Integrating multiple kidney function markers to predict all-cause and cardiovascular disease mortality: prospective analysis of 366,758 UK Biobank participants

Ryosuke Fujii1,2,3, Roberto Melotti1, Anna Köttgen4,5, Alexander Teumer6,7, Daniele Giardiello1, Cristian Pattaro1

1. Institute for Biomedicine, Eurac Research, Via A. Volta 21, Bolzano, Italy.
2. Department of Preventive Medical Science, Fujita Health University School of Medical Sciences, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Japan
3. Department of Preventive Medicine, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya, Japan
4. Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Center–University of Freiburg, Freiburg, Germany
5. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
6. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany
7. DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany

Correspondence to:
Ryosuke Fujii; E-mail: rfujii@fujita-hu.ac.jp
ABSTRACT

Background & Hypothesis. Reduced kidney function is a risk factor of cardiovascular and all-cause mortality. This association was demonstrated for several kidney function markers, but it is unclear whether integrating multiple measured markers may improve mortality risk prediction.

Methods. We conducted an exploratory factor analysis (EFA) of serum creatinine- and cystatin C-based estimated glomerular filtration rate (eGFRcre and eGFRcys, derived by Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) and European Kidney Function Consortium (EKFC) equations), blood urea nitrogen (BUN), uric acid, and serum albumin among 366,758 participants of the UK Biobank without history of kidney failure. Fitting Cox-proportional hazard models, we compared ability of the identified latent factors
to predict overall mortality and mortality by cardiovascular disease (CVD), also considering
CVD-specific causes like coronary heart disease and cerebrovascular disease.

**Results.** During 12.5 years of follow-up, 26,327 deceased from any cause, 5,376 died from
CVD, 2,908 from CHD, and 1,116 from cerebrovascular disease. We identified two latent
factors, EFA1 and EFA2 both representing kidney function variations. When using the CKD-
EPI equations, EFA1 performed like eGFRcys, with EFA1 showing slightly larger hazard
ratios for overall and CVD-related mortality. At 10-years of follow-up, EFA1 and eGFRcys
showed moderate discrimination performance for CVD-related mortality, outperforming all
other kidney indices. eGFRcre was the least predictive marker across all outcomes. When
using the EKFC equations, eGFRcys performed better than EFA1, all other results remaining
similar.

**Conclusions.** While EFA is an attractive approach to capture the complex effects of kidney
function, eGFRcys remains the most practical and effective measurement for all-cause and
CVD mortality risk prediction.
KEY LEARNING POINTS

What was known:

- Reduced kidney function is a risk factor for all-cause and cardiovascular mortality.
- eGFRcys predicts incident CVD and mortality better than eGFRcre across different age groups and conditions.
- Although utility of single kidney-related biomarker was examined, it is unclear whether integrating multiple biomarkers improves mortality prediction.

This study adds:

- The present study applied exploratory factor analysis (EFA) to integrate five kidney-related biomarkers using the UK Biobank dataset.
- The main identified latent factor and eGFRcys are the best predictors of all-cause and cardiovascular mortality.
- Among the studied kidney function biomarkers, eGFRcre is the worst predictor for all-cause and CVD mortality.

Potential impact:

- eGFRcys remains the most practical and effective marker for all-cause and cardiovascular mortality risk prediction.
- EFA should be expanded to include specific molecular markers that may reflect the different dimensions of kidney function.

