scientific reports

OPEN

Comparison between the EKFC-equation and machine learning models to predict Glomerular Filtration Rate

Felipe Kenji Nakano[1,2✉], Anna Åkesson[3,4], Jasper de Boer[1,2], Klest Dedja[1,2], Robbe D'hondt[1,2], Fateme Nateghi Haredasht[1,2], Jonas Björk[3,4], Marie Courbebaisse[5], Lionel Couzi[6], Natalie Ebert[7], Björn O. Eriksen[8], R. Neil Dalton[9], Laurence Derain-Dubourg[10], Francois Gaillard[11], Cyril Garrouste[12], Anders Grubb[13], Lola Jacquemont[14], Magnus Hansson[15], Nassim Kamar[16], Christophe Legendre[17], Karin Littmann[18], Christophe Mariat[19], Toralf Melsom[8], Lionel Rostaing[20], Andrew D. Rule[21], Elke Schaeffner[7], Per-Ola Sundin[22], Arend Bökenkamp[23], Ulla Berg[24], Kajsa Åsling-Monemi[24], Luciano Selistre[25], Anders Larsson[26], Ulf Nyman[27], Antoine Lanot[28,29,30], Hans Pottel[1], Pierre Delanaye[31,32] & Celine Vens[1,2]

In clinical practice, the glomerular filtration rate (GFR), a measurement of kidney functioning, is normally calculated using equations, such as the European Kidney Function Consortium (EKFC) equation. Despite being the most general equation, EKFC, just like previously proposed approaches, can still struggle to achieve satisfactory performance, limiting its clinical applicability. As a possible solution, recently machine learning (ML) has been investigated to improve GFR prediction, nonetheless the literature still lacks a general and multi-center study. Using a dataset with 19,629 patients from 13 cohorts, we investigate if ML can improve GFR prediction in comparison to EKFC. More specifically, we compare diverse ML methods, which were allowed to use age, sex, serum creatinine, cystatin C, height, weight and BMI as features, in internal and external cohorts against EKFC. The results show that the most performing ML method, random forest (RF), and EKFC are very competitive where RF and EKFC achieved respectively P10 and P30 values of 0.45 (95% CI 0.44;0.46) and 0.89 (95% CI 0.88;0.90), whereas EKFC yielded 0.44 (95% CI 0.43; 0.44) and 0.89 (95% CI 0.88; 0.90), considering the entire cohort. Small differences were, however, observed in patients younger than 12 years where RF slightly outperformed EKFC.

[1]Department of Public Health and Primary Care, KU Leuven Campus Kulak Kortrijk, Kortrijk, Belgium. [2]Itec, Imec Research Group at KU Leuven, Kortrijk, Belgium. [3]Division of Occupational and Environmental Medicine, Lund University, Lund, Sweden. [4]Clinical Studies Sweden, Forum South, Skåne University Hospital, Lund, Sweden. [5]Physiology Department, Georges Pompidou European Hospital, Assistance Publique Hôpitaux de Paris, INSERM U1151-CNRS UMR8253, Paris Descartes University, Paris, France. [6]CNRS-UMR 5164 Immuno ConcEpT, CHU de Bordeaux, Nephrologie-Transplantation-Dialyse, Université de Bordeaux, Bordeaux, France. [7]Institute of Public Health, Charité Universitätsmedizin Berlin, Berlin, Germany. [8]Metabolic and Renal Research Group, UiT the Arctic University of Norway, Tromsö, Norway. [9]The Wellchild Laboratory, Evelina London Children's Hospital, London, UK. [10]Néphrologie, Dialyse, Hypertension et Exploration Fonctionnelle Rénale, Hôpital Edouard Herriot, Hospices Civils de Lyon, France. [11]Renal Transplantation Department, Assistance Publique–Hôpitaux de Paris (AP-HP), Paris, France. [12]Department of Nephrology, Clermont-Ferrand University Hospital, Clermont-Ferrand, France. [13]Department of Clinical Chemistry, Skåne University Hospital, Lund University, Lund, Sweden. [14]Renal Transplantation Department, CHU Nantes, Nantes University, Nantes, France. [15]Function Area Clinical Chemistry, Karolinska University Laboratory, Karolinska Institute, Karolinska University Hospital Huddinge and Department of Laboratory Medicine, Stockholm, Sweden. [16]Department of Nephrology, Dialysis and Organ Transplantation, CHU Rangueil, INSERM U1043, IFR–BMT, University Paul Sabatier, Toulouse, France. [17]Hôpital Necker, AP-HP and Université Paris Descartes, Paris, France. [18]Institute om Medicine Huddinge (Med H), Karolinska Institute, Solna, Sweden. [19]Service de Néphrologie, Dialyse et Transplantation Rénale, Hôpital Nord, CHU de Saint-Etienne, Saint-Priest-en-Jarez, France. [20]Service de Néphrologie, Hémodialyse, Aphérèses et Transplantation Rénale, Hôpital Michallon, CHU Grenoble-Alpes, Tronche, France. [21]Division of Nephrology and Hypertension, Mayo

