ORIGINAL RESEARCH

Development and Internal Validation of a Discrete Event Simulation Model of Diabetic Kidney Disease Using CREDENCE Trial Data

Michael Willis · Christian Asseburg · April Slee · Andreas Nilsson · Cheryl Neslusan

ABSTRACT

Introduction: The Canagliflozin and Renal Endpoints in Diabetes with Established Nephropathy Clinical Evaluation (CREDENCE) study showed that compared with placebo, canagliflozin 100 mg significantly reduced the risk of major cardiovascular events and adverse renal outcomes in patients with diabetic kidney disease (DKD). We developed a simulation model that can be used to estimate the long-term health and economic consequences of DKD treatment interventions for patients matching the CREDENCE study population.

M. Willis (✉) · A. Nilsson
The Swedish Institute for Health Economics, Lund, Sweden
e-mail: mw@ihe.se

C. Asseburg
ESiOR Oy, Kuopio, Finland

A. Slee
Axio Research Inc, Seattle, WA, USA

C. Neslusan
Janssen Scientific Affairs, LLC, Titusville, NJ, USA

Methods: The CREDENCE Economic Model of DKD (CREDEM-DKD) was developed using patient-level data from CREDENCE (which recruited patients with estimated glomerular filtration rate 30 to < 90 mL/min/1.73 m², urinary albumin to creatinine ratio > 300–5000 mg/g, and taking the maximum tolerated dose of a renin–angiotensin–aldosterone system inhibitor). Risk prediction equations were fit for start of maintenance dialysis, doubling of serum creatinine, hospitalization for heart failure, nonfatal myocardial infarction, nonfatal stroke, and all-cause mortality. A micro-simulation model was constructed using these risk equations combined with user-definable kidney transplant event risks. Internal validation was performed by loading the model to replicate the CREDENCE study and comparing predictions with trial Kaplan–Meier estimate curves. External validation was performed by loading the model to replicate a subgroup of the CANagliflozin cardioVascular Assessment Study (CANVAS) Program with patient characteristics that would have qualified for inclusion in CREDENCE.

Results: Risk prediction equations generally fit well and exhibited good concordance, especially for the placebo arm. In the canagliflozin arm, modest underprediction was observed for myocardial infarction, along with overprediction of dialysis, doubling of serum creatinine, and all-cause mortality. Discrimination was strong (0.85) for the renal outcomes, but weaker

for the macrovascular outcomes and all-cause mortality (0.60–0.68). The model performed well in internal and external validation exercises.

Conclusion: CREDEM-DKD is an important new tool in the evaluation of treatment interventions in the DKD population.

Trial Registration: ClinicalTrials.gov identifier, NCT02065791.

Keywords: Canagliflozin; Diabetic kidney disease; Economic simulation model; Renal; Risk prediction; Type 2 diabetes mellitus

---

Key Summary Points

Why carry out this study?

Currently available economic models are not suitable for predicting outcomes with treatments with cardio and renal protection like the sodium glucose co-transporter 2 (SGLT2) inhibitor class for patients with type 2 diabetes and diabetic kidney disease (DKD).

This study used patient-level data from the CREDENCE trial of canagliflozin to develop an economic simulation model suitable for estimating the health and economic consequences of DKD treatment interventions for patients matching the CREDENCE patient population.

What was learned from the study?

The CREDENCE Economic Model of DKD (CREDEM-DKD) was designed, constructed, populated with patient-level data, and validated in accordance with current guidelines.

Risk prediction equations for renal and cardiovascular events were fit using patient-level data from the CREDENCE trial, and an economic micro-simulation model was constructed to evaluate the impact of long-term treatment in individuals with type 2 diabetes and DKD.

CREDEM-DKD is an important new tool in the evaluation of treatment interventions in the DKD population.

DIGITAL FEATURES

This article is published with digital features to facilitate understanding of the article. To view digital features for this article go to https://doi.org/10.6084/m9.figshare.12888458.

INTRODUCTION

Chronic kidney disease (CKD) is characterized by gradual and permanent loss of kidney function and the most common cause is diabetes [1]. CKD affects some 750 million individuals worldwide [2], including 40 million individuals in the USA alone [3]. When CKD is unchecked, dialysis or kidney transplant may ultimately be required to maintain life. CKD is also an important risk modifier for cardiovascular (CV) disease and patients with estimated glomerular filtration rate (eGFR) < 15 mL/min/1.73 m² have almost triple the risk [4]. The resulting economic burden of CKD is substantial [5–7]. High-income countries spend around 2–3% of their healthcare budgets treating just end-stage kidney disease (ESKD), even though individuals with ESKD represent less than 0.03% of the population [8]. Adding the costs of treating the much larger group of individuals with less severe CKD adds to this burden considerably (e.g., from $34 billion to $100 billion in 2015 in the USA [8]). The impact of CKD on health-related quality of life is likewise significant, especially at later stages and for patients treated with dialysis [9, 10].

The Canagliflozin and Renal Endpoints in Diabetes with Established Nephropathy Clinical Evaluation (CREDENCE) study (ClinicalTrials.gov identifier, NCT02065791) was the first dedicated renal outcomes study to report results for a sodium glucose co-transporter 2 (SGLT2) inhibitor in individuals with diabetic kidney disease (DKD), comparing canagliflozin 100 mg versus placebo on top of standard of care (SoC),

including the maximal tolerated dose of a renin–angiotensin–aldosterone system inhibitor. The CREDENCE trial was stopped early at the recommendation of the Independent Data Monitoring Committee because pre-specified efficacy criteria had been satisfied. Canagliflozin 100 mg reduced the risk of the primary endpoint (composite of doubling of serum creatinine [DoSCr], ESKD, or renal death, or CV death) by 30% versus placebo, and the results were statistically significant (hazard ratio [HR] 0.70 [95% confidence interval (CI) 0.59–0.82]; $P = 0.00001$). The results were also favorable for secondary and exploratory endpoints, including a 20% risk reduction in 3-point major adverse cardiac events (CV death, nonfatal myocardial infarction [MI], or nonfatal stroke; HR 0.80 [95% CI 0.67–0.95]; $P = 0.01$) and a 39% risk reduction in hospitalization for heart failure (HHF; HR 0.61 [95% CI 0.47–0.80]; $P < 0.001$) [11]. There were no imbalances between canagliflozin 100 mg and placebo in rates of renal-related adverse events (AEs), acute kidney injury, elevations in serum potassium levels, amputations, or fractures [11, 12]. On the basis of the results from the CREDENCE study, a new efficacy indication was added to the canagliflozin US prescribing information to reduce the risk of ESKD, DoSCr, CV death, and HHF in adults with type 2 diabetes mellitus (T2DM) and diabetic nephropathy with albuminuria > 300 mg/day [12].

Efficient allocation of scarce healthcare resources requires that treatments be not only safe and effective but also provide good "value for money" (i.e., are cost-effective). For chronic and progressive diseases such as DKD, economic evidence is routinely generated using economic modeling methods that extrapolate trial outcomes to longer time horizons (thus, capturing the full long-term costs and benefits of intervention). When properly constructed and validated, economic models can provide a more complete accounting of the relevant cost and health trade-offs of alternative treatments than economic analyses of individual clinical trials [13–15].

A recent literature review identified 101 models that included CKD [16]. Unfortunately, none of the models were suitable for simulating the full clinical and economic outcomes in the CREDENCE trial. In particular, 33 of the models were limited to ESKD and did not capture the full CKD disease spectrum. An additional 48 models considered diabetes broadly, but contained only limited renal sub-models. The rest were not limited to subjects with diabetes, excluded CV disease, or did not capture economic outcomes.

The objective of this study was to use patient-level CREDENCE trial data to develop an economic simulation model suitable for estimating the health and economic consequences of DKD treatment interventions for patients matching the CREDENCE patient population.