Keywords: cardiovascular disease, cystatin C, eGFR, factor analysis, kidney function
INTRODUCTION

Chronic kidney disease (CKD) affects over 700 million people worldwide\(^1\) and is on the rise to become the fifth leading cause of death, surpassing diabetes and most non-communicable diseases.\(^2\) CKD is a risk factor for incidence of kidney failure and cardiovascular disease (CVD) as well as all-cause and CVD mortality.\(^3,4\) CKD is defined as a reduced kidney function, which is normally quantified by the glomerular filtration rate (GFR), estimated using serum creatinine (eGFRcre)\(^5\) or cystatin C (eGFRcys).\(^6\)

Previous studies suggested that eGFRcys predicts incident CVD and mortality better than eGFRcre across different age groups and conditions.\(^7–10\) However, it is unclear whether combining eGFRcys with other commonly measured biochemical parameters of kidney function may further improve the prediction of overall mortality as well as mortality by cardiovascular-specific causes. Multivariate statistical analysis techniques include exploratory factor analysis (EFA) and structural equation modeling, which aim to identify latent, unobservable structures underlying observable markers measured with error. In the specific case of kidney function assessment, such markers include eGFRcre, eGFRcys, blood urea nitrogen (BUN), uric acid (UA), and serum albumin. Each of them is informative about kidney function but also reflect variability caused by the marker-specific metabolism. Additionally, eGFRcre and eGFRcys are just approximations of the real, underlying GFR, which is not measurable in large population samples.

To address this issue, we conducted an integrated analysis of five kidney function markers, typically assessed in clinical practice, in 366,758 participants to the UK Biobank (UKBB), followed over 12.5 years. Through EFA, we estimated latent signatures of kidney function, which were compared against Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) and European Kidney Function Consortium (EKFC) GFR estimating equations in
terms of ability to predict all-cause mortality as well as cause-specific mortality, namely CVD-, coronary heart disease (CHD)-, and cerebrovascular-related mortality.

MATERIALS AND METHODS

Study participants

The UKBB is a population-based study involving general population individuals, approved by the North West – Haydock Research Ethics Committee (No. 16/NW/0274) and described in detail elsewhere. We acquired a dataset including biochemical and demographic data on 502,410 participants (application no. 20272). We excluded 605 participants with diagnosis of kidney failure at baseline, and 135,781 participants with missing values on any fundamental variable (Supplementary Table S1), leaving 366,629 participants for analysis (Supplementary Figure S1).

Outcome definition

Information on participation date and cause of death was obtained through linkage to the National Health System registry. We classified ICD-10 codes “I00-I99” as CVD mortality, “I20-I25” as CHD mortality, and “I60-I69” as cerebrovascular disease mortality. For each participant, time-to-death was defined from the date of participation until the date of death or censoring (31 October 2021 in England and Wales; 30 September 2021 in Scotland).

Biomarker definition

We collected related information of five markers: serum creatinine, serum cystatin C, UA, BUN (derived as 0.467×urea), and serum albumin (information at the UKBB Data Showcase,
eGFRcre, eGFRcys, and creatinine- and cystatin C-based eGFR (eGFRcrecys) were estimated with CKD-EPI 2021 formulas without a race variable.\textsuperscript{5,6} Given that the CKD-EPI formula might not be considered reliable among Europeans\textsuperscript{12}, we additionally derived eGFRcre\textsuperscript{13} and eGFRcys\textsuperscript{14} using the EKFC formulas.

All GFR estimations were obtained using the R package “nephro” ver.1.3 (https://cran.r-project.org/web/packages/nephro/index.html).\textsuperscript{15}

Exploratory factor analysis

EFA is a statistical method to estimate unobserved factors, called latent factors, underlying a set of measured variables, called manifest variables. The technique exploits the network of pairwise correlations between the manifest variables to infer latent factors that influence one or more manifest variables. Factor loadings measure the influence of a latent factor on manifest variables. We conducted EFA of the five, normally distributed kidney-related markers (eGFRcre, eGFRcys, UA, BUN, and albumin), to identify factor loadings based on a maximum likelihood approach. The number of relevant factors to retain was determined via the scree plot, which reflects the specific and cumulative standardized (returned by the respective eigenvalue) amount of variance explained by each ordered latent factor. We applied factor rotation to simplify the structure of the pattern matrix of factor loadings and aid factor interpretations. Specifically, we chose the oblique promax rotation, which builds on the orthogonal varimax rotation, in an attempt to further discriminate the factor loadings assigned to each latent factor for the same items, at the cost of allowing for correlation among the latent factors. EFA was performed with the R package “psych” (ver.2.2.3).
Statistical analysis