Clinic, Rochester, MN, USA. [22]Karla Healthcare Centre, Faculty of Medicine and Health, Örebro University, 70182 Örebro, SE, Sweden. [23]Department of Paediatric Nephrology, Emma Children's Hospital, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [24]Department of Clinical Science, Intervention and Technology, Division of Pediatrics, Karolinska Institutet, Karolinska University Hospital Huddinge, Stockholm, Sweden. [25]Mestrado Em Ciências da Saúde-Universidade Caxias do Sul Foundation CAPES, Caxias Do Sul, Brazil. [26]Department of Medical Sciences, Clinical Chemistry, Uppsala University, Uppsala, Sweden. [27]Department of Translational Medicine, Division of Medical Radiology, Lund University, Malmö, Sweden. [28]Normandie Université, Unicaen, CHU de Caen Normandie, Néphrologie, Caen, France. [29]Normandie Université, Unicaen, UFR de Médecine, 2 Rue Des Rochambelles, Caen, France. [30]ANTICIPE U1086 INSERM-UCN, Centre François Baclesse, Caen, France. [31]Department of Nephrology-Dialysis-Transplantation, University of Liège (ULg CHU), CHU Sart Tilman, Liège, Belgium. [32]Department of Nephrology-Dialysis-Apheresis, Hopital Universitaire Caremeau, Nimes, France. ✉email: felipekenji.nakano@kuleuven.be

The glomerular filtration rate (GFR) quantifies the functioning of kidneys and is used to diagnose chronic kidney disease (CKD). GFR is also employed to indicate the severity of kidney damage, for example patients whose GFR values are lower than 15 mL/min/1.73m^2 may require dialysis or even transplantation. Thus, accurately measuring and/or estimating GFR is essential for managing kidney health.

In clinical practice, GFR is commonly estimated using equations, such as the Chronic Kidney Disease Epidemiology (CKD-EPI)[1] and European Kidney Function Consortium (EKFC)[2] equations, which are based on age, sex, serum creatinine (SCr), and/or cystatin C (CysC)[3]. Both equations are recognized as valid by the recent Kidney Disease Improving Global Outcomes (KDIGO) guidelines[4]. EKFC, the most recently proposed equation, relies on rescaling SCr or CysC with the so-called Q-value, the median SCr or CysC of healthy people and is applicable across the full age spectrum (2–100 years), and adaptable to any population, making it the most general equation. Using two separate equations CKiD (Chronic Kidney Disease in children)[5] and CKD-EPI (for adults)[1] presents shortcomings, such as an implausible rise in estimated GFR (eGFR) in patients who are transitioning from adolescence to adulthood, despite no change in SCr[6]. Finally, EKFC does not show overestimation in young adults (18–25 years), as is the case for CKD-EPI[7,8].

Despite many advantages, the EKFC-equation, like all other eGFR-equations, still presents deficiencies[7,9]. For instance, its imprecision remains rather large, which limits its clinical applicability, namely most eGFR equations are only able to predict the GFR in about 80–85% of patients within 30% of measured GFR (P30 = 80–85%). The performance increases up to P30 = 90–95% when both SCr and cystatin C are included. However, when P10, the performance of patients with an estimated GFR within 10% of measured GFR, is evaluated, the performance of equations does not surpass 50% in most of the cases[1–3]. Improving precision of GFR estimations is an important clinical need in the diagnosis of individual patients.

In this context, ML methods have been investigated to improve GFR estimations. Nonetheless, recent studies using this methodology often present limitations. For instance, several studies are specifically designed for a population, namely, older people (above 65 years), adults only, ICU patients or they just present a limited sample size[10–14].

The aim of the current study is to explore to what extent established ML methods can improve GFR estimation compared with EKFC (both for children and adults), representing the state-of-the-art among currently available equations.

Methods
Design overview
The ML models have been obtained using the datasets used for developing and validating the SCr-based EKFC equation[2] and in case cystatin C was available, for the combined SCr/CysC-based EKFC-equation. We have limited the analysis to white patients since we are using the same cohorts as in[2]. More details about participants centers, measurement methods and patient characteristics are available in the Supplementary Material Tables S1–S4. Briefly, we have data from 19,629 patients from 13 cohorts for development, internal and external validation. These cohorts were divided into development-internal validation or external validation datasets according to their age, exogenous marker used to measure GFR (mGFR) and mGFR levels, as described in Tables S1–S3 (Supplementary material). For the models based on the single biomarker SCr, we used 13 cohorts, 7 for the development and internal validation, which were further randomly split into development (n = 8473; 25%) and internal validation (n = 2778; 25%) dataset, and the remaining 6 cohorts (n = 8378) were used for external validation, which are described in Table 1 and S4. For the models based on both SCr and cystatin C, we selected the patients from the same cohorts where both biomarkers were available (Table 2 and S4), leading to the following subsets: development (n = 4849; 41%), internal validation (n = 1603; 13.5%) and external validation (n = 5389; 45.5%).

All data were anonymized and the original study was approved by the Ethical Board at Lund University (Sweden) with amendment approved by the Swedish Ethical Review Agency.

We used the single biomarker SCr-based EKFC equation and the mean of the SCr-based EKFC and cystatin C based EKFC as benchmark for the current comparison. As cystatin C is not always available in clinical practice, we focused on the single biomarker SCr-based EKFC-equation in a first part of the analysis, but, as the combined equation has the highest accuracy and precision, in a second analysis, we also evaluated the difference between the combined SCr/CysC-based EKFC and the ML models.

