



Outcomes and Risk Factors of Widened Difference in Estimated Glomerular Filtration Rate Based on Creatinine or Cystatin C: Adjust Model Development With More Than 300,000 UK Biobank Participants

Min W. Kang, Jae-ik Oh, Jinsun Lee, Minsang Kim, Jung H. Koh, Jeong M. Cho, Seong G. Kim, Semin Cho, Soojin Lee, Yaerim Kim, Dong K. Kim, and Sehoon Park

Rationale & Objective: Recent studies have highlighted the clinical significance of the discrepancies between cystatin C–based CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) estimated glomerular filtration rate (eGFR) (eGFR_{cys}) and creatinine-based CKD-EPI eGFR (eGFR_{cr}). This study explores the implications of the differences between eGFR_{cys} and eGFR_{cr} (eGFR_{diff}) on clinical outcomes and aims to develop an adjustment model using eGFR_{cr} and associated clinical variables.

Study Design: Retrospective cohort study.

Setting & Participants: A total of 343,854 UK Biobank participants.

Exposure: eGFR_{diff} (mL/min/1.73 m²): lower (eGFR_{cys} - eGFR_{cr} < -15), middle (-15 ≤ eGFR_{cys} - eGFR_{cr} ≤ 15), and upper (eGFR_{cys} - eGFR_{cr} > 15).

Outcome: Death, myocardial infarction (MI), and ischemic stroke.

Analytical Approach: We analyzed the risks of death, MI, and ischemic stroke in groups with eGFR_{diff} < -15 and eGFR_{diff} > 15 using Cox proportional hazards models. Logistic regression analysis was used to identify variables associated with eGFR_{diff}. In addition, we developed and validated a regression model using eGFR_{cr} and

variables associated with eGFR_{diff} to approximate eGFR with optimal performance.

Results: Individuals with eGFR_{diff} < -15 were associated with increased risks of death (HR, 1.37 [1.30-1.44]), MI (HR, 1.23 [1.15-1.33]), and ischemic stroke (HR, 1.19 [1.08-1.31]). Variables associated with eGFR_{diff} < -15 included high waist/hip circumference, high body weight, low fat-free mass, high fat/carbohydrate intake, low protein intake, diabetes, and smaller kidney volumes. A simplified regression model approximating eGFR_{cys} was developed, using only eGFR_{cr}, sex, age, height, weight, and body mass index. In the validation, applying the simplified linear model to eGFR_{cr} significantly enhanced its correlation with eGFR_{cys} and improved clinical outcome prediction, as demonstrated by favorable net reclassification indices for death, MI, and ischemic stroke.

Limitations: The absence of true GFR data.

Conclusions: This study confirmed that eGFR_{diff} < -15 is associated with increased risks of clinical outcomes and identified diverse factors associated with eGFR_{diff}. We developed an adjusted model to approximate eGFR_{cr} to eGFR_{cys}, demonstrating superior clinical outcome association using the adjusted values.

Complete author and article information provided before references.

Correspondence to S. Park (mailto:mailofsehoon@gmail.com)

Kidney Med. 8(4):101288. Published online February 10, 2026.

doi: 10.1016/j.xkme.2026.101288

© 2026 The Authors. Published by Elsevier Inc. on behalf of the National Kidney Foundation, Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The estimation of glomerular filtration rate (eGFR) is essential not only for assessing current kidney function but also for predicting future clinical outcomes.¹ Current guidelines recommend the use of the CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) equation for clinical purposes.²⁻⁴ This equation typically uses estimates based on creatinine and cystatin C.³ Although creatinine-based estimates are widely accessible and cost-effective, they are significantly influenced by muscle mass, gender, age, and other nonkidney factors.^{5,6} In contrast, cystatin C–based estimates are less affected by these factors and are considered to provide a more accurate approximation to true GFR.^{3,7}

Recent studies have illuminated the clinical importance of the discrepancy between cystatin C–based CKD-EPI eGFR (eGFR_{cys}) and creatinine-based CKD-EPI eGFR (eGFR_{cr}), termed eGFR_{diff}. This difference has been recognized as an indicator of overall health status and frailty.⁸⁻¹⁰ Notably, a

negative eGFR_{diff} value is associated with elevated risks of atrial fibrillation, mortality, and cardiovascular diseases.⁸⁻¹² This suggests that eGFR_{cr} may be influenced by a range of nonkidney factors, underlining the importance of cystatin C–based estimates for more precise assessments of kidney function. However, there remains a need for further studies to uncover clinical characteristics associated with eGFR_{diff}, particularly using data from extensively phenotyped large-scale biobanks. Furthermore, access to cystatin C measurement is still limited in some regions,¹³ necessitating the development of a refined model to approximate eGFR_{cr} to eGFR_{cys}, especially for clinical outcomes in which eGFR_{cys} demonstrates superior predictive accuracy, such as cardiovascular disease, myocardial infarction (MI), and mortality.^{7,14,15}

In this study, we hypothesized that the previous association between eGFR_{diff} and clinical outcomes would be replicated in the large-scale UK Biobank data, which

PLAIN-LANGUAGE SUMMARY

Kidney function is often estimated using blood markers such as creatinine or cystatin C. However, large gaps between these estimates can signal hidden health risks. In 343,854 UK Biobank participants, we found that those whose cystatin C–based eGFR was much lower than their creatinine-based eGFR faced higher risks of death, heart attack, and stroke. We identified body size, diet, and kidney volume factors linked to these gaps. Using these factors and creatinine-based eGFR, we built a simple regression model that closely approximates cystatin C–based eGFR and improves risk prediction. This tool can support clinicians in settings in which cystatin C testing is not available.

include deeply phenotyped clinical information. With the database, we sought to identify diverse clinical variables associated with widened eGFRdiff. Using these results, we aimed to develop the adjust model for eGFRcr to improve the clinical predictability of the value.

METHODS**Ethical Considerations**

The study was conducted in accordance with the Declaration of Helsinki. Ethical approval was granted by the Institutional Review Boards of Seoul National University Hospital (approval number: E-2306-031-1437) and the UK Biobank consortium (application number: 53799). The requirement for informed consent was waived as the study involved the analysis of public databases.

