KI REPORTS
KIReports.org
CLINICAL RESEARCH

Global Validation of a Model to Predict Reduced Estimated GFR in People With Type 2 Diabetes Without Diagnosis of CKD

Camilla Sammut-Powell[1], Rose Sisk[1], Estefania Vazquez-Mendez[2], Hardik Vasnawala[2], Susana Goncalves[3], Mark Edge[1] and Rory Cameron[1]

[1]Gendius Limited, Alderley Edge, UK; [2]AstraZeneca, Cambridge, UK; and [3]AstraZeneca, Buenos Aires, Argentina

Introduction: A minimal-resource model for predicting reduced kidney function among people with type 2 diabetes and no diagnosis of chronic kidney disease (CKD) stages 3 to 5 was previously developed in a UK population to pre-screen for undiagnosed CKD. This study aims to evaluate the performance of the model on a global population and assess its adequacy with and without regional adjustment.

Methods: A retrospective observational study was performed using data collected from the iCaReMe global registry (NCT03549754) and the DISCOVER study (NCT02322762 and NCT02226822). Patients were grouped by their World Health Organization classified region. An estimated glomerular filtration rate (eGFR) <60 ml/min per 1.73 m^2 was the marker of reduced kidney function. A regional-intercept recalibration was applied to adjust for regional variation. Discrimination and calibration were evaluated for the UK-developed and recalibrated models.

Results: A total of 14,180 patients (46 countries, 6 regions) were identified with type 2 diabetes, no previous diagnosis of CKD stages 3 to 5, and had a serum creatinine measurement or eGFR recorded. The UK model underestimated risk when applied globally and was deemed inadequate. The model with regional adjustment achieved the target sensitivity (80.5%; 95% confidence interval [CI]: 78.8%–82.3%) and demonstrated a relative improvement of 51.5% (95% CI: 48.1%–55.1%) in the positive predictive value (PPV), compared to a screen-all approach.

Conclusion: The regional-adjusted model demonstrated adequate performance globally. Incorporating the model within practice could help clinicians to risk-stratify and prioritize patients at high risk. This could enable improved efficiency via risk-tailored screening, particularly in lower-middle-income countries (LMICs).

Kidney Int Rep (2024) **9,** 2047–2055; https://doi.org/10.1016/j.ekir.2024.04.005
KEYWORDS: chronic kidney disease; low- and middle-income countries; risk stratification; screening; type 2 diabetes

Type 2 diabetes mellitus is a well-known risk factor for CKD, with over 40% of patients with type 2 diabetes currently presenting with CKD.[1] Therefore, annual CKD screening of people with type 2 diabetes is recommended.[2] However, undiagnosed CKD remains a problem,[3] with particular impact in LMICs.

In LMICs, a lack of testing is a long-standing issue, despite the disproportionate burden of diabetes compared with high-income countries.[4] The need to improve CKD screening capacity and access to CKD screening in LMICs has been articulated[5,6] but solutions are yet to be implemented.

A minimal-resource CKD pre-screening model has been developed to identify which people with type 2 diabetes (without a diagnosis of CKD stages 3 to 5) are likely to exhibit an eGFR <60 ml/min per 1.73 m^2.[2,7] The premise of the model was to pre-screen the population to determine which patients should be prioritized for testing. Despite promising performance during internal validation, the model was developed using UK primary care data and it remains to be externally validated in other geographic regions. This is essential to determine the generalizability of the model before implementation.

AstraZeneca has been building global data sets (DISCOVER and iCaReMe) comprising people with type

Correspondence: Camilla Sammut-Powell, Gendius Limited, Glasshouse, Alderley Park, Alderley Edge SK10 4ZE, UK. E-mail: camilla.sammut@outlook.com

Received 13 August 2023; revised 31 March 2024; accepted 1 April 2024; published online 4 April 2024

2 diabetes. These data sets have collected information to investigate treatments and complications in a globally representative sample of people with type 2 diabetes.[8-11]

The aim of this study was to determine whether the minimal-resource model performs adequately when applied to external global data or whether regional adaptations are required.

METHODS

Design and Data Sources

This was a retrospective, observational study using 2 global data sets collected by AstraZeneca in the DISCOVER study and the iCaReMe global registry.

The DISCOVER study comprises 2 observational studies (NCT02322762 and NCT02226822) on patients with type 2 diabetes who were initiating second-line glucose-lowering therapy.[8] Patients were recruited across 38 countries from 6 regions between December 2014 and June 2016.

The iCaReMe (NCT03549754) global registry had not completed recruitment at the time of analysis. Consequently, an extract from the registry was provided covering the start of recruitment in February 2018 until April 2022. In this extract, 12,606 patients had been recruited across 21 countries from 6 regions.

Ethics

No ethical approval was required; this was secondary use of data with no intervention. Patients consented to the use of their anonymized data at recruitment into the studies, and International Review Board/International Ethics Committee approvals were obtained.

Inclusion and Exclusion Criteria

The inclusion criteria are adults (aged 18 and over) with a diagnosis of type 2 diabetes. The exclusion criteria were: a diagnosis of CKD stages 3 to 5, an invalid (outside clinically plausible range) serum creatinine measurement, missing eGFR, or eGFR not calculable due to missing age or biological sex assigned at birth.

