Clinical Kidney Journal, 2024, vol. 17, no. 8, sfae207

https://doi.org/10.1093/ckj/sfae207
Advance Access Publication Date: 5 July 2024
Original Article

ORIGINAL ARTICLE

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

Ryosuke Fujii [1,2,3], Roberto Melotti[1], Anna Köttgen[4,5], Alexander Teumer [6,7], Daniele Giardiello[1] and Cristian Pattaro[1]

[1]Institute for Biomedicine, Eurac Research, Bolzano/Bozen, Italy, [2]Department of Preventive Medical Science, Fujita Health University School of Medical Sciences, Toyoake, Japan, [3]Department of Preventive Medicine, Nagoya University Graduate School of Medicine, 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 and [7]German Centre for Cardiovascular Research, Partner Site Greifswald, Greifswald, Germany

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

ABSTRACT

Background. 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.

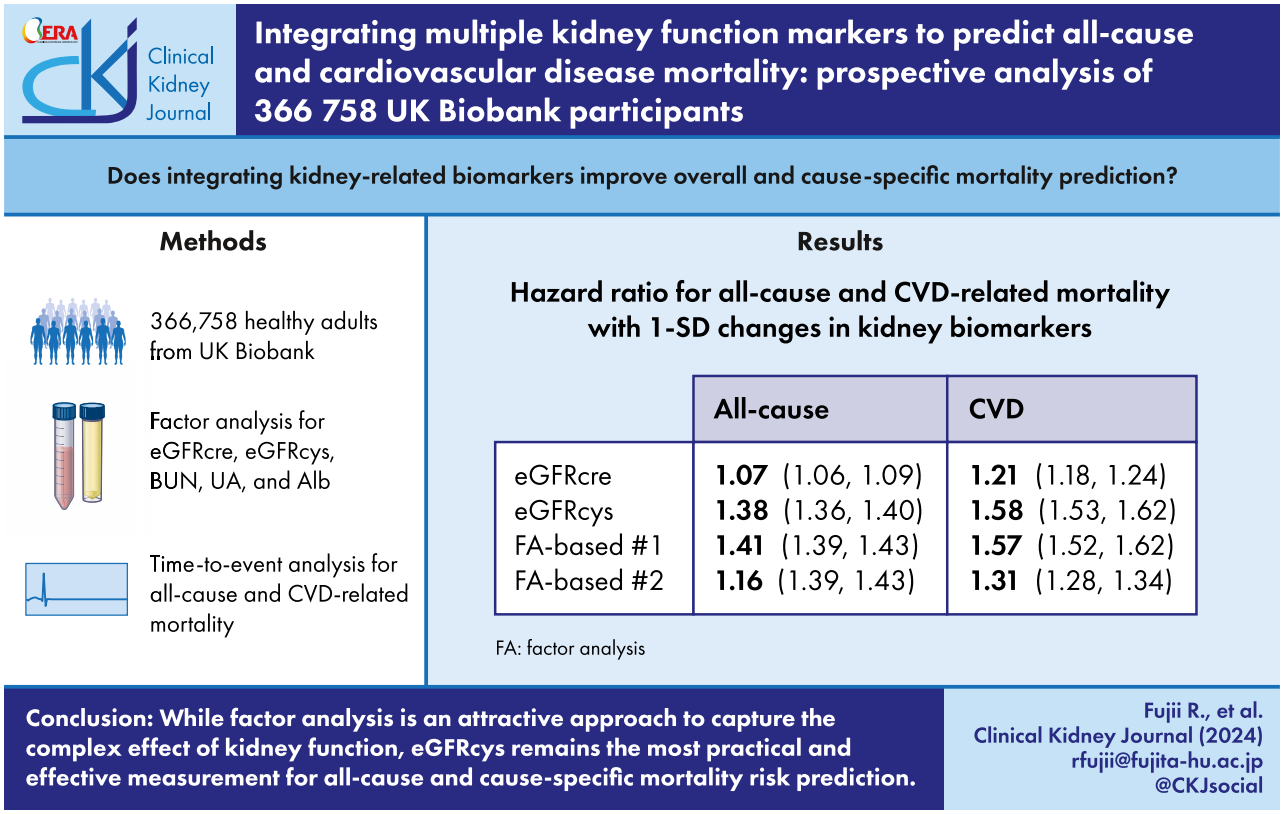
Methods. We conducted an exploratory factor analysis (EFA) of serum creatinine– and cystatin C–based estimated glomerular filtration rate [eGFRcre and eGFRcys; derived by the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) and European Kidney Function Consortium (EKFC) equations], blood urea nitrogen (BUN), uric acid and serum albumin among 366 758 participants in the UK Biobank without a history of kidney failure. Fitting Cox proportional hazards models, we compared the 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 (CHD) and cerebrovascular disease.

