diabetesresearchandclinicalpractice 233 (2026) 113137

Contents lists available at ScienceDirect

Diabetes Research and Clinical Practice

journal homepage: www.journals.elsevier.com/diabetes-research-and-clinical-practice

RenoTrue: A diabetes-specific machine learning model to estimate glomerular filtration rate for people with diabetes

Rodney Kwok [a,b,c], Kartik Kishore [b,c], Tina Zafari [b], Digsu N. Koye [c,d], Mariam Hachem [a,b,c], Ian H. de Boer [e], Tae-Dong Jeong [f], Won-Ki Min [g], Esteban Porrini [h], Petter Bjornstad [i,j,k], Yih Chung Tham [l], Richard J. MacIsaac [c,m,n], Leonid Churilov [c,o], Elif I. Ekinci [a,b,c,*]

[a] Department of Endocrinology, Austin Health, Melbourne, Australia
[b] Department of Medicine Austin Health, Melbourne Medical School, The University of Melbourne, Australia
[c] Australian Centre for Accelerating Diabetes Innovations, The University of Melbourne, Melbourne, Australia
[d] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia
[e] Kidney Research Institute, University of Washington, Seattle, United States of America
[f] Department of Laboratory Medicine, Ewha Womans University College of Medicine, Seoul, South Korea
[g] Department of Laboratory Medicine, University of Ulsan College of Medicine and Asan Medical Center, Seoul, South Korea
[h] Hospital Universitario de Canarias, Tenerife, Spain
[i] Department of Pediatrics and Department of Medicine, University of Washington, Seattle, WA, United States of America
[j] University of Washington Medicine Diabetes Institute, University of Washington, Seattle, WA, United States of America
[k] Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States of America
[l] Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[m] Department of Medicine, St Vincent's Hospital Melbourne, The University of Melbourne, Melbourne, Australia
[n] Department of Endocrinology and Diabetes, St Vincent's Hospital Melbourne, Australia
[o] Department of Medicine, Royal Melbourne Hospital, Melbourne Medical School, The University of Melbourne, Australia

ARTICLE INFO

Keywords:
Glomerular Filtration Rate
Diabetic Nephropathies
Diabetes Mellitus, Type 2
Diabetes Mellitus, Type 1
Creatinine
Renal Insufficiency
Chronic
Algorithm
Neural Networks
Computer
Machine Learning

ABSTRACT

Background: Existing methods for estimating GFR in people with diabetes have shown inaccuracies when compared to mGFR measurements. We developed and validated an artificial neural network – RenoTrue to improve estimating GFR in people with diabetes.

Methods: 5,619 individuals from five international cohorts with type 1 and type 2 diabetes was split into training (70%), validation (10%) and test (20%) datasets. RenoTrue was developed to estimate GFR using age, sex, and serum creatinine. The performance was evaluated in the test dataset by estimating agreement, bias (mean difference), and accuracy (p30), and compared to CKD-EPI estimates through a multi-level mixed effect regression model.

Findings: Median mGFR was 75 ml/ min per 1.73 m^2 [IQR: 49, 100] and median age was 59 years [IQR: 38, 69]. RenoTrue demonstrated high agreement (ICC: 0.87 (95% CI: 0.78, 0.93)), low bias (−0.57 (95% CI: -1.59, 0.46) ml/min per 1.73 m^2) and p30 of 81% (95% CI: 79%, 83%) compared to mGFR measurements. The 2009 CKD-EPI equation had an ICC of 0.86 (95% CI: 0.77, 0.92), bias of 4.17 (95% CI: 3.14, 5.20) ml/min per 1.73 m^2 and p30 of 74% (95% CI: 72%, 77%).

Conclusion: For people with diabetes, RenoTrue demonstrated better performance compared to the 2009 CKD-EPI equation in terms of estimating GFR across the full range of GFR.

1. Introduction

There is a wealth of evidence-based approaches and therapies that are available to address the risk factors driving the progression of diabetes-related kidney disease (DKD) [1–3]. However, in clinical practice, the detection and monitoring of kidney dysfunction, and the implementation of protective therapies may be suboptimal as the current methods used to estimate GFR lack diagnostic precision. [4,5].

* Corresponding author at: Australian Centre for Accelerating Diabetes Innovations, Level 8, 700 Swanston St, Melbourne Connect (Building 290), Carlton, VIC 3053, Australia.
E-mail address: elif.ekinci@unimelb.edu.au (E.I. Ekinci).

https://doi.org/10.1016/j.diabres.2026.113137
Received 5 August 2025; Received in revised form 14 January 2026; Accepted 31 January 2026
Available online 2 February 2026
0168-8227/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Currently, the most common method to estimate kidney function in people with and without diabetes is the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI 2009 & CKD-EPI 2021) equation (based on serum creatinine, age, and sex) [6]. Both CKD-EPI 2009 and CKD-EPI 2021 are equations recommended by international kidney guidelines such as the Kidney Disease Improving Global Outcomes (KDIGO) [7] and American Diabetes Association (ADA) [6]. However, recent studies have shown that estimating equations such as the CKD-EPI-2009 equation are less accurate in people with diabetes and have the substantial potential for misclassifying CKD stages [8–10]. In 2021, a "race-free" CKD-EPI equation (CKD-EPI-2021) [11] has been proposed as an alternate approach to estimating kidney function. However, studies evaluating the usefulness of this equation for estimating kidney function in people with diabetes are limited [12]. KDIGO guidelines similarly recognise that both CKD-EPI equations have limitations [7]. Therefore, more accurate measurements of true kidney function are required to help improve the detection and management of kidney disease.

Inaccuracies of estimating equations in diabetes have resulted in a call to develop better methods for estimating GFR in people with diabetes [7,9]. Previously, in our international individual participant pooled analysis, there was limited agreement, high bias, low precision and accuracy between the CKD-EPI equations and mGFR measurements [12]. This calls for methods to improve on current methods for estimating GFR in people with diabetes. One possible approach is integration with artificial intelligence and machine learning algorithms specifically an artificial neural network (ANN) which has shown promise in much smaller cohorts [9,13].

We therefore developed a GFR estimating algorithm (RenoTrue). We hypothesised that estimates of glomerular filtration rate derived from RenoTrue would outperform those derieved from the CKD-EPI-2009 and 2021 equations.

2. Methods

2.1. Data source and participants

The complete dataset for RenoTrue was compiled using contributions from five international studies [8,9,14–16] which are part of the Improving Diagnosis of DKD (ID:DKD) consortium which we formed from a previously conducted systematic review [9]. The studies identified during the review were asked to contribute individual participant data (n = 11,138 datapoints) for analysis. All the identified studies met the inclusion criteria which were: participants being at least 18 years; having established diabetes, a creatinine-based CKD-EPI measurement and an mGFR undertaken using the clearance/plasma disappearance of inulin, iohexol, iothalamate, diethylene triamine pent-acetic acid (DPTA), chromium edetate (Cr-EDTA) or other. Participants were excluded if they already had a kidney transplant, chronic corticosteroid use, chronic inflammatory disease(s), pregnancy, known non-diabetes-related kidney disease, thalassaemia, heart failure, and potential kidney donors, and/or if critically ill to reduce potential confounding variables. Those with excessively high mGFR values (>200 ml/min/1.73 m^2) possibly due to measurement error were excluded in the analysis (n = 1).