Covariates
ML models were allowed to use age, sex, SCr (and CysC), height, weight and BMI, as these data were also available for most of the participants. SCr was measured using assays traceable to the gold standard isotope

Characteristic	Development (7 cohorts, $n = 8473$)	Internal validation (7 cohorts, $n = 2778$)	External validation (6 cohorts, $n = 8378$)
Age (years) (mean $\pm$ SD) [range]	42.4 $\pm$ 25.2 [2.0–96.7]	42.4 $\pm$ 25.1 [2.0–94.6]	50.9 $\pm$ 22.3 [2.0–97.0]
mGFR (mL/min/1.73m^2) (mean $\pm$ SD) [range]	76.9 $\pm$ 33.1 [4–244]	77.8 $\pm$ 32.9 [5–254]	78.9 $\pm$ 32.2 [3–354]
SCr/Q (mean $\pm$ SD) Median [P25–P75] [range]	1.43 $\pm$ 0.91 1.15 [0.94–1.60] [0.11–19.8]	1.43 $\pm$ 0.93 1.13 [0.94–1.55] [0.27–10.6]	1.37 $\pm$ 0.97 1.09 [0.93–1.37] [0.08–12.8]
% Males	55.8	55.9	47.2
BMI	24.35 $\pm$ 6.16 23.94 [19.78—28.03] [9.47—71.43]	24.64 $\pm$ 6.02 23.92 [19.92—27.96] [9.67—57.73]	25.44 $\pm$ 5.76 25.09 [21.64—28.78] [8.56—63.28]

Table 1. Basic participant characteristics with only serum creatinine available in the development, internal and external validation datasets.

Characteristic	Development (7 cohorts, $n = 4849$)	Internal validation (7 cohorts, $n = 1603$)	External validation (6 cohorts, $n = 5389$)
Age (years) (mean $\pm$ SD) [range]	50.0 $\pm$ 24.0 [2.0–96.7]	48.8 $\pm$ 24.1 [2.0–94.6]	53.5 $\pm$ 24.3 [2.0–97.0]
mGFR (mL/min/1.73m^2) (mean $\pm$ SD) [range]	76.6 $\pm$ 30.9 [5.7 -221.0]	78.0 $\pm$ 31.7 [4.7–254.0]	70.0 $\pm$ 34.5 [3—354]
SCr/Q (mean $\pm$ SD) Median [P25-P75] [range]	1.36 $\pm$ 0.82 1.11 [0.92–1.53] [0.43–11.8]	1.37 $\pm$ 0.90 1.09 [0.92–1.50] [0.27–10.59]	1.54 $\pm$ 1.15 1.15 [0.95–1.67] [0.08–12.75]
CysC/Q' (mean $\pm$ SD) Median [P25-P75] [range]	1.31 $\pm$ 0.70 1.07 [0.87–1.47] [0.49–7.84]	1.30 $\pm$ 0.72 1.06 [0.87–1.47] [0.53–7.12]	1.62 $\pm$ 0.96 1.25 [0.99–1.92] [0.53–8.09]
% Males	56.2	55.8	51.0
BMI	24.35 $\pm$ 6.16 23.94 [19.78—28.03] [9.47—71.43]	24.33 $\pm$ 6.02 23.92 [19.92—27.96] [9.67—57.73]	25.44 $\pm$ 5.76 25.09 [21.64—28.78] [8.56—63.28]

Table 2. Basic participant characteristics with both serum creatinine and cystatin C available in the development, internal and external validation datasets.

dilution mass spectrometry method (results from the CRIC study were recalibrated) as described in[2]. Cystatin C assays were standardized to the international reference material (ERM-DA471/IFCC)[3].

Outcomes

Measured GFR was obtained using two methods: plasma clearance and urinary clearance. As previously described[2], GFR was measured with different markers, but they all are recognized as reference methods[15–17]. For more details see Supplementary Material Tables S1–S3.

Machine learning models

Several different ML models were evaluated: multi-layer perceptron (neural network), support vector machines, k nearest neighbors, linear regression, Random Forest regression, and XGBoost regression. We selected the best performing ML models regarding their performance on the internal validation set: linear regression, XGBoost and random forest.

- Linear regression: A linear regression model with L1 (lasso regression) and L2 (ridge regression) regularization. Lasso is an acronym for least absolute shrinkage and selection operator. Lasso regression adds the 'absolute value of magnitude' of the coefficient as a penalty term to the loss function, so the cost of outliers increases linearly. Ridge regression adds the 'squared magnitude' of the coefficient as the penalty term to the loss function, so the cost of outliers increases exponentially. This method has been added as a baseline comparison.
- Random forest regression: Random forest is an ensemble supervised ML algorithm made up of decision trees, which is used for both classification and regression problems. A random forest model with 100 trees was used. The model is constructed using bootstrapping, i.e., by constructing multiple datasets of the same size as the original dataset, created using resampling with replacement. Furthermore, candidate variables at each split are randomly selected to maximize diversity among the decision trees[18,19].
- XGBoost for regression: XGBoost is another powerful approach for building supervised regression models. An XGBoost model with 100 trees which employs mean squared error as its loss function was used. Boosting is a popular ensemble method, that sequentially builds models, in this case decision trees, where each model learns from the errors of the previous one[20].

Random Forest creates decision trees independently and combines their outputs, whereas XGBoost builds trees sequentially to correct errors. As their base models, the random forests used predictive clustering trees[19], a variant of decision trees which employ variance reduction as their heuristic. Furthermore, these trees can naturally handle missing covariate values. The library scikit-learn (version 1.4.2) was used for the linear regression and the random forest, whereas XGBoost was implemented using the homonymous library XGBoost (version 2.0).

