly sampled variables were used to perform each of the splits because there were only three predictors in total. In contrast to the MFP and GAM models, RF models are purely designed for

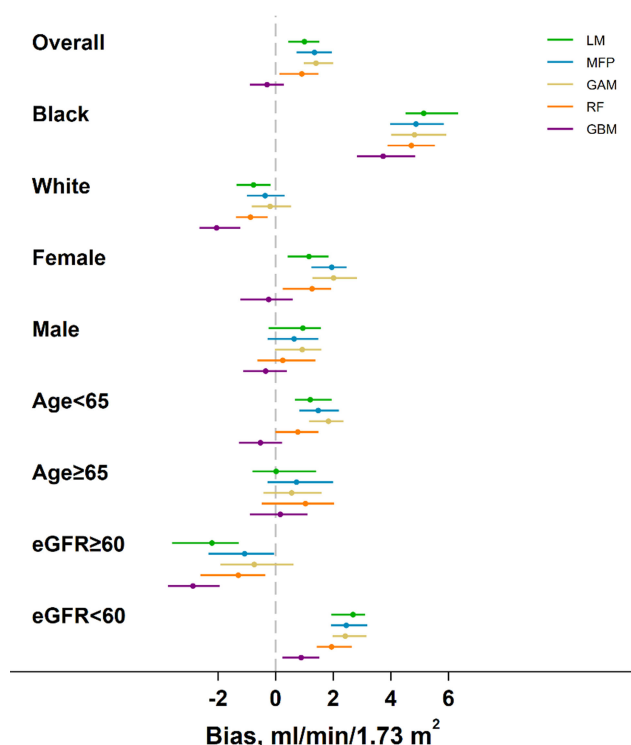


Figure 1. Bias of equations overall and by subgroups in the external validation data set. Shows the bias of all equations overall and across subgroups. The dots are point estimates and the horizontal lines are 95% confidence intervals. The vertical dashed line represents the unbiased reference line, with estimates closer to 0 indicating better performance. eGFR based on the 2021 CKD-EPI SCr equation was used to define the subgroups with eGFR < 60 ml/min/1.73 m² and eGFR ≥ 60 ml/min/1.73 m².

Note: Bias was defined as the median of the differences between mGFR and eGFR for each individual in the sample (mGFR minus eGFR); GFR = glomerular filtration rate; eGFR = estimated GFR; mGFR = measured GFR; CKD-EPI = Chronic Kidney Disease Epidemiology Collaboration; SCr = serum creatinine; LM = the refitted 2021 CKD-EPI SCr equation using a linear regression model; MFP = the equation based on the multivariable fractional polynomial model; GAM = the equation based on the generalized additive model; RF = the equation based on random forests; GBM = the equation based on gradient boosted machines.

prediction and thus do not provide interpretable equations. Supplemental Figure 3 shows how RF-based eGFR changes with SCr or age.

GBM equation

The GBM equation was developed based on a boosted regression model. A grid search, similar to that used for RF models, was performed to determine the parameters for the final model, including the total number of trees, the maximum depth of each tree, and the shrinkage parameter, based on the RMSE. A total of 1386 trees were included in the final model.

Like the RF equation, predictions can be easily generated from the GBM model, although interpretable equations are not available. Supplemental Figure 4 illustrates how the eGFR changes with SCr or age.

Refitted CKD-EPI SCr 2021 (LM) equation

The 2021 CKD-EPI SCr equation was refitted using our development data set. The refitted equation, referred to as the LM equation, was defined as follows:

$$135 * \left(\frac{SCr}{0.7}\right)^{-0.135} * 0.9940^{Age} \text{ if } SCr \leq 0.7, \text{ female}$$

$$135 * \left(\frac{SCr}{0.7}\right)^{-1.159} * 0.9940^{Age} \text{ if } SCr > 0.7, \text{ female}$$

$$139 * \left(\frac{SCr}{0.9}\right)^{-0.110} * 0.9940^{Age} \text{ if } SCr \leq 0.9, \text{ male}$$

$$139 * \left(\frac{SCr}{0.9}\right)^{-1.159} * 0.9940^{Age} \text{ if } SCr > 0.9, \text{ male}$$

Supplemental Figure 5 demonstrates that the refitted equation produced predictions generally consistent with the original CKD-EPI SCr equation across various SCr values.

