

RESEARCH

Open Access



# Risk prediction modeling for cardiorenal clinical outcomes in patients with non-diabetic CKD using US nationwide real-world data

Christoph Wanner<sup>1\*</sup>, Johannes Schuchhardt<sup>2</sup>, Chris Bauer<sup>2</sup>, Meike Brinker<sup>3</sup>, Frank Kleinjung<sup>4</sup> and Tatsiana Vaitiakhovich<sup>4</sup>

## Abstract

**Background** Chronic kidney disease (CKD) is a global health problem, affecting over 840 million individuals. CKD is linked to higher mortality and morbidity, partially mediated by higher cardiovascular risk and worsening kidney function. This study aimed to identify risk factors and develop risk prediction models for selected cardiorenal clinical outcomes in patients with non-diabetic CKD.

**Methods** The study included adults with non-diabetic CKD (stages 3 or 4) from the Optum® Clinformatics® Data Mart US healthcare claims database. Three outcomes were investigated: composite outcome of kidney failure/need for dialysis, hospitalization for heart failure, and worsening of CKD from baseline. Multivariable time-to-first-event risk prediction models were developed for each outcome using swarm intelligence methods. Model discrimination was demonstrated by stratifying cohorts into five risk groups and presenting the separation between Kaplan–Meier curves for these groups.

**Results** The prediction model for kidney failure/need for dialysis revealed stage 4 CKD (hazard ratio [HR] = 2.05, 95% confidence interval [CI] = 2.01–2.08), severely increased albuminuria-A3 (HR = 1.58, 95% CI = 1.45–1.72), metastatic solid tumor (HR = 1.58, 95% CI = 1.52–1.64), anemia (HR = 1.42, 95% CI = 1.41–1.44), and proteinuria (HR = 1.40, 95% CI = 1.36–1.43) as the strongest risk factors. History of heart failure (HR = 2.42, 95% CI = 2.37–2.48), use of loop diuretics (HR = 1.65, 95% CI = 1.62–1.69), severely increased albuminuria-A3 (HR = 1.55, 95% CI = 1.33–1.80), atrial fibrillation or flutter (HR = 1.53, 95% CI = 1.50–1.56), and stage 4 CKD (HR = 1.48, 95% CI = 1.44–1.52) were the greatest risk factors for hospitalization for heart failure. Stage 4 CKD (HR = 2.90, 95% CI = 2.83–2.97), severely increased albuminuria-A3 (HR = 2.30, 95% CI = 2.09–2.53), stage 3 CKD (HR = 1.74, 95% CI = 1.71–1.77), polycystic kidney disease (HR = 1.68, 95% CI = 1.60–1.76), and proteinuria (HR = 1.55, 95% CI = 1.50–1.60) were the main risk factors for worsening of CKD stage from baseline. Female gender and normal-to-mildly increased albuminuria-A1 were found to be associated with lower risk in all prediction models for patients with non-diabetic CKD stage 3 or 4.

**Conclusions** Risk prediction models to identify individuals with non-diabetic CKD at high risk of adverse cardiorenal outcomes have been developed using routinely collected data from a US healthcare claims database. The models may have potential for broad clinical applications in patient care.

\*Correspondence:

Christoph Wanner  
wanner\_c@ukw.de

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

**Keywords** Non-diabetic chronic kidney disease, End-stage kidney disease, Kidney failure, Dialysis, Hospitalization for heart failure, Hypertension, Risk prediction, Real-world evidence

## Background

Chronic kidney disease (CKD) is a major global health problem with numerous etiologies [1]. In 2016, CKD was the sixteenth leading cause of years of life lost worldwide, and it is expected to be the fifth leading cause of years of life lost worldwide by 2040 [2]. In 2017, 1.2 million people died from CKD globally, with 697.5 million cases of all-stage CKD recorded and a global prevalence of 9.1% [3]. Furthermore, CKD resulted in 35.8 million disability-adjusted life years, for which one-third of cases was attributed to diabetic nephropathy [3]. Although diabetes (types 1 and 2) is recognized as one of the leading causes of CKD, 50–70% of *current* CKD cases are estimated to have a non-diabetic cause [4].

CKD is linked to higher mortality and morbidity, partially mediated by higher cardiovascular (CV) risk, and worsening kidney function [4, 5]. Patients with CKD are 5–10 times more likely to die at earlier stages of disease from other causes than progress to end-stage kidney disease (ESKD) [4, 6]. A meta-analysis of >1.5 million individuals from general, high-risk, and CKD populations showed that decreasing estimated glomerular filtration rate (eGFR) and increasing albuminuria, which define the severity of CKD, were associated with increases in mortality from any cause, CV mortality, acute kidney injury incidence, and kidney disease progression [7]. Patients with CKD and low eGFR have an increased risk of major adverse CV events, specifically stroke, congestive heart failure (HF), fatal and nonfatal myocardial infarction, and sudden cardiac death at low eGFR [8]. Patients with CKD stages 3 or higher, denoted by an eGFR < 60 ml/min/1.73 m<sup>2</sup>, have been reported to have a 2- to-16-fold increased risk of major adverse CV events compared with those with eGFR > 60 ml/min/1.73 m<sup>2</sup> [8].

Development and progression of non-diabetic CKD have been found to be associated with hypertension, with blood pressure control being a key point of intervention to decrease the risk of decline in renal function and CV mortality [9]. Current therapies include renin–angiotensin system inhibitors, sodium-glucose co-transporter-2 inhibitors, and mineralocorticoid receptor antagonists. However, even though treatments such as dapagliflozin and ramipril are indicated in non-diabetic CKD, studies that have evaluated their effectiveness have been conducted primarily on patients with diabetic CKD [10, 11].

There are some theories that outcomes for patients with non-diabetic CKD are different compared with patients with diabetic CKD; however, it is generally

hypothesized that the benefits observed in the diabetes-specific (types 1 and 2) studies extend to patients with a non-diabetic CKD etiology [12]. This suggests that there is not only a need for further innovative therapies that will reduce cardiorenal risk for patients with non-diabetic CKD, but also for an improved understanding of risk factors and outcomes for patients with non-diabetic CKD.