Study Population

The UK Biobank is a prospective cohort study that includes 502,460 adults aged 40–69 years, enrolled from 2006–2010 across 22 assessment centers throughout the United Kingdom. The data for this cohort were collected and structured as described in previous research.^{16,17} From the UK Biobank population, we included participants with available anthropometric data (eg, fat-free mass) to study the association of these variables with eGFRdiff. As we studied incident outcomes, we excluded those with underlying MI or ischemic stroke. In addition, dialysis patients at baseline were excluded, as eGFR values in these patients are significantly deviated and have limited clinical relevance.

Participants with available food intake data and kidney volume data were included for subset analysis. For the development and validation of regression models, the entire population was randomly divided into train and test data sets at a ratio of 7:3. The train data set was used to develop the models, whereas the test data set was employed to validate their performance.

Data Collection and Definitions

Baseline demographic variables and medical history of interest were selected based on cardiovascular disease and mortality risk factors identified in previous studies.^{18–20} These variables included age, sex, waist circumference, hip circumference, height, body weight, body mass index (BMI), smoking status, systolic and diastolic blood pressure, self-reported history of cardiocerebrovascular diseases, diabetes, and cancer, and use of medications for hypertension, dyslipidemia, and insulin. Laboratory data included measurements of creatinine and cystatin C. CKD-EPI equations were employed to calculate eGFR using creatinine alone (eGFRcr) and cystatin C alone (eGFRcys), and a combined equation incorporating both creatinine and cystatin C (eGFRcr-cys) was also used.^{3,21} Fat-free mass data were derived using bioimpedance analysis. In addition, the total kidney volume adjusted for body surface area and the bilateral kidney volume ratio were also analyzed. Smoking statuses were categorized as past or current smoking and never smoking. Baseline medical conditions, medication use, smoking history, diabetes status, and dietary history were all self-reported during nurse-led interviews.

Participants were stratified into 3 groups based on eGFRdiff as follows: (1) upper eGFRdiff (eGFRcys - eGFRcr > 15), (2) middle eGFRdiff ($-15 \leq \text{eGFRcys} - \text{eGFRcr} \leq 15$), and (3) lower eGFRdiff (eGFRcys - eGFRcr < -15) to analyze the associated risk of outcomes. The threshold of 15 mL/min/1.73 m² was selected based on previous studies demonstrating its clinical relevance in predicting adverse outcomes.^{11,22}

Clinical outcomes were defined as death or the first recorded occurrence of MI or ischemic stroke. Diagnoses were recorded using the World Health Organization's International Classification of Diseases, Tenth Revision, codes. Data on MI and ischemic stroke were derived from algorithmic combinations of coded information from the UK Biobank data collection, including self-reported medical conditions, surgical procedures, and medications at baseline, as well as linked data from hospital admissions and death registries.

Statistical Analyses

Baseline characteristics were reported with categorical variables presented as numbers and percentages, and continuous variables as means with standard deviations. The risk associated with the 3 categories of eGFRdiff was analyzed using Cox proportional hazards models. Spline curves were examined to explore how hazard ratios (HRs) for clinical outcomes changed across different eGFRdiff levels. The models included a univariable model using only eGFRdiff and multivariable models in which model 1 was adjusted for sex, age, waist circumference, hip circumference, height, body weight, and fat-free mass, and model 2 included all variables from model 1 plus additional adjustments for systolic blood pressure, diastolic

blood pressure, histories of cardiocerebrovascular disease, use of medications for hypertension, dyslipidemia, and insulin, as well as histories of cancer and diabetes, and smoking status.

Logistic regression was employed to investigate clinical variables associated with $\text{eGFR}_{\text{diff}} < -15$, with participants having $\text{eGFR}_{\text{diff}} \geq -15$ serving as the reference group. Before performing logistic regression analysis, all continuous variables were standardized, ensuring that the resulting odds ratios represent the change in odds per 1 standard deviation increase. The logistic models included were the following: multivariable model 1 incorporating variables such as age, sex, waist circumference, hip circumference, height, BMI, body weight, and fat-free mass, and multivariable model 2, which additionally considered systolic blood pressure, diastolic blood pressure, history of cardiocerebrovascular disease, use of hypertension and dyslipidemia medications, and insulin, cancer, diabetes, and smoking status. Although BMI was included in the analysis, height and weight—which largely reflect lean body mass—were also adjusted for, because both height and lean body mass independently affect kidney function or creatinine levels, thereby providing a more comprehensive adjustment.^{5,23} Further models analyzed associations between $\text{eGFR}_{\text{diff}}$ and dietary data (including variables from model 2 with food weight, protein, fat, and carbohydrate intake) and kidney volume (including variables from model 2 with body surface area–

adjusted total kidney volume and total kidney volume ratio).

To assess the predictive performance of eGFR_{Cr} , eGFR_{Cys} , and $\text{eGFR}_{\text{Cr-Cys}}$ for clinical outcomes, we used the C-index and compared these indices at a 5-year time point using the net reclassification index (NRI). Receiver operating characteristic (ROC) curves and the area under the ROC curve were also analyzed to predict the occurrence of outcomes within 5 years for each eGFR measure. The NRI values were calculated with 95% confidence intervals using bootstrapping methods.

A regression model was developed, integrating clinical variables and eGFR_{Cr} , to predict the eGFR value with the highest predictive performance among eGFR_{Cys} and $\text{eGFR}_{\text{Cr-Cys}}$. On the train data set, we constructed both linear regression and machine learning models (including the light gradient boosting machine, CatBoost, and an ensemble model combining the light gradient boosting machine and CatBoost), with variable selection guided by significant associations identified in prior logistic regression analyses as well as by the clinical relevance and ease of collection of the variables. Model accuracy was then evaluated in the test data set using the mean absolute error and adjusted R^2 . Furthermore, the predictive performance of the linear regression models was assessed by calculating the C-index and NRI at a 5-year time point, along with ROC curves and the area under the ROC curve for additional outcome evaluation.