Performance of equations in the external validation data set

Overall, all equations underestimated mGFR with small biases ranging from 0.9 to 1.4, except for the GBM equation, which resulted in a bias that was not significantly different from zero (−0.3, 95%CI: −0.9, 0.3) (Table 2). The LM, MFP, and GAM equations demonstrated similar accuracy, as measured by both P10 and P30, while the GBM and RF equations showed lower accuracy (Table 2). Imprecision, RMSE, and MAE followed a similar pattern as P10 and P30, with GBM and RF demonstrating larger imprecision and errors than the other equations (Table 2). Despite these small differences, the overall performance of the equations was generally similar, as the 95% confidence intervals overlapped for most of the metrics.

The performance of these equations across subgroups of race, sex, age, and eGFR (calculated using the CKD-EPI SCr equation) (Supplemental Table 2) mirrored the overall trends. All equations, including the LM equation, exhibited the greatest underestimation biases in the Black subgroup, followed by the eGFR < 60 ml/min/1.73 m² subgroup (Supplemental Table 2, Figure 1). The GBM equation showed the smallest bias in both of these subgroups; however, it exhibited the largest overestimation bias in the White and eGFR ≥ 60 ml/min/1.73 m² subgroups. Other equations demonstrated relatively similar biases within each subgroup. The GBM and RF equations generally yielded lower accuracy, as measured by P10 and P30, within each subgroup when compared to other equations (Supplemental Table 2, Figures 2 and 3). The LM, MFP, and GAM equations performed similarly within most of the subgroups, but the MFP and GAM equations tended to show higher P10 and P30 than the LM equation in Black individuals and females (Figures 2 and 3). Other performance metrics, including imprecision, MAE, and RMSE, followed the pattern of P10 and P30, with GBM and RF showing slightly larger errors compared to the other equations in most subgroups (Supplemental Figure 6).

We further performed the analysis in subgroups based on combinations of race and age (<65 years), sex and age (<65 years), race and eGFR (<60 ml/min/1.73 m²), and sex and eGFR (<60 ml/min/1.73 m²) (Supplemental Tables 3 and 4, Supplemental Figure 7, Supplemental Figure 8). We found that the performances of GBM and RF were similar in these subgroups and comparable to the performance in the overall sample, with generally smaller biases but lower precision and accuracy. The GAM equation had higher P10 and P30 compared to the LM equation for Black individuals across age groups (Supplemental Figure 7B, Supplemental Figure 7C) and eGFR categories (Supplemental Figure 8B, Supplemental Figure 8C). Similarly, the MFP equation showed higher P30 than the LM equation for Black individuals, regardless of age group (Supplemental Figure 7C) or eGFR category (Supplemental Figure 8C), although 95% CIs overlapped for most of the equations across subgroups.

Table 2. Overall performance of estimating equations in the external validation data set

Equation	Bias (95% CI), ml/min/1.73 m ²	Imprecision (95% CI), ml/min/1.73 m ²	P10 (95% CI), %	P30 (95% CI), %	RMSE (95% CI), ml/min/1.73 m ² (log scale)	MAE (95% CI), ml/min/1.73 m ²
LM	1.0 (0.4, 1.5)	15.3 (14.3, 15.9)	38.7 (36.7, 40.8)	83.7 (82.3, 85.3)	0.23 (0.22, 0.24)	10.0 (9.6, 10.4)
MFP	1.3 (0.7, 1.9)	14.8 (14.0, 15.6)	39.3 (37.3, 41.3)	84.3 (82.9, 85.8)	0.23 (0.22, 0.24)	10.0 (9.6, 10.4)
GAM	1.4 (1.0, 2.0)	14.8 (13.9, 15.5)	39.4 (37.4, 41.4)	84.6 (83.2, 86.1)	0.23 (0.22, 0.24)	10.0 (9.6, 10.4)
RF	0.9 (0.1, 1.5)	16.2 (15.4, 17.0)	36.5 (34.6, 38.5)	82.3 (80.7, 83.9)	0.24 (0.23, 0.25)	10.4 (10.0, 10.8)
GBM	−0.3 (−0.9, 0.3)	15.8 (15.0, 16.7)	37.0 (35.1, 39.0)	81.3 (79.8, 82.9)	0.24 (0.23, 0.25)	10.4 (10.0, 10.8)