Results. During 12.5 years of follow-up, 26 327 participants died from any cause, 5376 died from CVD, 2908 died from CHD and 1116 died from cerebrovascular disease. We identified two latent factors, EFA1 and EFA2, both representing kidney function variations. When using the CKD-EPI equation, 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 equation, eGFRcys performed better than EFA1 while 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.

Received: 24.11.2023; Editorial decision: 5.6.2024

GRAPHICAL ABSTRACT



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

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.

INTRODUCTION

Chronic kidney disease (CKD) affects >700 million people worldwide [1] and is increasing to become the fifth leading cause of death, surpassing diabetes and most non-communicable diseases [2]. CKD is a risk factor for the 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 have 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 modelling, 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 the 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 [11]. We acquired a dataset including biochemical and demographic data for 502 410 participants (application no. 20272). We excluded 605 participants with a 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 Fig. S1).

Outcome definition

Information on participation date and cause of death was obtained through linkage to the National Health System registry. We classified International Classification of Diseases, Tenth Revision (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 for five markers: serum creatinine, serum cystatin C, UA, BUN (derived as 0.467 × urea) and serum albumin (information from the UKBB Data Showcase; http://biobank.ctsu.ox.ac.uk/crystal/). eGFRcre, eGFRcys and creatinine- and cystatin C–based eGFR (eGFRcrecys) were estimated with the 2021 CKD-EPI formulas without a race variable [5, 6]. Given that the CKD-EPI formula might not be considered reliable among Europeans [12], we additionally derived eGFRcre [13] and eGFRcys [14] using the EKFC formulas. All GFR estimations were obtained using the R package 'nephro' version 1.3 (https://cran.r-project.org/web/packages/nephro/index.html) [15].

EFA

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' (version 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' version 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 versus non-White), hypertension [defined as systolic blood pressure (BP) $\geq$140 mmHg or diastolic BP $\geq$90 mmHg or taking antihypertensive medication], type 2 diabetes (T2D, defined as blood glucose levels $\geq$7.0 mmol/l or haemoglobin A1c $\geq$6.5% or taking antidiabetic medication) and tobacco smoking (never versus ever). Hazard ratios (HRs) and 95% confidence intervals (CIs) were estimated per 1 standard deviation (SD) change of each kidney biomarker. Sensitivity analyses were performed including adjustment for baseline albumin:creatinine ratio (UACR) and C-reactive protein (CRP) levels.

Model discrimination ability was assessed using Harrell *et al.*'s [16] and Uno *et al.*'s [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 version 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 participants died from any cause, 5376 died from CVD, 2908 died from CHD and 1116 died from cerebrovascular disease (Table 1 and Supplementary Table S2). At baseline, the mean age of the alive group was 61.7 years (SD 6.4) and the mean age of the deceased group was 56.2 years (SD 8.1). 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 percentage of males among CHD-related compared with cerebrovascular-related deaths (79.8% versus 52.3%). At baseline, among individuals who died from CHD, we observed higher percentages of diabetes