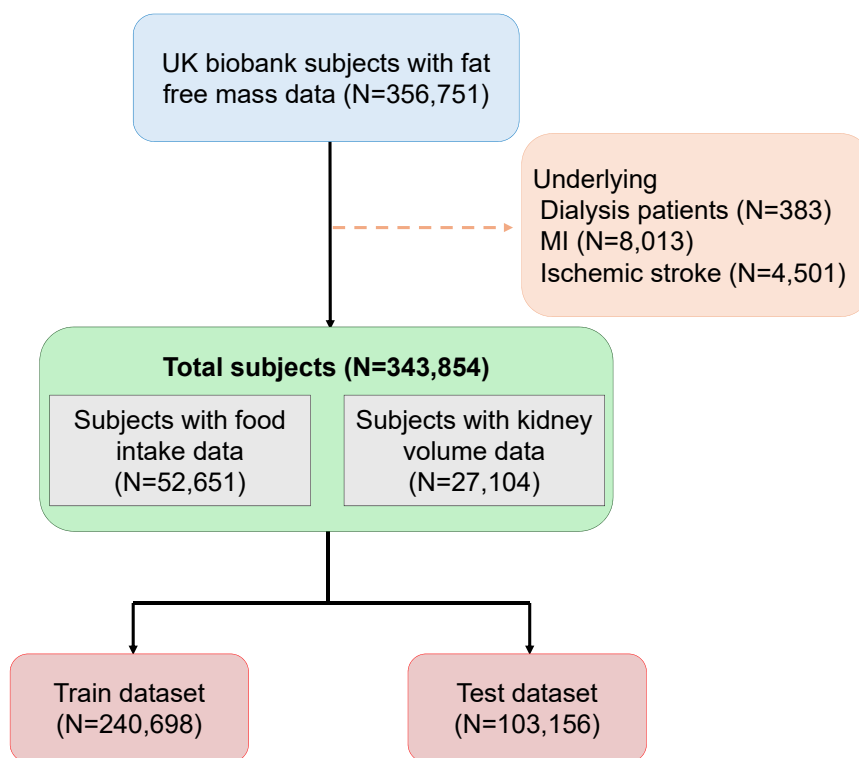


Figure 1. Study population. Abbreviations: GFR_{Cr} , creatinine-based Chronic Kidney Disease Epidemiology Collaboration estimated glomerular filtration; GFR_{Cys} , cystatin C–based Chronic Kidney Disease Epidemiology Collaboration estimated glomerular filtration; MI, myocardial infarction.

RESULTS

Study Population and Baseline Characteristics

From the UK Biobank population, a total of 356,751 individuals with available anthropometric data were initially selected. After excluding those with underlying conditions such as dialysis dependence, MI, or ischemic stroke, the final population comprised 343,854 participants for analysis (Fig 1). In addition, 52,651 participants with available food intake data and 27,104 with kidney volume data were included in subset analyses. For the train and evaluation of regression models, the population was randomly divided into a train data set of 240,698 individuals and a test data set of 103,156 individuals.

The distribution of eGFR differences within the population was as follows: 12.4% in the eGFRdiff > 15 group,

71.5% in the $-15 \leq \text{eGFRdiff} \leq 15$ group, and 16.2% in the eGFRdiff < -15 group (Table 1). The cohort comprised 45.1% female participants, with an average age of 56.9 years. The means of eGFRcr, eGFRcys, and eGFRcr-cys across the entire population were similar at 91.1, 90.0, and 90.7 mL/min/1.73 m², respectively, with similar trends observed in the middle eGFRdiff group. However, in the eGFRdiff < -15 group, values diverged significantly, with eGFRcys notably lower at 94.7, 72.4, and 83.0 mL/min/1.73 m².

Risk of Clinical Outcomes According to eGFRdiff

A trend of decreasing HR with increasing eGFRdiff was identified in both univariable and multivariable models (Fig 2). The univariable Cox proportional hazards model

Table 1. Baseline Characteristics

Variables	Total Population (N = 343,854)	eGFRcys - eGFRcr > 15 (N = 42,479)	-15 ≤ eGFRcys - eGFRcr ≤ 15 (N = 245,830)	eGFRcys - eGFRcr < -15 (N = 55,545)
Female (%)	155,204 (45.1%)	16,431 (38.7%)	112,845 (45.9%)	25,928 (46.7%)
Age (y)	56.9 ± 8.1	53.6 ± 8.0	57.1 ± 8.0	58.5 ± 7.7
Waist circumference (cm)	90.0 ± 13.3	84.5 ± 11.5	89.5 ± 12.7	96.5 ± 14.5
Hip circumference (cm)	103.3 ± 9.0	100.4 ± 7.3	102.9 ± 8.4	107.3 ± 11.5
Height (cm)	168.6 ± 9.3	168.3 ± 8.8	168.7 ± 9.3	168.1 ± 9.6
BMI (kg/m ²)	27.3 ± 4.7	25.7 ± 3.7	27.1 ± 4.3	29.7 ± 5.8
Weight (kg)	77.9 ± 15.7	73.0 ± 13.4	77.3 ± 15.0	83.9 ± 18.3
Diastolic blood pressure (mm Hg)	82.3 ± 10.1	80.4 ± 9.9	82.3 ± 10.0	83.8 ± 10.3
Systolic blood pressure (mm Hg)	137.8 ± 18.6	133.4 ± 18.0	138.0 ± 18.5	140.6 ± 18.6
Cardiocerebrovascular (%)	7,826 (2.3%)	598 (1.4%)	5,452 (2.2%)	1,776 (3.2%)
Hypertension (%)	65,119 (18.9%)	4,641 (10.9%)	45,082 (18.3%)	15,396 (27.7%)
Insulin (%)	3,215 (0.9%)	280 (0.7%)	2,017 (0.8%)	918 (1.7%)
Dyslipidemia (%)	51,678 (15.0%)	4,482 (10.6%)	36,635 (14.9%)	10,561 (19.0%)
Cancer (%)	25,919 (7.5%)	2,510 (5.9%)	18,302 (7.4%)	5,107 (9.2%)
Diabetes (%)	15,966 (4.6%)	1,144 (2.7%)	10,320 (4.2%)	4,502 (8.1%)
Smoking (%)	153,325 (44.6%)	16,141 (38.0%)	107,368 (43.7%)	29,816 (53.7%)
Cr (μmol/L)	71.9 ± 15.3	76.9 ± 15.9	72.0 ± 15.6	67.7 ± 11.9
Cystatin C (mg/L)	0.9 ± 0.2	0.8 ± 0.1	0.9 ± 0.1	1.1 ± 0.1
eGFR CKD-EPI (Cr) (mL/min/ 1.73 m ²)	91.1 ± 13.1	86.4 ± 14.6	91.1 ± 13.1	94.7 ± 10.5
eGFR CKD-EPI (cystatin C) (mL/min/1.73 m ²)	90.0 ± 17.6	109.8 ± 15.6	90.5 ± 15.0	72.4 ± 11.4
eGFR CKD-EPI (Cr + cystatin C) (mL/min/1.73 m ²)	90.7 ± 13.9	97.1 ± 13.7	91.4 ± 13.7	83.0 ± 11.1
Fat-free mass (kg)	53.2 ± 11.5	51.6 ± 10.8	53.2 ± 11.5	54.6 ± 12.1
Food weight intake ^a (g)	3,263.3 ± 930.6	3,276.6 ± 922.5	3,263.8 ± 919.4	3,247.1 ± 994.3
Protein intake ^a (g)	81.8 ± 31.0	83.0 ± 31.3	81.9 ± 30.4	80.5 ± 33.5
Fat intake ^a (g)	76.9 ± 36.2	73.5 ± 35.0	77.0 ± 35.7	79.9 ± 39.4
Carbohydrate intake ^a (g)	257.9 ± 107.3	246.3 ± 107.0	258.0 ± 104.0	269.0 ± 122.1
Kidney volume (BSA adjusted) ^b	0.14 ± 0.02	0.14 ± 0.02	0.15 ± 0.02	0.14 ± 0.02
Kidney volume difference ratio ^b	1.12 ± 0.14	1.12 ± 0.14	1.11 ± 0.12	1.13 ± 0.15

