



Universiteit
Leiden
The Netherlands

Accounting for the competing risk of death to predict kidney failure in adults with stage 4 chronic kidney disease

Al-Wahsh, H.; Tangri, N.; Quinn, R.; Liu, P.; Ferguson, T.; Fiocco, M.; ... ; Ravani, P.

Citation

Al-Wahsh, H., Tangri, N., Quinn, R., Liu, P., Ferguson, T., Fiocco, M., ... Ravani, P. (2021). Accounting for the competing risk of death to predict kidney failure in adults with stage 4 chronic kidney disease. *Jama Network Open*, 4(5). doi:10.1001/jamanetworkopen.2021.9225

Version: Publisher's Version
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)
Downloaded from: <https://hdl.handle.net/1887/3277370>

Note: To cite this publication please use the final published version (if applicable).



Accounting for the Competing Risk of Death to Predict Kidney Failure in Adults With Stage 4 Chronic Kidney Disease

Huda Al-Wahsh, PhD; Navdeep Tangri, MD, PhD; Rob Quinn, MD; Ping Liu, PhD; Thomas Ferguson, MS; Marta Fiocco, PhD; Ngan N. Lam, MD, MSc; Marcello Tonelli, MD; Pietro Ravani, MD, PhD

Abstract

IMPORTANCE Kidney failure risk prediction has implications for disease management, including advance care planning in adults with severe (ie, estimated glomerular filtration rate [eGFR] category 4, [G4]) chronic kidney disease (G4-CKD). Existing prediction tools do not account for the competing risk of death.

OBJECTIVE To compare predictions of kidney failure (defined as estimated glomerular filtration rate [eGFR] <10 mL/min/1.73 m² or initiation of kidney replacement therapy) from models that do and do not account for the competing risk of death in adults with G4-CKD.

DESIGN, SETTING, AND PARTICIPANTS This prognostic study linked population-based laboratory and administrative data (2002-2017) from 2 Canadian provinces (Alberta and Manitoba) to compare 3 kidney risk models: the standard Cox regression, cause-specific Cox regression, and Fine-Gray subdistribution hazard model. Participants were adults with incident G4-CKD (eGFR 15-29 mL/min/1.73 m²). Data analysis occurred between July and December 2020.

MAIN OUTCOMES AND MEASURES The performance of kidney risk models at prespecified times and across categories of baseline characteristics, using calibration, reclassification, and discrimination (for competing risks). Predictive characteristics were age, sex, albuminuria, eGFR, diabetes, and cardiovascular disease.

RESULTS The development and validation cohorts included 14 619 (7070 [48.4%] men; mean [SD] age, 74.1 [12.8] years) and 2295 (1152 [50.2] men; mean [SD] age, 71.9 [14.0] years) adults, respectively. The 3 models had comparable calibration up to 2 years from entry. Beyond 2 years, the standard Cox regression overestimated the risk of kidney failure. At 4 years, for example, risks predicted from standard Cox were 40% for people whose observed risks were less than 30%. At 2 years (risk cutoffs 10%-20%) and 5 years (risk cutoffs 15%-30%), 788 (5.4%) and 2162 (14.8%) people in the development cohort were correctly reclassified into lower- or higher-risk categories by the Fine-Gray model and incorrectly reclassified by standard Cox regression (the opposite was observed in 272 patients [1.9%] and 0 patients, respectively). In the validation cohort, 115 (5.0%) individuals and 389 (16.9%) individuals at 2 and 5 years, respectively, were correctly reclassified into lower- or higher-risk categories by the Fine-Gray model and incorrectly reclassified by the standard Cox regression; the opposite was observed in 98 (4.3%) individuals and 0 individuals, respectively. Differences in discrimination emerged at 4 to 5 years in the development cohort and at 1 to 2 years in the validation cohort (0.85 vs 0.86 and 0.78 vs 0.8, respectively). Performance differences were minimal during the entire follow-up in people at lower risk of death (ie, aged ≤65 years or without cardiovascular disease or diabetes) and greater in those with a higher risk of death. At 5 years, for example, in people aged 65 years or older, predicted risks from standard Cox were 50% where

(continued)

Key Points

Question Does accounting for the competing risk of death make a difference when predicting kidney failure in adults with stage 4 chronic kidney disease?

Findings In this external prognostic study with 14 619 people in the development cohort and 2295 in the validation cohort, models that did and did not account for the competing risk of death provided comparable 2-year predictions of kidney failure. Differences in model predictions emerged after 2 years and increased with longer prediction times, especially among participants aged 65 years or older and those with more comorbidity.

Meaning These findings suggest that accounting for the competing risk of death when making predictions about kidney failure becomes increasingly important with longer follow-up time, older age, and the presence of more comorbidities.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Abstract (continued)

observed risks were less than 30%. Similar miscalibration was observed at 5 years in people with albuminuria greater than 30 mg/mmol, diabetes, or cardiovascular disease.

CONCLUSIONS AND RELEVANCE In this study, predictions about the risk of kidney failure were minimally affected by consideration of competing risks during the first 2 years after developing G4-CKD. However, traditional methods increasingly overestimated the risk of kidney failure with longer follow-up time, especially among older patients and those with more comorbidity.