External validation cohort (n = 8378)					
Methods	Median bias (95%CI)	IQR (P25-P75)	P10 (95%CI)	P30 (95%CI)	MSE
EKFC	−0.90 [−1.30; −0.60]	17.15 (−9.20, 7.50)	0.43 [0.42; 0.44]	0.87 [0.86; 0.87]	296.49
RF	−0.39 [−0.73; −0.01]	17.13 (−9.51, 8.20)	0.41 [0.40; 0.42]	0.85 [0.85; 0.86]	294.24
LR	−3.06 [−3.51; −2.50]	22.21 (−15.49, 9.77)	0.31 [0.30; 0.32]	0.76 [0.75; 0.77]	509.88
XGBoost	−0.79 [−1.19; −0.41]	17.86 (−10.71, 8.43)	0.38 [0.37; 0.39]	0.84 [0.83; 0.84]	320.56

Table 3. Performance statistics of the different machine learning models, and EKFC$_{crea}$-equation, in the external validation dataset using only serum creatinine as biomarker. EKFC = European Kidney Function Consortium equation; RF = Random Forest model; LR = Linear Regression model; XGBoost = eXtreme Gradient Boosting model; IQR = Interquartile range; P10/P30 = fraction of patients with eGFR within 10%/30% of mGFR; MSE = Mean Square Error.

External validation cohort (n = 5389)					
Methods	Median bias (95%CI)	IQR (P25–P75)	P10 (95%CI)	P30 (95%CI)	MSE
EKFC	−0.60 [−0.90; −0.30]	15.60 (−7.9,5.4)	0.45 [0.44; 0.46]	0.89 [0.88; 0.90]	247.98
RF	−0.71 [−1.09; −0.27]	15.01 (−8.03,5.65)	0.44 [0.43; 0.45]	0.88 [0.87; 0.89]	227.3
LR	2.73 [2.06; 3.31]	23.49 (−10.4,14.47)	0.30 [0.28; 0.31]	0.68 [0.67; 0.69]	551.78
XGBoost	−1.38 [−1.83; −1.08]	15.58 (−9.02,5.9)	0.4 [0.39; 0.41]	0.86 [0.85; 0.87]	245.83

Table 4. Performance statistics of the different machine learning models, and EKFC$_{crea+cys}$-equation, in the external validaton dataset. EKFC = European Kidney Function Consortium equation; RF = Random Forest model; LR = Linear Regression model; XGBoost = eXtreme Gradient Boosting model; IQR = Interquartile range; P10/P30 = fraction of patients with eGFR within 10%/30% of mGFR; MSE = Mean Square Error.

Statistical analysis

The following usual metrics were used to compare the performance of ML methods with the EKFC equations[2,3].

The median bias is the difference between the estimated GFR and the measured GFR. Values close to 0 are desired for this measure, but an absolute bias less than 5 mL/min/1.73m^2 may be considered clinically acceptable.

The Interquartile Range (IQR) is the range of values between the 25th percentile and the 75th percentile of the difference between the estimated GFR and the measured GFR and represents the precision expressed in mL/min/1.73m^2. Smaller values are associated with better precision.

The percentage of patients whose GFR was estimated within 10 or 30% of the measured GFR is the accuracy within 10 or 30% (P10 and P30). The goal for P30 is 100%, but P30 > 75% has been considered as "sufficient for good clinical decision making" by the Kidney Disease Outcomes Quality Initiative (K/DOQI), although their goal was to reach a P30 > 90%[21]. The Mean-square error (MSE) is the average of the squared differences between estimated and measured GFR. Smaller values are associated with better precision.

Median quantiles for bias across the age spectrum were graphically presented using fractional polynomials (linear, square and logarithmic). Likewise, accuracy P30 (%) was graphically presented across the age spectrum using cubic splines with two free knots and using 3rd degree polynomials. Bland–Altman plots (difference versus average) were also used to comprehend the differences in performance between ML models and EKFC.

Median bias, P10 and P30 are reported with 95% CIs. To test if an equation is different from another equation in the same population, we did not use statistical tests to avoid numerous p-value calculations, but the reader may consider an equation as different when the 95% CI between equations was not overlapping (which is a more conservative criterion). We made sub-analyses according to age (younger than 6, 6 to 12, 12 to 18, 18 to 40, 40 to 65 and > 65 years), body mass index (BMI) (< 20, 20 to 25, 25 to 30, 30 to 35 and > 35 kg/m^2), measured GFR (mGFR) (< 30, 30 to 60, 60 to 90, 90 to 120, > 120 mL/min/1.73m^2), and sex. SHAP (Shapley additive explanations) values were used to better understand the variable importance in the ML models[22]. SHAP values were generated using the homonymous library SHAP, version (0.43.0).

Results

In the external validation dataset, participants had a mean ± SD age of 50.9 ± 22.3 years (range [2–97]), mGFR was 78.9 ± 32.2 (range [3–354]) mL/min/1.73m^2, SCr/Q was 1.37 ± 0.97 (range [0.08–12.8] and 47.2% were male. During the learning phase, we had 2056 children aged 2–18, with 1031 aged 2–12 years (see Table S4 (development cohort)).