Abbreviations: BMI, body mass index; BSA, body surface area; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration; Cr, creatinine; eGFR, estimated glomerular filtration rate; eGFRcr, creatinine-based Chronic Kidney Disease Epidemiology Collaboration eGFR; eGFRcys, cystatin C-based Chronic Kidney Disease Epidemiology Collaboration eGFR.

^aNo. of participants: 52,651, representing those with available food intake data.

^bNo. of participants: 27,104, representing those with available kidney volume data.

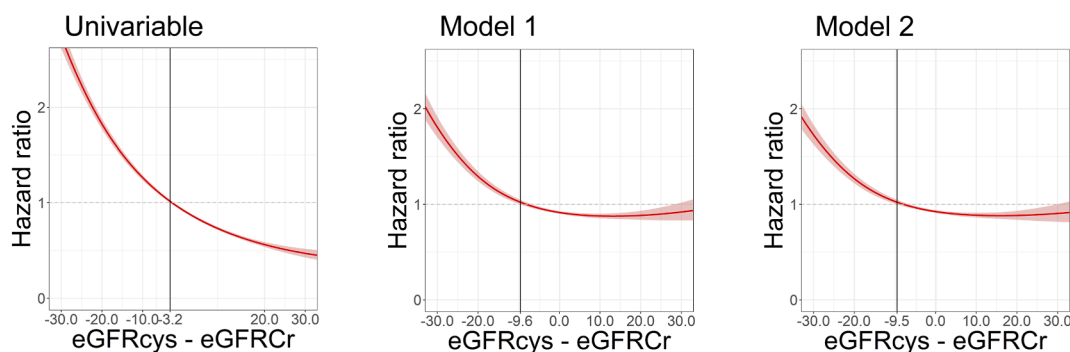
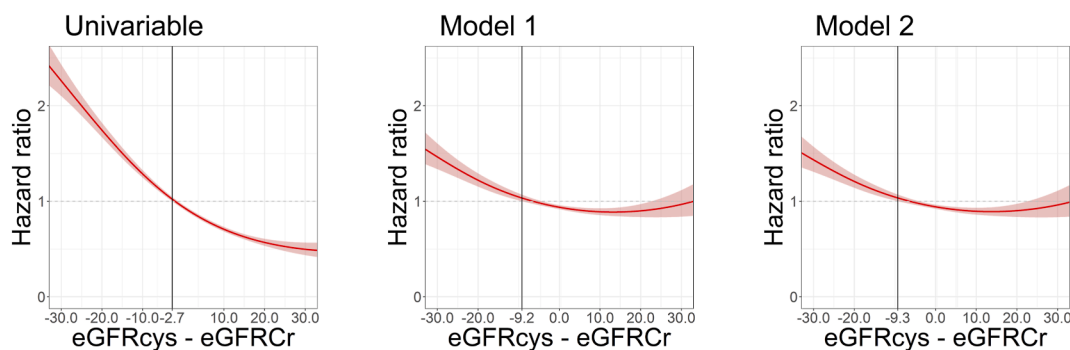
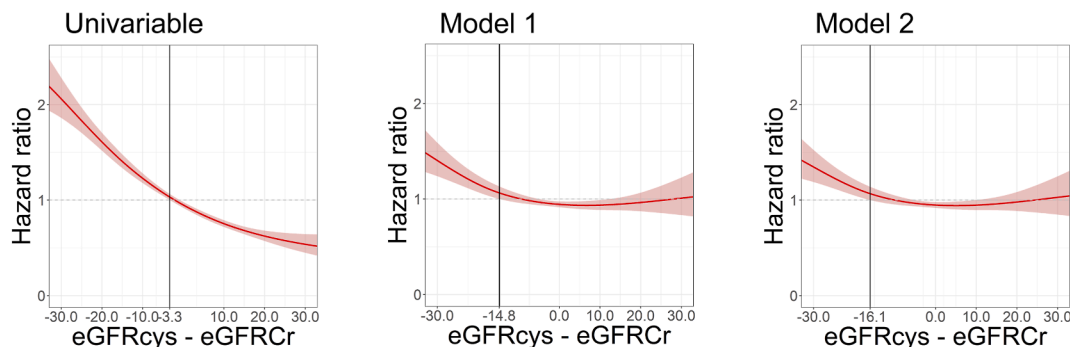
A. Death**B. MI****C. Ischemic stroke**