Prediction Model

We implemented the minimal-resource CKD pre-screening model developed by Gendius.[7] The predictors were age, sex assigned at birth, duration of type 2 diabetes, body mass index, systolic blood pressure, and diastolic blood pressure. The cutoff selected in the model development was used to categorize the predicted outcome as "high risk" or not. This model is referred to as the "UK model."

Primary Outcome

The eGFR was calculated from the serum creatinine measurements provided in the data, using the 2009 CKD-Epidemiology Collaboration equation.[12] The binary outcome then corresponded to whether the eGFR was <60 ml/min per 1.73 m^2 or not, consistent with the definition in the minimal-resource CKD pre-screening model.

Secondary Outcome

Patients with type 2 diabetes and an eGFR between 25 and 75 ml/min per 1.73 m^2 may be eligible for initiation of a sodium-glucose cotransporter 2 inhibitor (NICE Guideline TA775).[13] Therefore, our secondary outcome corresponds to whether the eGFR was 75 ml/min per 1.73 m^2 or less, that is, a composite indicator of follow-up action required to diagnose and/or review for initiation of preventative therapy.

Missing Data

When calculating the eGFR from a serum creatinine measurement, if a patient was missing their ethnicity, they were assumed not to be black for the CKD-EPI equation. Missing predictor information was imputed using single regression imputation, by data set, using only the variables within the model (age, sex assigned at birth, body mass index, duration of diabetes, systolic blood pressure, diastolic blood pressure) and the patient's country.

Statistical Methods

Summary statistics were evaluated to compare the populations across regions and against the original model development cohort. Continuous variables that were assumed to be normally distributed were summarized using the mean and SD; skewed continuous data were summarized using the median and interquartile range. Categorical variables were summarized as counts and percentages. Chi-square tests were performed to assess the statistical difference between distributions for categorical variables.

Model Application and Updating

The UK model was applied to the data and the distribution of the predicted risks were visualized by data set and region. A regional-intercept recalibration of the model was applied to estimate regional adjustments of the model using the iCaReMe registry, that is, a new intercept was estimated such that the recalibrated linear predictor for patient i in region R is

$$LP_{i,R}^{(recalibrated)} = \beta_R + LP_{i,R}^{UK}$$

where $LP_{i,R}^{(recalibrated)}$ is the recalibrated linear predictor for patient i in region R, β_R is the intercept adjustment for region R, and $LP_{i,R}^{UK}$ is the linear predictor for patient i in region R using the UK model. An external validation of the regionally recalibrated models was then performed in the DISCOVER data. Because intercept recalibration is equivalent to a change

in cutoff for discrimination, the cutoff was not altered by region to avoid overfitting and enable the validation the cutoff as a predetermined threshold.

Performance Metrics

The sensitivity and relative improvement in the PPV at the prespecified cutoff (identified during the model development) compared to a screen-all approach were the target performance metrics. The relative improvement in the PPV is defined as:

$$PPV_{rel\ imp}(\theta) = \frac{PPV(\theta)}{Prevalence} - 1$$

where θ is the cutoff for categorizing the outcome as "high risk" or not, that is, if the predicted probability is $\geq \theta$, the patient is predicted to be high risk, and $PPV(\theta)$ is the proportion of high risk patients (defined using θ) that had an eGFR <60 ml/min per 1.73 m^2. The screen-all approach is equivalent to $PPV(0) = Prevalence$, thus, $PPV_{rel\ imp}(0) = 0$. The estimates and their bootstrapped 95% CIs were evaluated overall and by region. The original target profile of the model was an increased PPV over a screen-all approach and a sensitivity of 80% (95% CI: 70%–100%) in alignment with the acceptable criteria proposed by the Medicines and Healthcare products Regulatory Agency for a comparable pre-screening tool.[14] The probability cutoff identified in the model development was selected at a level that achieves the target profile. Bayesian probabilities and confidence intervals of the sensitivity (α) were evaluated by region, where the posterior distribution of the sensitivity was defined as

$$\alpha \sim Beta(TP + 1, FN + 1)$$

where TP is the number of true positives and FN is the number of false negatives identified for a given model. The posterior distribution was used to determine the probability that the sensitivity was at least 80% (corresponding to the target performance point estimate). Similarly, the probability that the sensitivity was >70% was evaluated, corresponding to the chance of acceptable model performance with respect to the lower threshold of sensitivity in the target product profile.

Calibration and discrimination were assessed using calibration plots and the C-statistic, respectively. Linear predictors from the UK model and the recalibrated model were calculated and visualized. The linear predictor represents the estimated risk score on the log-odds scale, before transformation to a probability between 0 and 1.

A subgroup analysis was performed based upon World Health Organization region, determining the variation between the regional performances.

Sample Size Requirements

The minimum sample size for external validation of the minimal-resource model was calculated as 3306 patients.[15] For evaluating the sensitivity, the literature suggests that a sample size of 160 positive cases (i.e., patients with an eGFR <60 ml/min per 1.73 m^2) is sufficient for a target product profile of 80% sensitivity (95% CI: 70%–100%).[16]

All analyses were performed in R version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

There were 5612 and 8568 patients identified as having type 2 diabetes, a valid serum creatinine measurement or eGFR recorded, and no previous diagnosis of CKD stages 3 to 5 in the iCaReMe and DISCOVER data sets, respectively (Supplementary Figure S1, Table 1, and Supplementary Table S1). In total, 1940 (13.7%) had an eGFR <60 ml/min per 1.73 m^2.