Among all ML models, the best model was systematically the random forest regression model (RF). In the main manuscript, we only report the overall results for the SCr-based (Table 3) and SCr/CysC based (Table 4) eGFR models on the external validation set. More detailed results (according to subgroups) on the internal and external datasets obtained using EKFC, the random forest model, linear regression and XGBoost models are available in the Supplementary Material Tables S5–S24. First, we present results considering the patients with only SCr available (Table 3 and S5–S14), followed by the results for the patients with both SCr and cystatin C (Table 4 and S15–S24).
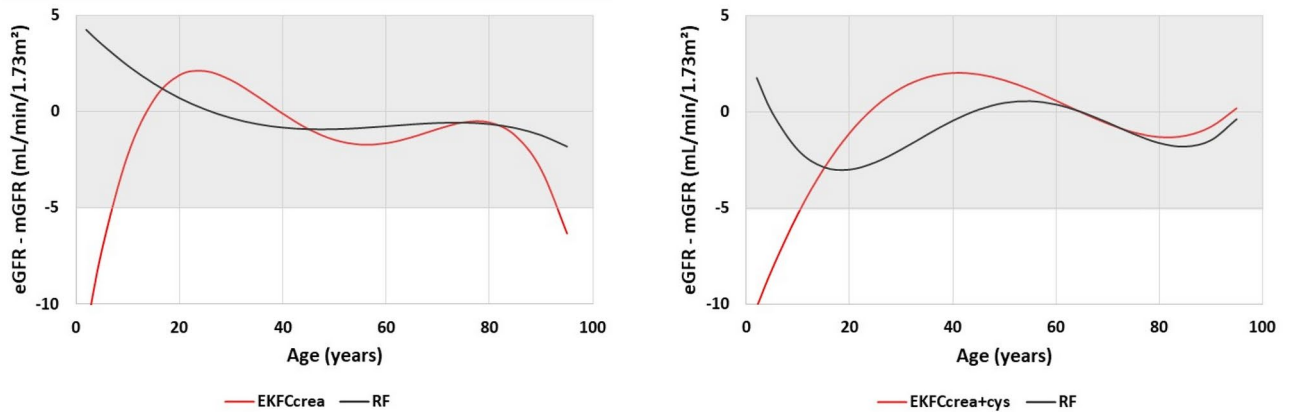
Fig. 1. Bias versus age for the SCr-based (left panel) and SCr/CysC-based (right panel) EKFC and RF models.
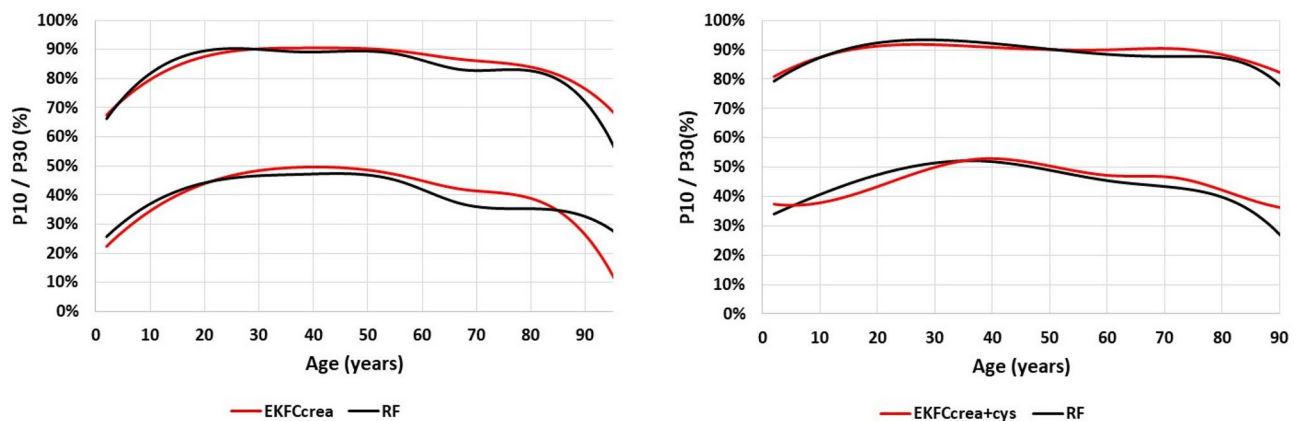

Fig. 2. P10/P30 accuracy versus age for the SCr-based (left panel) and SCr/CysC-based (right panel) EKFC and RF models.

Serum creatinine-based results

Overall, regarding cohorts with SCr available, random forest (RF) and EKFC are competitive with each other. Table 3 shows the overall performance results of the EKFC-equation and the three best performing ML models, in the external validation dataset. More specifically, RF reached a P10 and P30 of 0.41 (95% CI 0.40;0.42) and 0.85 (95% CI 0.85;0.86), whereas EKFC yielded 0.43 (95% CI 0.42; 0.44) and 0.87 (95% CI 0.86; 0.87), respectively.

In Figs. 1 and 2, we plotted bias and P10/P30 against age for EKFC and the best ML model (Random Forest). Figure 3 shows the Bland–Altman plots to evaluate the bias according to GFR-level.

The subgroup analyses according to age, mGFR, BMI and sex are shown in Supplement Tables S11–S14. The results based on age-subgrouping (Table S11, Figs. 1 and 2) show very comparable results between the RF model and EKFC. For subjects younger than 6–12 years, the two models are both less performant. According to Fig. 1 (and Table S11), there is a trend for a poorer bias for the EKFC model in young children, while RF stays within the clinically acceptable range. In all other subgroups, the results were very similar between the EKFC and the RF models. The Bland–Altman (Fig. 3) plots show that both methods (EKFC and RF) are overall unbiased and have similar imprecision (the 95% limits of agreement are nearly equal).