2.2. mGFR measurements

Studies participating in the consortium had used different exogenous filtration markers to measure kidney function including either iohexol (15%) [8,15], iothalamate (26%) [16], chromium-labelled ethylenediaminetetraacetic acid (Cr-EDTA) (17%) [14] and diethylenetriamine-pentaacetic acid (DTPA) (42%) [9]. Each study had multiple plasma samples collected at various time points over the course of 3 h – 8 h after injection to calculate GFR.

2.3. Model Predictors and outcome

The inputs for the model were serum creatinine, age and sex which were similar to the current CKD-EPI equations [11,17] and for ease of use in clinical practice. Serum creatinine was measured using the Jaffe method in DCCT, Spain and Korea; and using enzymatic method in Australia and Germany). The output from the model was a continuous eGFR estimation value (ml/min per 1.73 m^2).

2.4. Model development

Patient data were split randomly into either training (70%, n = 3,933), validation (10%, n = 557) and testing (20%, n = 1,129) data sets. The training dataset was used to train the initial model; validation data were used to fine tune the model's hyperparameters whilst the testing dataset was used to evaluate the final fully trained model's performance.

2.4.1. Artificial neural networks

Artificial neural networks are a method in artificial intelligence and machine learning. Artificial neural networks uses interconnected neurons in a layered structure that teaches computers to process data in a way that is inspired by the human brain [18]. Hence, several neural network architectures (2^n neurons (n: 0 to 6), up to 3 hidden layers) were considered when developing RenoTrue using the training dataset. The final model after hyperparameter tuning in the validation dataset, was constructed with a single hidden layer of 4 neurons. This is to improve transparency/explainability of the model and to reduce the number of free parameters for the model to learn.

Furthermore, during the training and validation of the neural network model, we used batch normalization and monitored training/validation loss curves to avoid over-fitting [19]. Batch normalization is a method used to stabilise training and improves generalisation of a model. The inputs to all models were standardized to facilitate better convergence of the algorithm and steps were taken to ensure that there is no data leakage from the training to validation or test datasets. For the neural network, the activation function used was Rectified Linear Unit (ReLU). Shapley Addictive Explanations (SHAP) [20] beeswarm plot were produced to provide model interpretability, whilst loss of epoch curve represents the models training progress.

2.5. Missing data

There were no missing data of independent features (e.g., serum creatinine, age and sex) as these variables were complete in the international dataset. Log transformations were also applied to serum creatinine and age to approximately conform any skewed data.

2.6. Statistical analysis

Using the test dataset, the performance of RenoTrue was compared to that of the 2009 CKD-EPI equation and the 2021 CKD-EPI equations with mGFR measurements through a mixed effect linear regression model. Mixed effect variance component model was chosen to account for the hierarchy structure, estimated difference among methods used to measure glomerular filtration rate and possible systemic differences of serum creatinine measurement within the international dataset. The metrics considered for each evaluation calculated from the mixed effect model included bias by examining the mean difference between the RenoTrue estimated eGFR and the mGFR values (eGFR – mGFR). Precision of RenoTrue compared to mGFR values was measured as standard deviation (SD) and accuracy was estimated by the percentage of RenoTrue estimates within 10% (P_{10}), 15% (P_{15}) and 30% (P_{30}) of the mGFR of their respective mGFR measurements. Additionally intraclass correlation coefficient (ICC) [21] was used to assess the agreement between RenoTrue estimates and mGFR measurements. We have also explored

the performance of the RenoTrue algorithms in CKD-stages, stratified by mGFR categories (<30, 30–59, 60–89, ≥90 ml/min per 1.73 m^2). The calculations for the CKD-EPI equations are provided in the supplementary materials. McNemar's Test [22] was used to compare significance of accuracy between RenoTrue and 2009 CKD-EPI and 2021 CKD-EPI. Bland-Altman plots [23] were also created between mGFR measurements and RenoTrue, 2009 CKD-EPI and 2021 CKD-EPI estimates. Bland-Altman plots can be influenced by outliers or extremely high values and is unable to account for the hierarchy nature of the data, hence, these were provided as a supplementary analysis. Reduced Major Axis Plots were created to visualise the relationship between mGFR measurements and RenoTrue, 2009 CKD-EPI and 2021 CKD-EPI estimates.

We also conducted analyses of clinically relevant subgroups by age and sex. Python 3.12 was used for model development. Additional list of packages used are listed in the supplementary materials. Moreover, STATA ICv16 (TX: StataCorp LLC) was used for statistical analysis and Excel (Microsoft Corporation 2025) was used for figure developments. This study was approved by the Austin Health Human Research Ethics Committee (Project Number: HREC/69945/Austin-2020) and reported according to the TRIPOD + AI [24] guidelines. A study protocol was not prepared.

3. Results

Participant characteristics from the pooled dataset are shown in Table 1. Overall, there were 5,619 participants, most of whom had type 2 diabetes (60%) and 31% (N = 1,727) had type 1 diabetes. Baseline characteristics were similar and balanced across the training, validation and testing datasets (Table 1).

3.1. Performance of Reno true across full observed range of mGFR measurements

Fig. 1 shows the performance of RenoTrue compared to the 2009 CKD-EPI and 2021 CKD-EPI equations in the test dataset. In the test dataset, eGFR values derived from RenoTrue indicated similar agreement (ICC: 0.87 (95% CI: 0.78, 0.93)) with direct measurements of kidney function compared to the 2009 CKD-EPI equation (ICC: 0.86 (95% CI: 0.77, 0.92)) and the 2021 CKD-EPI equation (ICC: 0.78 (95% CI: 0.67, 0.86)) (Fig. 1). However, despite similar ICC, bias (mean difference between eGFR and mGFR measurements) was lower, −0.57 ml/min per 1.73 m^2 (95% CI: −1.59, 0.46) for RenoTrue, compared to 9.57 ml/min per 1.73 m^2 (95% CI: 8.53, 10.58) for the 2021 CKD-EPI equation and 4.17 ml/min per 1.73 m^2 (95% CI: 3.14, 5.20) for the 2009 CKD-EPI equation (Fig. 2). RenoTrue measurements had better precision (standard deviation) of 11.36 ml/min per 1.73 m^2 (95% CI: 10.9, 11.84) compared to the 2021 CKD-EPI equation 18.02 ml/min per 1.73 m^2 (95% CI: 17.29, 18.77) but similar precision to the 2009 CKD-EPI equation 12.08 ml/min per 1.73 m^2 (95% CI: 11.62, 12.59) respectively (Fig. 3). RenoTrue measurements had higher accuracy with 39% (95% CI: 37%, 41%), 55% (95% CI: 52%, 58%) and 81% (95% CI: 79%, 83%) of were within 10%, 15% and 30% of their respective mGFR measurements. The 2009 CKD-EPI equation had lower accuracy of 33% (95% CI: 30%, 36%) (p10), 48% (95% CI: 45%, 51%) (p15) and 74% (95% CI: 72%, 77%) (p30) of their respective mGFR measurements compared to RenoTrue measurements. Similarly, 28% (95% CI: 25%, 31%), 40% (95% CI: 37%, 43%) and 64% (95% CI: 61%, 67%) of 2021 CKD-EPI estimates within 10%, 15% and 30% of their respective mGFR measurements (Fig. 4). The results from McNemar's test suggest that the proportion of RenoTrue estimates within p10, p15 and p30 are statistically significantly greater ($p < 0.001$), indicating that the accuracy of RenoTrue estimates is significantly greater compared to 2009 CKD-EPI and 2021 CKD-EPI estimates when compared to their respective mGFR measurements. Therefore, the null hypothesis is rejected, concluding that there is significant difference in performance of RenoTrue for people with diabetes compared to the CKD-EPI equations (Supplementary Table S1–S6). Supplementary Table S7 presents the accuracy metrics of the eGFR estimation methods with their respective 95% confidence intervals. The confidence intervals for p10, p15, p30 does not overlap between the CKD-EPI equations and RenoTrue, demonstrating statistically significance difference in performance between RenoTrue and the CKD-EPI equations in accurately estimating glomerular filtration rate. Supplementary Fig. S1–S3 shows the Bland-Altman plots between mGFR measurements and RenoTrue (bias: -0.57 ml/min per 1.73 m^2), 2009 CKD-EPI (bias: 4.2 ml/min per 1.73 m^2) and 2021 CKD-EPI (bias: 9.6 ml/min per 1.73 m^2) estimates. Supplementary Figs. S4–S6 presents the Reduced Major Axis Plots. Supplementary Table S8 presents the analyses by clinically relevant subgroups by age, sex and type of diabetes.