Note: RMSE = root mean square error; MAE = mean absolute error; GFR = glomerular filtration rate; eGFR = estimated GFR; mGFR = measured GFR; CKD-EPI = Chronic Kidney Disease Epidemiology Collaboration; SCr = serum creatinine; LM = the refitted 2021 CKD-EPI SCr equation using a linear regression model; MFP = the equation based on the multivariable fractional polynomial model; GAM = the equation based on the generalized additive model; RF = the equation based on random forests; GBM = the equation based on gradient boosted machines. Bias is defined as the median of the differences between mGFR and eGFR for each individual in the sample (mGFR minus eGFR). Imprecision is the interquartile range (75th minus the 25th percentiles) of the differences. P10 is the percentage of eGFRs within 10% of mGFR, and P30 is the percentage of eGFRs within 30% of mGFR. RMSE is the root mean square error of log mGFR and log eGFR. MAE is calculated as the average of the absolute differences between each eGFR and the corresponding mGFR.

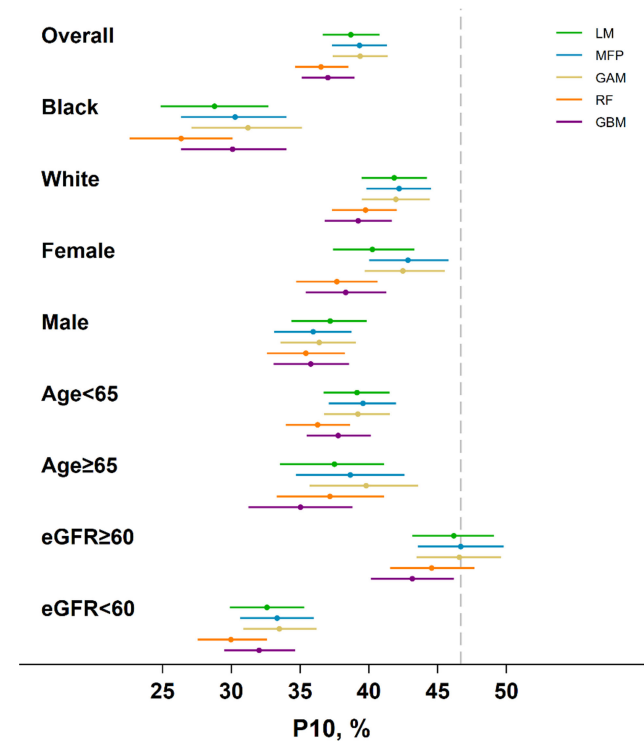


Figure 2. P10 of equations overall and by subgroups in the external validation data set. Shows accuracy measured by P10 of all equations overall and across subgroups. The dots are point estimates and the horizontal lines are 95% confidence intervals. The vertical reference line is positioned at the highest P10 value across all equations, with estimates closer to 100 indicating higher accuracy. eGFR based on the 2021 CKD-EPI SCr equation was used to define the subgroups with eGFR < 60 ml/min/1.73 m² and eGFR ≥ 60 ml/min/1.73 m².
Note: P10 is the percentage of eGFRs within 10% of mGFR; GFR = glomerular filtration rate; eGFR = estimated GFR; mGFR = measured GFR; CKD-EPI = Chronic Kidney Disease Epidemiology Collaboration; SCr = serum creatinine; LM = the refitted 2021 CKD-EPI SCr equation using a linear regression model; MFP = the equation based on the multivariable fractional polynomial model; GAM = the equation based on the generalized additive model; RF = the equation based on random forests; GBM = the equation based on gradient boosted machines.