JAMA Network Open. 2021;4(5):e219225. doi:10.1001/jamanetworkopen.2021.9225

Introduction

Chronic kidney disease (CKD) is defined by an estimated glomerular filtration rate (eGFR) less than 60 mL/min/1.73 m² and affects 10% to 16% of the general population globally.¹ CKD is associated with increasing risk of morbidity and mortality as kidney function declines. For this reason, people with severe (ie, eGFR category 4, [G4]) CKD (ie, an eGFR of 15-29 mL/min/1.73 m²) and a high predicted risk of kidney failure are advised to discuss treatment options for kidney failure, including kidney replacement therapy (ie, dialysis or kidney transplantation) or conservative (ie, palliative) care. However, most patients with G4-CKD are 70 years of age or older and are 3 to 10 times more likely to die than develop kidney failure.² Therefore, accounting for the competing risk of death could be an important consideration for clinical decision-making.

Kidney failure risk calculators are routinely used in clinical practice to risk-stratify patients and inform nephrology referral or treatment decisions.^{3,4} These tools were developed using methods that do not account for the competing risk of death and, thus, overestimate the risk of kidney failure.^{5,6} Censoring for a competing event is assumed to leave the risk of the event of interest unaltered, but the risk of kidney failure becomes 0 after death, and many people with CKD die without kidney failure.^{2,6} The extent of this bias is unknown, especially in short-term predictions when risk overestimation because of censoring for competing risks may be minimal. Understanding how failing to account for the competing risk of death may affect the predictions made by such kidney failure risk calculators is important because they are already implemented in many electronic clinical support systems.

In this population-based study, we identified adults with incident G4-CKD and examined the prognostic performance of standard Cox regression with 2 methods that account for competing risks—the cause-specific Cox regression and the Fine-Gray model. We developed and assessed the performance of models using data from 2 separate CKD cohorts at prespecified times and across categories of baseline characteristics to determine when, and in whom, differences between models that do and do not account for competing risks are most pronounced.

Methods

Study Design and Data Sources

Both the development cohort (from Alberta, Canada) and the validation cohort (from Manitoba, Canada) for this prognostic study were formed using population-based, linked, provincial administrative and laboratory data. The institutional ethics review boards at the University of Alberta, University of Calgary, and University of Manitoba approved this study with a waiver of participant consent because retrospective, deidentified data were used. We followed the Reporting of Studies Conducted Using Observational Routinely Collected Data (RECORD) reporting guideline and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline.^{7,8}

Population

We identified 2 provincial cohorts of residents from Alberta and Manitoba who were at least 18 years old and had G4-CKD (ie, an eGFR of 15-29 mL/min/1.73 m²). We calculated eGFR using the chronic kidney disease epidemiology equation, with serum creatinine values standardized to isotope dilution mass spectrometry traceable methods.⁹ Given the lack of information on race and the low prevalence of people who self-identify as Black in Alberta (3.3%)¹⁰ and Manitoba (2.4%),¹¹ we calculated eGFR assuming that all participants were White, as in our previous work.⁵ We applied a moving average eGFR method to identify newly documented G4-CKD cases using outpatient laboratory measurements between July 30, 2002, and March 31, 2014, for the development cohort, and between April 1, 2008, and March 31, 2014, for the validation cohort. This method minimizes the inclusion of people with acute kidney injury, unstable clinical conditions, or preexisting G4-CKD or kidney failure.¹² We determined the mean eGFR calculated using repeated measures within an individual recorded over a period of at least 90 days starting on the date of the first eGFR of less than 30 mL/min/1.73 m², provided that there were at least 2 measurements to calculate the mean. In addition, the first and the last eGFR were separated by more than 90 days, and all intervening measurements were within 90 days. We used the minimum value of eGFR when there were multiple measurements on the same day. Participants met the study entry criterion if the mean eGFR (index eGFR) during this period was 15 to 30 mL/min/1.73 m². We used the date of the last eGFR measurement included in the calculation of the mean eGFR to define cohort entry (index date). We excluded patients who had received kidney replacement therapy or had an eGFR of less than 15 mL/min/1.73 m² before the index date. We also excluded people without information on albuminuria because this variable has been included in all existing kidney failure risk calculators.^{3,4}

Independent Variables

We incorporated baseline covariates that are known to be associated with kidney failure in the models^{3,4}: age, sex, diabetes, cardiovascular disease, index eGFR, and albuminuria. We identified comorbid conditions using the *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* and *ICD-10* applied to physician claims and hospitalization data prior study entry (eTable 1 in the [Supplement](#)).¹³ We defined baseline albuminuria as the most recent albuminuria value on or within the 2 years preceding the index date.

Outcomes and Follow-up Time

The outcome was kidney failure,^{2,5} which we defined as the earliest initiation of kidney replacement therapy (dialysis or kidney transplantation) or having an eGFR of less than 10 mL/min/1.73 m². Dialysis initiation was ascertained by at least 1 inpatient or outpatient physician claim and from the provincial registry of chronic dialysis. The receipt of a kidney transplant was based on at least 1 physician claim or hospitalization (eTable 2 in the [Supplement](#)). For the eGFR criterion, we applied the same moving average method we used to define cohort entry starting on the date of the first inpatient or outpatient eGFR of less than 10 mL/min/1.73 m² over a period of more than 90 days and provided that there were at least 2 measurements.¹² The event date was defined as the date of the last eGFR measured during this period. In both cohorts, patients were followed from the date of cohort entry until kidney failure, death, study end (March 31, 2017), emigration from the province, or 10 years from cohort entry. To minimize bias in outcome ascertainment, we censored observations at 1.5 years from an eGFR if no subsequent measurement was available within 1.5 years of this eGFR, as in previous studies.^{2,5}