Table 1
Baseline characteristics of study participants.

		Overall	TRAINING	VALIDATION	TEST
		(n = 5,619)	(n = 3,933)	(n = 557)	(n = 1,129)
mGFR, median [Q1, Q3] (ml/min per 1.73m^2)		74.5 [48.5, 99.8]	74.0 [48.6,99.5]	75.6 [47.1,99.1]	75.6 [49.1,101.5]
2009 CKD-EPI, median [Q1, Q3]		84.7 [56.4,102.3]	84.4 [56.4,102.2]	85.1 [55.0,100.8]	84.8 [56.1,102.8]
2021 CKD-EPI, median [Q1, Q3]		88.9 [59.9,105.3]	88.8 [60.1,105.3]	89.5 [58.8,104.6]	89.4 [59.7,105.9]
Age (years), median [Q1, Q3]		59.0 [38.0,69.0]	59.0 [38.0,70.0]	60.0 [39.0,69.0]	57.5 [38.0,68.0]
Body Mass Index (kg/m^2), median [Q1, Q3]		27.7 [24.4,32.1]	27.8 [24.4,32.2]	27.7 [24.5,31.8]	27.6 [24.5,32.1]
HbA1c (%), median [Q1, Q3]		7.6 [6.7,8.7]	7.6 [6.7,8.7]	7.7 [6.9,8.8]	7.6 [6.7,8.7]
Triglycerides (mmol/L), median [Q1, Q3]		3.5 [1.4,107.1]	3.4 [1.4,107.0]	3.5 [1.4,95.5]	3.7 [1.4,115.0]
Cholesterol (mmol/L), median [Q1, Q3]		6.1 [4.2,162.0]	6.2 [4.2,163.0]	5.9 [4.2,156.0]	6.0 [4.2,163.0]
Serum creatinine (μmol/L), median [Q1, Q3]		80.0 [70.0,106.1]	80.0 [70.7,106.1]	79.6 [68.0,110.5]	81.0 [70.0,108.7]
Female, n (%)		2357 (41.9)	1629 (41.4)	241 (43.3)	487 (43.1)
Diabetes Type, n (%)	**Unknown**	535 (9.5)	371 (9.4)	57 (10.2)	107 (9.5)
	Type 1	1727 (30.7)	1209 (30.7)	155 (27.8)	363 (32.2)
	Type 2	3357 (59.7)	2353 (59.8)	345 (61.9)	659 (58.4)
mGFR category, n (%)	**<30**	583 (10.4)	408 (10.4)	58 (10.4)	117 (10.4)
	30–59	1449 (25.8)	1021 (26.0)	140 (25.1)	288 (25.5)
	60–89	1646 (29.3)	1161 (29.5)	174 (31.2)	311 (27.5)
	≥ 90	1941 (34.5)	1343 (34.1)	185 (33.2)	413 (36.6)
Cohort name/ Study Location, n (%)	**DCCT**	1339 (23.8)	946 (24.1)	113 (20.3)	280 (24.8)
	Australia	2652 (47.2)	1854 (47.1)	279 (50.1)	519 (46.0)
	German	173 (3.1)	130 (3.3)	6 (1.1)	37 (3.3)
	Korea	850 (15.1)	579 (14.7)	102 (18.3)	169 (15.0)
	Spain	605 (10.8)	424 (10.8)	57 (10.2)	124 (11.0)

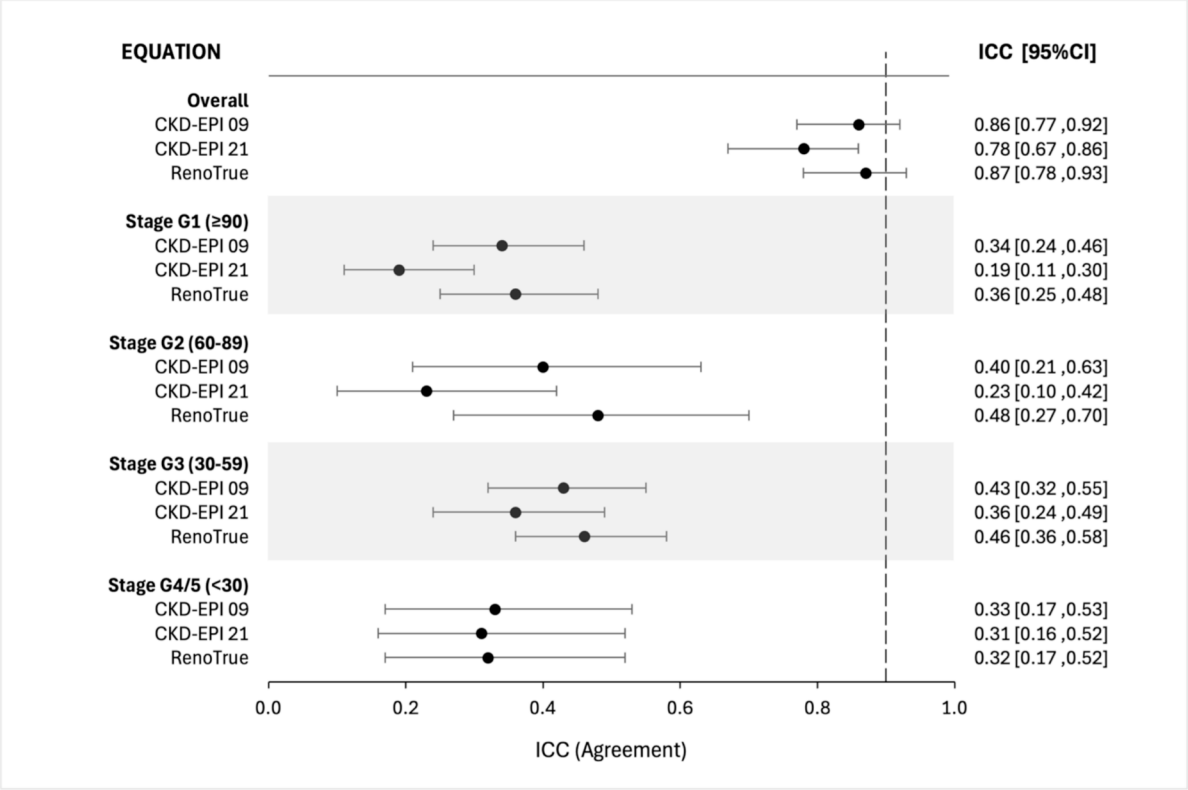


Fig. 1. Forest plot of performance of eGFR equations compared to direct measurements of kidney function. Overall and subgroups analyses of CKD stages are shown. ICC is estimated from the mixed effect regression model. Dotted line (ICC: 0.9) represents excellent agreement between eGFR estimates and direct measurements of kidney function. ICC: Intraclass correlation coefficient.