Cox regression models were fitted to assess associations of the five kidney markers with all-cause, CVD, CHD, and cerebrovascular mortality, using the R package “Survival” ver.3.5.5. For all outcomes, we fitted unadjusted and fully-adjusted models including sex and age, as well as baseline body mass index (BMI), self-reported ancestry (White vs. Non-White), hypertension (defined as systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg or taking antihypertensive drug treatment), type 2 diabetes (T2D, defined as blood glucose levels≥7.0 mmol/l or hemoglobin A1c≥6.5% or taking antidiabetic medication), and tobacco smoking (never vs. ever). Hazard ratios (HR) and 95% confidence intervals (CI) were estimated per 1 standard deviation (1-SD) change of each kidney biomarker. Sensitivity analyses were performed including adjustment for baseline albumin-to-creatinine ratio (UACR) and C reactive protein (CRP) levels.

Model discrimination ability was assessed using the Harrell’s\textsuperscript{16} and Uno’s\textsuperscript{17} C-statistics. Calibration at a fixed time point of 10 years was assessed by comparison of the observed outcome rate versus the expected survival probability using 10-fold cross-validation. All statistical analyses were performed using the R software ver.4.1.0 (http://www.R-project.org).
RESULTS

Study participants’ characteristics

Over a median follow-up of 12.5 years (interquartile range, 11.8-13.2), 26,327 deceased from any cause, 5,376 died from CVD, including 2,908 who died from CHD and 1,116 from cerebrovascular disease (Table 1 & Supplementary Table S2). At baseline, the mean age of the alive group was 61.7 (SD: 6.4) years and the mean age of the deceased group was 56.2 (SD: 8.1) year. The baseline levels of eGFR, UA, and BUN were substantially altered in those who died of any specific cause. Regarding the CVD-specific causes of death, we observed a higher percent of males among CHD-related compared to cerebrovascular-related deaths (79.8% versus 52.3%). At baseline, among individuals who died by CHD we observed higher percentages of diabetes (20.4%) and ever smokers (22.6%) than in those who died by cerebrovascular disease (12.9% and 17.7%, respectively).

Exploratory Factor Analysis

Scree plot inspection (Supplementary Figure S2) supported retention of two latent factors that we labeled EFA1 and EFA2. Both factors were compatible representations of kidney function, as they displayed direction-concordant factor loadings between eGFRcre and eGFRcys and between BUN and UA, and direction-discordant loadings of eGFR versus BUN and UA. For both factors, the loading of serum albumin was negligible. Specifically, EFA1 was dominated by eGFRcys, while EFA2 was characterized by a balanced combination of eGFRcre and both BUN and UA, with a lower loading on eGFRcys (Figure 1A). EFA1 and EFA2 were negatively correlated (Pearson’s $r=-0.65$; Figure 1B). EFA1 showed nearly perfect correlation with eGFRcys ($r=0.99$). EFA2 was positively correlated with BUN ($r=0.84$) and uric acid ($r=0.45$) and negatively correlated with eGFRcys ($r=-0.77$) and eGFRcre ($r=-0.75$). For interpretation, lower levels of EFA1 reflect lower kidney function, whereas higher levels
of EFA2 reflect lower kidney function. According to linear regression, a 1-SD lower EFA1 corresponded to 6.73 (95% CI: 6.70 to 6.76) ml/min/1.73m² lower eGFRcre and 16.00 (95% CI: 15.99 to 16.01) ml/min/1.73m² lower eGFRcys. A 1-SD larger EFA2 corresponded to 10.98 (95% CI: 10.95 to 11.01) ml/min/1.73m² lower eGFRcre, and 13.84 (95% CI: 13.80 to 13.88) ml/min/1.73m² lower eGFRcys.