Risk prediction models for worsening HF, as well as for progression of CKD in type 2 diabetes (T2D), have previously been published [13, 14]. A predictive model for progression of CKD to kidney failure, called the Kidney Failure Risk Equation [15], was developed for patients with CKD stages 3–5 using electronic health records from Canada (see also [16]). Kidney failure was defined [15] as need for dialysis or pre-emptive kidney transplantation. Two versions of the Kidney Failure Risk Equation are available; one includes eight variables [16, 17] and the other includes four variables [16, 18]. The eight-variable model includes age, sex, and routinely obtained laboratory tests: eGFR, albuminuria, serum calcium, serum phosphate, serum bicarbonate, and serum albumin. However, even large, nationwide administrative claims databases often do not contain laboratory data or have them available for only a part of the database members. This limits the application of the eight-variable Kidney Failure Risk Equation in observational studies using secondary data sources or in settings where the components of the equation cannot principally be collected. In turn, the four-variable version of the Kidney Failure Risk Equation requires the urine albumin-to-creatinine ratio (UACR), which is not commonly tested in routine clinical practice [19].

Several risk prediction models have recently been developed for CKD or, more specifically, for CKD associated with T2D. Most of these were developed on selected cohorts that included several thousand patients with a predefined set of risk predictors [15, 20, 21]. Some of these models used primary data, for example from randomized clinical trials or observational studies, while others were developed using secondary data. To the best of our knowledge, no risk prediction models exist to date for cardiorenal outcomes in non-diabetic CKD that are based on routinely collected data from more than 500,000 patients exploring thousands of potential predictors.

The paper reports the results of the real-world evidence Exploratory analysis of Longitudinal

patient-level Data for non-diabetic chronic kidney disease in a United States (US) claims database (FLIEDER) study that aimed to characterize patient demographics, clinical characteristics, treatment patterns, and clinical outcomes [22]. This study demonstrated that patients with non-diabetic CKD are at high risk of serious clinical outcomes, including kidney failure/need for dialysis, hospitalization for HF (HHF), and progression of CKD stage from baseline [22]. Furthermore, the FLIEDER study aimed to develop risk prediction models for kidney failure/need for dialysis, HHF, and worsening of CKD stage from baseline, for patients with non-diabetic CKD based on the data collected in the US routine clinical practice and used to reimburse the healthcare costs. The risk prediction models will be given here.

## Methods

### Study design

The study design and patient selection criteria have been published previously [22]. The FLIEDER study included longitudinal individual-level data from the Optum<sup>®</sup> Clinformatics<sup>®</sup> Data Mart (Optum CDM), collected between January 1, 2008, and December 31, 2018. The Optum CDM is one of the largest population-based claims databases and comprises 64 million patients from commercial health plan and Medicare Advantage data spanning all 50 states in the US. Optum CDM laboratory data are collected from several large laboratory vendors and are available for approximately 30% of the database members. In this study, the patient data analyzed were de-identified, and only aggregate results from the analysis are reported. Therefore, ethics approval and informed consent were not required (details provided in the ethics approval and consent to participate section). For patient data to be eligible for inclusion in the study cohort, individuals were required to be recorded as having non-diabetic moderate-to-severe CKD (stages 3 or 4) identified by an eGFR of 15–59 ml/min/1.73 m<sup>2</sup> and/or by the International Classification of Diseases (ICD) code and confirmed by a second eGFR value or the ICD code 90–365 days apart (index date). Individuals had to be  $\geq 18$  years old at index and must have had continuous enrollment in the insurance plan at least 1 year prior to index date (baseline period). Further patient inclusion and exclusion criteria have been published previously (Table S1) [22].

### Outcomes

Three primary outcomes were defined based on ICD-9/-10 diagnosis and procedure codes, Current Procedural Terminology-4/Healthcare Common Procedure Coding System (CPT-4/HCPCS) procedure codes or laboratory values observed in the post-index period: 1) a composite

measure of kidney failure/need for dialysis, 2) HHF, and 3) worsening of CKD stage from baseline. The worsening of CKD stage from baseline was defined based on eGFR values in the follow-up period or an observed diagnosis code for a more advanced CKD stage, where eGFR values were prioritized over diagnosis codes, if both were available. A change of stage had to be confirmed at least once, except if the change occurred in the last available data point.

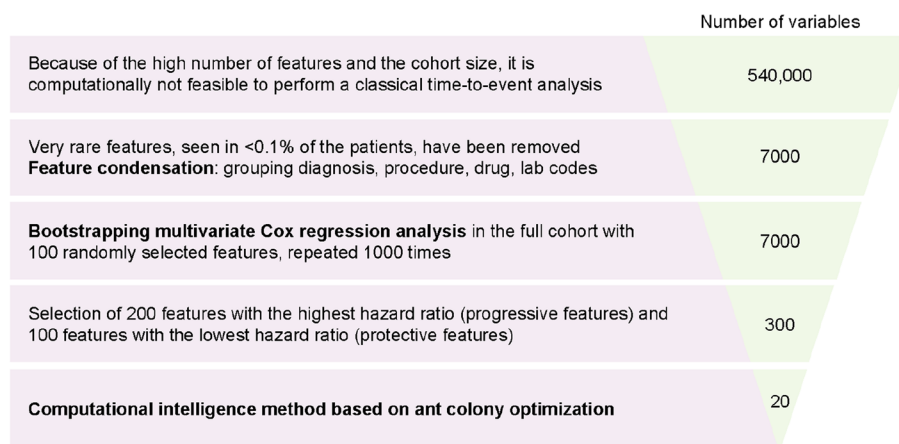
## Statistical analysis

### Risk prediction models

Data-driven time-to-first-event risk prediction models were developed to identify relevant risk factors and to estimate risk of clinical outcomes of interest. A detailed technical description of the methodological approach has been published previously [23]. The main analysis steps are outlined below.