The individual-level differences between mGFR and all eGFRs derived from all equations, including the LM equation, were large (Supplemental Table 5, Figure 4). 95% PIs were wide at each eGFR threshold used to define CKD diagnosis or staging. For example, at

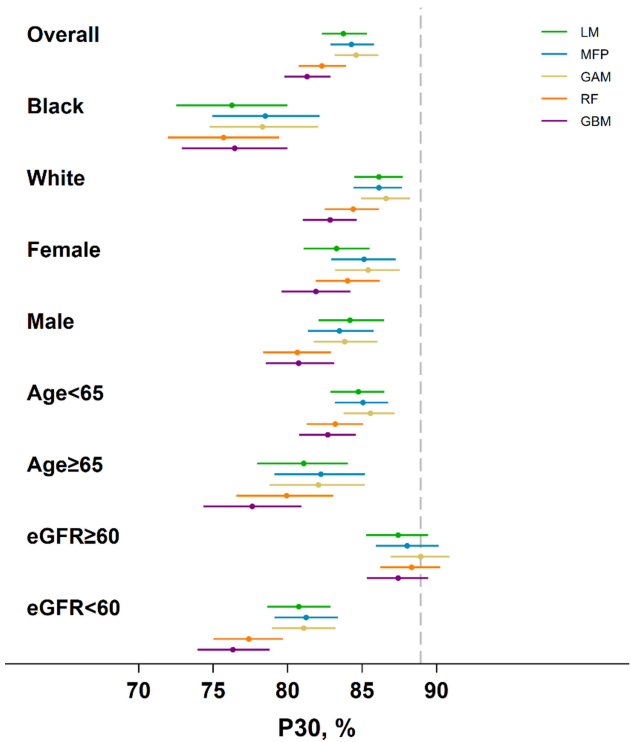


Figure 3. P30 of equations overall and by subgroups in the external validation data set. Shows accuracy measured by P30 of all equations overall and across subgroups; the dots are point estimates and the horizontal lines are 95% confidence intervals. The vertical reference line is positioned at the highest P30 value across all equations, with estimates closer to 100 indicating greater accuracy. eGFR based on the 2021 CKD-EPI SCr equation was used to define the subgroups with eGFR < 60 ml/min/1.73 m² and eGFR ≥ 60 ml/min/1.73 m².
Note: P30 is the percentage of eGFRs within 30% of mGFR; GFR = glomerular filtration rate; eGFR = estimated GFR; mGFR = measured GFR; CKD-EPI = Chronic Kidney Disease Epidemiology Collaboration; SCr = serum creatinine; LM = the refitted 2021 CKD-EPI SCr equation using a linear regression model; MFP = the equation based on the multivariable fractional polynomial model; GAM = the equation based on the generalized additive model; RF = the equation based on random forests; GBM = the equation based on gradient boosted machines.

eGFR of 60 ml/min/1.73 m², 95% PI of mGFR range from 38 to 89 ml/min/1.73 m² for the LM equation (i.e., mGFR could fall anywhere between 38 to 89 ml/min/1.73 m²), 36 to 88 ml/min/1.73 m² for both the MFP and GAM equation, 37 to 87 ml/min/1.73 m² for the