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

Sample characteristics	Alive (n = 340 302)	Deceased			
		All-cause (n = 26 327)	CVD (n = 5376)	CHD (n = 2908)	Cerebrovascular (n = 1116)
Age (years), mean (SD)	56.2 (8.1)	61.7 (6.4)	62.0 (6.3)	61.7 (6.3)	63.0 (5.5)
Male, n (%)	153 878 (45.2)	15 742 (59.8)	3791 (70.5)	2321 (79.8)	584 (52.3)
Non-White, n (%)	4911 (1.4)	240 (0.9)	45 (0.8)	17 (0.6)	11 (1.0)
Ever-smokers, n (%)	33 408 (9.8)	4928 (18.7)	1119 (20.8)	658 (22.6)	197 (17.7)
Hypertension, n (%)	176 863 (52.0)	18 394 (69.9)	4272 (79.4)	2333 (80.2)	865 (77.5)
T2D, n (%)	16 456 (4.8)	3144 (11.9)	918 (17.1)	593 (20.4)	144 (12.9)
BMI (kg/m^2), mean (SD)	27.3 (4.7)	28.3 (5.4)	29.2 (5.7)	29.5 (5.4)	28.0 (5.2)
Systolic BP (mmHg), mean (SD)	137.5 (18.5)	142.6 (19.6)	145.1 (20.8)	145.4 (20.9)	145.6 (20.9)
Diastolic BP (mmHg), mean (SD)	82.2 (10.1)	82.5 (10.6)	83.0 (11.5)	82.7 (11.4)	83.2 (11.4)
Serum creatinine (mg/dl), mean (SD)	0.80 (0.2)	0.86 (0.3)	0.91 (0.3)	0.93 (0.3)	0.86 (0.3)
Cystatin C (mg/l), mean (SD)	0.90 (0.1)	1.01 (0.3)	1.06 (0.3)	1.07 (0.3)	1.02 (0.3)
BUN (mg/dl), mean (SD)	15.1 (3.7)	16.0 (5.2)	16.9 (6.3)	17.0 (6.7)	16.6 (5.7)
UA (g/dl), mean (SD)	5.2 (1.3)	5.6 (1.5)	5.9 (1.6)	6.0 (1.5)	5.5 (1.5)
Serum albumin (g/l), mean (SD)	4.5 (0.3)	4.5 (0.3)	4.4 (0.3)	4.5 (0.3)	4.5 (0.3)
eGFRcre (ml/min/1.73 m^2), mean (SD)	95.0 (12.7)	90.3 (15.1)	88.3 (16.5)	88.6 (17.0)	88.6 (14.8)
eGFRcys (ml/min/1.73 m^2), mean (SD)	89.2 (15.6)	78.1 (17.9)	75.0 (18.7)	75.2 (19.2)	76.4 (17.5)
eGFRcrecys (ml/min/1.73 m^2), mean (SD)	95.6 (14.0)	86.9 (16.8)	84.0 (18.0)	84.2 (18.4)	85.2 (16.6)
EFA1, mean (SD)	0.05 (1.0)	−0.62 (1.1)	−0.78 (1.1)	−0.77 (1.1)	−0.71 (1.0)
EFA2, mean (SD)	−0.03 (0.9)	0.43 (1.2)	0.67 (1.4)	0.68 (1.4)	0.58 (1.2)

According to the ICD-10 codes, the mortality definitions of CVD, CHD and cerebrovascular disease were I00–I99, I20–I25 and I60–I69, respectively.