Baseline patient data, including patient demographic characteristics (age, sex), clinical diagnoses, procedures, and laboratory values, represented more than 540,000 variables, if any single code from the coding systems used to record information in the Optum CDM was considered a variable. For the data-driven analysis, variables were classified into several categories: single or aggregated ICD-9/-10 codes represented a specific disease; medical procedures were identified by ICD-9/-10, CPT-4 or HCPCS codes; and National Drug Codes were used to identify medications grouped by therapeutic classes. ICD-9/-10 codes for the same disease or procedure were merged. Laboratory data were not used to define risk factors for the data-driven analysis due to their limited availability in the Optum CDM (approximately 30% of database members). Except for demographic variables, each variable was assessed as present (yes/no) in the patient baseline period of 1 year. Variables with a frequency of appearance in  $< 0.1\%$  of patients in the study cohort were excluded. A multivariable Cox regression analysis was conducted to identify 200 variables with the highest hazard ratios (HRs) (top “progressive” risk factor candidates) and 100 variables with the lowest HRs (top “protective” risk factor candidates). This was carried out via bootstrapping by randomly selecting 100 variables and repeating the analysis 1000 times. The unequal number of progressive and protective risk factor candidates (200 vs 100) was dictated by a proportion of variables with the respective HRs in the overall set. The threshold of 300 variables was chosen based on the observed level of HRs and to secure reasonable computational time for the follow-up optimization procedure.

An ant colony optimization (ACO) method was utilized to identify a set of 20 risk factors, a desirable maximum amount of risk factors in the final model, among



**Fig. 1** Outline of data-driven methodological approach used to develop risk prediction models

300 candidates by optimizing the Bayesian information criterion of the corresponding Cox regression model (Fig. 1). The premise behind ACO is to identify variables that produce “traces” with a high “pheromone level,” indicating the greatest impact on the clinical outcome among all candidate variables. ACO mimics the behavior of ants who leave behind pheromone trails leading towards a food source for other ants [24]. After applying ACO, a free term  $\alpha$  was calculated, corresponding to the yearly hazard of the reference group (no risk factors, age zero, male gender, unknown race, and unknown albuminuria category).

The obtained data-driven risk prediction models were adjusted based on subject matter expertise, resulting in the final models for three clinical outcomes of interest. The chosen methodology allowed for the addition of well-known risk factors of cardiorenal outcomes in CKD into the predictive models. For example, albuminuria category (A1–A3), as defined based on the laboratory UACR test value, is a known clinical parameter that modifies individual risk of cardiorenal outcomes in CKD. UACR laboratory test results were available for only 6% of patients in the study cohort and like other laboratory data, were not used in the data-driven analysis. However, the effect of the albuminuria category as a risk factor was measured and added to the models “manually.” The possibility of estimating individual risk of clinical outcomes was preserved for patients with no reported UACR (i.e., with no assigned albuminuria category).

#### **Demonstrating model performance**

To demonstrate final models’ performance, all patients in the cohort were stratified into five risk groups based on the quintiles of hazards for a given outcome as estimated by the respective model: very low, low, medium, high,

and very high. Namely, each group comprised 20% of all patients (e.g., the 20% of patients with the lowest hazards were in the ‘very low’ group, and the 20% with the highest hazards were in the ‘very high’ group). To obtain a risk estimate for a given patient to experience the outcome, the patient’s baseline data were assessed for the presence or absence of the risk factors, and the risk was evaluated using the prediction model. Three Kaplan–Meier curves were built for the identified risk groups (Figs. 2B, 3B, S2B) using the respective outcome data in the database. Clear separation between the curves is an indicator of high discrimination of the model.

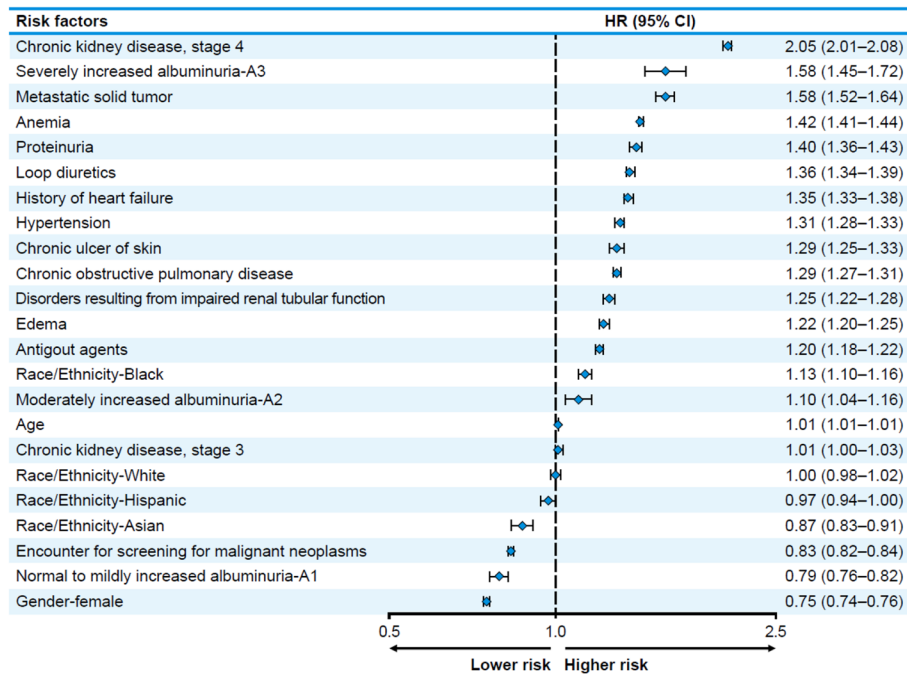
All analyses were performed using R (version 3.6.2). For survival analysis, the R packages survival (version 3.1–11), rms (version 5.1–4), and muhaz (version 1.2.6.1) were used. ACO was directly implemented in R. The R package randomForest (version 4.6–14) was used for the random forest algorithm.

## **Results**

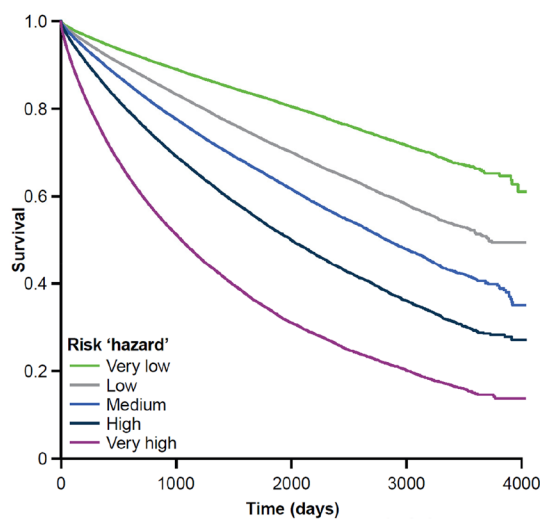
### **Patients**

The main study cohort of the FLIEDER study included 504,924 patients with non-diabetic CKD stages 3 or 4 (Fig. S1). Of these patients, 504,687 were available for clinical outcome analysis and had follow-up data. The baseline characteristics of the main cohort have been published previously [22] (Table S2). At baseline, eGFR values were available for 313,367 patients (62%); median (interquartile range) eGFR was 53.0 (47.1–57.0) ml/min/1.73 m<sup>2</sup>; and UACR was recorded for 30,793 patients (6%) of individuals, of whom 73%, 21%, and 6% had normal-to-mildly increased A1 (<30 mg/g), moderately increased A2 (30–≤300 mg/g), and severely increased A3 (>300 mg/g) albuminuria, respectively [22].