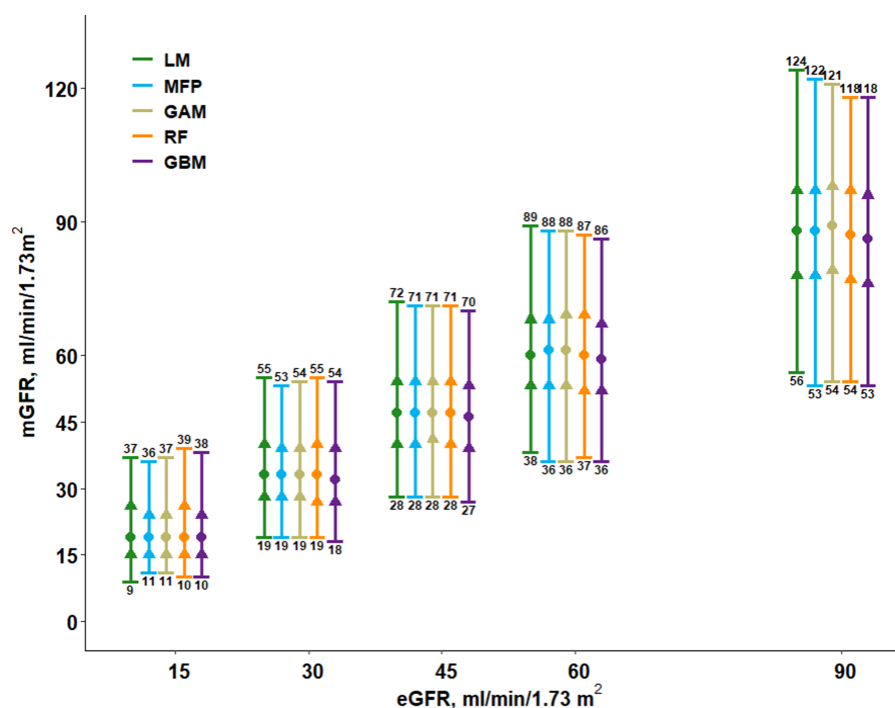


Figure 4. Comparison of 95% prediction intervals of mGFR among all equations in the external validation data set. Vertical lines represent prediction intervals of the new equations, with each equation represented by a different color. The numbers near the caps of vertical lines show the 2.5th and 97.5th percentiles of mGFR at given eGFR values. Symbols (arrows and dots) on the vertical lines identify the 25th and 75th percentiles, and median of mGFR at given eGFR values. The interpretation is that at a given eGFR, 95% of mGFRs range from the 2.5th to 97.5th percentiles. Similarly, 50% of mGFRs range from the 25th to 75th percentiles. For each equation, the percentile values of mGFR are obtained from separate quantile regression models (at the 2.5th, 25th, median, 75th, and 97.5th percentiles, respectively) of mGFR on eGFR.

Note: GFR = glomerular filtration rate; eGFR = estimated GFR; mGFR = measured GFR; CKD-EPI = Chronic Kidney Disease Epidemiology Collaboration; SCr = serum creatinine; LM = the refitted 2021 CKD-EPI SCr equation using a linear regression model; MFP = the equation based on the multivariable fractional polynomial model; GAM = the equation based on the generalized additive model; RF = the equation based on random forests; GBM = the equation based on gradient boosted machines.

RF equation, and 36 to 86 ml/min/1.73 m² for the GBM equation. Overall, all equations exhibited similar performance at the chosen eGFR thresholds.

Discussion

In this study, we developed and examined four new SCr-based GFR-estimating equations using advanced methods and compared their performance to that of the refitted linear regression-based 2021 CKD-EPI SCr equation (the LM equation). Our findings showed that the MFP and GAM equations performed very similarly to the refitted CKD-EPI SCr equation, with slightly improved accuracy as measured by P10 and P30 in certain subgroups, including Black individuals and females. The GBM and RF methods had smaller biases in the overall sample, as well as in some subgroups, including Black and eGFR < 60 ml/min/1.73 m² subgroups. However, they tended to show lower precision and accuracy compared to the other equations. Generally, differences among equations were modest across the entire external validation data set and subgroups defined by race, sex, age, and eGFR categories.

A few other studies have also looked into GFR estimation using machine learning methods. For example, Xunliu et al. employed an artificial neural network (ANN) model to estimate GFR using sex, age, and SCr as predictors based on data collected from a group of patients with CKD in China. The model did not outperform a linear regression model [27]. The authors later applied an ensemble approach, averaging predictions from ANN, support vector machines, and the regression model, to estimate GFR, which

improved precision but yielded bias and accuracy comparable to the regression-model-based approach [28].