Although both data sets covered the same regions, the majority of the patients in the iCaReMe population were from South-East Asia, whereas the DISCOVER population was more evenly distributed across the regions (Table 1). Regardless, age, body mass index, and blood pressure measurements were clinically similar between the data sets. The iCaReMe patients had on average a longer duration of diabetes (iCaReMe: 8 years vs. DISCOVER: 4.1 years; $P < 0.001$) and a higher proportion of patients with an eGFR <60 ml/min per 1.73 m^2 (iCaReMe: 19.4% vs. DISCOVER: 10.0%; $P < 0.001$). Despite this, the prevalence of CKD stage 5 in DISCOVER was double that in iCaReMe (DISCOVER: 1.8% vs. iCaReMe: 0.9%) and the distribution across stages 3 to 5 was significantly different ($P < 0.001$). Further, regional variation between the distributions of eGFR stage was observed within both data sets (Supplementary Table S2).

Overall Performance

The discrimination remained adequate in both data sets when applying the original model; the C-statistic was estimated as 0.737 (95% CI: 0.725–0.748) (iCaReMe: 0.744. 95% CI: 0.729–0.757; DISCOVER: 0.720, 95% CI: 0.705–0.737). However, the cutoff identified in the original model led to an overall sensitivity of 0.405 (95% CI: 0.380–0.425) which was lower than the prespecified acceptable threshold of 80%. This was consistent between the data sets (iCaReMe: 0.425, 95% CI: 0.398–0.451; DISCOVER: 0.378, 95% CI: 0.352–0.415) (Table 2).

When the UK model was recalibrated, the sensitivity at the cutoff increased to 0.843 (95% CI: 0.822–0.860) within the iCaReMe data. A relative improvement of 46.8% (95% CI: 42.6%–50.8%) in the PPV, compared to a screen-all approach (prevalence: 0.194; 95% CI: 0.182–0.207), was observed at the cutoff (PPV: 0.284;

Table 1. Patient baseline demographics

	Development	External validation		
Patient characteristics	**UK primary care (n = 9297)**	**iCaReMe (n = 5612)**	**DISCOVER (n = 8568)**	**Overall (N = 14,180)**
Region, n (%)				
Africa	0 (0.0)	134 (2.4)	309 (3.6)	443 (3.1)
Americas	0 (0.0)	497 (8.9)	933 (10.9)	1430 (10.1)
Eastern Mediterranean	0 (0.0)	293 (5.2)	1249 (14.6)	1542 (10.9)
Europe	9297 (100.0)	994 (17.7)	2135 (24.9)	3129 (22.1)
South-East Asia	0 (0.0)	2889 (51.5)	1317 (15.4)	4206 (29.7)
Western Pacific	0 (0.0)	805 (14.3)	2625 (30.6)	3430 (24.2)
Sex assigned at birth, n (%)				
Female	3919 (42.2)	2571 (45.8)	3680 (43)	6251 (44.1)
Male	5378 (57.8)	3041 (54.2)	4888 (57)	7929 (55.9)
Missing	0 (0)	0 (0)	0 (0)	0 (0)
Age (yr)				
Mean (SD)	65.5 (13.1)	57.3 (12.2)	57.4 (11.9)	57.4 (12.0)
Missing, n (%)	0 (0)	5 (0.1)	0 (0)	5 (0)
Body mass index (kg/m^2)				
Median (LQ–UQ)	30.0 (26.5–34.5)	28.5 (25.5–32.1)	28.2 (25.0–32.4)	28.3 (25.2–32.3)
Missing, n (%)	435 (4.7)	582 (10.4)	478 (5.6)	1060 (7.5)
Duration of type 2 diabetes (yr)				
Median (LQ–UQ)	6.6 (2.9–11.4)	8.0 (3.1–14.1)	4.1 (1.8–7.9)	5 (2.0–10.0)
Missing, n (%)	0 (0)	1240 (22.1)	2 (0)	1242 (8.8)
Systolic blood pressure (mm Hg)				
Mean (SD)	133.2 (15.1)	129.1 (17.7)	132.5 (16.6)	131.2 (17.1)
Missing, n (%)	1847 (19.9)	587 (10.5)	291 (3.4)	878 (6.2)
Diastolic Blood Pressure (mm Hg)				
Mean (SD)	75.9 (9.9)	77.9 (10.5)	79.9 (10.0)	79.1 (10.2)
Missing, n (%)	1849 (19.9)	593 (10.6)	297 (3.5)	890 (6.3)
CKD stage from eGFR, n (%)				
Stage G0–1 (eGFR $\geq$ 90)	3576 (38.5)	2307 (41.1)	4597 (53.7)	6904 (48.7)
Stage G2 (60 $\leq$ eGFR $<$ 90)	4316 (46.4)	2219 (39.5)	3117 (36.4)	5336 (37.6)
Stage G3a (45 $\leq$ eGFR $<$60)	917 (9.9)	593 (10.6)	522 (6.1)	1115 (7.9)
Stage G3b (30 $\leq$ eGFR $<$ 45)	372 (4.0)	316 (5.6)	147 (1.7)	463 (3.3)
Stage G4 (15 $\leq$ eGFR $<$ 30)	101 (1.1)	127 (2.3)	31 (0.4)	158 (1.1)
Stage G5 (eGFR $<$ 15)	15 (0.2)	50 (0.9)	154 (1.8)	204 (1.4)

CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; LQ, lower quartile, UQ, upper quartile.