Survival analysis

All five kidney indices were associated with all-cause and each cause-specific mortality without adjustment for potential confounders (Figure 2A). EFA1 and eGFRcys showed the largest effects. We observed HRs (95% CI) of 1.85 (1.83–1.87) and 1.87 (1.85–1.90) for all-cause mortality per each SD lower EFA1 and eGFRcys, respectively. eGFRcrecys showed smaller effects compared with EFA1 and eGFRcys. Then, eGFRcre displayed the smallest HR across all outcomes in this study. When adjusting for cardiovascular, metabolic and lifestyle factors, we observed an attenuation of all HRs (Figure 2B). This reflects the identification of relevant causal pathways that did not, however, nullify the effect of kidney function on mortality, pointing towards a partially independent effect. Relative to each other, the ranking performance of each marker reflected the same pattern observed in the unadjusted analysis, with slightly smaller effects of kidney function markers on cerebrovascular mortality compared to other types of CVD mortality. However, despite the demographic and clinical differences at baseline between cerebrovascular and CHD mortality cases, observed differences are minor.

Model performance

The same pattern observed for the HR was observed for the model’s discrimination performances (Figure 3A): in models using only biomarkers, EFA1 and eGFRcys performed
better than all other markers, and comparably well with respect to each: for CVD mortality, the C-statistics (95% CI) was 0.71 (0.71–0.72) and 0.71 (0.70–0.72), respectively. eGFRcre showed the worst performance for any cause of mortality. Discrimination performances improved when additionally accounting for cardiovascular, metabolic, and lifestyle factors, for every cause of mortality (Figure 3B). Similar results were obtained with the Uno’s C-statistic (Supplementary Figure S3). Prediction accuracy was also assessed with calibration for the observed and the expected survival probability at 10 years (Supplementary Table S3).

For all-cause mortality, there was almost perfect agreement between the predicted survival probabilities and the observed data, while there was weak agreement for predicting CVD mortality.

Analyses based on the European Kidney Function Consortium formulas

Instead of CKD-EPI formula, we re-assessed our analyses by estimating eGFRcre and eGFRcys with the EKFC formulas. EKFC-based eGFRcys (Figure 2D) showed larger HRs for all-cause and CVD mortality than CKD-EPI-based eGFRcys (Figure 2B). The effect sizes of eGFRcre were substantially equivalent. While results were essentially consistent with those observed using the CKD-EPI formulas, in the EKFC analysis EFA1 was not anymore equivalent to eGFRcys, leaving eGFRcys as the best mortality predictor. For discriminatory ability, there was no substantial difference of model performance between EKFC-based eGFRcys (Figure 3C) and CKD-EPI eGFRcys (Figure 3A). Similar to CKD-EPI formula (Figure 3B), including clinical risk factors into the model universally improved discriminatory performance across kidney indices (Figure 3D).
Sensitivity analyses

We incorporated UACR into the EFA based on the CKD-EPI equations (Supplementary Figure S5): this new analysis identified only one factor from five biomarkers. The standardized factor loading of UACR was almost null ($\lambda=-0.08$), while factor loadings of eGFRcre and eGFRcys were of 0.81 and 0.72. The estimated latent factor performed much worse than eGFRcys and eGFRcrecys to predict CVD mortality. Finally, when repeated the analysis including CRP, a marker of inflammation, as a covariate in all Cox-proportional hazard models, results did not change (right column; Supplementary Table S4).

DISCUSSION

Reduced kidney function is a powerful predictor of cardiovascular and all-cause mortality. While eGFRcys is known as the best mortality predictor among kidney function indices, it was unclear whether a combination of multiple kidney function markers could further improve mortality prediction models. Here we show that combining multiple markers provides negligible or no gain to using eGFRcys alone. In contrast, the commonly used index eGFRcre had the worst predictive performance across causes of death. Our results present a compendium of the relevance of kidney function on all-cause mortality and on mortality by specific causes such as CVD, CHD, and cerebrovascular disease.