METHODS

An economic model of DKD treatment was designed, constructed, populated with patient-level data, and validated in accordance with the recommendations of the International Society of Pharmacoeconomic and Outcomes Research-Society of Medical Decision Making (ISPOR-SMDM) Modeling Good Research Practices Task Force-2 [17], starting with a literature review of economic models of CKD and consultations with clinical experts. The model is intended for applications in different treatment settings, so a flexible modeling approach was adopted to ensure that the needs of different stakeholders could be satisfied (e.g., numeric parameters are user-definable).

Study Data

We used patient-level data from the CREDENCE trial—4401 subjects randomized to canagliflozin or placebo and followed for a median of 2.62 years [11, 18]—to fit risk prediction equations for renal and CV outcomes relevant to patients and economic stakeholders. Briefly, CREDENCE recruited patients with T2DM and DKD, defined as eGFR of 30–90 mL/min/1.73 m^2 and urinary albumin to creatinine ratio (UACR) of > 300 to 5000 mg/g, and taking a maximum dose of either an angiotensin-converting enzyme (ACE) inhibitor or angiotensin-

receptor blocker (ARB). After randomization, trial visits were conducted at weeks 3, 13, and 26 followed by alternating telephone calls and in-clinic visits at 13-week intervals. Biomarker values for eGFR and UACR were collected every 26 weeks up to week 234. All renal and CV components included in the primary and secondary composite endpoints were adjudicated. A thorough description of the CREDENCE trial can be found in Perkovic et al. [11] and baseline patient characteristics are presented in Table S1 in electronic supplementary material (ESM) 1. All analyses were performed using the intention-to-treat patient-level study data.

Health Outcomes

Risk prediction equations were fit for the individual components of the primary and secondary renal and CV outcomes in the CREDENCE trial, which included start of maintenance dialysis, DoSCr, HHF, nonfatal MI, nonfatal stroke, and all-cause mortality. Only the first events that occurred during the trial were considered (few subjects experienced multiple events of the same type). We also fit an equation to predict whether mortality events had CV disease as a cause. Trial definitions in CREDENCE were used [11].

Note that estimation of a risk equation for kidney transplant was planned but could not be fit because there were too few events. However, the model was constructed to support the inclusion of outcome via state-specific probabilities from other sources. An equation for attributing death to renal causes was also planned but could not be fit.

Retinopathy and neuropathy are also clinically and economically important outcomes for individuals with T2DM and are commonly included in economic models of the broader T2DM population [19–23]. However, they were not pre-specified outcomes in CREDENCE, and were excluded from the model to minimize uncertainty associated with including data from other sources.

Composite trial endpoints were not selected for the model because the individual components often have substantially different economic and prognostic consequences. Unlike these individual components, however, composite trial endpoints were pre-defined and powered in the trial. To provide a benchmark by which to evaluate the validity of the individual component risk equations, a risk equation was also fit for composite ESKD (consisting of eGFR < 15 mL/min/1.73 m^2, start of maintenance dialysis, or kidney transplant), though it was not included in the economic model.

Estimation and Validation of Risk Prediction Equations

Post hoc risk prediction equations were fit individually for start of maintenance dialysis, DoSCr, HHF, nonfatal MI, nonfatal stroke, all-cause mortality, and ESKD. Three parametric forms were considered—exponential, Weibull, and Gompertz—and the functional form with best goodness-of-fit was selected. The time scale for nonfatal events was disease duration (days) and the time scale for mortality was age (days). Equations were fit using data from the placebo study arm.

The risk equations were parsimonious to avoid overfitting. This included limiting the equations to crucial explanatory covariates, excluding biomarkers except for markers of kidney functioning (eGFR) and kidney damage (UACR), and limiting time-varying covariates to cases where ignoring a temporal relationship could not reasonably be avoided to minimize the risk of confounding with unobserved factors over the study follow-up. The covariates included in each risk equation are presented in Table S2 in ESM1, by event.

Age at diagnosis, sex, and smoking status at baseline were included in the covariate specifications for each outcome. Similarly, history at baseline of MI, stroke, and HF was included as an explanatory covariate for each outcome, as prior CV disease has been shown to be a predictor of renal and CV disease as well as mortality [24]. Although the minimum eGFR required at study entry was 30 mL/min/1.73 m^2 and the minimum UACR required at study entry was 300 mg/g, some patients may have experienced reductions in eGFR or UACR to

below required thresholds between recruitment and randomization; thus, coefficients for eGFR and UACR used values at randomization. Continuous values of baseline eGFR and ln(UACR) were included for the renal outcomes given their anticipated strong association. For the other outcomes, categorical baseline eGFR and ln(UACR) specifications were found to provide more stable fits. All-cause mortality could not be adequately modeled using only baseline patient characteristics, so time-varying covariates were included for start of dialysis, MI, stroke, HF, and an indicator for eGFR < 15 mL/min/1.73 m^2. There were almost no renal outcomes for the first year of CREDENCE (at baseline, subjects were not at immediate risk for dialysis based on eGFR recruitment criteria and DoSCr naturally requires times for creatinine levels to double), which required inclusion of an indicator variable for the 1st study year.

Goodness-of-fit was assessed on the basis of discrimination (C statistic) and calibration (ratio of predicted to observed cases at the study level, accounting for censoring on predictions using last outcome carried forward), separately by study arm as well as overall. Calibration for the canagliflozin study arm was adjusted to account for the trial-reported HRs. Kaplan–Meier curves for the observed cumulative incidence rates were also plotted against Kaplan–Meier curves based on predicted sample outcomes (assuming last outcome carried forward for subjects leaving the study) for each outcome, separately by study arm and adjusted by trial-reported HRs for the canagliflozin study arm.

To attribute death events to CV and non-CV causes, a logistic regression conditional on occurrence of death was fit. As an odds ratio for canagliflozin has not previously been estimated using CREDENCE study data, both the canagliflozin and placebo treatment arms were used. Goodness-of-fit for the CV death mortality equation was assessed by calculating the percentage correctly classified for the overall data set and for each study arm separately.

Model fitting was performed in Stata Statistical Software 14 (StataCorp LLC, College Station, TX, USA) and goodness-of-fit analysis was performed using R version 3.5.1 (The R Project, www.r-project.org).

The Economic Model

The CREDENCE Economic Model of DKD (CREDEM-DKD) model was implemented as a discrete event simulation (DES) micro-simulation model, which enables the model to capture important sources of heterogeneity in the DKD population as well as 1st and 2nd order uncertainty. The DES approach lends itself well to situations where there are a set of competing events, accelerating risks, and an availability of time-to-event risk prediction equations estimated using data with relatively long durations, as is the case with DKD and the CREDENCE trial. Time horizon is user-defined, with no maximum. CREDEM-DKD was implemented in Microsoft® Excel (Microsoft Corporation, Redmond, WA, USA) and Visual Basic for Applications (VBA). The model structure is presented in Fig. 1.

At simulation baseline cohorts of hypothetical patients are drawn at random from user-assigned distributions of patient characteristics, including age, sex, smoking status, diabetes duration, eGFR, UACR, MI history, stroke history, and heart failure (HF) history. Because these patient characteristics are not independent of each other, values are sampled with user-defined correlation to support risk factor clustering [25]. The distributions and correlation matrices from the CREDENCE trial are included as default values. eGFR and UACR are combined and discretized into National Kidney Foundation–Kidney Disease Outcomes Quality Initiative (NKF-KDOQI) disease states [25, 26]. While CREDENCE limited patient recruitment to stages 2 to 3B, all stages (1 to 5) were included for completeness and to enable the modeling of regression.