Serum creatinine and cystatin C based results

When considering both SCr and cystatin C as biomarkers, presented in Table 4, Supplement Tables S15–S24 and Figs. 1, 2 and 4, a very similar pattern is observed. That is, RF provided respectively P10 and P30 values of 0.45 (95% CI 0.44;0.46) and 0.89 (95% CI 0.88;0.90), whereas EKFC yielded 0.44 (95% CI 0.43; 0.44) and 0.89 (95% CI 0.88; 0.90). However, occasional small differences were observed in specific subgroups regarding median bias, as RF tends to perform better in patients younger than 12 years (Table S21 and Fig. 1) and patients with mGFR > 120 (Table S23).

Shapley values or SHAP plots help interpret prediction models as they show the contribution and the importance of each feature on the predictions. We present the SHAP diagram for the RF model using both creatinine and cystatin C in Fig. 5. As can be seen, the biomarkers are the most relevant features (with higher values associated with a lower GFR prediction), followed by age, weight (Wt), sex, height (Ht) and BMI.
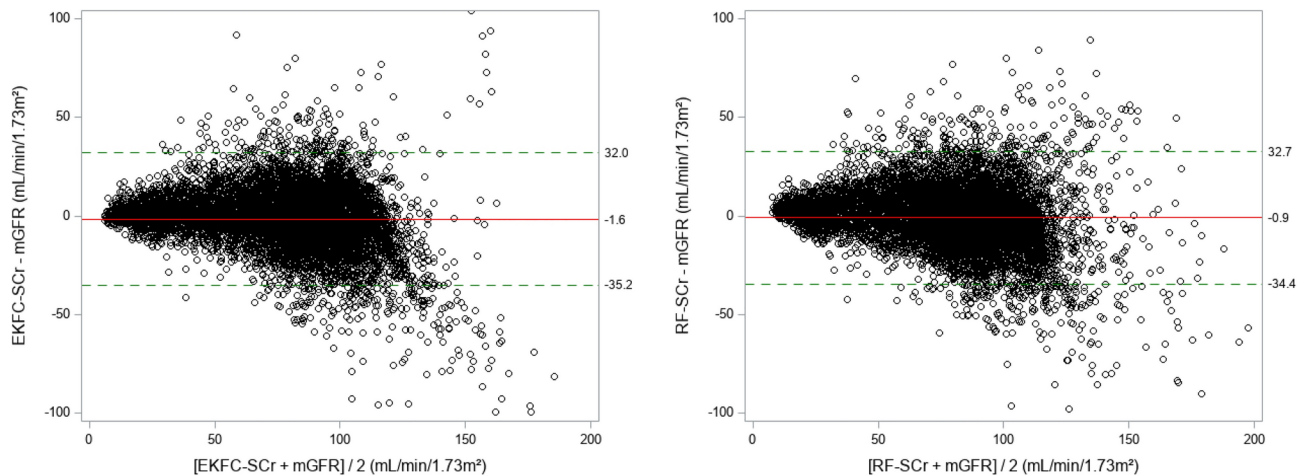
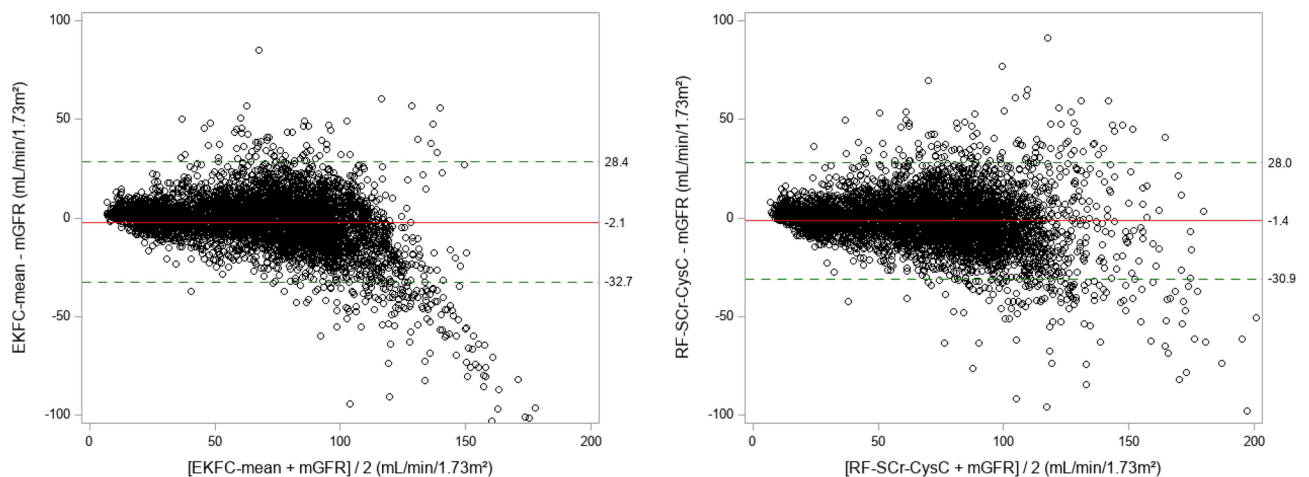**Fig. 3**. Bland–Altman plots for the SCr-based EKFC (left panel) and RF model (right panel).

Fig. 4. Bland–Altman plots for the SCr/CysC-based EKFC (left panel) and RF models (right panel).

Discussion

In this work, we have compared several ML algorithms, using age, sex, SCr (and CysC), height, weight and BMI as features, with the EKFC equation. According to the results, RF, the best performing ML method, managed to be competitive with EKFC in most of the cases. The only noticeable difference was observed in patients younger than 12 years where RF including both SCr and CysC slightly outperformed EKFC. We also performed experiments where height, weight and BMI were not included as features. In this case, all ML models performed systematically worse than their counterparts which could use all features.