Statistical Analysis

We estimated crude risks using the Kaplan-Meier method, which censors for competing events, and the Aalen-Johansen method, which accounts for competing events. We developed and validated 3 semiparametric models for the development and validation cohort, respectively. The standard Cox regression of kidney failure treated death as a noninformative censoring event, ie, an event that does

not alter the risk of the event of interest.¹⁴ The model assumes a 1-to-1 relationship between risk of the event of interest at a certain point in time and the cumulative hazard for that event. We used the following 2 models to account for the competing risk of death: the cause-specific hazard model of kidney failure and death and the Fine-Gray subdistribution hazard model. In the cause-specific hazard model, the standard Cox regression is used to model each competing event, and persons who move to another disease state (ie, death for the kidney failure model and kidney failure for the death model) are censored at their transition time. The cumulative incidence function for a specific cause k not only depends on the hazard of cause k , but also on the hazards of all other causes,¹⁵ and is obtained indirectly from the model parameters.^{16,17} The Fine-Gray model of kidney failure estimates the covariate effects on the cumulative incidence function of kidney failure in terms of subhazard ratios and links the cumulative incidence function to the subhazard function of kidney failure directly.^{18,19}

We included age, sex, eGFR, albuminuria, diabetes, and presence of cardiovascular disease in all models and tested all possible first-order interactions among these variables, retaining interaction terms that improved the goodness-of-fit of the model. We used Martingale residuals to assess the association between continuous covariates and outcome considering the use of quadratic terms or log-transformations to model nonlinear relationships. We used residual analyses to identify deviations from the proportionality assumption, influential observations, and outliers and assess the goodness-of-fit.^{20,21} During model building, we checked that the results were consistent across study time.

We assessed model performance using different methods. In calibration analysis, we evaluated the graphical agreement between observed and predicted risks at 1 to 8 years. In an ideal model, pairs of the observed and predicted risks lie on a 45° angle line. In reclassification analysis, we used the cutoffs at 2 years of 0% to less than 10%, 10% to less than 20%, and 20% or more at 2 years, and at 5 years of 0% to less than 15%, 15% to less than 30%, and 30% or more to define low-, intermediate-, and high-risk categories of kidney failure. These cutoffs were used in previous studies³ or existing recommendations (10% is the 2-year risk threshold used for multidisciplinary kidney care referral).²² We obtained individual risk predictions and populated a 3 × 3 table according to the predictions of rival models. We calculated the observed risk of the members assigned to each cell using the Aalen-Johansen function and obtained the total number and percentage of people who were incorrectly and correctly classified by each model. This approach addresses some limitations with net reclassification²³ and has been proposed for the competing risks settings.²⁴ We also used model concordance (C statistic) with methods for competing risks to assess the model's ability to separate individuals with kidney failure from those without kidney failure, ranging from 0.5 (no prediction ability beyond chance) to 1 (perfect discrimination).^{23,25,26} In general, a C index of 0.5 suggests no discrimination, 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding.²⁷ We also plotted time-dependent Brier scores accounting for censoring. Smaller scores indicated better performance in terms of both discrimination and calibration.²⁸

We repeated these analyses across categories defined by the following baseline characteristics: patients 65 years or younger vs patients older than 65 years, sex, albumin-to-creatinine ratio of 30 mg/mmol or less vs greater than 30 mg/mmol (to convert albumin to grams per liter, multiply by 10; creatinine to micromoles per liter, multiply by 88.4), eGFR of 27 or less vs greater than 27 mL/min/1.73 m² (the median eGFR in previous studies),^{2,5} diabetes status, and cardiovascular disease. We also repeated the reclassification analysis considering different thresholds,²⁹ including those used in general practice for nephrology referral (5% at 5 years) or preparation for dialysis (40% at 2 years).³⁰ In 1 sensitivity analysis, we assessed the consistency of results in subgroups defined by index date on or before vs after the median. In another sensitivity analysis, we defined kidney failure by the initiation of kidney replacement or the occurrence of sustained (as opposed to moving average) eGFR of less than 10 mL/min/1.73 m². We defined sustained eGFR of less than 10 mL/min/1.73 m² by the occurrence of 2 or more consecutive eGFR values of less than 10 mL/min/1.73 m² for more than

90 days. We used the date of the last eGFR measurement as the event date. We used the packages `prodlm`, `cmprsk`, and `riskRegression` in R version 4.0.3 for all analyses (R Project for Statistical Computing). We performed all analyses between July and December 2020.

Results

Cohort Description

We identified 14 619 people who met the study eligibility criteria in the development and 2295 who met the eligibility criteria in the validation cohorts (eFigure 1 in the [Supplement](#)). The 2 cohorts had a comparable distribution of sex and index eGFR (7070 men [48.4%] in the development cohort and 1152 men [50.2%] in the validation cohort) and index eGFR (median [IQR] 27.6 [25.1-29] mL/min/1.73 m² vs 25 [22-27] mL/min/1.73 m²). The development cohort was slightly older than the validation cohort (mean [SD] age, 74.1 [12.8] years vs 71.9 [14] years), had lower albuminuria (median [IQR] albumin-creatinine ratio, 7.6 [1.7-56.2] mg/mmol vs 19.7 [3.3-136.1] mg/mmol), and included more people with diabetes (9886 [67.6%] vs 1448 [63.1%]) and cardiovascular disease (8285 [56.7%] vs 796 [34.7%]) (**Table**). Baseline characteristics of the development cohort (which had longer accrual time) did not change over time (eTable 3 in the [Supplement](#)). The incidence rate of kidney failure was lower in the development cohort than in the validation cohort, while mortality rates were similar in both cohorts (incidence rate of kidney failure, 6.1 [95% CI, 5.9-6.4] events per 100 person-years vs 10.3 [95% CI, 9.5-11.0] events per 100 person-years; mortality rate, 12.3 [95% CI, 12.0-12.6] events per 100 person-years vs 12.4 [95% CI, 11.6-13.3] events per 100 person-years).