95% CI: 0.267–0.302). Similarly, when this recalibrated model was applied to the DISCOVER data, the sensitivity at the cutoff increased to 0.758 (95% CI: 0.732–0.785) compared to the UK model, and the relative improvement in the PPV was 50.5% (95% CI: 44.3%–55.6%). On average, 41.3% of the high risk patients were considered actionable (iCaReMe: 51.0%, DISCOVER: 34.2%; Table 2).

The regional-intercept recalibration improved the calibration in both data sets for predicted probabilities of an eGFR <60 ml/min per 1.73 m^2 that were below 0.1 (Figure 1). All regional-intercepts were positive, indicating that the UK model underestimates the risk when applied to global data. Within the iCaReMe data, the smallest adjustment was in the Americas (median risk adjustment: 0.048, interquartile range: 0.025–0.100; 92/497 changed risk group) and the largest was in the Eastern Mediterranean region (median risk adjustment: 0.154, interquartile range: 0.074–0.260; 137/293 changed risk group; Supplementary Figure S2; Table 3).

Regional Variation

Differences were observed across the regions and between data sets (Supplementary Tables S3 and S4); the prevalence of patients with an eGFR <60 ml/min per 1.73 m^2 was lowest in the Americas (0.171; 95% CI: 0.135–0.201) and highest in the Eastern Mediterranean (0.263; 95% CI: 0.208–0.324) in iCaReMe. However, the regions with the higher prevalences in iCaReMe had the lower prevalences within DISCOVER.

The C-statistic was acceptable across all regions within iCaReMe (range: 0.695–0.783) but decreased to 0.646 in the Eastern Mediterranean region within DISCOVER (Supplementary Table S4). Despite this, the sensitivity remained satisfactory at the cutoff with a relative increase of 31.9% in the PPV. The model performed best in the Americas within iCaReMe (sensitivity: 0.824, 95% CI: 0.747–0.912; relative improvement in PPV: 72.7%, 95% CI: 56.1%–94.0%) but the sensitivity decreased in this region within DISCOVER (sensitivity: 0.674, 95% CI: 0.610–0.756).

Table 2. Model performance metrics (bootstrapped 95% CIs), where 'overall' is combines the iCaReMe and DISCOVER datasets

Dataset	Number of patients	Prevalence	PPV at cutoff	Relative improvement in PPV at cutoff	Sensitivity at cutoff	Specificity at cutoff	C-statistic	Proportion high risk	Proportion actionable within high risk
Development bootstrap internal validation									
UK primary care	9297	0.151 (0.144–0.157)	0.282 (0.271–0.294)	0.863 (0.787–0.972)	0.823 (0.806–0.841)	0.627 (0.601–0.656)	0.803 (0.791–0.815)	Not reported	Not reported
External validation									
UK model									
iCaReMe	5612	0.194 (0.182–0.207)	0.421 (0.394–0.450)	1.174 (1.040–1.288)	0.425 (0.398–0.451)	0.859 (0.851–0.869)	0.744 (0.729–0.757)	0.196 (0.184–0.205)	0.643 (0.614–0.670)
DISCOVER	8568	0.100 (0.093–0.106)	0.247 (0.228–0.270)	1.479 (1.304–1.713)	0.378 (0.352–0.415)	0.872 (0.865–0.880)	0.720 (0.705–0.737)	0.153 (0.145–0.159)	0.483 (0.464–0.510)
Overall	14,180	0.137 (0.132–0.142)	0.326 (0.309–0.345)	1.386 (1.269–1.487)	0.405 (0.380–0.425)	0.868 (0.862–0.873)	0.737 (0.725–0.748)	0.170 (0.164–0.175)	0.556 (0.538–0.574)
Recalibrated									
iCaReMe	5612	As above	0.284 (0.267–0.302)	0.468 (0.426–0.508)	0.843 (0.822–0.860)	0.490 (0.477–0.506)	0.751 (0.736–0.764)	0.575 (0.562–0.585)	0.510 (0.491–0.528)
DISCOVER	8568	As above	0.151 (0.140–0.159)	0.505 (0.443–0.556)	0.758 (0.732–0.785)	0.524 (0.515–0.534)	0.706 (0.689–0.721)	0.504 (0.495–0.513)	0.342 (0.328–0.357)
Overall	14,180	As above	0.207 (0.199–0.217)	0.515 (0.481–0.551)	0.805 (0.788–0.823)	0.512 (0.504–0.521)	0.733 (0.722–0.744)	0.532 (0.524–0.539)	0.413 (0.403–0.424)

CI, confidence interval; PPV, positive predictive value.

The lowest acceptable level of sensitivity (70%) was within the 95% Bayesian CIs of sensitivity across all regions (Supplementary Table S5). Africa was the only region that did not meet the sample size requirement (at least 160 positive cases). All remaining regions had a high probability that the sensitivity was 70% or higher (regional probability range: 0.841–1.000). For the European and Western Pacific regions, the probability that the sensitivity was 80% or higher remained high (1.000 and 0.940, respectively).

DISCUSSION

Within this study, we have utilized large global data sets to demonstrate that the minimal-resource model adequately identifies patients at high risk of currently experiencing reduced kidney function (eGFR <60 ml/min per 1.73 m^2), globally, when regional adjustments are applied. The minimal-resource model reduced the screening population by over 50% to a subgroup of patients who were likely to be experiencing reduced kidney function. This subgroup contained over 80% of those with an eGFR <60 ml/min per 1.73 m^2 and therefore corresponded to a significant increase in the efficiency of eGFR screening to detect patients with an eGFR <60 ml/min per 1.73 m^2.