It has been suggested that ML methods could be superior to equations resulting from traditional statistical methods in estimating glomerular filtration rate[23]. Moreover, most ML methods are classification models to predict chronic kidney disease outcomes (risk of end-stage renal disease)[10,12,13], rather than regression models to predict glomerular filtration rate[23]. In case GFR was the outcome of interest, the ML models were limited to specific populations (e.g. intensive care unit patients[14], US population[23]) or were intended to choose the most optimal estimating GFR equation[11]. However, these ML methods were trained and validated on other datasets, making them difficult to compare with the performance of the ML models we obtained. In the current study, we used exactly the same data to develop and validate (internally and externally) an ML prediction model for estimating GFR as the data we used to develop and validate the SCr-based EKFC equation[2]. Many different models were evaluated (multi-layer perceptron (neural network), support vector machines, k nearest neighbors, linear regression, Random Forest regression, XGBoost regression) to come up with the best performing ML model as possible, with measured GFR as the outcome, and age, sex, SCr (and CysC), height, weight and BMI as features. As the EKFC equation is a full age range equation, we included data of both children and adults in this study.

The best performing model was the RF model, based on the overall performance in terms of bias, IQR, MSE and P10/P30 accuracy. This was the case both for the SCr-based RF model and SCr/CysC-based RF model. However, the results of our analysis (bias vs age, P10/P30 vs age, and Bland–Altman plots) and sub analysis
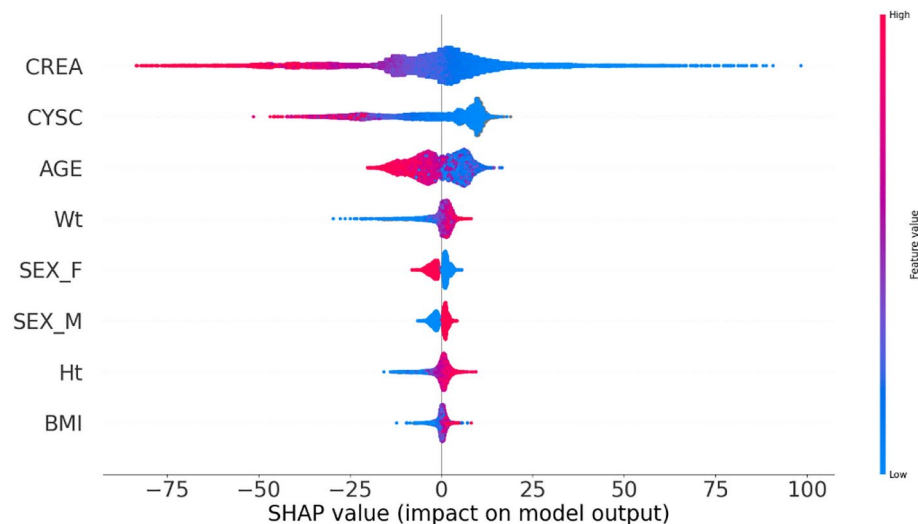
Fig. 5. SHAP plot to show feature importance for the RF model. That is, serum creatinine and BMI are the most and the least relevant features, respectively. Feature values (e.g., higher or lower serum creatinine values) are represented using the colors blue and red, where blue values are associated with lower values and red with higher ones. The placement of the dots is related to their impact on the output. Dots to the left side of ($x=0$) reduce the output and dots to the right increase the output.

(according to age, mGFR, sex and BMI subgroups) show that these RF models were not superior to the SCr-based and SCr/CysC-based EKFC-equations, respectively, in the external validation dataset. Furthermore, we noticed that using weight and height as features for the ML did not improve the results. The conclusion of our analysis is that ML models could not consistently outperform EKFC. EKFC and the random forest model slightly outperform each other in some subsets of patients, but not systematically. Accordingly, the RF model was slightly less biased in patients younger than 12 years and when mGFR was exceeding 120 mL/min/1.73m^2. However, especially in the latter subgroup, the performance of all equations was relatively poor, and the best recommendation would probably be to measure GFR.

One advantage of the EKFC-equation over the RF model is the ease with which you can implement it in the software of a clinical laboratory, which is not straightforward for a ML model. The EKFC-equation can also be explained in an intuitive manner: a person with a normal SCr-value (i.e. a SCr/Q value close to '1') will have an estimated GFR close to 107.3 for people younger than 40 years and an eGFR equal to $107.3 \times 0.990^{(Age-40)}$ for people older than 40 years, as renal decline naturally begins around the age of 40 years. Such interpretation is not as intuitively made from a ML model that only gives a prediction from a black box algorithm. However, random forests may use post-hoc methods, such as SHAP (Fig. 5), which can help to provide a clear explanation of the importance of the variables responsible for the prediction.

Limitations
Limitations of this study include the presence of only white patients. Further studies should investigate the performance of ML methods in cohorts with a mixture of ethnicities cohorts. Also, to make them directly comparable, ML methods were allowed to use the same features as the EKFC-equation, which might have limited their performance. Using more covariates, whenever applicable, such as albuminuria, diabetes status, blood pressure as input to the ML models, could theoretically lead to superior results for ML models, even when these features are not available, as missing features can be handled without imputation. One of the important advantages of ML models is that introducing new variables in the model is much easier than in the development of classical statistical equations like EKFC, as ML methods can directly incorporate new features, whereas the equations would require a functional form of the new features.