Crude Risks

Figure 1 shows the crude risks in the 2 cohorts estimated by the naïve Kaplan-Meier method and the competing risk method. The naïve Kaplan-Meier method results in risk overestimation, which begins to be observed before year 2. The sum of the 2 Kaplan-Meier failure functions corresponding to the 2 competing events (kidney failure and death) was more than 1 before year 8 in the development cohort and year 6 in the validation cohort (which is impossible because probabilities or risks range between 0 and 1).

Table. Baseline Characteristics of Participants in the Development and Validation Cohorts

Characteristic	No. (%)	
	Development cohort (n = 14 619)	Validation cohort (n = 2295)
Age, mean (SD), y	74.1 (12.8)	71.9 (14.0)
Men	7070 (48.4)	1152 (50.2)
Women	7549 (51.6)	1143 (49.8)
eGFR, median (IQR), mL/min/1.73 m ²	27.6 (25.1-29.0)	25.0 (22.0-27.0)
ACR, median (IQR), mg/mmol	7.6 (1.7-56.2)	19.7 (3.3-136.1)
Diabetes	9886 (67.6)	1448 (63.1)
Cardiovascular disease		
Any	8285 (56.7)	796 (34.7)
Myocardial infarction	1791 (12.3)	147 (6.4)
Congestive heart failure	5735 (39.2)	602 (26.2)
Stroke or TIA	3594 (24.6)	213 (9.3)
Peripheral vascular disease	1217 (8.3)	52 (2.3)
Follow-up time, median (IQR), y	3.2 (1.5-5.2)	2.9 (1.3-4.3)
Competing events		
Kidney failure	3265 (22.3)	722 (37.5)
Incidence per 100 person-years (95% CI)	6.1 (5.9-6.4)	10.3 (9.5-11.0)
Death without kidney failure	6528 (44.7)	875 (38.1)
Incidence per 100 person-years (95% CI)	12.3 (12.0-12.6)	12.4 (11.6-13.3)

Abbreviations: ACR, albumin-to-creatinine ratio; eGFR, estimated glomerular filtration rate; IQR, interquartile range; TIA, transient ischemic attack.

SI conversions: To convert albumin to grams per liter, multiply by 10; creatinine to micromoles per liter, multiply by 88.4.

Model Development

The final models are summarized in eTable 4 in the Supplement. The association between age and kidney failure was nonlinear, and the proportional assumption was satisfied for the hazard and subdistribution hazard scales by introducing a quadratic term for age. We found significant interactions between age (and its squared term) and cardiovascular disease and between log-albuminuria and cardiovascular disease.

Model Prediction Performance in Development and Validation Cohorts

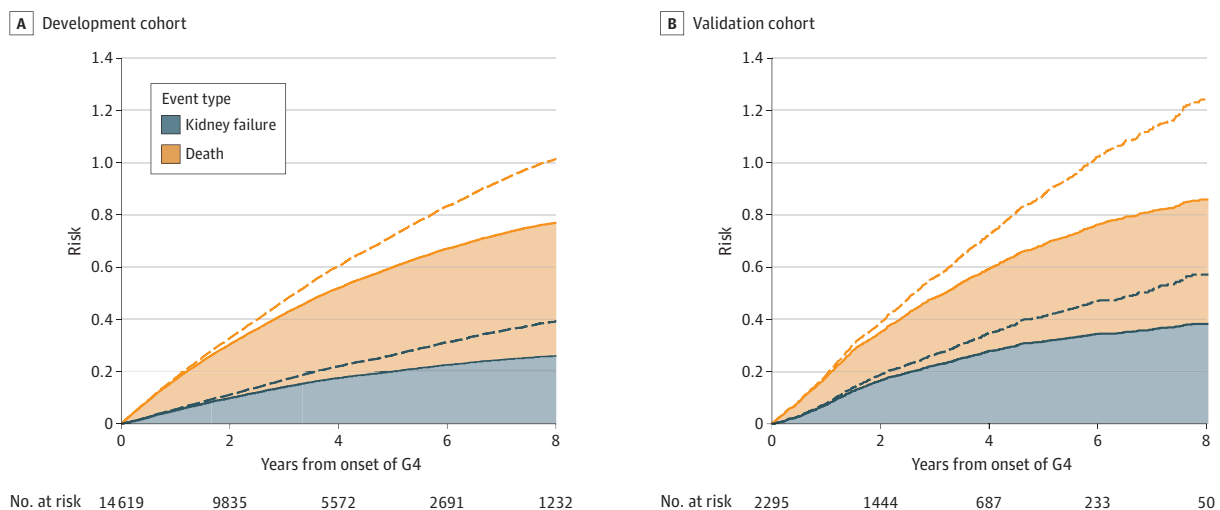
Calibration

Calibration of the 3 models was similar in short-term predictions (years 1 to 2). Beyond 2 years, the risks predicted by the standard Cox regression exceeded observed risks, while those from the competing risks models were closer to the 45° angle line for almost the entire risk range (Figure 2; eFigure 2 to 4 in the Supplement). At 4 years, for example, risks predicted from standard Cox were 40% for people whose observed risks were less than 30% (Figure 2).