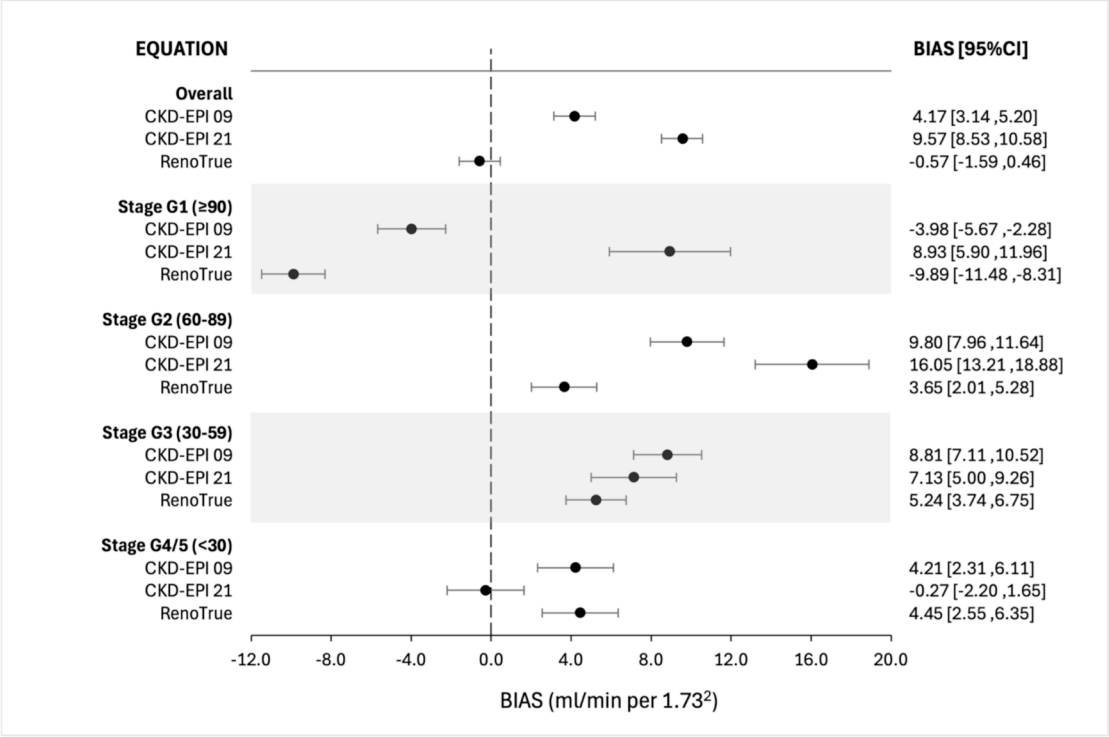


Fig. 2. Forest plot of performance of eGFR equations compared to direct measurements of kidney function. Overall and subgroups analyses of CKD stages are shown. Bias (mean difference) is estimated from the mixed effect regression model.

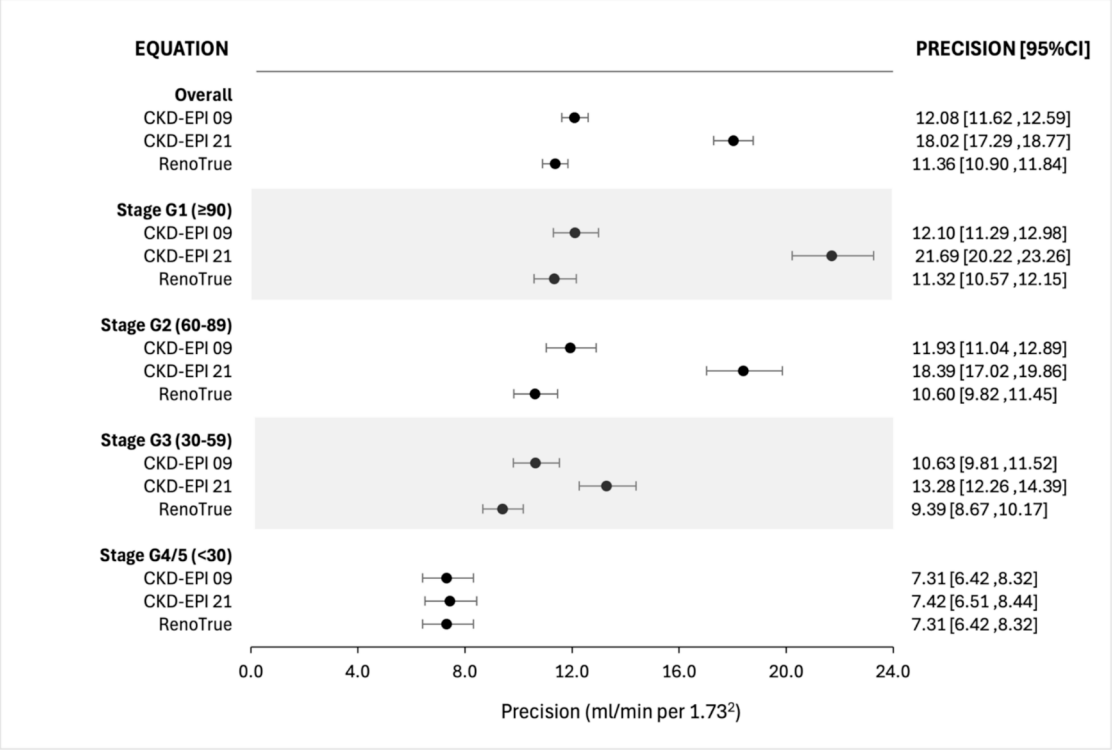


Fig. 3. Forest plot of performance of eGFR equations compared to direct measurements of kidney function. Overall and subgroups analyses of CKD stages are shown. Precision (standard deviation) is estimated from the mixed effect regression model.