Despite the numerous advantages of RF or GBM, our study did not demonstrate the clear benefits of these techniques for SCr-based GFR estimation. One possible explanation is that variable selection is one of the key strengths of these machine learning methods. However, in our study, the predictors were pre-selected according to the standard clinical practice. As a result, the variable selection capability of machine learning was not fully utilized, which may explain why GBM and RF did not significantly outperform the traditional linear regression approach in GFR estimation. In addition, we suspect that our data did not exhibit complex nonlinear relationships between SCr and mGFR. As a result, the usual benefits of advanced methods – such as their ability to handle intricate nonlinear relationships – may not be fully demonstrated in our study sample.

All equations, including the LM-based refitted 2021 CKD-EPI SCr equation, exhibited suboptimal performance in the external validation data set. The greatest underestimation biases were observed in Black individuals across all subgroups defined by race, sex, age, and eGFR (Figure 1). The largest biases were seen in the Black subgroup with eGFR > 60 ml/min/1.73 m², followed by the Black subgroup with eGFR < 60 ml/min/1.73 m² (Supplemental Table 4, Figure 8A). P30 was lowest for Black individuals with eGFR < 60 ml/min/1.73 m² subgroup (Supplemental Table 4, Figure 8C), ranging between 70% and 80%, while 90% is typically considered good. Our findings indicate that the limitations of SCr-based equations persist even when advanced statistical and machine learning methods are employed, highlighting the

difficulty of improving upon established SCr-based equations such as the CKD-EPI SCr equation.

Some limitations of the current study bear mentioning. First, the sample sizes of both our development and external validation data sets were relatively small compared to those used to develop the CKD-EPI equations. The external validation data set included fewer Black participants ($n = 535$, 24%) than White participants, resulting in wide confidence intervals in certain analyses, such as the subgroup analyses. This may have reduced the ability to detect true differences between White and Black participants across equations. Future research with a larger sample of Black participants would enable more robust and valid comparisons.

Moreover, the equations generated using the methods we employed are more difficult to understand than those based on linear regression models. While we were able to explicitly formulate the MFP equation, we could only visualize the GAM, RF, and GBM equations. User-friendly tools, such as online calculators and R packages, will be essential to facilitate the application of these advanced methodologies in clinical practice. However, the goal of this study is to assess whether advanced methods can improve SCr-based GFR-estimating equations without incorporating other markers rather than to advocate for any of these equations.

In conclusion, our results suggest that advanced methodologies, including MFP, GAM, RF, and GBM, may have limited utility in enhancing GFR estimation using SCr as the main predictor. Future research should aim to integrate novel biomarkers to enhance GFR estimation and to improve the clinical feasibility of mGFR measurements, especially for subgroups where the current eGFR equations show less optimal performance, such as Black individuals.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/cts.2025.10057>

Acknowledgments. The data from the CRIC, MDRD, AASK, CRISP, and ALTOLD studies reported here were supplied by the NIDDK Central Repository. Data for GENOA is housed at the University of Mississippi Medical Center. We thank Dr Andrew Rule from the Mayo Clinic for the help in obtaining data from the ECAC cohort study.

Author contributions. Xiaoqian Zhu: Conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, writing – original draft, writing – review and editing; Tariq Shafi: conceptualization, supervision, writing – review and editing; Keith Norris: supervision, writing – review and editing; Jeannette Simino: supervision, writing – review and editing; Srishti Shrestha: writing review and editing; Thomas Mosley: resources, writing – review and editing; Michael Griswold: supervision, writing – review and editing; Seth Lirette: conceptualization, methodology, supervision, writing – review and editing.