Reclassification

At 2 and 5 years, 788 people (5.4%) and 2162 people (14.8%) were correctly reclassified into lower- or higher-risk categories by the Fine and Gray model and incorrectly reclassified by standard Cox regression (the opposite was observed in 272 people [1.9%] and 0 people, respectively) in the development cohort (Figure 3). For the comparison between the cause-specific Cox regression and the standard Cox regression, the corresponding figures were 569 people (3.9%) and 2025 people (13.9%) (the opposite was observed 0 people in either cohort) (eFigure 5 in the Supplement). In the validation cohort, 115 (5.0%) and 389 (16.9%) of people at 2 and 5 years, respectively, were correctly reclassified into lower- or higher-risk categories by the Fine and Gray model and incorrectly reclassified by standard Cox regression (the opposite was observed in 98 people [4.3%] and 0 people, respectively). For the comparison between cause-specific Cox regression and standard Cox regression, the corresponding figures were 135 people (5.9%) and 393 people (17.1%) (the opposite was observed in 0 individuals in both cohorts).

Figure 1. Cumulative Incidence Functions vs Kaplan-Meier Failure Functions



Dashed lines indicate Kaplan-Meier estimator functions; solid lines, cumulative incidence functions; G4, severe CKD. All curves are stacked. The sum of the 2 Kaplan-Meier failure functions corresponding to the 2 competing events (kidney failure and death) is more

than 1 before year 8 in the development cohort (A) and year 6 in the validation cohort (B).

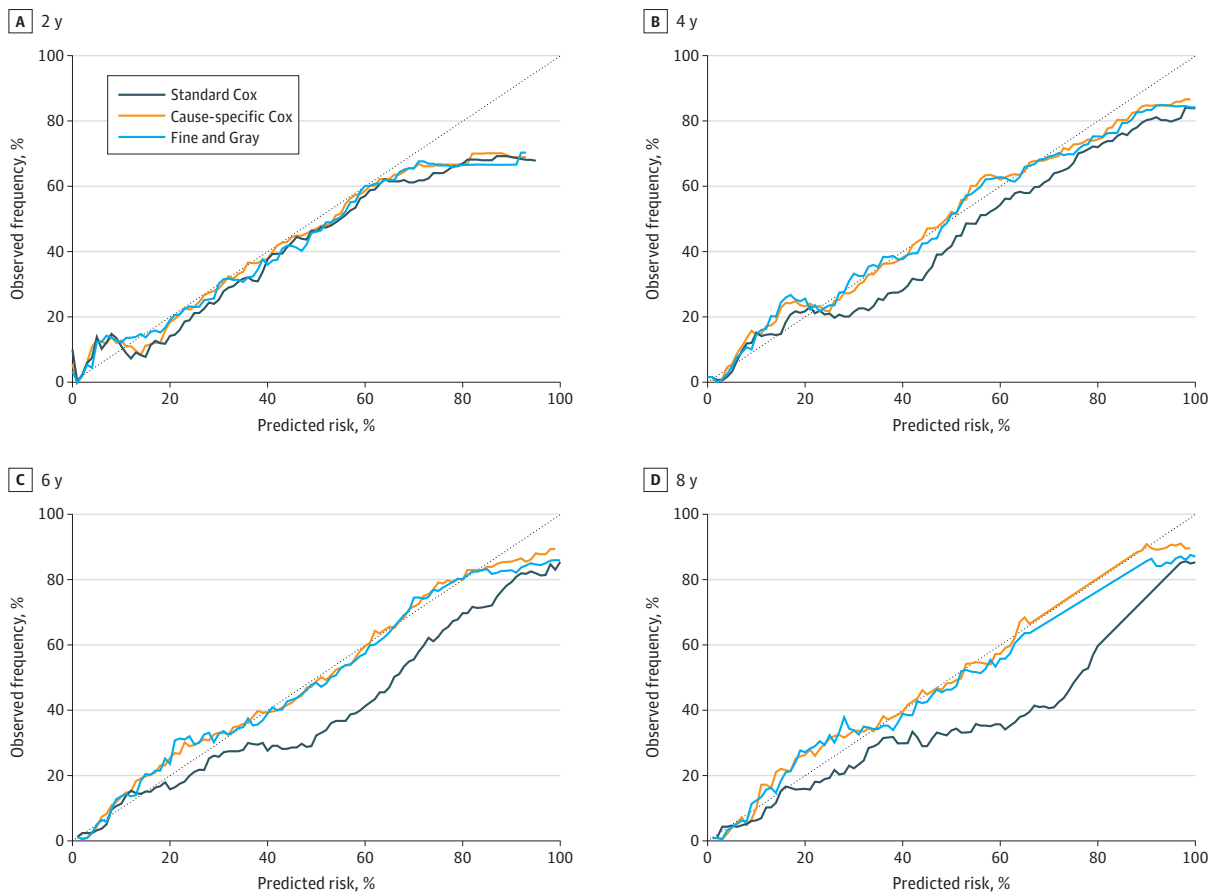
Brier Score and C Index

The time-dependent Brier scores of the 3 models started to separate at 4 to 5 years in the development cohort and at 3 to 4 years in the validation cohort (Figure 4; eTable 5 in the Supplement). The time-dependent C index of the 3 models ranged between 0.84 and 0.87 in the development cohort and between 0.74 and 0.84 in the validation cohort (Figure 4; eTable 5 in the Supplement). The curves of the c index started to separate at 4 to 5 years in the development cohort and at 2 years in the validation cohort (0.85 vs 0.86 and 0.78 vs 0.8, respectively).