A)



B)



**Fig. 2** Kidney failure/need for dialysis in patients with non-diabetic CKD **A)** Risk prediction model; **B)** Kaplan–Meier plots for non-diabetic CKD population stratified across five risk categories. CI, confidence interval; CKD, chronic kidney disease; HR, hazard ratio

**Clinical outcomes**

Results from the primary FLIEDER analysis are already published [22]. Briefly, over a median follow-up of 744 days, 24% of patients experienced the composite primary outcome of kidney failure/need for dialysis, with an incidence rate of 10.3 events per 100 patient-years. Furthermore, approximately 11% of patients experienced the HHF outcome (incidence rate: 4.0 events per 100

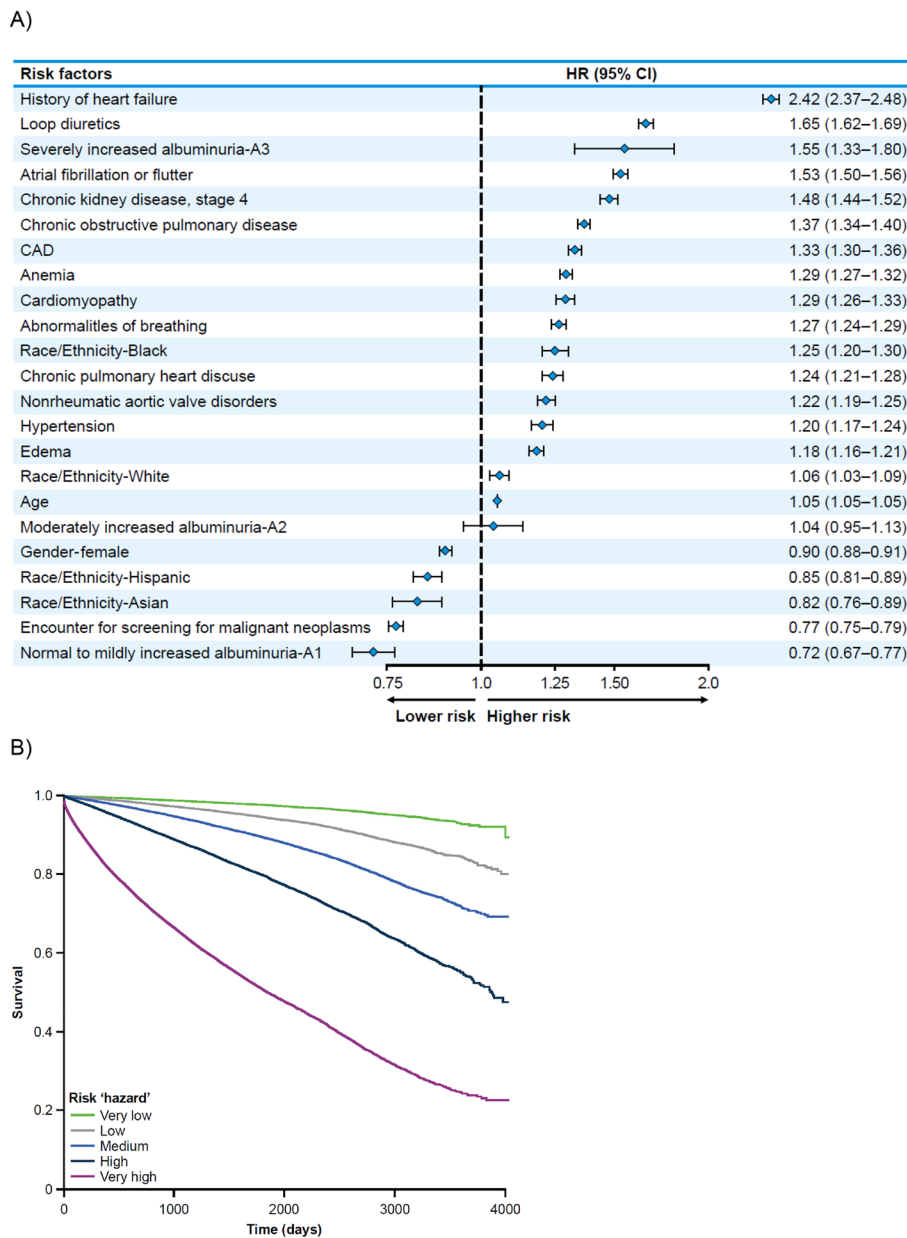
patient-years) and 11% of patients experienced the worsening of CKD stage outcome (incidence rate: 4.4 events per 100 patient-years).

**Predictive models**

**Kidney failure/need for dialysis**

The risk prediction model for the kidney failure/need for dialysis outcome in non-diabetic CKD is shown





**Fig. 3** HHF in patients with non-diabetic CKD stage 3 or 4 **A**) Risk prediction model; **B**) Kaplan–Meier plots for non-diabetic CKD population stratified across five risk categories. CI, confidence interval; CAD, coronary artery disease; CKD, chronic kidney disease; HHF, hospitalization for heart failure; HR, hazard ratio

in Fig. 2A. The five strongest risk factors identified were stage 4 CKD (HR=2.05, 95% confidence interval [CI]=2.01–2.08), severely increased albuminuria-A3 (HR=1.58, 95% CI=1.45–1.72), metastatic solid tumor (HR=1.58, 95% CI=1.52–1.64), anemia (HR=1.42, 95% CI=1.41–1.44), and proteinuria (HR=1.40, 95% CI=1.36–1.43). The influential protective factors against development of kidney failure/need for dialysis included female gender (HR=0.75,

95% CI=0.74–0.76), normal-to-mildly increased albuminuria-A1 where categories were defined in Murton et al. [25] (HR=0.79, 95% CI=0.76–0.82), screening for malignant neoplasms (HR=0.83, 95% CI=0.82–0.84), Asian (HR=0.87, 95% CI=0.83–0.91) and Hispanic (HR=0.97, 95% CI=0.94–1.00) race, where the reference group was unknown race. The Kaplan–Meier curves to display the model performance across different predicted risk groups are depicted in Fig. 2B, suggesting

good model performance based on clear separation between the curves. The reference hazard rate  $\alpha$  in the risk prediction model for kidney failure/need for dialysis is 0.063 per year.