Figure 2. Hazard ratios of outcomes based on cubic spline curves. Model 1 was adjusted for sex, age, waist circumference, hip circumference, height, body weight, and fat-free mass. Model 2 was adjusted for sex, age, waist circumference, hip circumference, height, body weight, fat-free mass, systolic blood pressure, diastolic blood pressure, histories of cardiocerebrovascular disease, use of medications for hypertension, dyslipidemia, and use of insulin, histories of cancer and diabetes, and smoking status. Abbreviations: GFRcr, creatinine-based Chronic Kidney Disease Epidemiology Collaboration estimated glomerular filtration; GFRcys, cystatin C-based Chronic Kidney Disease Epidemiology Collaboration estimated glomerular filtration; MI, myocardial infarction.

indicated that, compared with the reference middle eGFRdiff group, the risk of MI and ischemic stroke was lower in the eGFRdiff > 15 group and higher in the eGFRdiff < -15 group. The risk of death was higher in the eGFRdiff < -15 group (Table 2). In multivariable model 1, only the eGFRdiff < -15 group showed significantly higher risks for death (HR, 1.42 [1.35-1.49]), MI (HR, 1.25 [1.16-1.35]), and ischemic stroke (HR, 1.22 [1.10-1.34]). Multivariable model 2 showed similar trends, with significantly increased risks for death (HR, 1.37 [1.30-1.44]), MI (HR, 1.23 [1.15-1.33]), and ischemic

stroke (HR, 1.19 [1.08-1.31]) observed exclusively in the eGFRdiff < -15 group.

Clinical Variables Associated With eGFRdiff < -15

Given the significantly elevated risks of death, MI, and ischemic stroke observed in the eGFRdiff < -15 group, logistic regression was employed to identify the variables associated with this group. In multivariable model 1, higher waist circumference, weight, and lower fat-free mass were associated with eGFRdiff < -15 (Table 3). This trend was also observed in model 2, in which a higher

Table 2. Cox Proportional Hazards Model

Outcome	eGFR Difference	Univariable			Model 1 ^a			Model 2 ^b		
		Hazard Ratio	P Value		Hazard Ratio	P Value		Hazard Ratio	P Value	
Death	eGFRcys - eGFRcr > 15	1.01 (0.92-1.10)	0.84		1.00 (0.92-1.09)	0.99		0.99 (0.90-1.08)	0.75	
	-15 ≤ eGFRcys - eGFRcr ≤ 15	1 (reference)			1 (reference)			1 (reference)		
	eGFRcys - eGFRcr < -15	1.23 (1.17-1.29)	<0.001		1.42 (1.35-1.49)	<0.001		1.37 (1.30-1.44)	<0.001	
MI	eGFRcys - eGFRcr > 15	0.56 (0.49-0.63)	<0.001		1.00 (0.88-1.13)	0.97		0.99 (0.88-1.13)	0.93	
	-15 ≤ eGFRcys - eGFRcr ≤ 15	1 (reference)			1 (reference)			1 (reference)		
	eGFRcys - eGFRcr < -15	1.84 (1.72-1.96)	<0.001		1.25 (1.16-1.35)	<0.001		1.23 (1.15-1.33)	<0.001	
Ischemic stroke	eGFRcys - eGFRcr > 15	0.58 (0.50-0.67)	<0.001		0.99 (0.84-1.16)	0.86		1.00 (0.85-1.17)	0.97	
	-15 ≤ eGFRcys - eGFRcr ≤ 15	1 (reference)			1 (reference)			1 (reference)		
	eGFRcys - eGFRcr < -15	1.69 (1.55-1.85)	<0.001		1.22 (1.10-1.34)	<0.001		1.19 (1.08-1.31)	0.001	

Abbreviations: eGFR, estimated glomerular filtration rate; eGFRcr, creatinine-based Chronic Kidney Disease Epidemiology Collaboration eGFR; eGFRcys, cystatin C–based Chronic Kidney Disease Epidemiology Collaboration eGFR; MI, myocardial infarction.
^aAdjusted variables: eGFRcys, sex, age, waist circumference, hip circumference, height, weight, and fat-free mass.
^bAdjusted variables: eGFRcys, sex, age, waist circumference, hip circumference, height, weight, fat-free mass, systolic blood pressure, diastolic blood pressure, cardiocerebrovascular history, hypertension medication, insulin medication, dyslipidemia medication, cancer history, diabetes history, and smoking.

hip circumference was additionally significantly associated. Furthermore, higher fat and carbohydrate intake, along with lower total food weight and protein intake, was related to eGFRdiff < -15. Moreover, a lower body surface area–adjusted total kidney volume was significantly associated with eGFRdiff < -15, whereas the kidney volume ratio was not.

Comparing the Performance of Predicting Clinical Outcomes Using Various eGFR

Among eGFRcr, eGFRcys, and eGFRcr-cys, we found that eGFRcys demonstrated the best predictive performance for death, MI, and ischemic stroke across all subjects, with C-index values of 0.661, 0.653, and 0.642, respectively, and the area under the ROC curve values for predicting events within 5 years of 0.663, 0.653, and 0.644, respectively (Table S1 and Fig S1). NRI analysis indicated that eGFRcys had significantly higher C-index values for death, MI, and ischemic stroke compared with eGFRcr and eGFRcr-cys. This trend was particularly pronounced in the eGFRdiff < -15 subgroup, in which eGFRcys also showed higher C-index values in NRI analysis compared with eGFRcr and eGFRcr-cys.