Because eGFR and UACR together are used to classify DKD stages [26, 27] and because they are important determinants of mortality, the model includes the Perkovic et al. [11] linear mixed model for eGFR and for ln(UACR) evolution equations to track their values at each point in time. The eGFR equation included acute (first 3 weeks) and chronic phases (thereafter) and treatment effects for canagliflozin are supported.
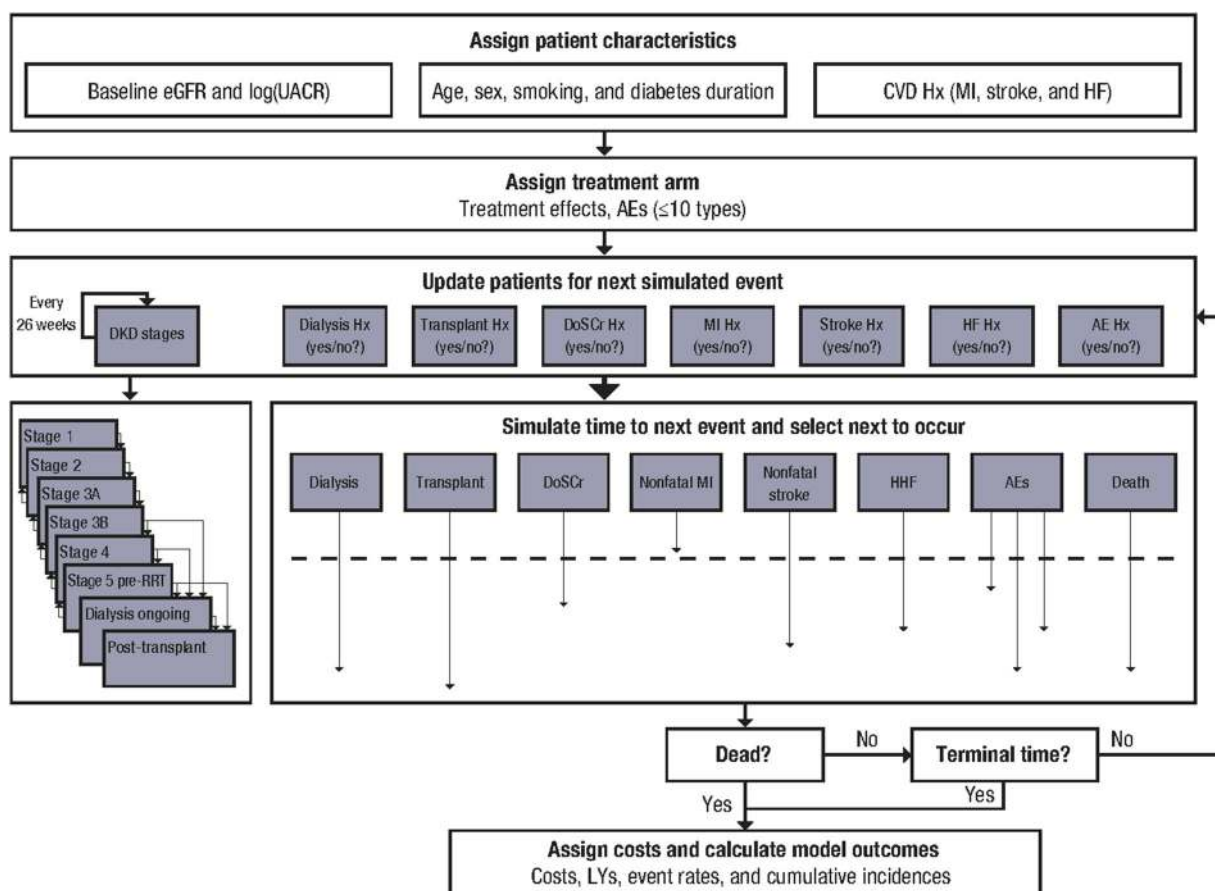
Fig. 1 CREDEM-DKD model structure. AE adverse event, CV cardiovascular, CVD cardiovascular disease, CREDEM-DKD CREDENCE Economic Model of DKD, CREDENCE Canagliflozin and Renal Endpoints in Diabetes with Established Nephropathy Clinical Evaluation, DoSCr doubling of serum creatinine, eGFR estimated glomerular filtration rate, HF heart failure,

HHF hospitalization for heart failure, Hx history, ICER incremental cost-effectiveness ratio, MI myocardial infarction, LY life-year, RRT renal replacement therapy, UACR urine albumin to creatinine ratio

eGFR and UACR are updated every 26 weeks (the frequency of data collection in CREDENCE) on the basis of the linear mixed model evolution equations and CKD status is recalculated. The risk equations are used to calculate expected time-to-event for each model outcome individually for each patient (on the basis of their personal health history), using common random numbers to avoid confusing true differences between simulated treatment arms with artificial uncertainty resulting from random number generation [28]. The event with the shortest expected time until occurrence is applied and the model jumps to that time. Risks

are recomputed on the basis of the updated patient history and the process is repeated until the time horizon is reached or the hypothetical patient dies.

Separation for outcomes was limited during the first year of the CREDENCE trial, so the user can optionally assign different treatment effects for the first year and the user can optionally set a maximum duration of treatment effects. Renal and cardiovascular health states are modelled independently. The mutually exclusive and exhaustive renal health states include traditional DKD stages (1, 2, 3A, 3B, 4, or 5 pre-dialysis [26, 27]), dialysis, and post-kidney

transplant. DKD status is classified using biomarker evolution curves reflecting current values of eGFR and UACR (in logarithmic form). Timing of start of maintenance dialysis is modeled using the CREDENCE risk prediction equation and is determined largely by eGFR and UACR at baseline but can optionally be initiated automatically at a user-defined eGFR threshold. Patients with at least stage 3B DKD or on dialysis are assigned user-defined state-specific risks for kidney transplant. Transplant survival (i.e., durability of the transplanted kidney) can optionally be defined by the user. DoSCr is simulated with the CREDENCE risk equation. However, clinical effects are captured directly via eGFR and UACR, and DoSCr is not prognostic.

The macrovascular health states (not mutually exclusive) include no history and history of MI, no history and history of stroke, and no history and history of HF. Macrovascular events are modeled with the CREDENCE risk prediction equations. All patients are at risk for CV events, including those with baseline history of that complication.

Death is the final health state. By default, the timing of death is simulated using the CREDENCE all-cause mortality risk equation. Optionally, the user can also specify a maximum age after which death will be applied automatically, a risk of procedural mortality associated with kidney transplantation, and/or a minimum eGFR consistent with human life. Alternatively, user-defined life tables can be applied for hypothetical patients undergoing dialysis or after a kidney transplant (separately by sex and different age categories). Conditional on mortality, CV or other causes of death are assigned using the CREDENCE cause of death logistic regression equation.

Two treatment arms are supported. The model allows for both comparison versus SoC and an active comparator. Treatment effects for eGFR and ln(UACR) are entered as initial mean change while treatment effects for DKD, macrovascular endpoints, and mortality are entered as HRs. AEs can be modeled via user-defined event rates (which can be different for male and female subjects) for up to 10 different events. The model supports inclusion of

acquisition and administration costs, separately by treatment arm, as well as event and state costs for the renal and macrovascular complications and by cause of death. Treatment-related AEs can be assigned fixed event costs and per day costs. A user-defined discount rate can be applied. Final model outcomes include event rates, numbers needed to treat, life years, incremental costs, incremental cost-effectiveness ratios, cost-effectiveness scatter plots, and measures of model convergence to ensure adequate sample sizes.

Model Validation

The conceptual model was designed and structured following a literature review of economic models of CKD, and best practices were identified and borrowed. A thorough model specification was prepared and then updated on the basis of comments from expert health economists and clinicians. The final model has been presented to the CREDENCE Steering Committee and to an expert nephrologist and many of their comments have been incorporated.

CREDEM-DKD was then thoroughly debugged and tested to ensure correct implementation, which consisted of more than 50 artificial simulations designed to reveal errors in both logic and programming (i.e., "stress testing"). Idiosyncratic results were thoroughly investigated and any identified errors in programming or logic were corrected and documented. Internal validation of CREDEM-DKD was then performed by loading the model to replicate the CREDENCE study (patient populations, treatment effects, and time horizon) and comparing model predictions of the cumulative incidence of the outcomes with the Kaplan–Meier curves from the CREDENCE trial.[1] External validation of CREDEM-DKD was performed by loading the model to replicate a subgroup of the CANVAS Program [29] with patient characteristics matching those of the CREDENCE study and comparing model predictions of the cumulative incidence of the outcomes with the Kaplan–Meier curves. The subgroup included a total of 567 patients and the baseline patient characteristics for the CANVAS Program subgroup

with characteristics matching the CREDENCE trial are presented in Table S14 in ESM1 and the treatment effects are presented in Table S15 in ESM1. The mean (standard error [SE]) Kaplan–Meier follow-up times for this subgroup were 46.2 (1.67) months and the median follow-up time was 48.8 months. Note that as the CANVAS Program was designed to evaluate cardiovascular outcomes, the start of maintenance dialysis was not recorded.