Subgroup and Sensitivity Analysis

Differences in predictions between standard Cox regression and competing risks models were reduced in people at lower risk of death and vice versa (eFigure 6-11 in the Supplement). At 5 years, for example, in people aged 65 years or older, predicted risks from standard Cox were 50% where observed risks were less than 30%. Similar mis-calibration was observed at 5 years in people with albuminuria greater than 30 mg/mmol, diabetes or cardiovascular disease. These differences varied in magnitude at 2 years depending on the thresholds for risk categories used for reclassification but were consistent in magnitude at 5 years across different thresholds (eFigure 12-15, eTable 6 and 7 in the Supplement). For example, at 5 years 17% of people aged 65 years or older were correctly reclassified into lower- or higher-risk categories by competing risk models and incorrectly reclassified by standard Cox regression (vs 5% of people younger than 65 years), and the opposite was observed

Figure 2. Calibration Plots at 2, 4, 6, and 8 Years in the Validation Cohort



Calibration plots summarize the graphical agreement between observed and predicted risks at years 1 through 8. In an ideal model, pairs of the observed and predicted risks lie on a 45-degree angle line. Curves falling under the 45-degree angle line indicate that

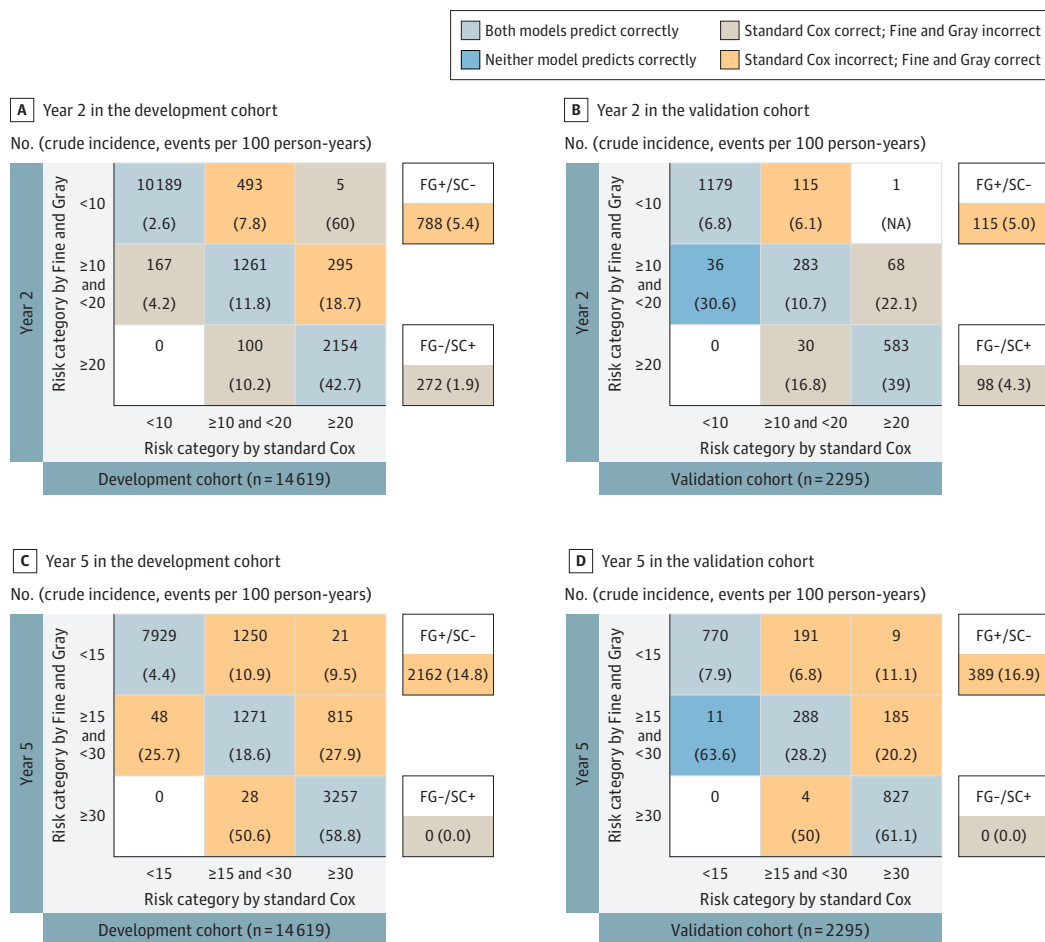
predicted risks overestimate (are higher than) observed risks. Corresponding plots at years 1 through 8 for the development cohort and for years 1, 3, 5, and 7 for the validation cohort are provided in the eFigures 2-4 in the Supplement.

in 0 people. Reclassification improvement was even more pronounced in people with albuminuria greater than 30 mg/mmol, diabetes or cardiovascular disease (18% to 24%). We obtained similar results in other sensitivity analyses (eFigure 16-18 in the Supplement).