Performance of Regression Models

Given the superior clinical outcome predictive performance of eGFRcys, regression models were developed using eGFRcr and associated clinical variables to approximate eGFRcys. The simplified models, including the linear regression model and machine learning regression models, included variables such as eGFRcr, sex, age, height, weight, and BMI, whereas the full models included all variables from the simplified models plus fat-free mass, systolic blood pressure, diastolic blood pressure, cardiocerebrovascular history, hypertension medication, insulin, dyslipidemia medication, cancer history, diabetes history, and smoking status. In the test data set, our simplified linear regression model more accurately approximated eGFRcys than unadjusted eGFRcr, as reflected by improved performance metrics. Moreover, both the simplified and full models demonstrated consistently enhanced predictive performance for clinical outcomes—including death, MI, and ischemic stroke—relative to unadjusted eGFRcr, as evidenced by superior area under the ROC curve, C-index, and NRI measures (Tables S2-S4 and Fig S2). Although the machine learning models showed marginally better performance than the linear regression models, the improvement was minimal.

DISCUSSION

This study, consistent with prior research, confirmed that more negative values of eGFRdiff between eGFRcys and eGFRcr are associated with increased risks of death, MI, and ischemic stroke. In addition, factors such as higher waist and hip circumference, higher body weight, lower fat-free mass, and food intake—particularly increased fat

Table 3. Standardized Odds Ratio for eGFRcys - eGFRcr < -15

Variables	Odds Ratio	P Value	Odds Ratio	P Value
Female	1.44 (1.40-1.47)	<0.001	1.45 (1.41-1.49)	<0.001
Age (y)	1.15 (1.13-1.16)	<0.001	1.13 (1.11-1.14)	<0.001
Waist circumference	1.68 (1.64-1.72)	<0.001	1.63 (1.59-1.67)	<0.001
Hip circumference	1.01 (0.98-1.03)	0.63	1.03 (1.01-1.05)	0.008
Height	0.85 (0.80-0.91)	<0.001	0.85 (0.80-0.91)	<0.001
BMI	0.87 (0.80-0.95)	0.002	0.87 (0.80-0.95)	0.003
Weight	2.09 (1.87-2.33)	<0.001	2.05 (1.84-2.29)	<0.001
Diastolic blood pressure			1.01 (1.00-1.03)	0.16
Systolic blood pressure			0.99 (0.98-1.01)	0.41
Cardiocerebrovascular			1.01 (1.00-1.02)	0.13
Hypertension			1.08 (1.07-1.09)	<0.001
Insulin			1.02 (1.02-1.03)	<0.001
Dyslipidemia			0.90 (0.89-0.91)	<0.001
Cancer			1.05 (1.04-1.06)	<0.001
Diabetes			1.05 (1.04-1.06)	<0.001
Smoking			1.19 (1.18-1.20)	<0.001
Fat-free mass	0.40 (0.38-0.42)	<0.001	0.40 (0.38-0.42)	<0.001
Food weight intake ^a			0.94 (0.92-0.97)	<0.001
Protein intake ^a			0.73 (0.70-0.76)	<0.001
Fat intake ^a			1.16 (1.12-1.20)	<0.001
Carbohydrate intake ^a			1.30 (1.26-1.35)	<0.001
Total kidney volume (BSA adjusted) ^b			0.81 (0.77-0.85)	<0.001
Total kidney volume ratio ^b			1.00 (0.96-1.04)	0.98

Abbreviations: BMI, body mass index; BSA, body surface area; eGFR, estimated glomerular filtration rate; eGFRcr, creatinine-based Chronic Kidney Disease Epidemiology Collaboration eGFR; eGFRcys, cystatin C-based Chronic Kidney Disease Epidemiology Collaboration eGFR.

^aNo. of participants: 52,651, representing those with available food intake data.

^bNo. of participants: 27,104, representing those with available kidney volume data.

and carbohydrate consumption with reduced protein intake—were associated with lower eGFRcys -eGFRcr values. These significant variables, along with eGFRcr, were used to develop a regression model to approximate eGFRcys, which was validated and proven effective for predicting clinical outcomes.

Numerous studies have indicated that negative eGFRcys - eGFRcr values are linked to elevated risks of cardiovascular outcomes and all-cause mortality.^{7-10,12} Our results are consistent with these findings, demonstrating higher risks for death, MI, and ischemic stroke associated with lower eGFRcys values compared with eGFRcr. The underlying pathophysiological mechanisms proposed to explain this phenomenon are diverse. Advanced age, reduced muscle mass, and low physical activity are known to contribute to the overestimation of eGFRcr compared with true GFR.^{5,24} In addition, cystatin C levels can be influenced by thyroid function, inflammation, and obesity, potentially resulting in a lower eGFRcys value compared with true GFR.²⁵⁻²⁷ Furthermore, an eGFRcys to eGFRcr ratio below 0.6, known as shrunken pore syndrome, is characterized by decreased elimination of medium-sized molecules. This leads to the accumulation of proteins that promote atherosclerosis and inflammation, factors known to adversely affect cardiovascular outcomes.^{28,29}

The main strength of this study is the identification of various clinicodemographic factors associated with negative eGFRcys - eGFRcr values using the deeply phenotyped UK

Biobank data. Variables associated with lower eGFRcys - eGFRcr values included high waist/hip circumference, high body weight, low fat-free mass, high fat/carbohydrate intake, low protein intake, and diabetes. These characteristics, indicative of poor metabolic status, can also elevate the risk of cardiovascular outcomes and mortality.³⁰ Notably, low fat-free mass combined with high fat and carbohydrate intake is generally associated with higher inflammation³¹⁻³⁴ and a body composition more likely to include excessive adipose tissue, possibly leading to higher cystatin C levels and, consequently, more negative eGFRcys - eGFRcr values.³⁵ Moreover, low fat-free mass often indicates reduced muscle mass, which can result in lower creatinine values, thus presenting as negative eGFRcys - eGFRcr values.⁵ In a similar vein, older age is associated with a decline in muscle mass, which may further contribute to these observed discrepancies in eGFR.³⁶ In addition, the influences of sex on creatinine levels were found to be associated with lower eGFRcys - eGFRcr values, whereas factors such as male gender and smoking were correlated with higher cystatin C levels, aligning with the study findings.³⁷ Furthermore, this study found that a smaller kidney volume was associated with lower eGFRdiff, revealing correlations that had not been previously identified.