We completed an Assessment of the Validation Status of Health-Economic Decision Models evaluation [30], which is included in ESM2.

Compliance with Ethics Guidelines

This article is based on previously conducted studies and does not contain any studies with human participants or animals performed by any of the authors.

RESULTS

CREDENCE Risk Prediction Equations

CREDENCE patient characteristics at baseline have been described previously [11, 18] and are reproduced along with correlation coefficients in Tables S2 and S3 of ESM1. There were 2199 subjects in the placebo arm and 2202 patients in the canagliflozin arm followed for a median follow-up of 2.62 years. Numbers of events during follow-up are presented in Table S4 in ESM1. In the placebo arm for which most analyses were performed, 100 subjects started maintenance dialysis, 186 subjects experienced a DoSCr, and 81, 65, 129, and 201 subjects experienced nonfatal MI, nonfatal stroke, HHF, and all-cause mortality events, respectively. Of the 201 deaths, 140 were attributed to CV causes.

Parametric forms and the regression coefficients are presented in Tables 1, 2, and 3 for the

---

[1] Note: this differs from validation of the risk prediction equations in two important ways: (1) the analysis used randomly sampled hypothetical patients (as opposed to actual CREDENCE subjects) and (2) risk prediction equations are evaluated collectively as a system (rather than individually).

renal, CV, and all-cause mortality outcomes, respectively. To promote model reproducibility, the covariance matrices for the coefficient estimates are presented in Tables S5 to S11 in ESM1 (enabling modeling of 2nd order uncertainty).

The Weibull parametric form provided the best fit for start of maintenance dialysis (the Gompertz functional form failed to converge), as determined by Akaike information criterion (AIC). Baseline values of eGFR and ln(UACR) were statistically significant ($P < 0.001$) and strong predictors of dialysis and age at T2DM diagnosis was statistically protective ($P = 0.001$). The other covariates were not statistically significant, but collectively improved the predictive accuracy of the risk equation. Except for stroke, these covariates had HRs less than 1. Because there were almost no events in the first year of the trial, an indicator for the first year was required to achieve satisfactory fit. The resulting C statistics were 0.86 for the placebo study arm and 0.84 for the canagliflozin study arm, values widely considered good. Calibration for the placebo study arm was 0.998. Calibration was 1.065 for the canagliflozin study arm after adjustment for the trial-reported HR, indicating that the approach tends to underpredict the canagliflozin treatment effect. Graphically, predicted starts of dialysis matched observed trial starts well for both treatment arms (Fig. 2a), at least until about 3.5 years when the sample size was low (only approx. 30% of participants had more than 3 years of follow-up). A description of the results for ESKD, which was intended for assessing the robustness of the dialysis equation, is presented in Table S12 in ESM1. The results are in line with the equation for start of maintenance dialysis with the exception that stroke had a protective effect for the risk of ESKD.

The Weibull parametric form also provided the best fit for DoSCr. Baseline values of eGFR, ln(UACR), and age at T2DM diagnosis were statistically significant ($P < 0.001$) and strong predictors of DoSCr. The other covariates were not statistically significant, but collectively improved the predictive accuracy of the risk equation. Like dialysis, an indicator for the first year was required to achieve satisfactory fit. The resulting C statistics were 0.85 for both study

Table 1 Risk equations: start of dialysis and doubling of serum creatinine

Parameter	Start of maintenance dialysis		Doubling of serum creatinine	
	Weibull		Weibull	
	N = 2012		N = 2013	
	100 events		186 events	
	HR (SE)	P	HR (SE)	P
Intercept	0.001 (0.003)	0.013	6.48E−06 (1.33E−05)	< 0.001
ρ	0.371 (0.182)	< 0.001	0.664 (0.145)	< 0.001
Age at T2DM diagnosis	0.969 (0.009)	0.001	0.974 (0.007)	< 0.001
Female	0.707 (0.158)	0.121	0.920 (0.142)	0.589
Smoking	0.463 (0.186)	0.056	0.777 (0.187)	0.295
MI Hx	0.693 (0264)	0.355	0.847 (0.212)	0.506
Stroke Hx	1.193 (0.350)	0.549	0.741 (0.199)	0.265
HF Hx	0.903 (0.266)	0.728	0.752 (0.168)	0.203
eGFR	0.938 (0.009)	< 0.001	0.970 (0.005)	< 0.001
ln (UACR)	3.327 (0.508)	< 0.001	3.719 (0.438)	< 0.001
First-year indicator	0.206 (0.050)	< 0.001	0.254 (0.042)	< 0.001
AIC	190.025		183.129	
C statistic				
Placebo	0.861		0.846	
Canagliflozin	0.837		0.847	
Calibration				
Placebo	0.998		0.998	
Canagliflozin[a]	1.065		1.048	

AIC Akaike information criterion, *eGFR* estimated glomerular filtration rate, *HR* hazard ratio, *Hx* history, *MI* myocardial infarction, *SE* standard error, *UACR* urine albumin to creatinine ratio

[a] Adjusted for trial-reported HR for canagliflozin

arms, which is widely considered good. Calibration for the placebo study arm was 0.998. Calibration was 1.048 for the canagliflozin study arm after adjustment for the trial-reported HR, indicating that the approach tends to modestly underpredict modestly the canagliflozin treatment effect. Graphically, predicted DoSCr incidence matched observed trial cases reasonably well for both treatment arms (Fig. 2b) for the first 3 years, though the

predictions do not exhibit the stair step of observed incidence (caused by the periodicity of laboratory testing).

The exponential parametric form provided the best fits for nonfatal MI and nonfatal stroke as judged by AIC, while Weibull provided the best fit for HHF. Categorical baseline eGFR covariates were not statistically different from the reference category (eGFR < 30 mL/min/ 1.73 m^2) for any of the outcomes, but the HRs

Table 2 Risk equations: nonfatal MI, nonfatal stroke, and HHF

Parameter	Nonfatal MI		Nonfatal stroke		HHF	
	Exponential		Exponential		Weibull	
	$N = 2199$		$N = 2199$		$N = 2199$	
	81 events		65 events		129 events	
	HR (SE)	P	HR (SE)	P	HR (SE)	P
Intercept	2.97E−05 (2.36E−05)	<0.001	2.95E−05 (2.77E−05)	< 0.001	1.45E−07 (3.37E−07)	< 0.001
ρ					1.424 (0.204)	< 0.001
Age at T2DM diagnosis	1.003 (0.010)	0.761	0.995 (0.011)	0.654	1.023 (0.010)	0.027
Female	1.276 (0.304)	0.306	0.662 (0.193)	0.157	0.996 (0.192)	0.985
Smoking	1.932 (0.538)	0.018	0.520 (0.243)	0.162	1.416 (0.343)	0.150
MI Hx	3.584 (0.904)	< 0.001	2.028 (0.623)	0.021	2.161 (0.464)	< 0.001
Stroke Hx	1.124 (0.382)	0.73	2.733 (0.793)	0.001	1.064 (0.286)	0.819
HF Hx	0.732 (0.250)	0.362	1.372 (0.433)	0.317	2.253 (0.448)	< 0.001
eGFR ≥ 60	0.512 (0.252)	0.174	0.682 (0.422)	0.536	0.756 (0.332)	0.525
eGFR < 60, ≥ 45	0.564 (0.283)	0.254	0.880 (0.544)	0.836	0.799 (0.356)	0.613
eGFR < 45, ≥ 30	0.981 (0.476)	0.969	0.762 (0.478)	0.665	1.137 (0.496)	0.768
eGFR < 30	Reference		Reference		Reference	
UACR > 300	1.213 (0.456)	0.607	1.648 (0.771)	0.286	2.347 (0.917)	0.029
UACR ≤ 300	Reference		Reference		Reference	
AIC	391.205		340.738		480.995	
C statistic						
Placebo	0.661		0.662		0.656	
Canagliflozin	0.597		0.600		0.693	
Calibration						
Placebo	1.000		1.000		1.003	
Canagliflozin[a]	0.944[a]		1.003		0.964	