In addition, this validation has demonstrated the feasibility of applying the model globally into practice; overall, the level of missing data across the input variables was low. The most poorly recorded variable was the duration of diabetes; however, most missing cases were in the iCaReMe data. This is likely due to the DISCOVER studies capturing data only on patients with type 2 diabetes, and consequently this may have been perceived to be of greater importance to collect across all patients, whereas iCaReMe studies patients with type 2 diabetes, hypertension, heart failure or CKD.

A key strength of this study was the use of large, global data sets; the model performance was evaluated in patients from 46 countries, with significant representation from LMICs. The data were collected during routine care and are therefore representative of data collected within practice. The model performance across the data sets was similar, indicating that these results are generalizable. In total, 5 of 6 regions were adequately represented to assess the performance metrics; only Africa did not have a sufficient sample size to reliably estimate the performance.

Differences between the patient characteristics from iCaReMe and DISCOVER can be explained by several factors. The DISCOVER recruitment targeted patients with type 2 diabetes moving from a first-line to second-line therapy, likely contributing to the difference observed in the duration of diabetes between the data

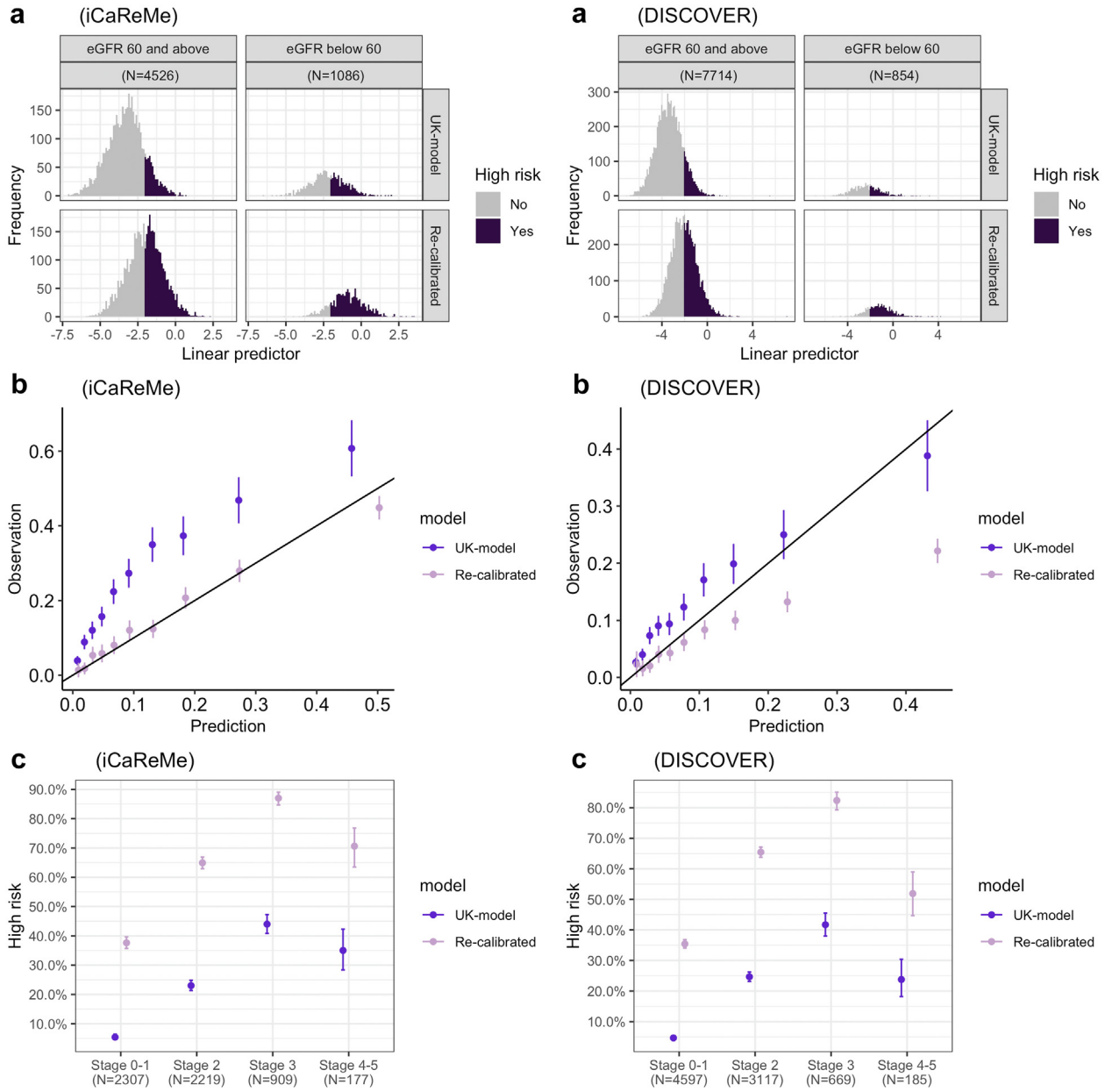

Figure 1. Performance plots for the (left:) iCaReMe and (right:) DISCOVER data, demonstrating (a) the distribution of the linear predictor, indicating the population that were high risk (dark purple shading); (b) the calibration of the UK model (purple) and recalibrated model (pink); and (c) the proportion of patients within G-staged CKD groups that were identified as being high risk evaluated under the UK model (purple) and the recalibrated model (pink). CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate.

sets. In addition, some of the data provided to the DISCOVER study was extracted from patients' electronic health record and predated their recruitment. Lastly, the dates of recruitment vary between the data sets; the data captured within iCaReMe coincides with the COVID-19 pandemic which may have biased patient recruitment.