Furthermore, the predictive performance for clinical outcomes was superior with eGFRcys compared with eGFRcr and eGFRcr-cys for all outcomes. This is attributed to the lesser impact of nonkidney factors on cystatin C than

on creatinine, enhancing its prognostic performance. The regression models developed using demographic and medical history data related to $eGFR_{diff} < -15$ demonstrated superior predictive performance for clinical outcomes compared with $eGFR_{cr}$. Notably, even regression models using only simple, easily measurable variables such as sex, age, height, weight, and BMI outperformed $eGFR_{cr}$. The linear regression model, compared with other machine learning regression models, showed no notable disparities in errors when compared with $eGFR_{cys}$, suggesting it is a simple model with acceptable accuracy. Using the models developed in this study could aid outpatient services in centers where cystatin C measurement is unavailable, allowing for close approximation to $eGFR_{cys}$ and superior predictive ability for outcomes.^{13,38}

The study has several limitations. First, the absence of true GFR data means it is unclear whether the measured values of $eGFR_{cr}$ and $eGFR_{cys}$ are higher or lower than the true GFR values. Without true GFR values, it is also challenging to assert that the performance of the regression model accurately approximates true GFR. Second, as baseline creatinine and cystatin C were measured only once, inaccuracies owing to the state at the time of measurement may exist. Last, external validation of the constructed adjustment model is necessary for its widespread application.

In conclusion, this study corroborated previous findings that a lower $eGFR_{cys}$ - $eGFR_{cr}$ value is associated with increased risks of death, MI, and ischemic stroke. Furthermore, it identified clinical variables associated with lower $eGFR_{cys}$ - $eGFR_{cr}$ values and developed an adjustment model using these variables. Using this model can assist in prognosis prediction without measuring cystatin C, thus potentially benefiting outpatient care.

SUPPLEMENTARY MATERIALS

Supplementary File (PDF)

Figure S1: Receiver operating characteristic curves and the area under the receiver operating characteristic curve of creatinine, cystatin C, and creatinine-cystatin C based estimated glomerular filtration.

Figure S2: Receiver operating characteristic curves and the area under the receiver operating characteristic curve of adjust models.

Table S1: C-index of univariable Cox proportional hazards model for outcomes.

Table S2: Performances of regression models in test data.

Table S3: Equations of linear regression models.

Table S4: C-index of univariable Cox proportional hazards model using linear regression models for outcomes in test data.

ARTICLE INFORMATION

Authors' Full Names and Academic Degrees: Min W. Kang, MD, Jae-ik Oh, MD, Jinsun Lee, MD, Minsang Kim, MD, Jung H. Koh, MD, Jeong M. Cho, MD, Seong G. Kim, MD, Semin Cho, MD, Soojin Lee, MD, PhD, Yaerim Kim, MD, PhD, Dong K. Kim, MD, PhD, and Sehoon Park, MD, PhD.

Authors' Affiliations: Department of Internal Medicine (MWK), Korea University Guro Hospital, Seoul, Korea; Department of

Internal Medicine (JO, JL, MK, JHK, DKK, SP), Seoul National University Hospital, Seoul, Korea; Department of Internal Medicine (JMC, SC), Chung-Ang University Gwangmyeong Hospital, Gyeonggi-do, Korea; Department of Internal Medicine (SGK), Inje University Sanggye Paik Hospital, Seoul, Korea; Department of Internal Medicine (SL), Uijeongbu Eulji University Medical Center, Uijeongbu, Korea; and Department of Internal Medicine (YK), Keimyung University School of Medicine, Daegu, Korea.

Address for Correspondence: Sehoon Park, MD, PhD, Department of Internal Medicine, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul, 03080, Korea. Email: mailofsehoon@gmail.com

Authors' Contributions: MWK and SP designed the study. MWK and SP collected, generated, and analyzed data. MWK, JO, JL, MK, JHK, JMC, SGK, SC, SL, YK, and DKK interpreted data. Each author contributed important intellectual content during manuscript drafting or revision and accepts accountability for the overall work by ensuring that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved.

Support: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (grant number: RS-2024-00345867). The authors independently performed this study.

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

Data Sharing: The data sets analyzed in the current study are publicly available and can be accessed through the UK Biobank.

Acknowledgments: This study was based on data provided by the UK Biobank Consortium (application number: 53799).

Peer Review: Received September 06, 2024. Evaluated by 1 external peer reviewer, with direct editorial input from the Statistical Editor, an Associate Editor, and the Editor-in-Chief. Accepted in revised form May 08, 2025.

REFERENCES

- Skali H, Uno H, Levey AS, Inker LA, Pfeffer MA, Solomon SD. Prognostic assessment of estimated glomerular filtration rate by the new Chronic Kidney Disease Epidemiology Collaboration equation in comparison with the Modification of Diet in Renal Disease Study equation. *Am Heart J*. 2011;162(3):548-554.
- Inker LA, Astor BC, Fox CH, et al. KDOQI US commentary on the 2012 KDIGO clinical practice guideline for the evaluation and management of CKD. *Am J Kidney Dis*. 2014;63(5):713-735.
- Inker LA, Eneanya ND, Coresh J, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N Engl J Med*. 2021;385(19):1737-1749.
- Stevens PE, Ahmed SB, Carrero JJ, et al. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int*. 2024;105(4S):S117-S314.
- Baxmann AC, Ahmed MS, Marques NC, et al. Influence of muscle mass and physical activity on serum and urinary creatinine and serum cystatin C. *Clin J Am Soc Nephrol*. 2008;3(2):348-354.
- Salive ME, Jones CA, Guralnik JM, Agodoa LY, Pahor M, Wallace RB. Serum creatinine levels in older adults: relationship with health status and medications. *Age Ageing*. 1995;24(2):142-150.
- Shlipak MG, Matsushita K, Ärnlöv J, et al. Cystatin C versus creatinine in determining risk based on kidney function. *N Engl J Med*. 2013;369(10):932-943.