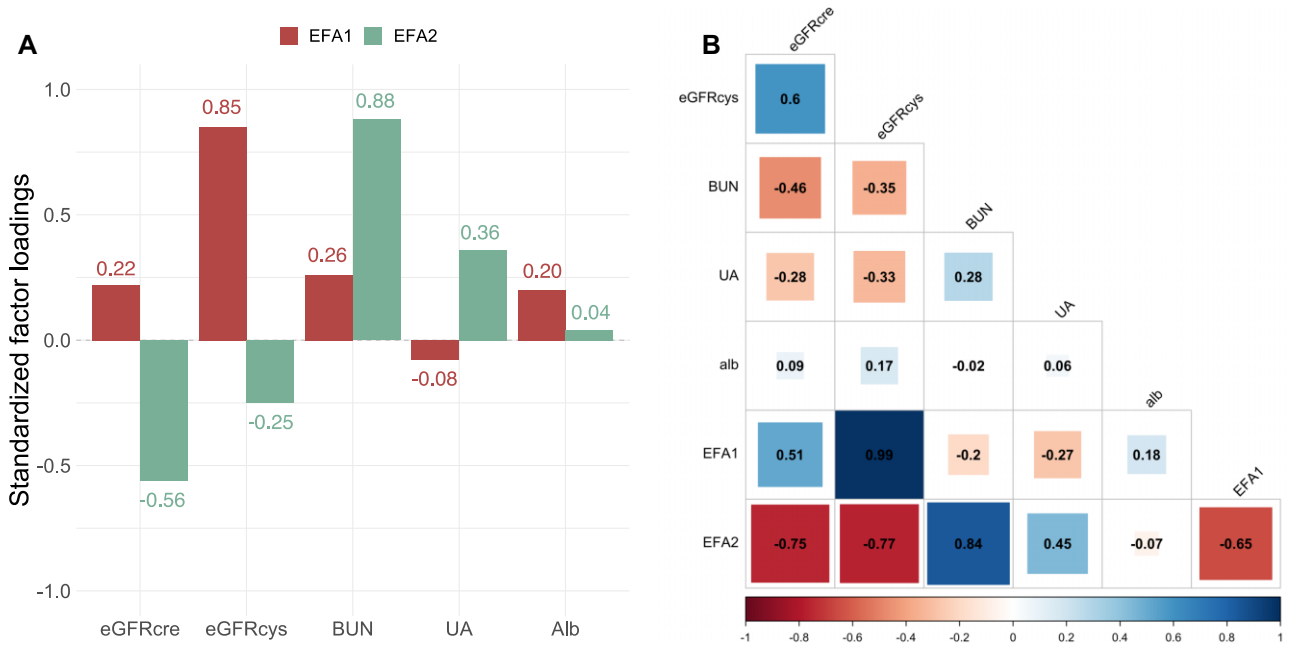


Figure 1: The results from EFA. **(A)** Standardized factor loadings of five kidney indices for EFA1 and EFA2. **(B)** Correlation heatmap of EFA-derived variables (EFA1 and EFA2) with eGFR values (eGFRcre, eGFRcys and eGFRcrecys). We applied maximum likelihood estimation with promax rotation to estimate factor loadings (see Methods for details).

(20.4%) and ever-smokers (22.6%) than in those who died from cerebrovascular disease (12.9% and 17.7%, respectively).

Exploratory factor analysis (EFA)

Scree plot inspection (Supplementary Fig. S2) supported retention of two latent factors that we labelled 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 (Fig. 1A). EFA1 and EFA2 were negatively correlated (Pearson's r = −0.65; Fig. 1B). EFA1

showed nearly perfect correlation with eGFRcys ($r = 0.99$). EFA2 was positively correlated with BUN ($r = 0.84$) and UA ($r = 0.45$) 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 ml/min/1.73 m^2 (95% CI 6.70–6.76) lower eGFRcre and 16.00 ml/min/1.73 m^2 (95% CI 15.99–16.01) lower eGFRcys. A 1-SD larger EFA2 corresponded to 10.98 ml/min/1.73 m^2 (95% CI 10.95–11.01) lower eGFRcre and 13.84 ml/min/1.73 m^2 (95% CI 13.80–13.88) lower eGFRcys.

Survival analysis

All five kidney indices were associated with all-cause and each cause-specific mortality without adjustment for potential confounders (Fig. 2A). EFA1 and eGFRcys showed the largest effects. We observed HRs of 1.85 (95% CI 1.83–1.87) and 1.87 (95% CI 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. 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 (Fig. 2B). This reflects the identification of relevant causal pathways that did not 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 with other types of CVD mortality. However, despite the demographic and clinical differences at baseline between cerebrovascular and CHD mortality cases, observed differences were minor.

Model performance

The same pattern observed for the HR was observed for the model's discrimination performance (Fig. 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, with C-statistics of 0.71 (95% CI 0.71–0.72) and 0.71 (95% CI 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 (Fig. 3B). Similar results were obtained with Uno *et al.*'s C-statistic (Supplementary Fig. S3). Prediction accuracy was also assessed with calibration for the observed and 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 EKFC equations

Instead of the CKD-EPI equations, we reassessed our analyses by estimating eGFRcre and eGFRcys with the EKFC equations. EKFC-based eGFRcys (Fig. 2D) showed larger HRs for all-cause and CVD mortality than CKD-EPI-based eGFRcys (Fig. 2B). The effect sizes of eGFRcre were substantially equivalent. While results were essentially consistent with those observed using the CKD-EPI equations, in the EKFC analysis, EFA1 was no longer equivalent to eGFRcys, leaving eGFRcys as the best mortality predictor. For discriminatory ability, there was no substantial difference in model performance between EKFC-based eGFRcys (Fig. 3C) and CKD-EPI eGFRcys (Fig. 3A). Similar to the CKD-EPI equation (Fig. 3B), including clinical risk factors into the model universally improved discriminatory performance across kidney indices (Fig. 3D).

Sensitivity analyses

We incorporated UACR into the EFA based on the CKD-EPI equations (Supplementary Fig. 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 the analysis was repeated including CRP (a marker of inflammation) as a covariate in all Cox proportional hazards models, the 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 ≈360 000 adults from the general population, with more than 10 years of 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 with the kidney functionality captured by the other markers net of cystatin, regardless of the type of equation used [18].