The minimal-resource model focuses only on predicting that a patient's current eGFR is <60 ml/min per 1.73 m^2; however, an increased urine albumin-to-creatinine ratio can also be an early indicator of CKD. Using our model does not guarantee accuracy of predicting those with an increased urine albumin-to-creatinine ratio as high risk, but urine albumin-to-creatinine ratio measurement and recording have been reported to be poor;[9] thus, an eGFR-focused outcome is more aligned with current CKD screening practices. Similarly, for a CKD diagnosis, a patient is required to a have evidence of a sustained reduction in kidney function; therefore, patients who are high risk and have a resulting eGFR <60 ml/min per 1.73 m^2 may not be diagnosed with CKD if their eGFR recovers.

The CKD-Epidemiology Collaboration equation used to estimate the eGFR in the modeling may not be the

Table 3. Recalibration adjustments to the estimated probability of an eGFR less than 60 ml/min per 1.73 m^2 and the proportion patients that changed risk category

			Risk adjustment			Change in prediction category		
Region	Data	Number	Median	Lower quartile	Upper quartile	Number changed to high risk	Number changed to not high risk	Proportion changed (%)
Africa	iCaReMe	134	0.096	0.044	0.166	49	0	36.6
	DISCOVER	309	0.070	0.038	0.132	108	0	35.0
Americas	iCaReMe	497	0.048	0.025	0.100	92	0	18.5
	DISCOVER	933	0.036	0.018	0.066	160	0	17.1
Eastern Mediterranean	iCaReMe	293	0.155	0.074	0.259	138	0	47.1
	DISCOVER	1249	0.095	0.052	0.166	551	0	44.1
Europe	iCaReMe	994	0.086	0.045	0.150	319	0	32.1
	DISCOVER	2135	0.076	0.041	0.132	686	0	32.1
South-East Asia	iCaReMe	2889	0.099	0.046	0.178	1178	0	40.8
	DISCOVER	1317	0.059	0.033	0.113	426	0	32.3
Western Pacific	iCaReMe	805	0.120	0.056	0.202	351	0	43.6
	DISCOVER	2625	0.107	0.052	0.201	1078	0	41.1

eGFR, estimated glomerular filtration rate.

same as the eGFR calculation used in clinic, because a new version of the CKD-Epidemiology Collaboration formula has been released[17] and alternative equations such as the Modification of Diet in Renal Disease formula exist.[18] However, we do not expect that variations in calculations would significantly affect the performance due to the categorization of the outcome.

Existing models for CKD risk stratification within patients with type 2 diabetes have focused on risk of future development of CKD.[19-24] Although this is important to assess, such models often rely on the availability of test results obtained via CKD screening. Consequently, when screening is not routinely performed, these risk models remain unusable. Therefore, when patients are determined to be high risk by our model but do not have an eGFR <60 ml/min per 1.73 m^2, they can still benefit from their eGFR result being fed into other models, which may highlight their future risk of developing CKD.

Lifestyle modifications and pharmaceutical therapies are common interventions for managing CKD and risk of complications. Recent clinical trials studying the efficacy of sodium-glucose cotransporter 2 inhibitors have demonstrated that CKD progression can be slowed in its early stages and CKD complications can be prevented.[25-28] Therefore, early identification of CKD within primary care is important to initiate these therapies when they can have the greatest benefit. Consequently, we believe that implementing our model within primary care could support clinicians and nurses to provide more timely and targeted interventions, particularly when they may not have specialist knowledge of CKD. In LMICS where sodium-glucose cotransporter 2 inhibitors may not be affordable, early identification can still support CKD management through emphasizing the need for blood pressure control to prevent poor outcomes.

Prioritization tools are not a new concept; during the COVID-19 pandemic, the provision of routine services in UK primary care was severely disrupted.[29] A risk stratification tool was proposed by National Health Service England to prioritize patients for delivering diabetes care,[30] but it did not encompass the patient's risk of undiagnosed CKD. When CKD screening resource is limited, the minimal-resource model offers a practical and globally equitable solution to identify a subgroup of patients with type 2 diabetes who should be prioritized for screening. Where data from electronic health records is available, recent data can be extracted and fed into the model to risk-stratify the patients at little to no burden. Integrating these risk scores back into health systems could trigger proactive review and management of patients that are at high risk, ultimately improving the quality of care and prognosis for patients[31] and supporting more efficient use of health care resources.

In countries where CKD screening is performed regularly, as per guidelines, CKD diagnosis remains imperfect.[32,33] For example, an audit of CKD diagnosis coding performed across 1039 general practitioner practices in the UK indicated that 30% of patients that qualified for a CKD stages 3 to 5 diagnosis did not have a corresponding diagnosis code within their record.[32] We expect that our model would identify such patients as high risk and highlighting this to clinicians could facilitate targeted clinical review and improved CKD identification.[34] Given that a patient's awareness of having CKD is associated with having a diagnosis,[33] improvements in diagnosis could empower earlier lifestyle intervention.