3.2. Performance of Reno true within CKD stages by clinically relevant mGFR cut-offs

i) CKD stage G4/5: mGFR $<$ 30 ml/min per 1.73 m^2

The intraclass correlation coefficient between mGFR measurements and 2009 CKD-EPI, 2021 CKD-EPI and RenoTrue at eGFR less than 30 ml/min per 1.73 m^2 were relatively similar (Fig. 1). Bias was 4.45 ml/min per 1.73 m^2 (95% CI: 2.55, 6.35) for RenoTrue (Fig. 2) with the highest precision (7.31 ml/min per 1.73 m^2 (95% CI: 6.42, 8.32)) (Fig. 3) respectively. Accuracy, p10, p15 and p30, were all similar between the three estimates and mGFR measurements (Fig. 4). Within CKD stage G4/5, RenoTrue had similar performance compared to the 2009 CKD-EPI equation (bias: 4.21 ml/min per 1.73 m^2 (95% CI: 2.31, 6.11)) and 2021 CKD-EPI equation (bias: −0.27 ml/min per 1.73 m^2 (95% CI: −2.20, 1.65)).

ii) CKD stage G3: mGFR 30–59 ml/min per 1.73 m^2

The intraclass correlation coefficient was 0.46 (95% CI: 0.36, 0.58) for RenoTrue estimates. Bias was found to be 5.24 ml/min per 1.73 m^2 (95% CI: 3.74, 6.75) and precision (SD) was 9.39 ml/min per 1.73 m^2 (95% CI: 8.67, 10.17). RenoTrue had 32% (95% CI: 27%, 38%), 46% (95% CI: 40%, 52%), and 70% (95% CI: 65%, 75%), of estimates within 10%, 15% and 30% of their respective mGFR measurements (Fig. 4). In CKD stage G3, RenoTrue had better agreement, lower bias, better precision and higher accuracy compared to 2009 CKD- EPI [Bias: 8.81 ml/min per 1.73 m^2 (95% CI: 7.11, 10.5); p30: 60% (95% CI: 54%, 65%)] & 2021 CKD-EPI equations [Bias: 7.13 ml/min per 1.73 m^2 (95% CI: 5.00, 9.26); p30: 61% (95% CI: 55%, 66%)].

iii) CKD stage G2: mGFR 60–89 ml/min per 1.73 m^2

When comparing agreement between mGFR measurements, RenoTrue at eGFR 60–89 ml/min per 1.73 m^2 was 0.48 (95% CI: 0.27, 0.70). Bias was 3.65 ml/min per 1.73 m^2 (95% CI: 2.01, 5.28) and precision was 10.6 ml/min per 1.73 m^2 (95% CI: 9.82, 11.45). In terms of accuracy, 41% (95% CI: 35%, 47%), 59% (95% CI: 54%, 65%) and 87% (95% CI: 83%, 90%) of RenoTrue estimates were within 10%, 15% and 30% of their respective mGFR measurements (Fig. 4). In CKD stage G2, RenoTrue had better agreement, lower bias, better precision and higher accuracy compared to 2009 CKD-EPI [Precision: 11.9 ml/min per 1.73 m^2 (95% CI: 11.0, 12.9%); p30: 74% (95% CI: 69%, 78%)] and 2021 CKD-EPI equation [Precision: 18.4 ml/min per 1.73 m^2 (95% CI: 17.0, 19.9); p30: 59% (95% CI: 54%, 65%)].

iv) CKD Stage G1: mGFR $\geq$ 90 ml/min per 1.73 m^2

Within CKD stage G1, RenoTrue had similar performance compared to the 2009 and 2021 CKD-EPI equation. The intraclass correlation coefficient was 0.36 (95% CI: 0.25, 0.48) for RenoTrue and bias was found to be −9.89 ml/min per 1.73 m^2 (95% CI: −11.5, −8.31). Precision was 11.3 ml/min per 1.73 m^2 (95% CI: 10.6, 12.2) (Fig. 3). RenoTrue had 48% (95% CI: 43%, 53%), 66% (95% CI: 61%, 70%) and 94% (95% CI: 92%, 96%) of estimates within 10%, 15% and 30% of their respective mGFR measurements. In comparison, RenoTrue had similar performance to the 2009 CKD-EPI equation [Bias: −3.98 ml/min per 1.73 m^2 (95% CI: −5.67, −2.28); p30: 93% (95% CI: 92%, 96%)] but better performance compared to the 2021 CKD-EPI equation [Precision: 21.7 ml/min per 1.73 m^2 (95% CI: 20.2, 23.3); p30: 77% (95% CI: 73%, 81%)].

3.3. Model Specification

From the SHAP beeswarm plot (Supplementary Fig. S7), sex (male) is the more important input feature followed by age and then creatinine. The loss of epoch curve (Supplementary Fig. S8) highlights that there was no presence of overfitting nor underfitting of the RenoTrue algorithm during the model training and development.

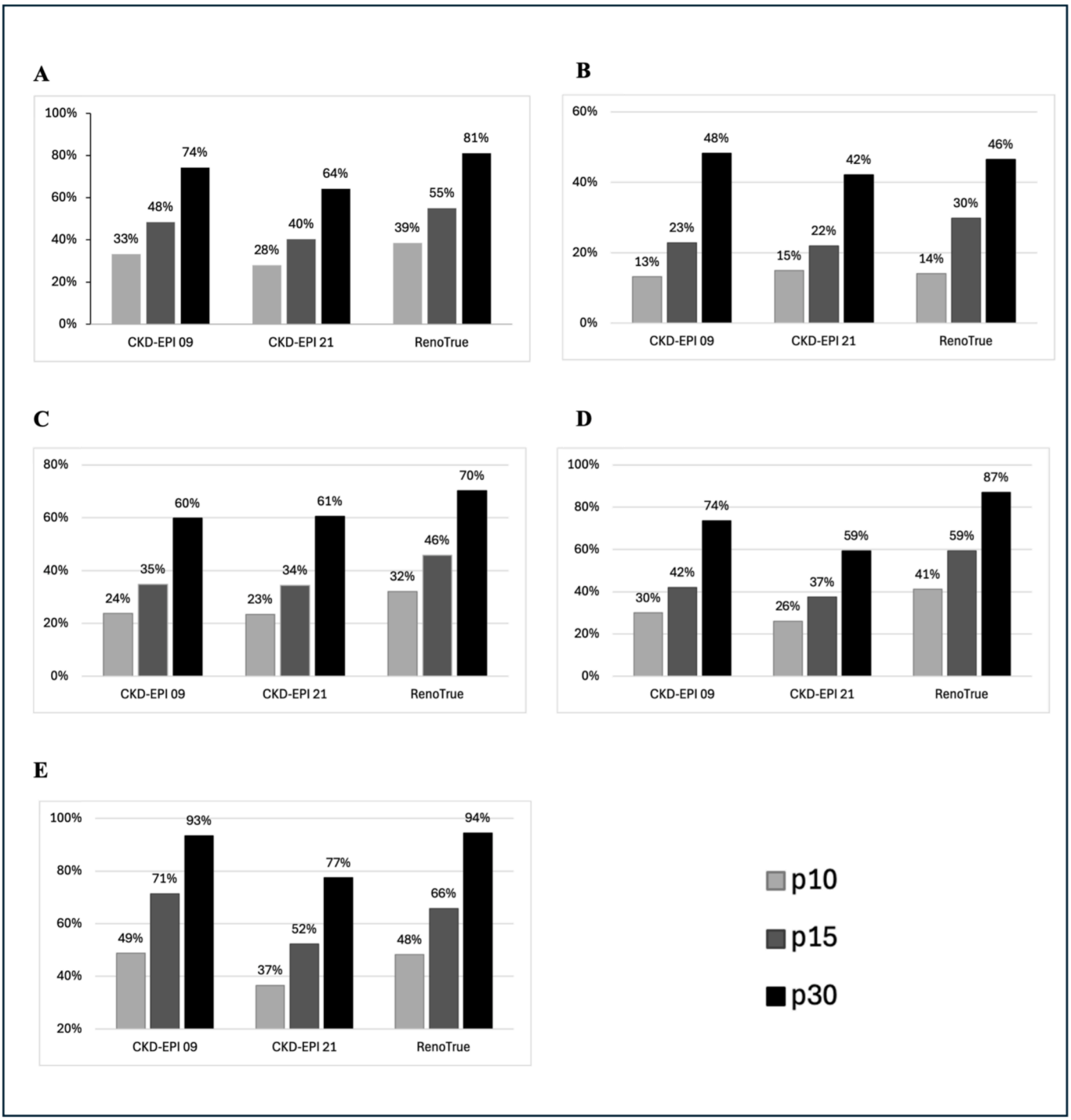


Fig. 4. Bar plots of accuracy of eGFR equations compared to direct measurements of kidney function. P10 is estimates within 10% of their respective mGFR measurements, P15 is estimates within 15% of their respective mGFR measurements and P30 is estimates within 30% of their respective mGFR measurements. A: Overall. B: CKD stage G4/5 (<30). C: CKD stage G3 (30–59). D: CKD stage G2 (60–89). E: CKD stage G1 (≥90).