### **HHF**

The risk prediction model for the HHF outcome in non-diabetic CKD is shown in Fig. 3A. According to the model, history of HF (HR=2.42, 95% CI=2.37–2.48), use of loop diuretics (HR=1.65, 95% CI=1.62–1.69), severely increased albuminuria-A3 (HR=1.55, 95% CI=1.33–1.80), atrial fibrillation or flutter (HR=1.53, 95% CI=1.50–1.56), and stage 4 CKD (HR=1.48, 95% CI=1.44–1.52) were the greatest risk factors that led to HHF. The influential protective factors against HHF included normal-to-mildly increased albuminuria-A1 (HR=0.72, 95% CI=0.67–0.77), screening for malignant neoplasms (HR=0.77, 95% CI=0.75–0.79), Asian (HR=0.82, 95% CI=0.76–0.89) and Hispanic (HR=0.85, 95% CI=0.81–0.89) race, where the reference group was unknown race, and female gender (HR=0.90, 95% CI=0.88–0.91). Clear separation between the risk-stratified Kaplan–Meier curves suggests good performance of the model (Fig. 3B). The reference hazard rate  $\alpha$  value for HHF is 0.033 per year.

### **Worsening of CKD stage from baseline**

The risk prediction model for worsening of CKD stage outcome in non-diabetic CKD is shown in Fig. S2A; the risk-stratified Kaplan–Meier curves are depicted in Fig. S2B. For details see Supplementary Appendix.

## **Discussion**

In this real-world evidence study of patients with non-diabetic CKD, risk prediction models were successfully developed to estimate risk of three cardiorenal clinical outcomes: kidney failure/need for dialysis, HHF, and worsening of non-diabetic CKD based on data collected in the US routine clinical practice and used for reimbursement purposes.

It is well recognized [15] that predictions based on eGFR and albuminuria laboratory tests alone may not be sufficiently powerful, and adding further parameters into risk prediction models may substantially improve quality of risk estimation. Models often need to be adjusted or newly developed depending on the setting in which they are going to be used. No unique model can fit the universal purpose of predicting individual risk of an outcome.

A well-known Kidney Failure Risk Equation [15] represents a high-performing model to predict kidney failure in patients with CKD stages 3–5. The model includes age, sex, and routinely obtained laboratory test results. However, access to administrative claims databases is often restricted among database members to only a few,

and may not even contain laboratory data. For example, in our study using the Optum CDM database, with approximately 30% of database members having available laboratory data, the number of patients with non-diabetic CKD stage 3 or 4, for whom the laboratory tests from the Kidney Failure Risk Equation were performed and the results were recorded during the 1-year baseline period, was very low (approximately 400 patients out of more than 500,000 in the main study cohort). This makes application of the Kidney Failure Risk Equation unfeasible in observational studies using secondary data sources similar to the Optum CDM or in settings where the components of the equation cannot be collected.

To the best of our knowledge, no risk prediction models exist for cardiorenal outcomes in non-diabetic CKD based on claims data collected in routine clinical practice and used for reimbursement purposes. Development of such models was the aim of the present study. Variables to define risk factors were classified into several categories representing a specific disease, medical procedure, or medication, and were assessed as present (yes/no) in the patient baseline period of 1 year. Laboratory data were not used to define risk factors in the data-driven part of the analysis due to their limited availability. However, the chosen methodology allowed for the addition of well-known risk factors for cardiorenal outcomes in CKD into the predictive models. For example, albuminuria category (A1–A3), as defined based on the UACR lab test, was added to the models “manually”. The possibility of estimating individual risk of clinical outcomes was preserved for patients with no reported UACR (i.e., with no assigned albuminuria category). A similar approach could be used for adding further laboratory-based risk factors (e.g., those used in the Kidney Failure Risk Equation) and is the subject of future research.

In the newly developed risk prediction model for kidney failure/need for dialysis in non-diabetic moderate-to-severe CKD, a diagnosis code for stage 4 CKD and severely increased albuminuria-A3 by UACR measurement were identified as the strongest predictors. These findings support previous literature that indicate that patients with CKD who have severely increased albuminuria or fall within the KDIGO high-risk or very high-risk groups experience the highest burden of the disease [25].

Hypertension has been shown to be a predictive risk factor for the incidence and severity of HF, with albuminuria also being a strong risk factor for HHF [26–29]. This aligns with findings from our risk prediction model for HHF in patients with non-diabetic CKD, where the three strongest risk factors were history of HF, loop diuretic use (used for the treatment of hypertension), and severely increased albuminuria-A3. Research indicates that one or more incidences of

HHF in patients with CKD increases the risk of ESKD, CKD progression, and mortality resulting from altered renal hemodynamics or comorbid CV disease [30, 31]. Our risk prediction models showed that a history of HF was the strongest risk factor for HHF, and also a risk factor associated with kidney failure/need for dialysis, suggesting that treatment of HF could be important in improving cardiorenal outcomes in the non-diabetic CKD population.

In our study, the strongest risk factors leading to worsening of CKD stage from baseline were a diagnosis code for stage 4 CKD, increased albuminuria (A2 or A3) by UACR measurement, other kidney diseases (i.e., polycystic kidney disease, proteinuria, renal tubulo-interstitial disease, and nephritic syndrome), anemia, hypertension, and HF. This result is in line with previous research demonstrating a link between CKD and cardiorenal events [3, 4, 25, 31].

Most observed protective factors across the three prediction models included female gender, normal-to-mildly increased albuminuria-A1, as well as Asian or Hispanic race. The findings about gender and race shall be validated further in a later study. The protective effect of the albuminuria-A1 can be explained by a lower risk of non-diabetic CKD patients with UACR of A1 category experiencing severe clinical outcome as compared with patients with UACR of categories A2 or A3.