Comparison of unadjusted versus adjusted models highlights 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 behavioural determinants of mortality, adjusted analysis results enable appreciation of 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

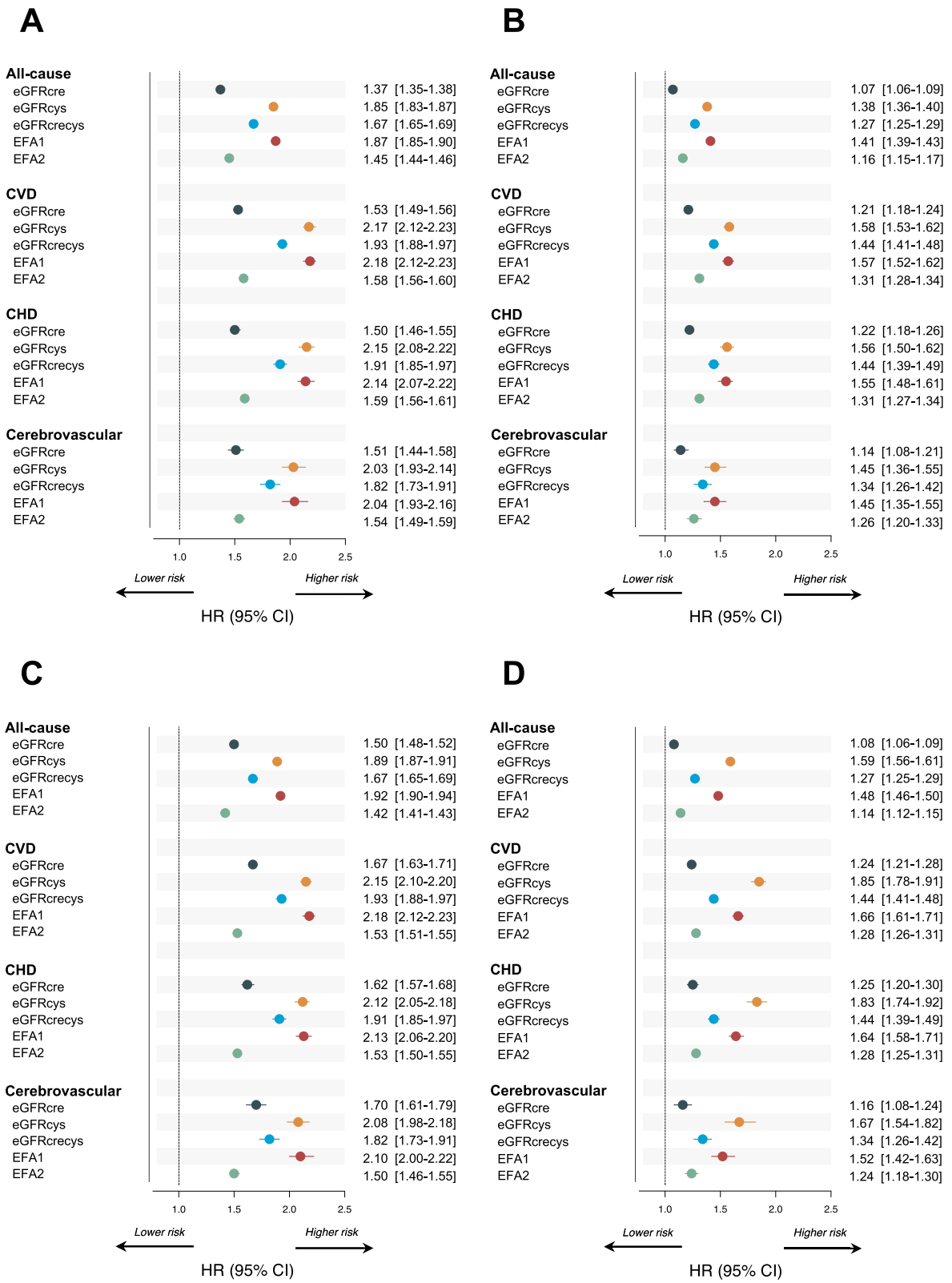


Figure 2: HRs and 95% CIs for all-cause and cause-specific mortality from **(A)** unadjusted and **(B)** fully adjusted models using CKD-EPI eGFRcre and eGFRcys equations and **(C)** unadjusted and **(D)** adjusted models using EKFC eGFRcre and eGFRcys equations. The HRs and 95% CIs are expressed with a 1-SD change in each kidney index. A 1-SD change in eGFRcre, eGFRcys and eGFRcrecys corresponded to 13.0, 16.0 and 14.4 ml/min/1.73 m^2, respectively. Estimated by the linear regression model, a 1-SD lower EFA1 corresponded to 6.73 ml/min/1.73 