AIC Akaike information criterion, *eGFR* estimated glomerular filtration rate, *HF* heart failure, *HHF* hospitalization for heart failure, *HR* hazard ratio, *Hx* history, *MI* myocardial infarction, *SE* standard error, *UACR* urine albumin to creatinine ratio

[a] Adjusted for trial-reported HR for canagliflozin

were numerically most protective for eGFR > 60 mL/min/1.73 m^2 and the degree of protection decreased generally with decreasing eGFR values. Patients with UACR exceeding 300 mg/g at baseline had numerically greater event risks than the reference category

Table 3 Risk equations: all-cause mortality and CV cause of death

Parameter	All-cause mortality		CV cause of death		
	Exponential		Logistic		
	N = 2199		140 CV deaths		
	201 events				
	HR (SE)	P	Maximum likelihood estimates	Standard error	P
Intercept	2.23E−05 (7.52E−06)	< 0.001	0.615	0.683	0.368
Age at T2DM diagnosis	1.020 (0.006)	0.002	0.005	0.011	0.663
Female	1.046 (0.166)	0.778			
Smoking	1.137 (0.233)	0.53			
Diabetes duration		–	− 0.006	0.011	0.602
Sex (female)		–	0.093	0.270	0.731
MI Hx	1.111 (0.216)	0.587			
Stroke Hx	1.414 (0.280)	0.081			
HF Hx	1.374 (0.243)	0.072			
MI Hx[a]	3.580 (0.873)	< 0.001	0.513	0.291	0.078
Start of maintenance dialysis[a]	5.987 (1.850)	< 0.001	− 1.028	0.418	0.014
HF Hx[a]	2.792 (0.642)	< 0.001	0.550	0.276	0.046
eGFR < 15[a]	1.317 (0.415)	0.382			
Stroke Hx[a]	7.656 (1.693)	< 0.001	0.357	0.300	0.234
eGFR < 60, ≥ 45[a]			− 0.180	0.374	0.630
eGFR < 45, ≥ 30[a]			− 0.382	0.321	0.234
eGFR < 30[a]			0.788	0.428	0.066
UACR > 300[a]			0.287	0.300	0.339
Canagliflozin			− 0.084	0.123	0.491
AIC	133.5836				
C statistic			AUC		
Placebo	0.671		Placebo	0.714	
Canagliflozin	0.670		Canagliflozin	0.651	
Calibration			Concordance (%)		
Placebo	1.001		Placebo	71.2	

Table 3 continued

Parameter	All-cause mortality		CV cause of death		
	Exponential		Logistic		
	N = 2199		140 CV deaths		
	201 events				
	HR (SE)	P	Maximum likelihood estimates	Standard error	P
Canagliflozin[b]	1.080		Canagliflozin	65.0	

AIC Akaike information criterion, *AUC* area under the curve, *CV* cardiovascular, *eGFR* estimated glomerular filtration rate, *H*F heart failure, *HR* hazard ratio, *Hx* history, *MI* myocardial infarction, *SE* standard error, *UACR* urine albumin to creatinine ratio

[a] Time-varying

[b] Adjusted for trial-reported HR for canagliflozin

(UACR < 300 mg/g), which was statistically significant only for HHF. The strongest driver of CV risk was CV history. Baseline history of each event type was a statistically and numerically significant determinant for that outcome. Baseline history of MI was also a statistically and numerically significant determinant of risk for nonfatal stroke and HHF. C statistics for each CV outcome were about 0.65 for the placebo study arm and ranged between 0.60 and 0.68 for the canagliflozin study arm, which are generally considered borderline discrimination. Calibration was 1.00 for all three outcomes in the placebo study arm (for which the prediction equations were fit). Adjusted for the trial-reported HRs, calibration in the canagliflozin study arm was 0.94, 1.00, and 0.96 for nonfatal MI, nonfatal stroke, and HHF, respectively, indicating underprediction of MI risk for canagliflozin-treated patients. Graphically, the predicted cumulative incidence of nonfatal MI matched observed trial outcomes well for both treatment arms (Fig. 2c) for the first 3 years, though some benefit associated with canagliflozin during the first year was not captured. The prediction equations for nonfatal stroke fit the observed placebo arm and the observed canagliflozin arm after 1.5 years closely (Fig. 2d), though it underpredicted the observed

canagliflozin data early on (i.e., the canagliflozin treatment effect emerged first after 1 year in the CREDENCE trial). The prediction equations for HHF fit the observed placebo arm and the observed canagliflozin arm after 1.5 years (Fig. 2e), though both arms were underpredicted during the initial period.

The Weibull parametric form narrowly provided the best fit, as judged by AIC, for all-cause mortality. The intercept term was extremely small, however, which when logged as required led to model instability and poor internal validation. The exponential parameter was selected instead as the default risk equation on the basis of best replication of the CREDENCE trial survival curve. Time-varying occurrences of MI, stroke, HF, and start of maintenance dialysis were statistically significant ($P < 0.001$) and numerically large predictors of mortality risk. Age at diagnosis of T2DM was associated with increased mortality risk ($P = 0.02$). Reaching eGFR < 15 mL/min/1.73 m^2 was associated with a 32% increased risk of death, though this was not statistically significant in the final specification. The C statistic was 0.67 for the placebo study arm and the canagliflozin study arm, which are generally considered borderline. Calibration was 1.00 for the placebo study arm (for which the prediction equation was fit) and
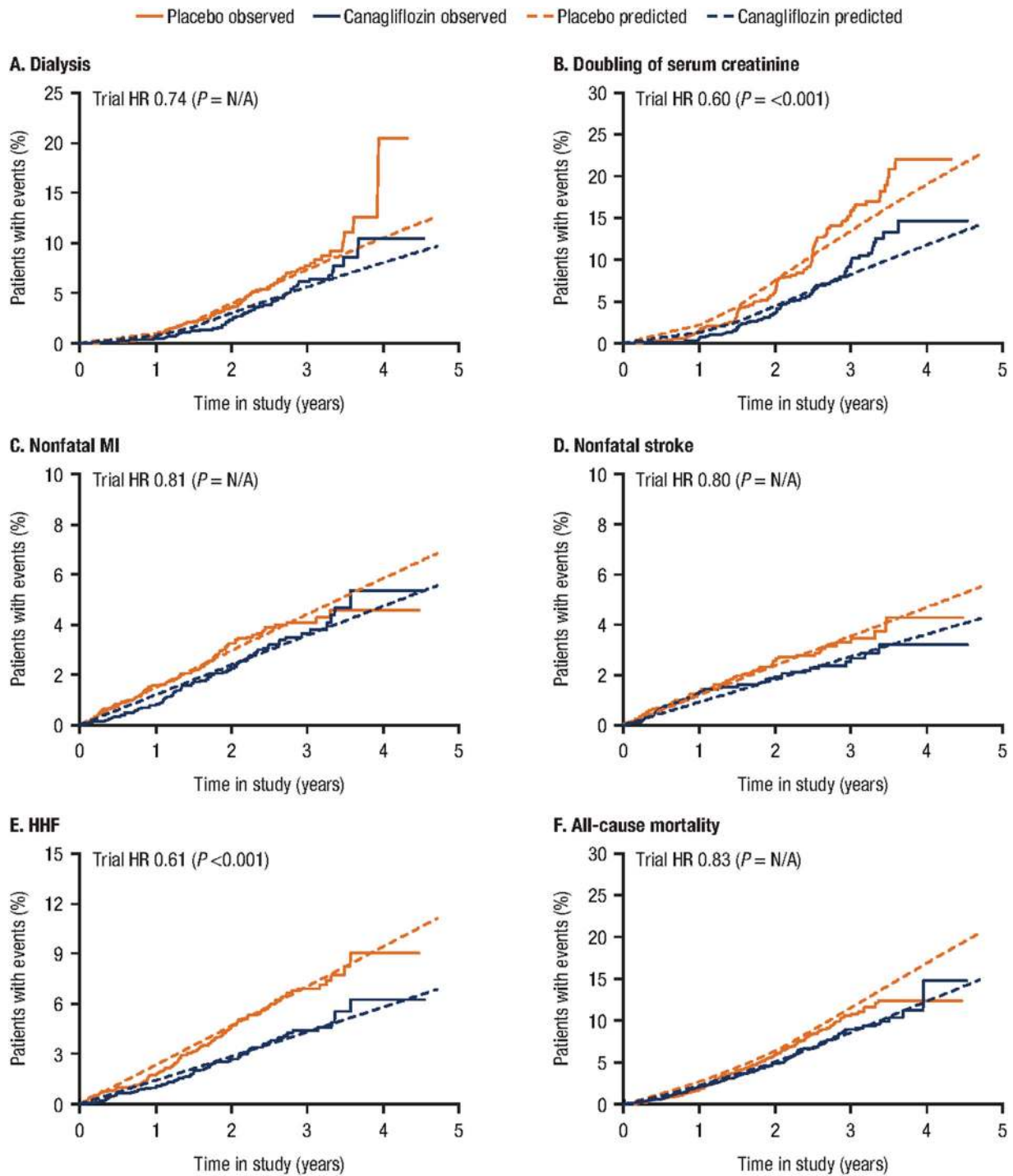
Fig. 2 Kaplan–Meier cumulative incidence curves for risk equation predictions and observed CREDENCE values, by outcome. CREDENCE Canagliflozin and Renal Endpoints in Diabetes with Established Nephropathy Clinical Evaluation, HHF hospitalization for heart failure, HR hazard ratio, MI, myocardial infarction