Discussion

We developed and validated 2 models to predict kidney failure in adults with G4-CKD accounting for the competing risk of death. The risk of kidney failure varies in adults with severe CKD depending on age, sex, index eGFR, degree of albuminuria,^{3,4} and comorbidities.^{5,31} The 2 competing risk models we built with these predictors demonstrated excellent accuracy (defined as well-calibrated models with C index 0.8-0.9) for the entire risk range and prediction time. The standard Cox regression, which does not account for the competing risk of death, provided similar performance during the first 2 years after onset of G4-CKD, but differences began to emerge thereafter, especially among people older than 65 years or with more comorbidities. The traditional approach overestimates the risk of kidney failure for people with a higher risk of death. We built these models using comorbidity and laboratory data as in previous studies.^{3,4} If confirmed in other settings, our findings suggest that

Figure 3. Predicted vs Observed Risk of Kidney Failure at Years 2 and 5 in the Development and Validation Cohorts



Risk (%) was predicted for each member of the development (A, C) and validation (B, D) cohort according to the Fine and Gray subdistribution hazard model and the standard Cox model at years 2 and 5 from study entry. People were then assigned to each cell of a 3 × 3 table corresponding to the combination of the model predictions. Each cell of the 3 × 3 table includes the number of people (top) and their actual observed risk (crude

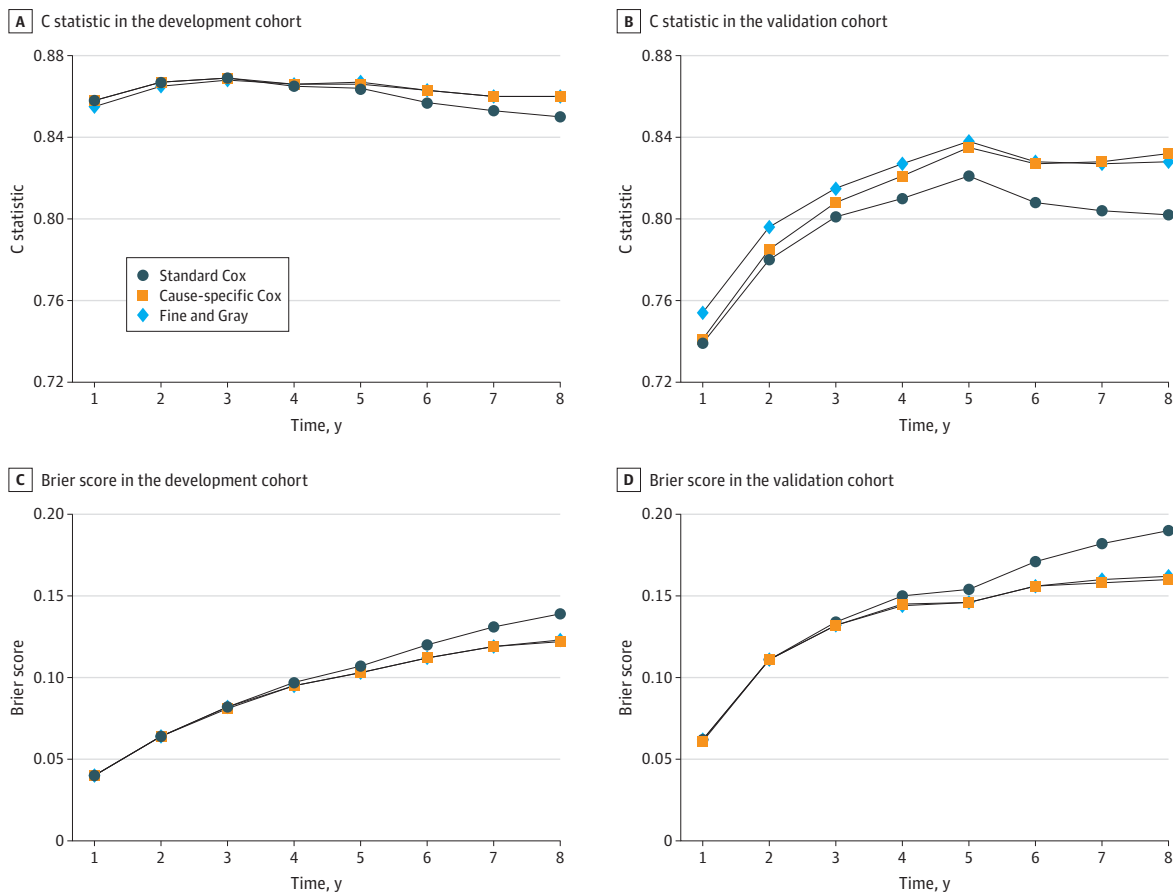
cumulative incidence function) at 2 or 5 years (bottom, bold). FG-/SC+, total No. (%) of people incorrectly classified by Fine-Gray model and correctly classified by standard Cox regression with respect to the actual observed risk; FG+/SC-, total No. (%) of people correctly classified by the Fine-Gray model and incorrectly classified by standard Cox regression with respect to the actual observed risk; NA indicates not available.

existing clinical risk calculators could be updated to use similar methods and perhaps improve the accuracy of predictions about the risk of kidney failure.