m^2 (95% CI 6.70–6.76) lower eGFRcre and 16.00 ml/min/1.73 m^2 (95% CI 15.99–16.01) lower eGFRcys. A 1-SD larger EFA2 corresponded to 10.98 ml/min/1.73 m^2 (95% CI 10.95–11.01) lower eGFRcre and 13.84 ml/min/1.73 m^2 (95% CI 13.80–13.88) lower eGFRcys. Fully adjusted models included sex, age, self-reported ancestry, BMI, hypertension, T2D and tobacco smoking as potential confounders.

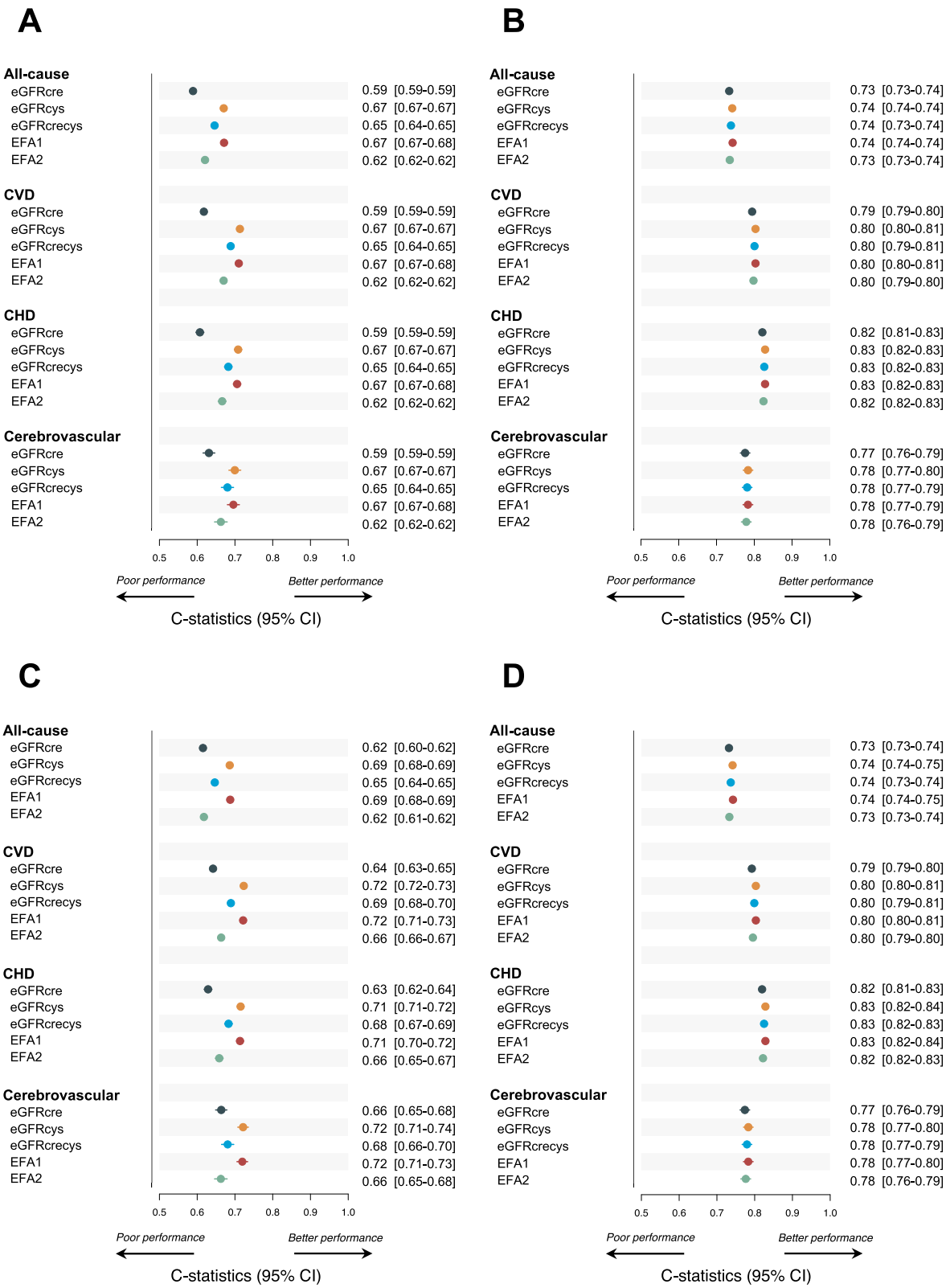


Figure 3: C-statistics and 95% CIs for all-cause and cause-specific mortality from **(A)** unadjusted and **(B)** fully adjusted models using CKD-EPI eGFRcre and eGFRcys equations and **(C)** unadjusted and **(D)** adjusted models using EKFC eGFRcre and eGFRcys equations. Fully adjusted models included sex, age, self-reported ancestry, BMI, hypertension, T2D and tobacco smoking as potential confounders.