4. Discussion

In this study we developed, validated and tested RenoTrue, a new approach to estimating GFR, in an international cohort of people with diabetes [8,9,14–16] across a wide spectrum of kidney function and with direct measurements of kidney function. We found RenoTrue to have good agreement and reliability (ICC > 0.8) [21] with direct measurements of kidney function, lower bias and higher accuracy in estimating GFR compared to commonly used eGFR equations such as the 2009 CKD-EPI and 2021 CKD-EPI equations. Our findings were consistent both across a wide spectrum of GFR values and within CKD classification stages defined by clinically relevant GFR cut-offs.

Unlike the CKD- EPI equations' development (n = 1581, 29%) and validation (n = 1089, 28%) datasets that had a limited number of people with diabetes [25], the RenoTrue cohort consisted entirely of people with diabetes. The entire CKD-EPI equation epidemic cohort was also much younger (47 years versus 65 years) and had a lower mean direct measurement of kidney function (67.6 versus 88.9 ml/min per 1.73 m^2) compared to our international dataset. Hence, the CKD- EPI cohort characteristics are not truly reflective of people living with diabetes.

4.1. Clinical Implications of the study findings

Measurement of glomerular filtration rate plays a vital role in the management of diabetes-related kidney disease. However, previous studies have highlighted concerns with the use of the CKD-EPI equation in people with diabetes which underestimates directly measured GFR values as well as early decline in GFR to a comparable degree [9,12,26,27]. This is clinically important as estimated glomerular filtration rate is used by the KDIGO CKD classification system for assessing diabetes-related kidney disease and determining the severity the of disease [7]. Moreover, it is important to know that medication recommendation and dosage adjustments are dependent on GFR stages. These includes medications such as renin-angiotensin-system (RAS)-acting agents, sodium glucose cotransporter 2 inhibitors (SGLT-2 s), spironolactone, finerenone, glucagon-like peptide-1 receptor agonists (GLP1-RA) and gastric inhibitory peptide (GIP) combination therapies which all can help slow the development and progression of diabetes-related kidney disease [7]. Henceforth, with RenoTrue providing more accurate estimations compared to current estimating equations has the potential to enable more timely therapeutic interventions for people with diabetes and to delay or slow the progression of kidney disease.

4.2. Comparison to other studies

Previous ANN models for estimating GFR have been proved to be more accurate and precise than current eGFR equations such as the CKD-EPI equation [13,28,29]. Machine learning models can capture complex, non-linear relationships and interactions between variables that traditional statistical methods might miss. They also have the ability to easily incorporate additional variables into the model in the future unlike current equations used in current practice. However, similarly to the CKD-EPI equation cohort, these ANN models were developed and validated with study populations that only included approximately 40% of people with diabetes. Hence, there are only a limited number of previous studies that developed and validated ANN models to estimate glomerular filtration rate in cohorts that consisted entirely of people with diabetes [30].

In 2015, Chen et al. [30] developed multiple new GFR-estimating models for Chinese people with type 2 diabetes using ANN algorithms. The simplest three variable model (age, sex, serum creatinine) produced a mean bias of −2.70 ml/min per 1.73 m^2, precision (IQR of the difference) of 21.37 units, and p30 of 87.6% in the above study. The performance of this ANN model had similar performance to RenoTrue (bias: 0.39 ml/min per 1.73 m^2, precision: 16.23 ml/min per 1.73 m^2, p30: 81.77%) which could be explained by the two cohorts used to derive the respective ANNs having closely matching clinical and physiological characteristics. However, the ANN model by Chen and his colleagues was only developed in a cohort of 519 people with diabetes and the algorithm was validated in an external dataset of 105 patients. These sample sizes are very small for ANN development and validation and the algorithm for the above study may only be only specific to a Chinese population.

Furthermore, we considered and tested several machine learning models and methods before continuing with the specific ANN architecture format that was used to develop RenoTrue. Ordinary Least Square and Random Forests are common machine learning methods used to develop algorithms, like ANNs. Firstly, ordinary least square (OLS) is a parametric method that assumes a linear relationship between the independent and dependent variables. Although it is simple approach to explain and computationally efficient, it is sensitive to outliers and several assumptions need to be met before using this model (e.g., could not use repeated GFR measurements from the same patient) [31]. Secondly, Random Forest is an ensemble technique that combines multiple decision trees to make predictions. As these models are non-parametric, they are negligent in terms of assumptions around the data, but they are known to overfit if not tuned properly [32]. Unlike Neural Networks, they are designed to recognise patterns and can handle complex relationships between variables. ANNs can also deal well with high dimension data and are robust to noise [33]. Hence, the ANN model was selected as the algorithmic baseline for the development of RenoTrue as it was the best performing model compared to OLS and Random Forest.

Additionally, the international dataset had heterogeneity between the studies, especially with the Diabetes Control and Complication Trial (DCCT) cohorts (younger population and mainly people with type 1 diabetes) compared to the rest (older population, mainly people with type 2 diabetes). Moreover, different measurement techniques were used in each individual cohort to measure kidney function. Hence, to increase generalisability, robustness and applicability of the algorithm and to account for different techniques used to directly measure kidney function, it was opted to pool the international cohorts together and then split into training, validation and testing. This approach also allows to minimise potential biases introduced by a single centre's data characteristics and account for less tailoring when the model is applied to different settings and populations [34,35]. Furthermore, our results are also reflective of the ICC's ability to effectively contextualise the extent of variances between mGFR measurements and eGFR estimates by method of estimation based on the measurement scale of interest [12]. The same magnitude of absolute disagreement (e.g. a difference of 4.17, 9.57 an −0.52 ml/min per 1.73 m^2) naturally results in a better measure of agreement across a broader range of GFR (i.e. between 0–100 ml/min per 1.73 m^2) compared to a narrower range within the specific GFR groups (for example between 30–59 or between 60 and 89 ml/min per 1.73 m^2). This is because, in the broader range, the magnitude of the difference between mGFR measurements and eGFR estimates by method of estimation is relatively small compared to the overall observed mGFR range. Conversely, in the narrow ranges, the same the magnitude of the difference is quite comparable to the width of the mGFR range (30 ml/min per 1.73 m^2).