8. Potok OA, Ix JH, Shlipak MG, et al. The difference between cystatin C- and creatinine-based estimated GFR and associations with frailty and adverse outcomes: a cohort analysis of the Systolic Blood Pressure Intervention Trial (SPRINT). *Am J Kidney Dis.* 2020;76(6):765-774.
9. Potok OA, Katz R, Bansal N, et al. The difference between cystatin C- and creatinine-based estimated GFR and incident frailty: an analysis of the Cardiovascular Health Study (CHS). *Am J Kidney Dis.* 2020;76(6):896-898.
10. Purde MT, Nock S, Risch L, et al. Ratio of cystatin C and creatinine-based estimates of the glomerular filtration rate predicts mortality in healthy seniors independent of kidney function. *Scand J Clin Lab Invest.* 2016;76(4):341-343.
11. Heo GY, Koh HB, Jung CY, et al. Difference between estimated GFR based on cystatin C versus creatinine and incident atrial fibrillation: a cohort study of the UK Biobank. *Am J Kidney Dis.* 2024;83(6):729-738.e1.
12. Kim H, Park JT, Lee J, et al. The difference between cystatin C- and creatinine-based eGFR is associated with adverse cardiovascular outcome in patients with chronic kidney disease. *Atherosclerosis.* 2021;335:53-61.
13. Teaford HR, Rule AD, Mara KC, et al. Patterns of cystatin C uptake and use across and within hospitals. *Mayo Clin Proc.* 2020;95(8):1649-1659.
14. Ristiniemi N, Lund J, Tertti R, et al. Cystatin C as a predictor of all-cause mortality and myocardial infarction in patients with non-ST-elevation acute coronary syndrome. *Clin Biochem.* 2012;45(7-8):535-540.
15. Svensson-Färbom P, Ohlson Andersson M, Almgren P, et al. Cystatin C identifies cardiovascular risk better than creatinine-based estimates of glomerular filtration in middle-aged individuals without a history of cardiovascular disease. *J Intern Med.* 2014;275(5):506-521.
16. Allen N, Sudlow C, Downey P, et al. UK Biobank: current status and what it means for epidemiology. *Health Policy Techn.* 2012;1(3):123-126.
17. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779.
18. Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation.* 2014;129(25)(suppl 2):S49-S73.
19. Lissner L, Björkelund C, Heitmann BL, Seidell JC, Bengtsson C. Larger hip circumference independently predicts health and longevity in a Swedish female cohort. *Obes Res.* 2001;9(10):644-646.
20. Parker ED, Pereira MA, Stevens J, Folsom AR. Association of hip circumference with incident diabetes and coronary heart disease: the Atherosclerosis Risk in Communities study. *Am J Epidemiol.* 2009;169(7):837-847.
21. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med.* 2009;150(9):604-612.
22. Chen DC, Shlipak MG, Scherzer R, et al. Association of intraindividual difference in estimated glomerular filtration rate by creatinine vs cystatin C and end-stage kidney disease and mortality. *JAMA Netw Open.* 2022;5(2):e2148940.
23. Park S, Lee S, Kim Y, et al. Causal linkage between adult height and kidney function: an integrated population-scale observational analysis and Mendelian randomization study. *PLoS One.* 2021;16(7):e0254649.
24. Stevens LA, Coresh J, Greene T, Levey AS. Assessing kidney function—measured and estimated glomerular filtration rate. *N Engl J Med.* 2006;354(23):2473-2483.
25. Okura T, Jotoku M, Irita J, et al. Association between cystatin C and inflammation in patients with essential hypertension. *Clin Exp Nephrol.* 2010;14(6):584-588.
26. Fricker M, Wiesli P, Brändle M, Schwegler B, Schmid C. Impact of thyroid dysfunction on serum cystatin C. *Kidney Int.* 2003;63(5):1944-1947.
27. Muntner P, Winston J, Uribarri J, Mann D, Fox CS. Overweight, obesity, and elevated serum cystatin C levels in adults in the United States. *Am J Med.* 2008;121(4):341-348.
28. Grubb A, Lindström V, Jonsson M, et al. Reduction in glomerular pore size is not restricted to pregnant women. Evidence for a new syndrome: 'Shrunken pore syndrome'. *Scand J Clin Lab Invest.* 2015;75(4):333-340.
29. Almén MS, Björk J, Nyman U, et al. Shrunken pore syndrome is associated with increased levels of atherosclerosis-promoting proteins. *Kidney Int Rep.* 2018;4(1):67-79.
30. Malik S, Wong ND, Franklin SS, et al. Impact of the metabolic syndrome on mortality from coronary heart disease, cardiovascular disease, and all causes in United States adults. *Circulation.* 2004;110(10):1245-1250.
31. Oliveira MC, Menezes-Garcia Z, Henriques MCC, et al. Acute and sustained inflammation and metabolic dysfunction induced by high refined carbohydrate-containing diet in mice. *Obesity.* 2013;21(9):E396-E406.
32. Oliveira G, Oliveira C, Gaspar I, et al. Fat-free mass depletion and inflammation in patients with bronchiectasis. *J Acad Nutr Diet.* 2012;112(12):1999-2006.
33. Gregersen S, Samocha-Bonet D, Heilbronn LK, Campbell LV. Inflammatory and oxidative stress responses to high-carbohydrate and high-fat meals in healthy humans. *J Nutr Metab.* 2012;2012:238056.
34. Duan Y, Zeng L, Zheng C, et al. Inflammatory links between high fat diets and diseases. *Front Immunol.* 2018;9:2649.
35. Naour N, Fellahi S, Renucci JF, et al. Potential contribution of adipose tissue to elevated serum cystatin C in human obesity. *Obesity (Silver Spring).* 2009;17(12):2121-2126.
36. Wilkinson DJ, Piasecki M, Atherton PJ. The age-related loss of skeletal muscle mass and function: measurement and physiology of muscle fibre atrophy and muscle fibre loss in humans. *Ageing Res Rev.* 2018;47:123-132.
37. 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.
38. Lasserson DS, Shine B, O'Callaghan CA, James T. Requirement for cystatin C testing in chronic kidney disease: a retrospective population-based study. *Br J Gen Pract.* 2017;67(663):e732-e735.