In this study, metastatic solid tumor was among the strongest risk factors for kidney failure/need for dialysis. Patients with solid tumors are at risk of developing acute kidney injury and CKD due to the nephrotoxicity associated with many cancer therapies, and kidney disease may subsequently complicate cancer treatment [32]. This highlights the need for kidney monitoring in patients with solid tumors as well as early referral to a nephrology clinic. This study also found that screening for malignant neoplasms was associated with a reduction in the risk of kidney failure/need for dialysis and HHF. Studies suggest that individuals who attend cancer screenings have more trust in healthcare providers, a better relationship with their provider, and may be more proactive in managing their health [33]. Consequently, these individuals may be more likely to be appropriately monitored and screened for existing conditions, contributing to a reduced risk of poor clinical outcomes. Moreover, screening for malignant neoplasms may have been performed if a patient showed signs of malignancy during a visit, and some screened patients may have subsequently developed cancer and died. Death as a competing risk would reduce the number of observed outcome events and the feature “screening for malignant neoplasms” would appear to be protective for the respective outcome.

Despite an upward trend in recognition of CKD burden and consequences on patient wellbeing, nearly 50% of patients with low eGFR remain undiagnosed with CKD [34]. In addition, there is insufficiency in UACR screening in clinical practice, despite severely increased albuminuria being reported to be associated with a high burden of CKD [25]. A previously developed model for predicting risk of CKD onset and its progression in individuals with T2D names albuminuria and eGFR as the most important risk factors; however, the predictive ability of the model was found to be modest [13]. The findings from the present study also highlight the need for CKD screening through eGFR and UACR testing, as stage 4 CKD and/or severely increased albuminuria were identified as strong risk factors in all three prediction models in non-diabetic CKD. The risk factors for stage 3 or stage 4 CKD were defined by ICD-9/-10 codes. If eGFR measurements are available, the models can be applied to estimate the risk of the clinical outcomes by matching the eGFR value with the respective CKD stage and setting it as yes/no in the models accordingly.

Large volumes of individual patient data and advances in data analytics and technology have created unprecedented opportunities for emerging risk prediction models in various patient population and outcomes. The models might be implemented into electronic health records, enabling automatic calculation of patient risks during interactions with healthcare systems and encouraging treating physicians or patients to take action. Precision medicine applications with tailored treatments such as these may lead to improved clinical outcomes and higher quality of patient life.

This study generated real-world evidence on patients with non-diabetic CKD using one of the largest administrative claims databases in the US. Besides clinical characteristics and rates of cardiorenal outcomes, three risk prediction models were developed containing risk factors that are collected routinely at the point of care of non-diabetic CKD patients. The generated results may be critical for healthcare quality improvement in clinical practice, as well as in clinical research and healthcare decision-making [35, 36]. Furthermore, the models may be applied to individual risk calculation to provide patients with a better understanding of their disease. They may serve as bases for novel risk scores to predict worsening of kidney function or HF outcome in non-diabetic CKD; in order to find an application area, these models shall be validated in external data sources.

#### **Study strengths and limitations**

Unlike many of the previously reported studies focusing on diabetic CKD patients, this study focused on the



development of risk prediction models for kidney failure/need for dialysis, HHF, and worsening of CKD stage in non-diabetic CKD stages 3 or 4. The study investigated a large patient cohort comprising more than 500,000 patients treated in routine clinical practice in the US for whom the individual-level claims data used for reimbursement purposes were available.

Certain limitations of this study are relative to the data source, to the nature of the study, and to analytical methods. The study cohort was restricted to the US, and the source population was comprised of individuals in the Optum CDM database. Individuals enrolled in the Optum CDM database are largely representative of the insured US population in terms of age, sex, and region. Therefore, generalizability of the results of this study to the entire US population should be considered acceptable. However, it should be noted that data from uninsured patients is not available in Optum CDM, so these patients were not investigated. Furthermore, our study is limited to patient groups with access to the US healthcare system. Optum CDM laboratory data are collected from several large laboratory vendors and are available for a fraction (approximately 30%) of the database members, as is common in many claims databases. Laboratory examinations done in a hospital setting or directly in the physician's office are underrepresented. However, there is no reason to believe that selection bias is of concern because participation of laboratories in the Optum agreement is assumed to be random. Laboratory results for eGFR and UACR tests are the gold standard for defining CKD, although the majority of patients in the study cohort were included based on ICD codes for CKD. While this approach has limitations, it has been shown that the code-based definition of CKD stages 3–4 using claims databases has a positive predictive value of >80% [37]; therefore, it can be assumed that the findings of the study are not dependent on the method used for defining CKD.

Considerable effort was made in our study to define a non-diabetic CKD cohort as close as possible to the existing clinical definition of the disease while accounting for the limitations of the available patient data in administrative claims. "Target trial emulation" principles were followed, and the definition of non-diabetic CKD from the main contemporary randomized clinical trials was used. These trials define CKD based on eGFR or eGFR plus UACR laboratory tests and often include patients with CKD of various etiologies. In this study, patients were defined with CKD stage 3 or 4 based on CKD diagnostic codes (ICD-10-CM N18.3X, N18.4) or eGFR laboratory values (G3, G4). Moreover, eGFR values were prioritized over diagnosis codes, if both were available. No ICD codes indicating specific kidney diseases from the ICD ontology were used to build the study cohort. However,

additional kidney diseases (as per diagnostic codes) were reported for study patients during the baseline period.

The typical course of CKD progression from stage 1 to 5 lasts more than 10 years. Because of limited patient follow-up time in Optum CDM, with an average time in the database of approximately 3 years, full progression of CKD from stage 1 to kidney failure is not feasible to investigate. Therefore, this study focused on CKD progression beginning with moderate to severe stages. Both incident and prevalent cases of CKD were included and, subsequently, patients may have had a different duration of disease when entering the study cohort. To build risk prediction models, patient data were used as recorded in the baseline period of 1 year prior to the index event of CKD stage 3 or 4 diagnosis. Therefore, some risk factors may have been measured 1 year prior to index date. This shall be accounted for when interpreting the results of the study.