1.08 for the canagliflozin study arm, thus understating risk of death in the placebo arm and overstating it in the canagliflozin arm (and thus underpredicting the treatment effect). Graphically, the predicted cumulative incidence of death matched observed trial outcomes well for both treatment arms (Fig. 2f) for the first 3 years, though it showed some overprediction for the placebo arm during the first year.

For the subset of patients with a death event, only start of maintenance dialysis (protective) and HF history (promotive) were statistically significant determinants of CV as a cause of death. MI history and stroke were also promotive, but not statistically significant ($P = 0.078$ and $P = 0.234$, respectively). Canagliflozin is associated with a log-odds ratio of -0.084, which corresponds to a 15.5% reduction in the odds ratio (i.e., it is less likely that a mortality event for canagliflozin-treated patients is CV related). Area under the curve (AUC) was 0.71 and 0.65 for the placebo and canagliflozin study arms, respectively, and 64.3% and 68.4% of subjects were correctly classified.

Validation of CREDEM-DKD

Model testing and debugging included 67 artificial simulations designed to reveal errors in logic and programming. The tests included extreme value testing of all model compartments. All baseline characteristics, treatment inputs, and risk calculations were tested individually. Stress-test results are presented in Table S13 in ESM1. Two minor bugs were identified and fixed. The results of each of the 67 scenarios were generally consistent with expectations when the simulations were rerun.

Kaplan–Meier cumulative incidence curves describing internal model validation (i.e., replication of the CREDENCE trial) are presented in Fig. 3. The model tended to overpredict the start of maintenance dialysis over the duration of the CREDENCE trial for both study arms, but the difference between arms was similar to that in the trial. By year 3, however, model predictions matched closely with the CREDENCE cumulative incidence. The model

prediction of DoSCr was visually better, though the stair step pattern, which was an artifact related to the frequency of laboratory measurements, and a sharp jump for the placebo arm after 2.5 years could not be replicated. The fit for the macrovascular and mortality outcomes were visually generally good, with the greatest discrepancy being overestimation of the positive treatment effect for HHF during the first year.

The Kaplan–Meier results of the limited external validation using the subgroup of 567 patients in the CANVAS Program that met CREDENCE eligibility criteria are presented in Fig. 4. CREDEM-DKD overpredicted the incidence of DoSCr in the CANVAS Program. Predictions for nonfatal MI and HHF visually matched CANVAS Program outcomes closely, however, at least until the sample of at-risk patients was diminished because of trial-related factors. The cumulative incidence of nonfatal stroke in this CANVAS Program subgroup was sparse, with curves crossing after 3 years of follow-up, and the model generally underpredicted absolute events and treatment effects in the first 2 years but overpredicted treatment effects thereafter. The model also underpredicted all-cause mortality, especially after 1.5 years, and underpredicted the treatment effect.

DISCUSSION

We fit a set of pragmatic and parsimonious risk prediction equations for renal and CV events using patient-level data from the CREDENCE trial, the first dedicated renal outcomes study for any agent in the class of SGLT2 inhibitors, and constructed an economic micro-simulation model for evaluating the cost-effectiveness of long-term treatment in individuals with T2DM and DKD following best economic modeling practices [17, 31–33]. The risk prediction equations generally fit well and exhibited good concordance, excellent for the placebo study arm from which they were estimated and with modest underprediction (MI) or overprediction (dialysis, DoSCr, and all-cause mortality) for the canagliflozin study arm. Discrimination was
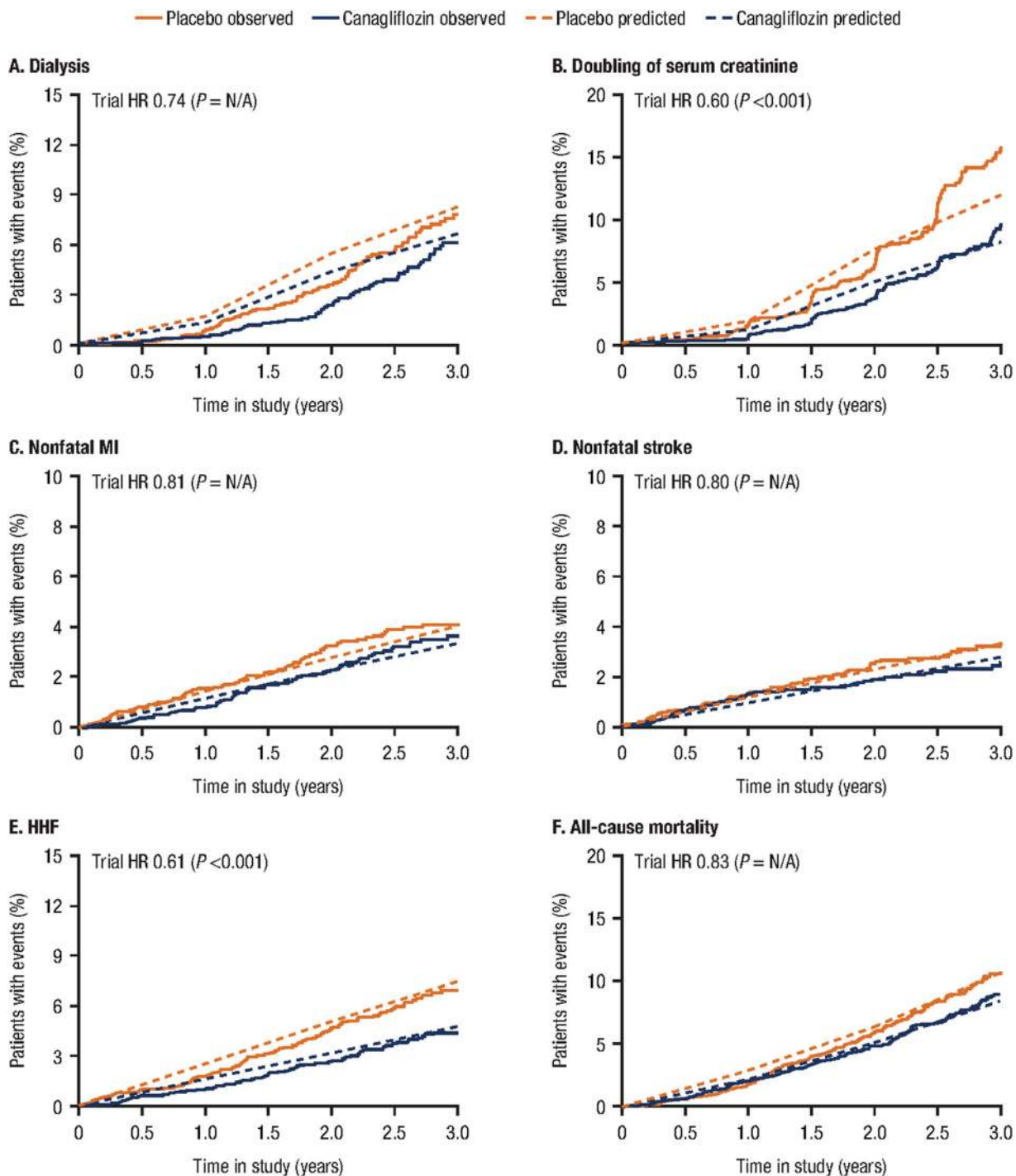
Fig. 3 Kaplan–Meier cumulative incidence curves for predicted and observed CREDENCE trial values, by outcome. CREDENCE Canagliflozin and Renal Endpoints in Diabetes with Established Nephropathy Clinical Evaluation, HHF hospitalization for heart failure, HR hazard ratio, MI myocardial infarction
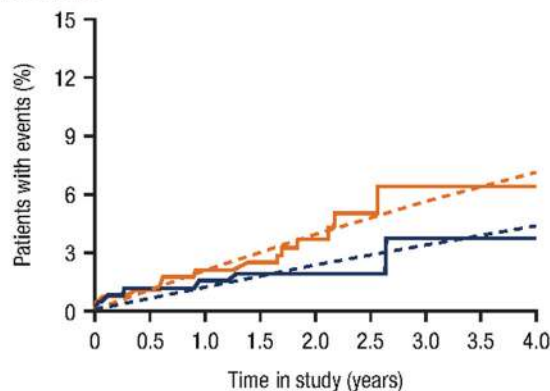
Placebo observed —— Canagliflozin observed – – Placebo predicted – – Canagliflozin predicted