Our analysis involved five kidney markers and was conducted on about 360,000 adults from the general population, with more than ten-year follow-up, enabling the observation of several events linked to different causes of death. In the absence of objectively measured kidney function, which is unfeasible in large population studies, the EFA provided
an opportunity to derive a latent signature of the true kidney function, overcoming the
disadvantages related to each single marker. We identified two correlated factors: EFA1,
essentially corresponding to eGFRcys, and EFA2, representing a balanced covariation of
eGFRcre, BUN, and urate. Both factors are consistent with the identification of kidney-
function related variability. However, EFA1 predicted any type of cardiovascular and overall
mortality better than EFA2. This may reflect either a relatively stronger relation of cystatin C
with CVD-related factors, or the more detrimental effect of kidney function on mortality,
compared to the kidney functionality captured by the other markers net of cystatin, regardless
of the type of equation used\textsuperscript{18}.

Comparison of unadjusted versus adjusted models highlight the extent to which
kidney-function-based prediction of all-cause and cause-specific mortality is attenuated by
the included confounders, across markers. After removing the effect of cardiometabolic and
behavioral determinants of mortality, adjusted analysis results enable to appreciate the
independent and non-negligible effect of kidney function on mortality. When looking at the
C-statistics, the unadjusted model shows the discriminatory ability of each individual marker
in isolation and allowing for confounding. However, the adjusted C-statistic does not show
the unconfounded discrimination ability of the individual markers but rather the joint effect
of the markers and all confounders to predict the outcome. That is why the C-statistics of the
adjusted models were nearly identical across kidney function markers for all outcomes.

Our complementary analysis with EKFC instead of CKD-EPI formulas showed that
eGFRcys was the best mortality predictor overall, with larger effects than EFA1 across all
outcomes. The HR for eGFRcys estimated by EKFC was larger than that of eGFRcys using
CKD-EPI formula, in line with recent evidence from non-black population samples reporting
superior discriminatory ability of EKFC formula for all-cause and CVD mortality compared
to the CKD-EPI formula\textsuperscript{19}. 
Our study had several limitations. The young age and selection towards a healthy status of UKBB participants, recommends that findings’ transportability is tested in alternative population and patients’ samples. However, our findings are in line with previous results obtained from both different clinical status and in different age groups. Additionally, the results are in line with our recent investigation of a structural equation model for eGFRcys, eGFRcre, uric acid, and BUN in the longitudinal evaluation of incident CVD risk in an Alpine community, where the identified latent factor showed comparable discrimination performance for CVD risk as eGFRcre alone. Contrary to naive principal component analysis, which is a descriptive dimensionality reduction technique centered on variables, factor analysis is a model-based technique prone to measurement error components centered on observations, which exploits the correlation structure among them to identify latent variables that may help explaining this structure in a simplified way. Despite the limited number of kidney function markers available in the present study, there is no preclusion on their number to perform EFA. While correlated due to oblique rotation for ease of interpretation, the two factors identified, EFA1 and EFA2, are both compatible with a multi-facet representation of kidney function. EFA1 almost exclusively loads on eGFRcys, which is known to capture a portion of non-GFR determinants in contrast to other markers, whereas EFA2 highlights ‘general’ kidney function reflected in consistent levels of BUN, UA and eGFRcre. Ideally, if available, but this was not the case, we would have included additional biomarkers relevant to kidney function such as to capture aspects of filtration, excretion, or damage and compare our results to alternative approaches, which used multiple kidney biomarkers, including for example FGF23 and KIM-1. Further including proteomics or metabolomic markers could be another line of investigation. Such omics data are not typically available in routine biochemical examinations, but they certainly offer an important perspective to the effort to identify the underlying, unobserved signatures of kidney function.
alterations. We did not consider applying factor analysis directly on serum creatinine and
cystatin C, to separate the effect of age and sex embedded within the eGFR equations, as we
have already shown that factor analysis based on eGFRcre and eGFRcys instead of crude
serum creatinine and cystatin C has better predictive performance\textsuperscript{24}. In addition, the use of
age and sex in the GFR estimating equations is instrumental in obtaining the best possible
approximation of the real GFR,\textsuperscript{27} so that eGFRcre and eGFRcys should be considered as the
best possible approximations of kidney function. In our analysis, a moderate proportion
(~27\%) of the cohort with missing data was excluded. Among those who were included
(n=366,629) and those who were excluded from the analysis (n=135,781), we compared the
proportion of all-cause mortality cases (7.2\% vs. 8.2\%) and the mean levels of eGFRcre (94.7
vs. 94.7) and eGFRcys (88.1 vs. 88.1), resulting in similar kidney related health risk profiles.
In view of these considerations, we are confident that the adjusted complete case analysis is
sufficiently robust to the possible effects of selection. Finally, competing causes of deaths
beyond those listed in the manuscript were not considered. However, the comparison between
overall mortality and cause-specific mortality somehow includes the possibility of
competitive events that should factor into the overall mortality. The performance of the
markers on cause-specific mortality follows the same order as for the overall mortality.
Furthermore, the five markers of kidney function and the two estimated latent factors are
closely representative of kidney function itself, making it unlikely that any competitive event
may have influenced the markers differentially.