There are existing models that predict worsening of CKD to ESKD or HHF, but these address patient populations that are different to non-diabetic CKD. Moreover, despite large volumes of emerging literature on risk prediction models across different therapy areas and population types, analytical techniques used for building the models often consider outcomes as binary events occurring within a fixed time period, for instance 1 year, rather than applying time-to-event analysis. The latter takes censoring events and varying individual time under risk into account and is largely accepted as the method of choice in event-based clinical and observational studies. In the present study, a data-driven time-to-event-based approach accomplished by the optimization method and subject matter expertise was used to develop risk prediction models. They predict individual risk for varying future time intervals that can be plugged into the model. One limitation of the used approach is that there is no opportunity to apply standard techniques to validate model performance. Well-known methods to estimate area-under-the-curve, sensitivity, and specificity are not easily applicable for the time-to-event analysis. The developed models demonstrated high discrimination ability to separate patient risk groups, although they need to be validated in other patient populations and datasets. This represents an important limitation and requires further research.

## Conclusions

Results of the study allow prediction of individual risk for adverse cardiorenal outcomes in non-diabetic CKD and support identification of patients with this disease at high risk of such outcomes. Stage 4 CKD, severely increased albuminuria-A3, history of HF, and use of loop diuretics were identified as the most influential risk factors within the three prediction models

for kidney failure/need for dialysis, HHF, and worsening of CKD stage from baseline. The most commonly observed protective factors across the three prediction models included female gender, normal-to-mildly increased albuminuria-A1, as well as Asian or Hispanic race. The risk prediction models developed in this study have potential broad clinical applications in patient care because they include risk factors routinely collected by healthcare providers. The use of risk prediction models in clinical practice may aid healthcare decision-making and improve patient outcomes in the non-diabetic CKD population. The next steps would be to validate these models in external data sources.

#### Abbreviations

|           |  |
|-----------|--|
| ACO       | Ant colony optimization  |
| CI        | Confidence interval  |
| CKD       | Chronic kidney disease   |
| CPT-4     | Current Procedural Terminology-4   |
| CV        | Cardiovascular   |
| eGFR      | Estimated glomerular filtration rate   |
| ESKD      | End-stage kidney disease   |
| FLIEDER   | Exploratory analysis of Longitudinal patient-level Data for non-diabetic chronic kidney disease in a United States claims database |
| HCPCS     | Healthcare Common Procedure Coding System  |
| HF        | Heart failure  |
| HHF       | Hospitalization for heart failure  |
| HR        | Hazard ratio   |
| ICD       | International Classification of Diseases   |
| Optum CDM | Optum Clinformatics® Data Mart   |
| T2D       | Type 2 diabetes  |
| UACR      | Urine albumin-to-creatinine ratio  |
| US        | United States  |

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12882-024-03906-2>.

Supplementary Material 1.

#### Acknowledgements

A portion of the results reported here was presented at the European Society of Cardiology – The Digital Experience meeting, 27 to 30 August 2021 – Virtual Meeting, and a detailed technical description of the methodological approach was published in November 2024 [23]. Medical writing assistance was provided by Rosa Banuelos, PhD, of Healthcare Consultancy Group, and was funded by Bayer AG.

#### Authors' contributions

Conceptualization: all authors. Formal analysis and investigation: all authors. Writing, review, and editing: all authors. Final approval: all authors.

#### Funding

The study was funded by Bayer AG.

#### Data availability

Data sharing underlying the findings described in this manuscript is in accordance with Bayer AG data sharing policy described at link: <https://clinicaltrials.bayer.com/transparency-policy/> and can be organized via the corresponding author upon reasonable request. However, restrictions apply to these data, which were used under license to Bayer AG for the current study by a third party, Optum®, and are not publicly available. For any questions, please contact Optum, Inc.

#### Declarations

##### Ethics approval and consent to participate

Data included in the Optum® CDM database are de-identified and are in compliance with the Health Insurance Portability and Accountability Act of 1996 to preserve participant anonymity and confidentiality, and as such this study followed the principles of the Declaration of Helsinki without the requirement for review from a formal ethics review committee. The use of the provided Optum® data was determined by the New England Institutional Review Board to not constitute research involving human subjects and was therefore exempt from board oversight.

##### Consent for publication

Not applicable.

##### Competing interests

CW reports advisory board and lecture fees from AstraZeneca, Bayer, Boehringer Ingelheim, Eli-Lilly, Gilead, GSK, and MSD. CB and JS are employees of MicroDiscovery GmbH, Berlin, Germany. MB is an employee of Bayer AG, Wuppertal, Germany. FK is an employee of Bayer AG, Berlin, Germany. TV was a full-time employee of Bayer AG, Berlin, Germany at the time the study was performed and own shares in Bayer AG. TV is now an employee of Boehringer Ingelheim Pharma GmbH & Co. KG.

##### Author details

<sup>1</sup>Medizinische Klinik Und Poliklinik 1, Schwerpunkt Nephrologie, Universitätsklinik Würzburg, Würzburg, Germany. <sup>2</sup>MicroDiscovery GmbH, Berlin, Germany. <sup>3</sup>Bayer AG, Wuppertal, Germany. <sup>4</sup>Bayer AG, Berlin, Germany.

Received: 20 February 2023 Accepted: 5 December 2024

Published online: 07 January 2025

#### References

- Jager KJ, Kovesdy C, Langham R, Rosenberg M, Jha V, Zoccali C. A single number for advocacy and communication-worldwide more than 850 million individuals have kidney diseases. *Kidney Int.* 2019;96:1048–50.
- Foreman KJ, Marquez N, Dolgert A, Fukutaki K, Fullman N, McGaughey M, et al. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. *Lancet.* 2018;392:2052–90.
- Bikbov B, Purcell CA, Levey AS, Smith M, Abdoli A, Abebe M, et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* 2020;395:709–33.
- Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. *Lancet.* 2017;389:1238–52.
- Eckardt KU, Bärthlein B, Baid-Agrawal S, Beck A, Busch M, Eitner F, et al. The German Chronic Kidney Disease (GCKD) study: design and methods. *Nephrol Dial Transplant.* 2012;27:1454–60.
- 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;351:1296–305.
- Levey AS, de Jong PE, Coresh J, El Nahas M, Astor BC, Matsushita K, et al. The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. *Kidney Int.* 2011;80:17–28.
- Mathew RO, Bangalore S, Lavelle MP, Pellikka PA, Sidhu MS, Boden WE, et al. Diagnosis and management of atherosclerotic cardiovascular disease in chronic kidney disease: a review. *Kidney Int.* 2017;91:797–807.
- Tsai WC, Wu HY, Peng YS, Yang JY, Chen HY, Chiu YL, et al. Association of intensive blood pressure control and kidney disease progression in nondiabetic patients with chronic kidney disease: a systematic review and meta-analysis. *JAMA Intern Med.* 2017;177:792–9.
- Shunan F, Jiqing Y, Xue D. Effects of angiotensin-converting enzyme inhibitors and angiotensin receptor blockers on cardiovascular events in patients with diabetes and overt nephropathy: a meta-analysis of