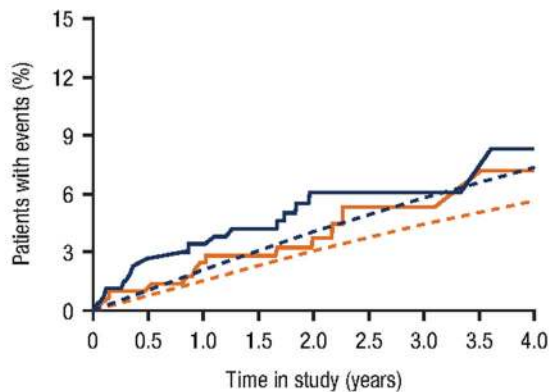**A. Doubling of serum creatinine**
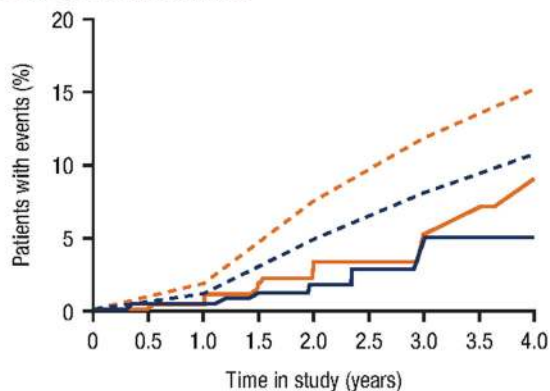
Placebo	296	289	278	267	177	78	51	50	48
Canagliflozin	271	266	263	255	164	70	45	42	40

B. Nonfatal MI

Placebo	296	286	273	266	177	73	50	49	48
Canagliflozin	271	263	259	253	167	71	43	41	39

C. Nonfatal stroke

Placebo	296	285	273	265	179	77	52	51	49
Canagliflozin	271	260	255	246	159	69	45	42	40

D. HHF
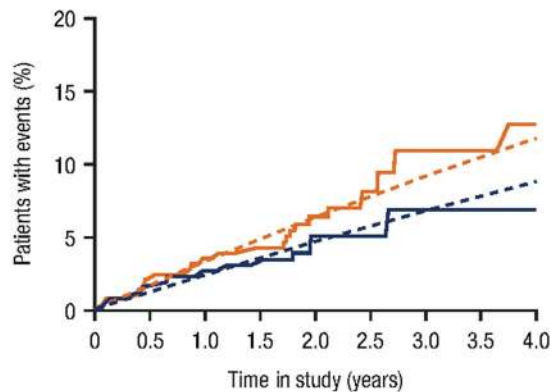
Placebo	296	283	270	262	176	76	51	50	48
Canagliflozin	271	263	258	250	157	67	42	39	38

E. All-cause mortality
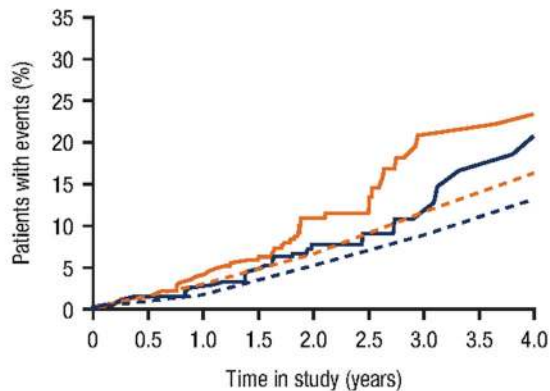
Placebo	296	293	285	278	190	87	59	59	58
Canagliflozin	271	267	264	259	169	72	46	43	41

◀**Fig. 4** Kaplan–Meier cumulative incidence curves for predicted and observed CANVAS Program subgroup values, by outcome. CANVAS CANagliflozin cardioVascular Assessment Study, HHF hospitalization for heart failure, HR hazard ratio, MI, myocardial infarction

strong (0.85) for the renal outcomes, but weaker for the macrovascular outcomes and all-cause mortality (0.60–0.68). Lower discrimination is common to macrovascular risk prediction equations and several important recent studies have reported discrimination between 0.6 and 0.7 [34–36]. Planned estimation of risk equations for kidney transplant and renal death could not be generated because of small event counts. We illustrate the intended usage of the risk prediction equations in a numerical example in ESM1.

Most of the risk equations fit the CREDENCE study well, providing evidence of internal validity. Start of dialysis was overpredicted compared with CREDENCE cumulative incidence over the first 2.5 years for both study arms, though predictions and observed values converged for both study arms by year 3. Unlike the other outcomes, start of dialysis is procedural (as opposed to clinical), which may explain part of the difficulties in fitting. DoSCr was visually the most discordant, but this can likely be explained by the periodicity of laboratory samples and by the special nature of the outcome (i.e., the magnitude of increase needed to qualify as a doubling is determined by the starting value).

A parsimonious DES micro-simulation model was constructed using these risk equations (with an option for combination with user-definable kidney transplant event risks) in order to fill an important gap in the CKD literature and enable a robust economic analysis of canagliflozin in the treatment of DKD. CREDEM-DKD was subjected to internal validation, in which the model performed well for the CREDENCE trial with a certain degree of overestimation for start of dialysis, associated presumably with the absence of dialysis events in the first year in CREDENCE. Ideally, the external validation would have been conducted comparing versus a renal outcomes trial, but none are available at

this time. Several trials will report in the upcoming years. Therefore, an external validation exercise was conducted using the results of a subgroup of patients from the CANVAS Program that would have been eligible for CREDENCE and the model performed reasonably well. While the DoSCr prediction equation, likely for reasons noted above, was unable to predict the CANVAS Program closely, prediction equations for the other outcomes were not excessively discordant and fit closely for non-fatal MI and HHF, providing additional confidence that the risk equations are robust.