Funding statement. XZ is partially supported by 75N92022D0004, HHSN268201800012i, and HIS-2020C1-19350. JS is supported by 75N92022D0004, 5P20GM144041, and 1RF1AG059421. KCN is partially supported by NIH research grants UL1TR001881, P30AG021684, U2CDK129496, P50MD017366, and OT2OD032581. TS is supported by R01NR017399, R01DK123062, U01DK127918, and R01HL153499. SL is partially supported by the Mississippi Center for Clinical and Translational Research and Mississippi Center of Excellence in Perinatal Research COBRE funded by the National Institute of General Medical Sciences of the National Institutes of Health under Award Numbers 5U54GM115428 and P20GM121334.

Competing interests. KCN is a Kidney Disease Quality Improvement Consultant for Atlantis Health, Inc.

References

1. Delgado C, Baweja M, Crews DC, et al. A unifying approach for GFR estimation: recommendations of the NKF-ASN task force on reassessing the inclusion of race in diagnosing kidney disease. *Am J Kidney Dis.* 2022;**79**(2):268–288.e1. doi: [10.1053/j.ajkd.2021.08.003](https://doi.org/10.1053/j.ajkd.2021.08.003).
2. Kashani K, Rosner MH, Ostermann M. Creatinine: from physiology to clinical application. *Eur J Intern Med.* 2020;**72**:9–14. doi: [10.1016/j.ejim.2019.10.025](https://doi.org/10.1016/j.ejim.2019.10.025).
3. Tio MC, Shafi T, Zhu X, Kalantar-Zadeh K, Chan A, Nguyen L. Traditions and innovations in assessment of glomerular filtration rate using creatinine to cystatin C. *Curr Opin Nephrol Hypertens.* 2023;**32**(1):89–97. doi: [10.1097/mnh.0000000000000854](https://doi.org/10.1097/mnh.0000000000000854).
4. Inker LA, Eneanya ND, Coresh J, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *New Engl J Med.* 2021;**385**(19):1737–1749. doi: [10.1056/NEJMoa2102953](https://doi.org/10.1056/NEJMoa2102953).
5. Lousa I, Reis F, Beirão I, Alves R, Belo L, Santos-Silva A. New potential biomarkers for chronic kidney disease management-a review of the literature. *Int J Mol Sci.* 2020;**22**(1):1–43. doi: [10.3390/ijms22010043](https://doi.org/10.3390/ijms22010043).
6. Benito S, Unceta N, Maciejczyk M, et al. Revealing novel biomarkers for diagnosing chronic kidney disease in pediatric patients. *Sci Rep.* 2024;**14**(1):11549. doi: [10.1038/s41598-024-62518-w](https://doi.org/10.1038/s41598-024-62518-w).
7. Raman M, Middleton R, Kalra P, Green D. Estimating renal function in old people: an in-depth review. *Int Urol Nephrol.* 2017;**49**:1979–1988.
8. Rule AD, Bailey KR, Lieske JC, Peyser PA, Turner ST. Estimating the glomerular filtration rate from serum creatinine is better than from cystatin C for evaluating risk factors associated with chronic kidney disease. *Kidney Int.* 2013;**83**(6):1169–1176. doi: [10.1038/ki.2013.7](https://doi.org/10.1038/ki.2013.7).
9. Appel LJ, Middleton J, Miller ERI, et al. The rationale and design of the AASK cohort study. *J Am Soc Nephrol.* 2003;**14**(suppl_2):S166–S172. doi: [10.1097/01.Asn.0000070081.15137.CO](https://doi.org/10.1097/01.Asn.0000070081.15137.CO).
10. 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. doi: [10.7326/0003-4819-130-6-199903160-00002](https://doi.org/10.7326/0003-4819-130-6-199903160-00002).
11. Chapman AB, Guay-Woodford LM, Grantham JJ, et al. Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): the consortium for radiologic imaging studies of polycystic kidney disease (CRISP) cohort ¹. *Kidney Int.* 2003;**64**(3):1035–1045. doi: [10.1046/j.1523-1755.2003.00185.x](https://doi.org/10.1046/j.1523-1755.2003.00185.x).
12. Rule AD, Bergstralh EJ, Slezak JM, Bergert J, Larson TS. Glomerular filtration rate estimated by cystatin C among different clinical presentations. *Kidney Int.* 2006;**69**(2):399–405. doi: [10.1038/sj.ki.5000073](https://doi.org/10.1038/sj.ki.5000073).
13. Kasiske BL, Anderson-Haag T, Ibrahim HN, et al. A prospective controlled study of kidney donors: baseline and 6-month follow-up. *Am J Kidney Dis.* 2013;**62**(3):577–586. doi: [10.1053/j.ajkd.2013.01.027](https://doi.org/10.1053/j.ajkd.2013.01.027).
14. Anderson AH, Yang W, Hsu CY, et al. Estimating GFR among participants in the chronic renal insufficiency cohort (CRIC) study. *Am J Kidney Dis.* 2012;**60**(2):250–261. doi: [10.1053/j.ajkd.2012.04.012](https://doi.org/10.1053/j.ajkd.2012.04.012).
15. 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–S165. doi: [10.1097/01.asn.0000070080.21680.cb](https://doi.org/10.1097/01.asn.0000070080.21680.cb).
16. Kwong YT, Stevens LA, Selvin E, et al. Imprecision of urinary iothalamate clearance as a gold-standard measure of GFR decreases the diagnostic accuracy of kidney function estimating equations. *Am J Kidney Dis.* 2010;**56**(1):39–49. doi: [10.1053/j.ajkd.2010.02.347](https://doi.org/10.1053/j.ajkd.2010.02.347).
17. Sauerbrei PRW. *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis based on Fractional Polynomials for Modelling Continuous Variables.* John Wiley & Sons Ltd, 2008: 322
18. Wood S. *Generalized Additive Models.* 2nd ed. Chapman and Hall/CRC, 2017: 496.
19. Breiman L. Random Forests. *Mach Learn.* 2001;**45**(1):5–32. doi: [10.1023/a:1010933404324](https://doi.org/10.1023/a:1010933404324).