4.3. Limitations and Strengths

RenoTrue was trained, validated and tested, from the largest and most diverse diabetes cohort analysed so far to derive an ANN model to estimate GFR. Nevertheless, our study has limitations. The datasets used for this study were predominately from high-income countries, thus the findings of this study may have limited application to low-and-middle income nations which typically have higher prevalence of people living with diabetes [36]. Additionally, we had relatively fewer numbers of participants in CKD stage G1 and CKD stage G4/5 which could potentially affect the performance of RenoTrue in those stages. Furthermore, RenoTrue is a diabetes-specific ANN which requires the clinical status of diabetes to be included as an input when estimating GFR. Moreover, the selection of the most suitable eGFR equation for people living with diabetes could present a significant challenge for clinicians as people with diabetes often have multiple comorbidities. Similarly to the CKD-EPI equations [11], RenoTrue was developed in a cohort excluding patients with kidney transplants, corticosteroid use, inflammatory disease etc. This is done to minimise confounding factors that could affect the relationship between serum creatinine and true kidney function. Likewise, the CamGFRv2 [37] equation for person with cancer or the Kidney Recipient-Specific (KRS) [38] GFR equation for kidney-transplant recipients each were developed in specific patient populations to improve GFR estimation in the presence of unique physiological conditions. Thus, this complexity and availability of multiple eGFR equations could pose a significant challenge for clinicians when determining the most appropriate equation to estimate GFR. Hence, our future research will incorporate machine learning techniques such as classification and regression trees (CART) to handle the intricate relationships between the various factors affecting kidney function and guiding clinicians to select the most suitable equation for GFR estimation (RenoTrue or other available equations) based on the specific clinical situation and presence of comorbidities. This also underscores

the need for a more comprehensive approach to GFR estimation which future research could focus on developing a unified model that accounts for multiple comorbidities and patient characteristics. The machine learning foundations of RenoTrue offer a comprehensive approach to continuously incorporate a wider range of participant data, working towards a more unified eGFR model.

Unfortunately, racial data was not consistently collected across all ID: DKD cohort included in our model. This limitation prevented us from conducting a comprehensive sub-analysis of the performance of RenoTrue compared to measured GFR by race in people with diabetes. However, it is crucial to note that race is increasingly recognised as a social construct rather than a biological construct in medicine. Hence, international guidelines have recommended to standardise the race coefficient as it may contribute to racial disparities in healthcare access and outcomes [6,7]. In addition, RenoTrue was not developed or evaluated with the inclusion of cystatin C as participants in our consortium did not have cystatin C measurements performed. It is worth noting that cystatin C is often less available and only in a few laboratories compared to creatinine [38]. Furthermore, although there are other equations such as the European Kidney Function Consortium Equation [39] (EKFC) being recognised for use to estimate GFR, RenoTrue was only compared with the CKD-EPI equations. The rationale behind this is that the EKFC equation is currently only being adopted and recommended for use in clinical practice throughout Europe. However, further comparative research could be conducted with other eGFR equations to confirm the clinical utility of RenoTrue.

5. Summary and Conclusions

This study developed, validated and tested a novel diabetes-specific machine learning model, RenoTrue, to estimate kidney function for people with diabetes. RenoTrue yielded more accurate estimates of GFR than current commonly used equations such as the 2009 and 2021 CKD-EPI equations. In the era of precision and personalized medicine, the integration and use of artificial intelligence, and machine learning algorithm approaches to estimate kidney function offers a new perspective to diabetes-related kidney disease management. It is envisaged that using RenoTrue in clinical practice by either embedding in electronic medical records or diabetes-related kidney disease screening platforms will improve patient outcomes and ensure more precise grading of chronic kidney disease stages, recording of kidney function trends, and the timely interventions to delay the progression of diabetes-related kidney disease. Nonetheless, selecting the most suitable eGFR equation for people with diabetes will always be determined on the specific clinical situation and presence of comorbidities. Further assessment and validation of RenoTrue in a prospective dataset is needed to confirm the accuracy of the present results.

6. Authors' contributions

RK and KK contributed to the study design and *RK, KK, TZ, DNK, RM, LC* and *EIE* initiated the study conception. *TDJ, WKM, EP & PB* provided the international datasets. *RK* and *MH* conducted the data collection. *RK* and *KK* conducted the study, data cleaning and statistical analysis. *RK* prepared the writing, editing and planning of initial manuscript. *DNK* and *LC* provided statistical expertise in study analysis plan. *RM* and *EIE* provided clinical expertise in outcome measurements. *RK, TZ, DNK, RM, LC and EIE* contributed to ethics of the study. All authors contributed to the review of the study and approved the final manuscript.

7. Disclosure

PB reports serving or having served as a consultant for AstraZeneca, Bayer, Bristol-Myers Squibb, Boehringer Ingelheim, Lilly, LG Chemistry, Sanofi, Novo Nordisk, and Amgen/Horizon Pharma. PB also serves or have served on the advisory boards and/or steering committees of AstraZeneca, Bayer, Boehringer Ingelheim, Novo Nordisk, Lilly, and XORTX.

EIE conducts unrelated clinical research with Eli Lilly, Novo Nordisk, Boehringer, Amgen, Versanis, Novartis, Endogenex, Bayer and funds are paid to EIE's institution. EIE is on advisory boards and gives presentations for Eli Lilly, Boehringer, Bayer and these funds are paid to EIE's institution and used for diabetes research.

RM conducts unrelated clinical research with Astra Zeneca, Novo Nordisk, Boehringer, Bayer, Servier and Insulet. RM also serves or have served on the advisory boards and/or steering committees of AstraZeneca, Bayer, Boehringer Ingelheim, Novo Nordisk and Servier.

8. Data and code sharing statement

Upon reasonable request, the summary statistics and the code used in the analysis can be made available.

9. Registration and patient/ public involvement

The study was not registered, and patient/public involvement was not involved with the code development.

CRediT authorship contribution statement

Rodney Kwok: Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kartik Kishore:** Writing – review & editing, Validation, Methodology, Investigation. **Tina Zafari:** Writing – review & editing, Conceptualization. **Digsu N. Koye:** Writing – review & editing, Supervision. **Mariam Hachem:** Writing – review & editing. **Ian H. de Boer:** Writing – review & editing, Data curation. **Tae-Dong Jeong:** Writing – review & editing, Data curation. **Won-Ki Min:** Writing – review & editing, Data curation. **Esteban Porrini:** Writing – review & editing, Data curation. **Petter Bjornstad:** Writing – review & editing, Data curation. **Yih Chung Tham:** Writing – review & editing. **Richard J. MacIsaac:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Leonid Churilov:** Writing – review & editing, Supervision, Conceptualization. **Elif I. Ekinci:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

- DCCT data provided by NIDDK CR, a program of the National Institute of Diabetes and Digestive and Kidney Diseases.
- The German data was provided by Natalie Ebert PhD and Elke Schaeffner PhD from the Institute of Public Health, Charité Universitätsmedizin, Berlin, Germany.
- This project was supported by ACADI funding. The Australian Centre for Accelerating Diabetes Innovations (ACADI) was established through MRFF funding from the Australian Government's Targeted Translation Research Accelerator program, delivered by MTPConnect.
- This research is also supported by an Australian Government Research Training Program (RTP) Scholarship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.

org/10.1016/j.diabres.2026.113137.