of the adjusted models were nearly identical across kidney function markers for all outcomes.

Our complementary analysis with EKFC instead of CKD-EPI equations showed that eGFRcys was the best mortality predictor overall, with larger effects than EFA1 across all outcomes. The HR for eGFRcys estimated by the EKFC equation was larger than that of eGFRcys using the CKD-EPI equation, in line with recent evidence from non-black population samples reporting superior discriminatory ability of the EKFC equation for all-cause and CVD mortality compared with the CKD-EPI equation [19].

Our study has several limitations. The young age and selection towards a healthy status of UKBB participants [20] recommends that the transportability of findings is tested in alternative population and patient samples. However, our findings are in line with previous results obtained from both different clinical status [21, 22] and in different age groups [9, 23]. Additionally, the results are in line with our recent investigation of a structural equation model for eGFRcre, eGFRcys, UA and BUN in the longitudinal evaluation of incident CVD risk in an Alpine community [24], where the identified latent factor showed comparable discrimination performance for CVD risk as eGFRcys alone. Contrary to naïve principal component analysis, which is a descriptive dimensionality reduction technique centred on variables, factor analysis is a model-based technique prone to measurement error components centred on observations, which exploits the correlation structure among them to identify latent variables that may help explain 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 multifaceted 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 [25], 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 with alternative approaches that used multiple kidney biomarkers, including e.g. fibroblast growth factor 23 and kidney injury molecule 1 [26]. 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 in 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 [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 [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 (n = 135 781) from the analysis, we compared the proportion of all-cause mortality cases (7.2% versus 8.2%) and the mean levels of eGFRcre (94.7 versus 94.7) and eGFRcys (88.1 versus 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 are 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.

SUPPLEMENTARY DATA

Supplementary data are available at *Clinical Kidney Journal* online.

FUNDING

The 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. We appreeciated that the Department of Innovation, Research University and Museums of the Autonomous Province of Bozen/Bolzano covered the open access publication costs.

AUTHORS' CONTRIBUTIONS

R.F. and C.P. conceived the study, drafted the manuscript and contributed to revision of 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. All the authors reviewed and approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

CONFLICT OF INTEREST STATEMENT

C.P. is a consultant for Quotient Therapeutics (UK). The remaining authors declare no conflicts of interest.