In conclusion, our results suggest that EFA is a valuable method to disentangle
different factors underlying the distribution of kidney function-related biochemical markers,
and uncover possible relevant latent signatures of kidney health. However, until additional
biomarkers may be integrated and explored further within a similar analytical framework,
eGFRcys may represent the index of choice for both all-cause mortality as well as
cardiovascular mortality, advocating for a better understanding and use of this marker in clinical practice.

CONFLICT OF INTEREST STATEMENT
Cristian Pattaro is a consultant for Quotient Therapeutics (UK).

AUTHORS’ CONTRIBUTIONS
R.F. and C.P. conceived the study and drafted the manuscript; R.F. performed the statistical analyses; A.T. prepared the datasets for analysis; R.M., A.K., and D.G. contributed to the interpretation of results; R.M. and C.P. contributed to the revision of the manuscript. All the authors reviewed and approved the final version of the manuscript.

FUNDING
This present research was supported by the Autonomous Province of Bozen/Bolzano – Department for Innovation, Research and University in the frame of the Seal of Excellence Programme (CUP/D55F20002560003); and the Uehara Memorial Foundation, Oversea Fellowship for Post-doc Students. This work was conducted within approved UK Biobank application number 20272.

DATA AVAILABILITY STATEMENT
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.
REFERENCES


Table 1. Baseline characteristics of UK Biobank participants by different causes of mortality