20. **Friedman JH.** Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*. 2001;**29**(5):1189–1232, 44.
21. **Friedman JH.** Stochastic gradient boosting. *Comput Stat Data Anal*. 2002;**38**(4):367–378. doi: [10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).
22. **Levey AS, Stevens LA, Schmid CH, et al.** A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;**150**(9):604–612. doi: [10.7326/0003-4819-150-9-200905050-00006](https://doi.org/10.7326/0003-4819-150-9-200905050-00006).
23. **Stevens LA, Zhang Y, Schmid CH.** Evaluating the performance of equations for estimating glomerular filtration rate. *J Nephrol*. 2008;**21**(6):797–807.
24. **Austin PC, Tu JV, Daly PA, Alter DA.** The use of quantile regression in health care research: a case study examining gender differences in the timeliness of thrombolytic therapy. *Stat Med*. 2005;**24**(5):791–816. doi: [10.1002/sim.1851](https://doi.org/10.1002/sim.1851).
25. **Shafi T, Zhu X, Lirette ST, et al.** Quantifying individual-level inaccuracy in glomerular filtration rate estimation : a cross-sectional study. *Ann Intern Med*. 2022;**175**(8):1073–1082. doi: [10.7326/m22-0610](https://doi.org/10.7326/m22-0610).
26. **Royston P, Sauerbrei W.** Bootstrap assessment of the stability of multivariable models. *Stata J*. 2009;**9**(4):547–570. doi: [10.1177/1536867X0900900403](https://doi.org/10.1177/1536867X0900900403).
27. **Liu X, Li NS, Lv LS, et al.** A comparison of the performances of an artificial neural network and a regression model for GFR estimation. *Am J Kidney Dis*. 2013;**62**(6):1109–1115. doi: [10.1053/j.ajkd.2013.07.010](https://doi.org/10.1053/j.ajkd.2013.07.010).
28. **Liu X, Li N, Lv L, et al.** Improving precision of glomerular filtration rate estimating model by ensemble learning. *J Transl Med*. 2017;**15**(1):231. doi: [10.1186/s12967-017-1337-y](https://doi.org/10.1186/s12967-017-1337-y).