References

[1] Fenta ET, Eshetu HB, Kebede N, Bogale EK, Zewdie A, Kassie TD, et al. Prevalence and predictors of chronic kidney disease among type 2 diabetic patients worldwide, systematic review and meta-analysis. Diabetol Metab Syndr 2023;15(1):245.
[2] MacIsaac RJ, Jerums G, Ekinci EI. Effects of glycaemic management on diabetic kidney disease. World J Diabetes 2017;8(5):172–86.
[3] Nauck MA, D'Alessio DA, Tirzepatide. A dual GIP/GLP-1 receptor co-agonist for the treatment of type 2 diabetes with unmatched effectiveness regrading glycaemic control and body weight reduction. Cardiovasc Diabetol 2022;21(1):169.
[4] Bakris GL, Agarwal R, Anker SD, Pitt B, Ruilope LM, Rossing P, et al. Effect of finerenone on chronic kidney disease outcomes in type 2 diabetes. N Engl J Med 2020;383(23):2219–29.
[5] Cheung AK, Chang TI, Cushman WC, Furth SL, Hou FF, Ix JH, et al. Executive summary of the KDIGO 2021 clinical practice guideline for the management of blood pressure in chronic kidney disease. Kidney Int 2021;99(3):559–69.
[6] de Boer IH, Khunti K, Sadusky T, Tuttle KR, Neumiller JJ, Rhee CM, et al. Diabetes management in chronic kidney disease: a consensus report by the american diabetes association (ADA) and kidney disease: improving global outcomes (KDIGO). Diabetes Care 2022;45(12):3075–90.
[7] Stevens PE, Ahmed SB, Carrero JJ, Foster B, Francis A, Hall RK, et al. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int 2024;105(4):S117–314.
[8] Luis-Lima S, Higueras Linares T, Henríquez-Gómez L, Alonso-Pescoso R, Jimenez A, López-Hijazo AM, et al. The error of estimated GFR in type 2 diabetes mellitus. J Clin Med 2019;8(10).
[9] Zafari N, Churilov L, Wong LY, Lotfaliany M, Hachem M, Kiburg KV, et al. Evaluation of the diagnostic performance of the creatinine-based chronic kidney disease epidemiology collaboration equation in people with diabetes: a systematic review. Diabet Med 2021;38(1):e14391.
[10] MacIsaac RJ, Ekinci EI, Premaratne E, Lu ZX, Seah J-m, Li Y, et al. The Chronic Kidney Disease-Epidemiology Collaboration (CKD-EPI) equation does not improve the underestimation of Glomerular Filtration Rate (GFR) in people with diabetes and preserved renal function. BMC Nephrol 2015;16(1):198.
[11] Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, et al. New creatinine- and cystatin C–based equations to estimate GFR without race. N Engl J Med 2021;385(19):1737–49.
[12] Kwok R, Kishore K, Zafari T, Koye DN, Hachem M, de Boer IH, et al. Comparative performance of CKD-EPI equations in people with diabetes: an international pooled analysis of individual participant data. Diabetes Res Clin Pract 2025;112104.
[13] Liu X, Pei X, Li N, Zhang Y, Zhang X, Chen J, et al. Improved glomerular filtration rate estimation by an artificial neural network. PLoS One 2013;8(3):e58242.
[14] Jeong TD, Lee W, Yun YM, Chun S, Song J, Min WK. Development and validation of the Korean version of CKD-EPI equation to estimate glomerular filtration rate. Clin Biochem 2016;49(9):713–9.
[15] Ebert N, Jakob O, Gaedeke J, van der Giet M, Kuhlmann MK, Martus P, et al. Prevalence of reduced kidney function and albuminuria in older adults: the Berlin Initiative Study. Nephrol Dial Transplant 2017;32(6):997–1005.
[16] Group* TDCaCTR. The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-Term Complications in Insulin-Dependent Diabetes Mellitus. New England Journal of Medicine. 1993;329(14):977-86.
[17] Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro 3rd AF, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med 2009;150(9):604–12.
[18] Krogh A. What are artificial neural networks? Nat Biotechnol 2008;26(2):195–7.
[19] Ioffe S. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:150203167. 2015.
[20] Lundberg S. A unified approach to interpreting model predictions. arXiv preprint arXiv:170507874. 2017.
[21] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155–63.
[22] McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947;12(2):153–7.
[23] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307–10.
[24] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024;385:e078378.
[25] Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. modification of diet in renal disease study group. Ann Intern Med 1999;130(6):461–70.
[26] MacIsaac RJ, Ekinci EI, Premaratne E, Lu ZX, Seah J-M, Li Y, et al. The Chronic Kidney Disease-Epidemiology Collaboration (CKD-EPI) equation does not improve the underestimation of Glomerular Filtration Rate (GFR) in people with diabetes and preserved renal function. BMC Nephrol 2015;16:198 -.
[27] Silva LCF, Rocha BM, Escott GM, Porgere IF, Tochetto LA, de Almeida BL, et al. Accuracy evaluation of 2021 chronic kidney disease epidemiology collaboration, full age spectrum and European kidney function consortium equations for estimating glomerular filtration rate in type 2 diabetes mellitus and healthy adults. Clin Chim Acta 2022;534:14–21.
[28] Jiang S, Li Y, Jiao Y, Zhang D, Wang Y, Li W. A back propagation neural network approach to estimate the glomerular filtration rate in an older population. BMC Geriatr 2023;23(1):322.
[29] Liu X, Li N-s, Lv L-s, Huang J-h, Tang H, Chen J-x, et al. A comparison of the performances of an artificial neural network and a regression model for GFR estimation. Am J Kidney Dis 2013;62(6):1109–15.
[30] Chen J, Tang H, Huang H, Lv L, Wang Y, Liu X, et al. Development and validation of new glomerular filtration rate predicting models for chinese patients with type 2 diabetes. J Transl Med 2015;13:317.
[31] Ugrinowitsch C, Fellingham GW, Ricard MD. Limitations of ordinary least squares models in analyzing repeated measures data. Med Sci Sports Exerc 2004;36(12):2144–8.
[32] Rigatti SJ. Random forest. J Insur Med 2017;47(1):31–9.
[33] Kriegeskorte N, Golan T. Neural network models and deep learning. Curr Biol 2019;29(7):R231–6.
[34] de Jong VMT, Moons KGM, Eijkemans MJC, Riley RD, Debray TPA. Developing more generalizable prediction models from pooled studies and large clustered data sets. Stat Med 2021;40(15):3533–59.
[35] Schinkel M, Bennis FC, Boerman AW, Wiersinga WJ, Nanayakkara PWB. Embracing cohort heterogeneity in clinical machine learning development: a step toward generalizable models. Sci Rep 2023;13(1):8363.
[36] Mendenhall E, Kohrt BA, Norris SA, Ndetei D, Prabhakaran D. Non-communicable disease syndemics: poverty, depression, and diabetes among low-income populations. Lancet 2017;389(10072):951–63.
[37] Williams EH, Flint TR, Connell CM, Giglio D, Lee H, Ha T, et al. CamGFR v2: a new model for estimating the glomerular filtration rate from standardized or non-standardized creatinine in patients with cancer. Clin Cancer Res 2021;27(5):1381–90.
[38] Raynaud M, Al-Awadhi S, Juric I, Divard G, Lombardi Y, Basic-Jukic N, et al. Race-free estimated glomerular filtration rate equation in kidney transplant recipients: development and validation study. BMJ 2023;381:e073654.
[39] Pottel H, Björk J, Courbebaisse M, Couzi L, Ebert N, Eriksen BO, et al. Development and validation of a modified full age spectrum creatinine-based equation to estimate glomerular filtration rate. Ann Intern Med 2020;174(2):183–91.