Although this study has highlighted the potential opportunity of implementing the model within practice, further investigation is needed to determine its impact on the identification and management of CKD, and patient outcomes.

CONCLUSION

We have demonstrated that a model using only easily collectible information can be used globally to effectively risk-stratify patients with type 2 diabetes for reduced kidney function. Applying the model offers a simple way for clinicians to understand their patient population and help them to prioritize patients who are most in need. This could enable improved efficiency in screening programs via targeted screening, particularly in LMICs, and act as a reminder to clinicians of the importance of CKD screening and diagnosis in patients with type 2 diabetes.

DISCLOSURE

SG, HV, and EVM are employees of AstraZeneca. EVM and SG are shareholders of AstraZeneca. CSP, RS, ME and RC are employees of Gendius Limited.

ACKNOWLEDGMENTS

Preliminary work has previously been presented at EASD 2022 and ASN Kidney Week 2022. This study was fully funded by AstraZeneca and the data analyzed were collected from AstraZeneca funded studies.

DATA AVAILABILITY STATEMENT

The data sets analyzed for the study are owned by AstraZeneca. Patients did not give written consent for their data to be shared publicly; thus, the data is not freely available. CSP is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

AUTHOR CONTRIBUTIONS

RC, ME, SG, and HV conceived the study, and all the authors were involved in the design of the study. EVM, SG, and HV provided the data for the study. CSP and RS developed the statistical analysis plan and performed the analyses. CSP drafted the manuscript which was reviewed by all coauthors.

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Figure S1. Patient selection.

Figure S2. Distribution of linear predictor by data set, World Health Organization region, outcome, and model.

Table S1. Patient numbers by country.

Table S2. Staging of estimated glomerular filtration rates among selected patients within the iCaReMe and DISCOVER data sets.

Table S3. Demographics and outcome by region and data set.

Table S4. Model performance metrics by World Health Organization region.

Table S5. Bayesian confidence intervals and probabilities for sensitivity by region.

STROBE Statement.

REFERENCES

1. Wu B, Bell K, Stanford A, et al. Understanding CKD among patients with T2DM: prevalence, temporal trends, and treatment patterns—NHANES 2007–2012. *BMJ Open Diabetes Res Care*. 2016;4:e000154. https://doi.org/10.1136/bmjdrc-2015-000154
2. American Diabetes Association. Standards of medical care in diabetes—2022 abridged for primary care providers. *Clin Diabetes*. 2022;40:10–38. https://doi.org/10.2337/cd22-as01
3. Middleton RJ, Foley RN, Hegarty J, et al. The unrecognized prevalence of chronic kidney disease in diabetes. *Nephrol Dial Transplant*. 2006;21:88–92. https://doi.org/10.1093/ndt/gfi163
4. Lam AA, Lepe A, Wild SH, Jackson C. Diabetes comorbidities in low- and middle-income countries: an umbrella review. *J Glob Health*. 2021;11:04040. https://doi.org/10.7189/jogh.11.04040
5. George C, Echouffo-Tcheugui JB, Jaar BG, Okpechi IG, Kengne AP. The need for screening, early diagnosis, and prediction of chronic kidney disease in people with diabetes in low- and middle-income countries—a review of the current literature. *BMC Med*. 2022;20:247. https://doi.org/10.1186/s12916-022-02438-6
6. Mbanya JC, Aschner P, Gagliardino JJ, et al. Screening, prevalence, treatment and control of kidney disease in patients with type 1 and type 2 diabetes in low-to-middle-income countries (2005–2017): the International Diabetes Management Practices Study (IDMPS). *Diabetologia*. 2021;64:1246–1255. https://doi.org/10.1007/s00125-021-05406-6
7. Sammut-Powell C, Sisk R, Budd J, Patel N, Edge M, Cameron R. Development of minimal resource pre-screening tools for chronic kidney disease in people with type 2 diabetes. *Future Healthc J*. 2022;9:305–309. https://doi.org/10.7861/fhj.2022-0020
8. Ji L, Bonnet F, Charbonnel B, et al. Towards an improved global understanding of treatment and outcomes in people with type 2 diabetes: rationale and methods of the DISCover observational study program. *J Diabetes Complications*. 2017;31:1188–1196. https://doi.org/10.1016/j.jdiacomp.2017.03.011
9. Khunti K, Chen H, Cid-Ruzafa J, et al. Glycaemic control in patients with type 2 diabetes initiating second-line therapy: results from the global DISCover study programme. *Diabetes Obes Metab*. 2020;22:66–78. https://doi.org/10.1111/dom.13866
10. Kosiborod M, Gomes MB, Nicolucci A, et al. Vascular complications in patients with type 2 diabetes: prevalence and associated factors in 38 countries (the DISCover study

program). *Cardiovasc Diabetol*. 2018;17:150. https://doi.org/10.1186/s12933-018-0787-8