Model strengths include use of risk prediction equations estimated with patient-level data from a landmark renal outcomes trial [11, 18] in individuals with DKD (large sample size, relatively long follow-up, and comprehensive prospectively collected data on patient- and payor-relevant DKD and CV outcomes), enabling modeling of features that were found absent in most existing models of CKD [16].

As with other studies, estimating risk prediction equations with randomized controlled trial (RCT) data risks incorporation of bias as study recruitment criteria frequently lead to nonrepresentative samples and protocols steering treatment (e.g., glycemic equipoise) can affect outcomes. In particular, causality is difficult to ensure (baseline smoking status, for example, was found protective for stroke). Analyses with post-baseline (time-varying) covariates, for example, can suffer from endogeneity bias, which affects extrapolation to patient groups that are quite different, so time-varying covariates were avoided where possible. By design, the model includes only renal and CV outcomes, so it does not capture the full set of outcomes relevant for all patients with T2DM. Future enhancements to this model could include incorporation of a treatment algorithm.

CONCLUSIONS

We used patient-level data from the first dedicated renal outcomes trial of an SGLT2 inhibitor to report renal and CV protection for an SGLT2 inhibitor to construct an economic simulation

model. The underlying risk equations and the model were subjected to extensive internal validation and the model underwent an external validation exercise. Therefore, the CREDEM-DKD model represents an important advance for the evaluation of treatment interventions in the DKD population.

ACKNOWLEDGEMENTS

REFERENCES

1. Jha V, Garcia-Garcia G, Iseki K, et al. Chronic kidney disease: global dimension and perspectives. Lancet. 2013;382(9888):260–72.

2. Bikbov B, Perico N, Remuzzi G. Disparities in chronic kidney disease prevalence among males and females in 195 countries: analysis of the global burden of disease 2016 study. Nephron. 2018;139(4):313–8.

3. Murphy D, McCulloch CE, Lin F, et al. Trends in prevalence of chronic kidney disease in the United States. Ann Intern Med. 2016;165(7):473–81.

4. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. N Engl J Med. 2004;351(13):1296–305.

5. McQueen RB, Farahbakhshian S, Bell KF, Nair KV, Saseen JJ. Economic burden of comorbid chronic kidney disease and diabetes. J Med Econ. 2017;20(6):585–91.

6. Thomas R, Kanso A, Sedor JR. Chronic kidney disease and its complications. Prim Care. 2008;35(2):329–44, vii.

7. Vanholder R, Annemans L, Brown E, et al. Reducing the costs of chronic kidney disease while delivering quality health care: a call to action. Nat Rev Nephrol. 2017;13(7):393–409.

8. Luyckx VA, Tonelli M, Stanifer JW. The global burden of kidney disease and the sustainable development goals. Bull World Health Organ. 2018;96(6):414-22D.

9. Gorodetskaya I, Zenios S, McCulloch CE, et al. Health-related quality of life and estimates of utility in chronic kidney disease. Kidney Int. 2005;68(6):2801–8.

10. Pagels AA, Söderkvist BK, Medin C, Hylander B, Heiwe S. Health-related quality of life in different stages of chronic kidney disease and at initiation of dialysis treatment. Health Qual Life Outcomes. 2012;10:71.

11. Perkovic V, Jardine MJ, Neal B, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. N Engl J Med. 2019;380(24):2295–306.

12. INVOKANA® (canagliflozin) tablets, for oral use [package insert]. Titusville, NJ: Janssen Pharmaceuticals; September 2019.

13. Farahani P. A perspective on principles of comparative cost-effectiveness studies for pharmacotherapy of chronic diseases. Clin Diabetes. 2012;30(2):54.

14. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices-Modeling Studies. Value Health. 2003;6(1):9–17.

15. Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation (handbooks in health economic evaluation). 1st ed. New York: Oxford University Press; 2006.

16. Sugrue DM, Ward T, Rai S, McEwan P, van Haalen HGM. Economic modelling of chronic kidney disease: a systematic literature review to inform conceptual model design. Pharmacoeconomics. 2019;37(12):1451–68.

17. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. Med Decis Making. 2012;32(5):678–89.

18. Mahaffey KW, Jardine MJ, Bompoint S, et al. Canagliflozin and cardiovascular and renal outcomes in type 2 diabetes and chronic kidney disease in primary and secondary cardiovascular prevention groups: results from the randomized CREDENCE trial. Circulation. 2019;140(9):739–50.

19. McEwan P, Foos V, Palmer JL, Lamotte M, Lloyd A, Grant D. Validation of the IMS CORE diabetes model. Value Health. 2014;17(6):714–24.

20. McEwan P, Ward T, Bennett H, Bergenheim K. Validation of the UKPDS 82 risk equations within the Cardiff Diabetes Model. Cost Eff Resour Alloc. 2015;13:12.

21. Willis M, Johansen P, Nilsson A, Asseburg C. Validation of the Economic and Health Outcomes Model of Type 2 Diabetes Mellitus (ECHO-T2DM). Pharmacoeconomics. 2017;35(3):375–96.

22. Lundqvist A, Steen Carlsson K, Johansen P, Andersson E, Willis M. Validation of the IHE Cohort Model of type 2 diabetes and the impact of choice of macrovascular risk equations. PLoS One. 2014;9(10):e110235.

23. Shao H, Fonseca V, Stoecker C, Liu S, Shi L. Novel risk engine for diabetes progression and mortality in USA: building, Relating, Assessing, and Validating Outcomes (BRAVO). Pharmacoeconomics. 2018;36(9):1125–34.

24. American Diabetes Association. Standards of medical care in diabetes—2020. Diabetes Care. 2020;43(Suppl 1):S1–212.

25. Brown JB, Russell A, Chan W, Pedula K, Aickin M. The global diabetes model: user friendly version 3.0. Diabetes Res Clin Pract. 2000;50(Suppl 3):S15–46.

26. National Kidney Foundation. 2018. https://www.kidney.org/professionals/guidelines/guidelines_commentaries. Accessed 14 Apr 2020.

27. National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease:

evaluation, classification, and stratification. Am J Kidney Dis. 2002;39(2 Suppl 1):S1–266.

28. Murphy DR, Klein RW, Smolen LJ, Klein TM, Roberts SD. Using common random numbers in health care cost-effectiveness simulation modeling. Health Serv Res. 2013;48(4):1508–25.

29. Neal B, Perkovic V, Mahaffey KW, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. N Engl J Med. 2017;377(7):644–57.

30. Vemer P, Corro Ramos I, van Voorn GA, Al MJ, Feenstra TL. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. Pharmacoeconomics. 2016;34(4): 349–61.

31. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices-overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. Value Health. 2012;15(6):796–803.

32. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling

Good Research Practices Task Force–7. Value Health. 2012;15(6):843–50.

33. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Moller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–4. Value Health. 2012;15(6):821–7.

34. Zethelius B, Eliasson B, Eeg-Olofsson K, Svensson AM, Gudbjornsdottir S, Cederholm J. A new model for 5-year risk of cardiovascular disease in type 2 diabetes, from the Swedish National Diabetes Register (NDR). Diabetes Res Clin Pract. 2011;93(2): 276–84.

35. Kengne AP, Patel A, Marre M, et al. Contemporary model for cardiovascular risk prediction in people with type 2 diabetes. Eur J Cardiovasc Prev Rehabil. 2011;18(3):393–8.

36. McEwan P, Bennett H, Ward T, Bergenheim K. Refitting of the UKPDS 68 risk equations to contemporary routine clinical practice data in the UK. Pharmacoeconomics. 2015;33(2):149–61.