Conclusions
Given that ML models are able to model non-linearity and were allowed to use all available covariates (including age, sex, height, weight, BMI, SCr and CysC), but could not outperform the EKFC-equation (which only uses age, sex, SCr and/or CysC), we may assume that we are reaching the limits of accuracy and precision in estimating GFR, with the current features available. It also shows that the proposed two-spline EKFC-equation is a very realistic model, optimally matching measured GFR within the current limitations of accuracy and precision.

Data availability
The datasets used and analyzed during the current study are available on reasonable request to author Hans Pottel.

References

1. Inker, L. A. et al. New creatinine- and cystatin c-based equations to estimate GFR without race. *N Engl. J. Med.* **385**, 1737–1749 (2021).
2. Pottel, H. et al. Development and validation of a modified full age spectrum creatinine-based equation to estimate glomerular filtration rate. *Ann. Int. Med.* **174**, 183–191 (2021).
3. Pottel, H. et al. Cystatin C-based equation to estimate GFR without the inclusion of race and sex. *N Engl. J. Med.* **388**, 333–343 (2023).
4. Stevens, P. E. et al. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int.* **105**, S117-314 (2024).
5. Schwartz, G. J. et al. New equations to estimate GFR in children with CKD. *J. Am. Soc. Nephrol.* **20**, 629–637 (2009).
6. Pottel, H. et al. Estimating glomerular filtration rate at the transition from pediatric to adult care. *Kidney Int.* **95**, 1234–1243 (2019).
7. Delanaye, P., Cavalier, E., Stehlé, T. & Pottel, H. Glomerular filtration rate estimation in adults: myths and promises. *Nephron.* **148**(6), 408–414 (2024).
8. Delanaye, P. et al. Estimating glomerular filtration in young people. *Clin. Kidney J.* https://doi.org/10.1093/ckj/sfae261 (2024).
9. Delanaye, P. et al. Diagnostic standard: Assessing glomerular filtration rate. *Nephrol. Dial. Transp.* **39**(7), 1088–1096. https://doi.org/10.1093/ndt/gfad241 (2023).
10. Schena, F. P., Anelli, V. W., Abbrescia, D. I. & Di Noia, T. Prediction of chronic kidney disease and its progression by artificial intelligence algorithms. *J. Nephrol.* **35**, 1953–71. https://doi.org/10.1007/s40620-022-01302-3 (2022).
11. Fan, Z. et al. Construct a classification decision tree model to select the optimal equation for estimating glomerular filtration rate and estimate it more accurately. *Sci. Rep. [Internet].* **12**, 1–13. https://doi.org/10.1038/s41598-022-19185-6 (2022).
12. Peng, H. et al. A two-stage neural network prediction of chronic kidney disease. *IET Syst. Biol.* **15**, 163–171 (2021).
13. Singh, V., Asari, V. K. & Rajasekaran, R. A deep neural network for early detection and prediction of chronic kidney disease. *Diagnostics* **12**, 1–22 (2022).
14. Woillard, J. B. et al. A machine learning approach to estimate the glomerular filtration rate in intensive care unit patients based on plasma Iohexol concentrations and covariates. *Clin. Pharmacok. [Internet].* **60**, 223–33. https://doi.org/10.1007/s40262-020-00927-6 (2021).
15. Soveri, I. et al. Measuring GFR: A systematic review. *Am. J. Kidney Dis.* **64**, 411–424 (2014).
16. Delanaye, P. et al. Iohexol plasma clearance for measuring glomerular filtration rate in clinical practice and research: A review. Part 1: How to measure glomerular filtration rate with Iohexol?. *Clin. Kidney J.* **9**, 1–18 (2016).
17. Delanaye, P. et al. Iohexol plasma clearance for measuring glomerular filtration rate in clinical practice and research: A review. Part 2: Why to measure glomerular filtration rate with Iohexol?. *Clin. Kidney J.* **9**, 700–4 (2016).
18. Breiman, L. RFRSF: Employee turnover prediction based on random forests and survival analysis. *Mach. Learn.* **45**, 5–32 (2001).
19. Kocev, D., Vens, C., Struyf, J. & Dzeroski, S. Ensembles for predicting structured outputs. *Patt. Recognit.* **46**, 817–33 (2013).
20. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* **13**(17), 785–94 (2016).
21. Early, A., Miskulin, D., Lamb, E. J., Levey, A. S. & Uhlig, K. Annals of internal medicine review estimating equations for glomerular filtration rate in the era. *Ann. Int. Med.* **156**, 785–95 (2012).
22. Lundberg, S. & Su-In, L. A unified approach to interpreting model predictions Scott. *Adv. Neural Inf. Process Syst.* **30** (2017).
23. Wang, H. et al. A deep learning approach for the estimation of glomerular filtration rate. *IEEE Trans. Nanobiosci.* **21**, 560–569 (2022).

Acknowledgements

Author contributions

FKN, AA, JdB, KD, RDh, FTN, AL and HP performed and validated the experiments. FKN and HP created the figures. MC, LC, NE, BOE, RND, LDD, FG, CG, AG, LJ, MG, NK, CL, KL, CM, TM, LR, ADL, ES, PS, AB, UB, KAM, LS, AL, UN performed data collection. JB, HP, PD and CV supervised the work. All authors were involved in the discussion and drafting of the manuscript.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-77618-w.

Correspondence and requests for materials should be addressed to F.K.N.

Reprints and permissions information is available at www.nature.com/reprints.