REFERENCES

1. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease

Study 2017. *Lancet* 2020;**395**:709–33. https://doi.org/10.1016/S0140-6736(20)30045-3
2. Foreman KJ, Marquez N, Dolgert A *et al.* Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. *Lancet* 2018;**392**:2052–90. https://doi.org/10.1016/S0140-6736(18)31694-5
3. Fox CS, Matsushita K, Woodward M *et al.* Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without diabetes: a meta-analysis. *Lancet* 2012;**380**:1662–73. https://doi.org/10.1016/S0140-6736(12)61350-6
4. Coresh J, Turin TC, Matsushita K *et al.* Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *JAMA* 2014;**311**:2518–31. https://doi.org/10.1001/jama.2014.6634
5. Inker LA, Schmid CH, Tighiouart H *et al.* Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med* 2012;**367**:20–9. https://doi.org/10.1056/NEJMoa1114248
6. 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**:1737–49. https://doi.org/10.1056/NEJMoa2102953
7. Shlipak MG, Sarnak MJ, Katz R *et al.* Cystatin C and the risk of death and cardiovascular events among elderly persons. *N Engl J Med* 2005;**352**:2049–60. https://doi.org/10.1056/NEJMoa043161
8. Astor BC, Levey AS, Stevens LA *et al.* Method of glomerular filtration rate estimation affects prediction of mortality risk. *J Am Soc Nephrol* 2009;**20**:2214–22. https://doi.org/10.1681/ASN.2008090980
9. 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**:506–21. https://doi.org/10.1111/joim.12169
10. Lees JS, Welsh CE, Celis-Morales CA *et al.* Glomerular filtration rate by differing measures, albuminuria and prediction of cardiovascular disease, mortality and end-stage kidney disease. *Nat Med* 2019;**25**:1753–60. https://doi.org/10.1038/s41591-019-0627-8
11. 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**:e1001779. https://doi.org/10.1371/journal.pmed.1001779CrossRef]
12. Delanaye P, Schaeffner E, Cozzolino M *et al.* The new, race-free, Chronic Kidney Disease Epidemiology Consortium (CKD-EPI) equation to estimate glomerular filtration rate: is it applicable in Europe? A position statement by the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM). *Clin Chem Lab Med* 2023;**61**:44–7. https://doi.org/10.1515/cclm-2022-0928
13. Pottel H, Björk J, Courbebaisse M *et al.* Development and validation of a modified full age spectrum creatinine-based equation to estimate glomerular filtration rate: a cross-sectional analysis of pooled data. *Ann Intern Med* 2021;**174**:183–91. https://doi.org/10.7326/M20-4366
14. Pottel H, Björk J, Rule AD *et al.* Cystatin C-based equation to estimate GFR without the inclusion of race and sex. *N Engl J Med* 2023;**388**:333–43. https://doi.org/10.1056/NEJMoa2203769
15. Pattaro C, Riegler P, Stifter G *et al.* Estimating the glomerular filtration rate in the general population using different equations: effects on classification and association. *Nephron Clin Pract* 2013;**123**:102–11. https://doi.org/10.1159/000351043
16. Harrell FE, Jr, Lee KL, Califf RM *et al.* Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984;**3**:143–52. https://doi.org/10.1002/sim.4780030207
17. Uno H, Cai T, Pencina MJ *et al.* On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med* 2011;**30**:1105–17. https://doi.org/10.1002/sim.4154
18. Fabian J, Kalyesubula R, Mkandawire J *et al.* Measurement of kidney function in Malawi, South Africa, and Uganda: a multicentre cohort study. *Lancet Glob Health* 2022;**10**:e1159–69. https://doi.org/10.1016/S2214-109X(22)00239-X
19. Lees JS, Rutherford E, Stevens KI *et al.* Assessment of cystatin C level for risk stratification in adults with chronic kidney disease. *JAMA Netw Open* 2022;**5**:e2238300. https://doi.org/10.1001/jamanetworkopen.2022.38300
20. Fry A, Littlejohns TJ, Sudlow C *et al.* Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol* 2017;**186**:1026–34. https://doi.org/10.1093/aje/kwx246
21. Muntner P, Mann D, Winston J *et al.* Serum cystatin C and increased coronary heart disease prevalence in US adults without chronic kidney disease. *Am J Cardiol* 2008;**102**:54–7. https://doi.org/10.1016/j.amjcard.2008.02.098
22. Helmersson-Karlqvist J, Lipcsey M, Ärnlöv J *et al.* Addition of cystatin C predicts cardiovascular death better than creatinine in intensive care. *Heart* 2022;**108**:279–84. https://doi.org/10.1136/heartjnl-2020-318860
23. Zappitelli M, Greenberg JH, Coca SG *et al.* Association of definition of acute kidney injury by cystatin C rise with biomarkers and clinical outcomes in children undergoing cardiac surgery. *JAMA Pediatr* 2015;**169**:583–91. https://doi.org/10.1001/jamapediatrics.2015.54
24. Fujii R, Melotti R, Gögele M *et al.* Structural equation modeling (SEM) of kidney function markers and longitudinal CVD risk assessment. *PLoS One* 2023;**18**:e0280600. https://doi.org/10.1371/journal.pone.0280600
25. Stevens LA, Schmid CH, Greene T *et al.* Factors other than glomerular filtration rate affect serum cystatin C levels. *Kidney Int* 2009;**75**:652–60. https://doi.org/10.1038/ki.2008.638
26. Lee AK, Katz R, Jotwani V *et al.* Distinct dimensions of kidney health and risk of cardiovascular disease, heart failure, and mortality. *Hypertension* 2019;**74**:872–9. https://doi.org/10.1161/HYPERTENSIONAHA.119.13339
27. Levey AS, Stevens LA, Schmid CH *et al.* A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;**150**:604–12. https://doi.org/10.7326/0003-4819-150-9-200905050-00006

Received: 24.11.2023; Editorial decision: 5.6.2024