11. Sundström J, Bodegard J, Bollmann A, et al. Prevalence, outcomes, and cost of chronic kidney disease in a contemporary population of 2·4 million patients from 11 countries: the CaReMe CKD study. *Lancet Reg Health Eur*. 2022;20:100438. https://doi.org/10.1016/j.lanepe.2022.100438
12. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150:604–612.
13. National Institute for Health and Care Excellence. Dapagliflozin for treating chronic kidney disease. Accessed March 8, 2023. https://www.nice.org.uk/guidance/ta775/chapter/1-Recommendations
14. GOV.UK. Target product profile: point of care SARS-CoV-2 detection tests. Accessed February 16, 2023. https://www.gov.uk/government/publications/how-tests-and-testing-kits-for-coronavirus-covid-19-work/target-product-profile-point-of-care-sars-cov-2-detection-tests#target-product-profile-point-of-care-sars-cov-2-detection-tests
15. Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med*. 2021;40:4230–4251. https://doi.org/10.1002/sim.9025
16. Sammut-Powell C, Reynard C, Allen J, et al. Examining the effect of evaluation sample size on the sensitivity and specificity of COVID-19 diagnostic tests in practice: a simulation study. *Diagn Progn Res*. 2022;6:12. https://doi.org/10.1186/s41512-021-00116-4
17. 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;385:1737–1749. https://doi.org/10.1056/NEJMoa2102953
18. Levey AS, Coresh J, Greene T, et al. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Ann Intern Med*. 2006;145:247–254. https://doi.org/10.7326/0003-4819-145-4-200608150-00004
19. Dunkler D, Gao P, Lee SF, et al. Risk prediction for early CKD in type 2 diabetes. *Clin J Am Soc Nephrol*. 2015;10:1371–1379. https://doi.org/10.2215/CJN.10321014
20. Allen A, Iqbal Z, Green-Saxena A, et al. Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus. *BMJ Open Diabetes Res Care*. 2022;10:e002560. https://doi.org/10.1136/bmjdrc-2021-002560
21. Nelson RG, Grams ME, Ballew SH, et al. Development of risk prediction equations for incident chronic kidney disease. *JAMA*. 2019;322:2104-2104. https://doi.org/10.1001/JAMA.2019.17379
22. Low S, Lim SC, Zhang X, et al. Development and validation of a predictive model for Chronic Kidney Disease progression in type 2 diabetes mellitus based on a 13-year study in Singapore. *Diabetes Res Clin Pract*. 2017;123:49–54. https://doi.org/10.1016/J.DIABRES.2016.11.008
23. Basu S, Sussman JB, Berkowitz SA, Hayward RA, Yudkin JS. Development and validation of Risk Equations for Complications of type 2 Diabetes (RECODe) using individual participant data from randomised trials. *Lancet Diabetes Endocrinol*. 2017;5:788–798. https://doi.org/10.1016/S2213-8587(17)30221-8
24. Lin CC, Niu MJ, Li CI, et al. Development and validation of a risk prediction model for chronic kidney disease among individuals with type 2 diabetes. *Sci Rep*. 2022;123AD:4794-4794. https://doi.org/10.1038/s41598-022-08284-z
25. Dagogo-Jack S, Cannon CP, Cherney DZI, et al. Cardiorenal outcomes with ertugliflozin assessed according to baseline glucose-lowering agent: an analysis from VERTIS CV. *Diabetes Obes Metab*. 2022;24:1245–1254. https://doi.org/10.1111/dom.14691
26. Perkovic V, Jardine MJ, Neal B, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *N Engl J Med*. 2019;380:2295–2306. https://doi.org/10.1056/NEJMoa1811744
27. Heerspink HJL, Stefánsson BV, Correa-Rotter R, et al. Dapagliflozin in patients with chronic kidney disease. *N Engl J Med*. 2020;383:1436–1446. https://doi.org/10.1056/NEJMoa2024816
28. Yau K, Dharia A, Alrowiyti I, Cherney DZI. Prescribing SGLT2 inhibitors in patients with CKD: expanding indications and practical considerations. *Kidney Int Rep*. 2022;7:1463–1476. https://doi.org/10.1016/j.ekir.2022.04.094
29. Carr MJ, Wright AK, Leelarathna L, et al. Impact of COVID-19 restrictions on diabetes health checks and prescribing for people with type 2 diabetes: a UK-wide cohort study involving 618 161 people in primary care. *BMJ Qual Saf*. 2022;31:503–514. https://doi.org/10.1136/bmjqs-2021-013613
30. Bakhai DC. Delivering diabetes care during the COVID-19 pandemic-the 'New Normal' guidance for general practice. NHS England and NHS Improvement. Published June 12, 2020. Accessed February 24, 2023. https://www.diabetes.org.uk/resources-s3/public/2020-06/Delivering%20Diabetes%20Care%20during%20the%20COVID-19%20Pandemic%20220620.pdf
31. Hull SA, Rajabzadeh V, Thomas N, et al. Improving coding and primary care management for patients with chronic kidney disease: an observational controlled study in East London. *Br J Gen Pract*. 2019;69:e454–e461. https://doi.org/10.3399/bjgp19X704105
32. Kim LG, Cleary F, Wheeler DC, et al. How do primary care doctors in England and Wales code and manage people with chronic kidney disease? Results from the national chronic kidney disease audit. *Nephrol Dial Transplant*. 2018;33:1373–1379. https://doi.org/10.1093/ndt/gfx280
33. Szczech LA, Stewart RC, Su HL, et al. Primary care detection of chronic kidney disease in adults with type-2 diabetes: the ADD-CKD study (awareness, detection and drug therapy in type 2 diabetes and chronic kidney disease). *PLoS One*. 2014;9:e110535. https://doi.org/10.1371/journal.pone.0110535
34. Litvin CB, Hyer JM, Ornstein SM. Use of clinical decision support to improve primary care identification and management of chronic kidney disease (CKD). *J Am Board Fam Med*. 2016;29:604–612. https://doi.org/10.3122/jabfm.2016.05.160020