- randomised controlled trials. *J Renin Angiotensin Aldosterone Syst.* 2018;19:1470320318803495.
11. Bakris GL, Agarwal R, Anker SD, Pitt B, Ruilope LM, Rossing P, et al. Effect of finerenone on chronic kidney disease outcomes in type 2 diabetes. *N Engl J Med.* 2020;383:2219–29.
  12. Heerspink HJL, Stefánsson BV, Correa-Rotter R, Chertow GM, Greene T, Hou FF, et al. Dapagliflozin in patients with chronic kidney disease. *N Engl J Med.* 2020;383:1436–46.
  13. Ferrero P, Iacovoni A, D'Elia E, Vaduganathan M, Gavazzi A, Senni M. Prognostic scores in heart failure — critical appraisal and practical use. *Int J Cardiol.* 2015;188:1–9.
  14. Dunkler D, Gao P, Lee SF, Heinze G, Clase CM, Tobe S, et al. Risk prediction for early CKD in type 2 diabetes. *Clin J Am Soc Nephrol.* 2015;10:1371–9.
  15. Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, et al. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA.* 2011;305:1553–9.
  16. Lerner B, Desrochers S, Tangri N. Risk prediction models in CKD. *Semin Nephrol.* 2017;37:144–50.
  17. Tangri N, Inker LA, Hiebert B, Wong J, Naimark D, Kent D, et al. A dynamic predictive model for progression of CKD. *Am J Kidney Dis.* 2017;69:514–20.
  18. Ferguson T, Ravani P, Sood MM, Clarke A, Komenda P, Rigatto C, et al. Development and external validation of a machine learning model for progression of CKD. *Kidney Int Rep.* 2022;7:1772–81.
  19. Stempniewicz N, Vassalotti JA, Cuddeback JK, Ciemins E, Storfer-Isser A, Sang Y, et al. Chronic kidney disease testing among primary care patients with type 2 diabetes across 24 U.S. health care organizations. *Diabetes Care.* 2021;44:2000–9.
  20. Tangri N, Kitsios GD, Inker LA, Griffith J, Naimark DM, Walker S, et al. Risk prediction models for patients with chronic kidney disease: a systematic review. *Ann Intern Med.* 2013;158:596–603.
  21. Lai TS, Tsao HM, Chou YH, Liang SL, Chien KL, Chen YM. A competing risk predictive model for kidney failure in patients with advanced chronic kidney disease. *J Formos Med Assoc.* 2023; <https://doi.org/10.1016/j.jfma.2023.11.010>.
  22. Wanner C, Schuchhardt J, Bauer C, Lindemann S, Brinker M, Kong SX, et al. Clinical characteristics and disease outcomes in non-diabetic chronic kidney disease: retrospective analysis of a US healthcare claims database. *J Nephrol.* 2023;36:45–54.
  23. Bauer C, Schuchhardt J, Vaitiakhovich T, Kleinjung F. Computational and human intelligence methods for constructing practical risk prediction models: An application to cardio-renal outcomes in non-diabetic CKD patients. *Int J Comput Intell Syst.* 2024;17:276.
  24. Dorigo M, Birattari M, Stützle T. Ant colony optimization. *IEEE Comput Intell Mag.* 2006;1:28–39.
  25. Murton M, Goff-Leggett D, Bobrowska A, Garcia Sanchez JJ, James G, Wittbrodt E, et al. Burden of chronic kidney disease by KDIGO categories of glomerular filtration rate and albuminuria: a systematic review. *Adv Ther.* 2021;38:180–200.
  26. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. *Nat Rev Cardiol.* 2011;8:30–41.
  27. Ziaieian B, Fonarow GC. Epidemiology and aetiology of heart failure. *Nat Rev Cardiol.* 2016;13:368–78.
  28. Hockensmith ML, Estacio RO, Mehler P, Havranek EP, Ecker ST, Lundgren RA, et al. Albuminuria as a predictor of heart failure hospitalizations in patients with type 2 diabetes. *J Card Fail.* 2004;10:126–31.
  29. Liang W, Liu Q, Wang QY, Yu H, Yu J. Albuminuria and dipstick proteinuria for predicting mortality in heart failure: a systematic review and meta-analysis. *Front Cardiovasc Med.* 2021;8:665831.
  30. Sud M, Tangri N, Pintilie M, Levey AS, Naimark DMJ. ESRD and death after heart failure in CKD. *J Am Soc Nephrol.* 2015;26:715–22.
  31. Bansal N, Zelnick L, Bhat Z, Dobre M, He J, Lash J, et al. Burden and outcomes of heart failure hospitalizations in adults with chronic kidney disease. *J Am Coll Cardiol.* 2019;73:2691–700.
  32. Yarandi N, Shirali AC. *Onconeurology: Core Curriculum 2023.* *Am J Kidney Dis.* 2023;82:743–61.
  33. Young B, Bedford L, Kendrick D, Vedhara K, Robertson JFR, das Nair R. Factors influencing the decision to attend screening for cancer in the UK: a meta-ethnography of qualitative research. *J Public Health (Oxf).* 2018;40:315–39.
  34. Tuttle KR, Alicic RZ, Duru OK, Jones CR, Daratha KB, Nicholas SB, et al. Clinical characteristics of and risk factors for chronic kidney disease among adults and children: an analysis of the CURE-CKD registry. *JAMA Netw Open.* 2019;2:e1918169.
  35. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiol Drug Saf.* 2017;26:1033–9.
  36. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - what is it and what can it tell us? *N Engl J Med.* 2016;375:2293–7.
  37. Paik JM, Patorno E, Zhuo M, Bessette LG, York C, Gautam N, et al. Accuracy of identifying diagnosis of moderate to severe chronic kidney disease in administrative claims data. *Pharmacoepidemiol Drug Saf.* 2022;31:467–75.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.