<table>
<thead>
<tr>
<th>Sample characteristics</th>
<th>Alive (n = 340,302)</th>
<th>All-cause (n = 26,327)</th>
<th>CVD (n = 5,376)</th>
<th>CHD (n = 2,908)</th>
<th>Cerebrovascular (n = 1,116)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs)</td>
<td>56.2 (8.1)</td>
<td>61.7 (6.4)</td>
<td>62.0 (6.3)</td>
<td>61.7 (6.3)</td>
<td>63.0 (5.5)</td>
</tr>
<tr>
<td>Male (n, %)</td>
<td>153,878 (45.2%)</td>
<td>15,742 (59.8%)</td>
<td>3,791 (70.5%)</td>
<td>2,321 (79.8%)</td>
<td>584 (52.3%)</td>
</tr>
<tr>
<td>Non-White (n, %)</td>
<td>4,911 (1.4%)</td>
<td>240 (0.9%)</td>
<td>45 (0.8%)</td>
<td>17 (0.6%)</td>
<td>11 (1.0%)</td>
</tr>
<tr>
<td>Ever smokers (n, %)</td>
<td>33,408 (9.8%)</td>
<td>4,928 (18.7%)</td>
<td>1,119 (20.8%)</td>
<td>658 (22.6%)</td>
<td>197 (17.7%)</td>
</tr>
<tr>
<td>Hypertension (n, %)</td>
<td>176,863 (52.0%)</td>
<td>18,394 (69.9%)</td>
<td>4,272 (79.4%)</td>
<td>2,333 (80.2%)</td>
<td>865 (77.5%)</td>
</tr>
<tr>
<td>Type 2 diabetes (n, %)</td>
<td>16,456 (4.8%)</td>
<td>3,144 (11.9%)</td>
<td>918 (17.1%)</td>
<td>593 (20.4%)</td>
<td>144 (12.9%)</td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td>27.3 (4.7)</td>
<td>28.3 (5.4)</td>
<td>29.2 (5.7)</td>
<td>29.5 (5.4)</td>
<td>28.0 (5.2)</td>
</tr>
<tr>
<td>Systolic blood pressure (mmHg)</td>
<td>137.5 (18.5)</td>
<td>142.6 (19.6)</td>
<td>145.1 (20.8)</td>
<td>145.4 (20.9)</td>
<td>145.6 (20.9)</td>
</tr>
<tr>
<td>Diastolic blood pressure (mmHg)</td>
<td>82.2 (10.1)</td>
<td>82.5 (10.6)</td>
<td>83.0 (11.5)</td>
<td>82.7 (11.4)</td>
<td>83.2 (11.4)</td>
</tr>
<tr>
<td>Serum creatinine (mg/dl)</td>
<td>0.80 (0.2)</td>
<td>0.86 (0.3)</td>
<td>0.91 (0.3)</td>
<td>0.93 (0.3)</td>
<td>0.86 (0.3)</td>
</tr>
<tr>
<td>Cystatin C (mg/l)</td>
<td>0.90 (0.1)</td>
<td>1.01 (0.3)</td>
<td>1.06 (0.3)</td>
<td>1.07 (0.3)</td>
<td>1.02 (0.3)</td>
</tr>
<tr>
<td>Blood urea nitrogen (mg/dl)</td>
<td>15.1 (3.7)</td>
<td>16.0 (5.2)</td>
<td>16.9 (6.3)</td>
<td>17.0 (6.7)</td>
<td>16.6 (5.7)</td>
</tr>
<tr>
<td>Uric acid (g/dl)</td>
<td>5.2 (1.3)</td>
<td>5.6 (1.5)</td>
<td>5.9 (1.6)</td>
<td>6.0 (1.5)</td>
<td>5.5 (1.5)</td>
</tr>
<tr>
<td>Serum albumin (g/l)</td>
<td>4.5 (0.3)</td>
<td>4.5 (0.3)</td>
<td>4.4 (0.3)</td>
<td>4.5 (0.3)</td>
<td>4.5 (0.3)</td>
</tr>
<tr>
<td>eGFRcre (ml/min/1.73m(^2))</td>
<td>95.0 (12.7)</td>
<td>90.3 (15.1)</td>
<td>88.3 (16.5)</td>
<td>88.6 (17.0)</td>
<td>88.6 (14.8)</td>
</tr>
<tr>
<td>eGFRcys (ml/min/1.73m(^2))</td>
<td>89.2 (15.6)</td>
<td>78.1 (17.9)</td>
<td>75.0 (18.7)</td>
<td>75.2 (19.2)</td>
<td>76.4 (17.5)</td>
</tr>
<tr>
<td>eGFRcrecys (ml/min/1.73m(^2))</td>
<td>95.6 (14.0)</td>
<td>86.9 (16.8)</td>
<td>84.0 (18.0)</td>
<td>84.2 (18.4)</td>
<td>85.2 (16.6)</td>
</tr>
<tr>
<td>EFA1</td>
<td>0.05 (1.0)</td>
<td>-0.62 (1.1)</td>
<td>-0.78 (1.1)</td>
<td>-0.77 (1.1)</td>
<td>-0.71 (1.0)</td>
</tr>
<tr>
<td>EFA2</td>
<td>-0.03 (0.9)</td>
<td>0.43 (1.2)</td>
<td>0.67 (1.4)</td>
<td>0.68 (1.4)</td>
<td>0.58 (1.2)</td>
</tr>
</tbody>
</table>

According to the ICD-10 code, the mortality definition of CVD, CHD, and cerebrovascular disease was I00-I99, I21-I25, and I60-I69, respectively.

BMI: body mass index; CVD: cardiovascular disease; eGFRcre: creatinine-based estimated glomerular filtration rate; eGFRcys: cystatin C-based estimated glomerular filtration rate; eGFRcrecys: creatinine- and cystatic C-based glomerular filtration rate; EFA: exploratory factor analysis.
Figure 1. The results from exploratory factor analysis (EFA). A) Standardized factor loadings of five kidney indices for EFA1 and EFA2 and B) Correlation heatmap of EFA-derived variables (EFA1 and EFA2) with eGFR values (eGFRcre, eGFRcys, and eGFRcrecys).

We applied maximum likelihood estimation with promax rotation to estimate factor loadings (see methods for details). eGFRcre: creatinine-based estimated glomerular filtration rate; eGFRcys: cystatin C-based estimated glomerular filtration rate; eGFRcrecys: creatinine- and cystatin C-based glomerular filtration rate; EFA: exploratory factor analysis.
Figure 2. Hazard ratio (HR) and 95% confidence intervals (CI) for all-cause and cause-specific mortality from A) unadjusted and B) fully adjusted models using CKD-EPI eGFRcre and eGFRcys formulas and C) unadjusted and D) adjusted models using EKFC eGFRcre and eGFRcys formulas. The HR and 95% CIs were expressed with a 1-SD
change in each kidney index. 1-SD change in eGFRcre, eGFRcys, and eGFRcrecys corresponded to 13.0, 16.0, and 14.4 ml/min/1.73m², respectively. Estimated by the linear regression model, a 1-SD lower EFA1 corresponded to 6.73 (95% CI: 6.70 to 6.76) ml/min/1.73m² lower eGFRcre and 16.00 (95% CI: 15.99 to 16.01) ml/min/1.73m² lower eGFRcys. A 1-SD larger EFA2 corresponded to 10.98 (95% CI: 10.95 to 11.01) ml/min/1.73m² lower eGFRcre, and 13.84 (95% CI: 13.80 to 13.88) ml/min/1.73m² lower eGFRcys. Fully adjusted models included sex, age, self-reported ancestry, BMI, hypertension, T2D, and tobacco smoking as potential confounders. BMI: body mass index; CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration; CVD: cardiovascular disease; EFA: exploratory factor analysis; eGFRcre: creatinine-based estimated glomerular filtration rate; eGFRcys: cystatin C-based estimated glomerular filtration rate; eGFRcrecys: creatinine- and cystatin C-based glomerular filtration rate; EKFC: European Kidney Function Consortium; T2D: type 2 diabetes.
Figure 3. C-statistics and 95% confidence intervals (CI) for all-cause and cause-specific mortality from A) unadjusted and B) fully adjusted models using CKD-EPI eGFRcre and eGFRcys formulas, and C) unadjusted and D) adjusted models using EKFC eGFRcre and eGFRcys formulas. Fully adjusted models included sex, age, self-reported ancestry, BMI, hypertension, T2D, and tobacco smoking as potential confounders. BMI: body mass index; CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration; CVD: cardiovascular disease; EFA: exploratory factor analysis; eGFRcre: creatinine-based estimated glomerular filtration rate; eGFRcys: cystatin C-based estimated glomerular
filtration rate; eGFRcrecys: creatinine- and cystatin C-based glomerular filtration rate; EKFC: European Kidney Function Consortium; T2D: type 2 diabetes