Using standard Cox regression to predict kidney failure results in risk overestimation because standard Cox regression treats death as a censoring event. Censoring is assumed to leave the risk of the event of interest unaltered, but the risk of kidney failure becomes 0 after death. However, the impact of censoring for death in an analysis of the risk of kidney failure may be minimal if the competing risk is infrequent.³² With longer prediction times or in the presence of a high risk of death relative to the risk of kidney failure, censoring for death could result in kidney failure risk predictions that are increasingly biased upward. This is especially true among people with severe CKD, many of whom are older and have multiple comorbidities. These limitations should be acknowledged if standard Cox regression will be used, for example, because of computational issues with competing risks modeling or because the calculator based on standard Cox regression has already been implemented in the decision support system to guide nephrology referral or inform advanced care planning.²

Our findings are consistent with previous studies that showed similar performance of methods that do and do not account for competing risks in short-term predictions and in people who have higher risk of kidney failure than death.^{3,4} Indeed, small differences between these methods may be explained by the inclusion of adults with mild to severe CKD, younger and more selected populations (patients referred to a nephrologist),⁶ and shorter follow-up times (2 to 4 years) in those studies.

Figure 4. C Statistics and Brier Score in the Development and Validation Cohort



The C statistic in the development cohort (A) and validation cohort (B) assesses the ability of separating people with kidney failure from those without kidney failure, ranging from 0.5 (no prediction ability beyond chance) to 1 (perfect discrimination). Smaller Brier

scores in the development cohort (C) and validation cohort (D) indicate better performance in terms of both discrimination and calibration.

Our study is novel, as it compares standard Cox regression with competing risks models over a longer follow-up time and using recommended measures of model performance.⁸ We have considered the limitations of Cox models in the competing risks setting by emphasizing the role of calibration²³ and using adapted reclassification methodology.²⁴ Considering that many adults with severe CKD are 75 years or older and have a higher risk of death than kidney failure, our findings suggest that predictions based on a competing risks model may be warranted to minimize the risk of unnecessary interventions or overtreatment.⁵

We used population-based data from a geographically defined area served by a universal health care system, a relatively large sample size, and a long follow-up to develop and validate competing risks models that demonstrated accurate predictions. The excellent external validation results support the use of these equations in diverse populations, including relatively younger and healthier populations in whom the competing risk of death may be lower than in the population we used to develop these models. These models are based on information that is routinely collected in clinical practice, which can be integrated into electronic support systems for clinical management and future decision aids. The use of population-based data suggests that our findings are broadly generalizable to people with severe CKD (rather than the select population who have been referred to nephrologists)⁶ at least in those who are White and reside in North America and Europe. We used rigorous methods, as in our previous studies,² for ascertaining the presence or absence of comorbidity, defining eligibility criteria to align sample characteristics to those of the target population with severe CKD, and maximizing the inclusion of incident patients with severe CKD. Furthermore, we extended our definition of kidney failure to include kidney failure with and without kidney replacement therapy. This is important considering that the decision to initiate dialysis or perform kidney transplantation is partially subjective and that older patients may opt for conservative management of kidney failure.

Limitations

Our study has limitations, including the use of routinely collected data from people who accessed medical services and the use of data from 2 Canadian provinces with a very low prevalence of people who self-identify as Black. Our study only included people with G4-CKD and had albuminuria information available, which may have contributed to the overrepresentation of patients with diabetes in our cohorts. Although our findings will require confirmation in other settings, we do not believe that these limitations pose a threat to the validity of our results or interpretations.

Conclusions

These findings suggest that accounting for the competing risk of death when making predictions about kidney failure among people with G4-CKD becomes increasingly important with longer follow-up time, older age, and the presence of more comorbidity. If these findings are confirmed in other settings, existing clinical risk calculators should be updated to use similar methods and may improve predictions about the risk of kidney failure.

ARTICLE INFORMATION

Accepted for Publication: March 17, 2021.

Published: May 4, 2021. doi:10.1001/jamanetworkopen.2021.9225

Open Access: This is an open access article distributed under the terms of the [CC-BY License](#). © 2021 Al-Wahsh H et al. *JAMA Network Open*.

Corresponding Author: Pietro Ravani, MD, PhD, Cumming School of Medicine, Department of Medicine, University of Calgary, 3230 Hospital Dr NW, Calgary, AB T2N 4Z6, Canada (pravani@ucalgary.ca).

Author Affiliations: Cumming School of Medicine, Department of Medicine, University of Calgary, Calgary,

Alberta, Canada (Al-Wahsh, Quinn, Liu, Lam, MD, MSc, Tonelli, Ravani); Department of Medicine, Department of Community Health Sciences, Seven Oaks General Hospital, University of Manitoba, Winnipeg, Manitoba, Canada (Tangri, Ferguson, MS); Mathematical Institute, Leiden University, Medical Statistics Section, Department of Biomedical Data Science, Leiden University Medical Centre, Leiden, the Netherlands (Fiocco).

Author Contributions: Drs Al-Wahsh and Ravani had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Al-Wahsh, Tangri, Quinn, Lam, Ravani.

Acquisition, analysis, or interpretation of data: Al-Wahsh, Tangri, Quinn, Liu, Ferguson, Fiocco, Tonelli, Ravani.

Drafting of the manuscript: Al-Wahsh, Tangri, Fiocco, Ravani.

Critical revision of the manuscript for important intellectual content: Al-Wahsh, Tangri, Quinn, Liu, Ferguson, Lam, Tonelli, Ravani.

Statistical analysis: Al-Wahsh, Ferguson, Fiocco, Ravani.

Obtained funding: Ravani.

Administrative, technical, or material support: Tangri, Liu, Tonelli, Ravani.

Supervision: Tangri, Quinn, Ravani.

Conflict of Interest Disclosures: Dr Tangri reported receiving personal fees from Pulsedata, receiving personal fees and having equity in ClinPredict/Palpate Health, which uses kidney failure risk equations (KFRE), during the conduct of the study, and receiving grants and personal fees from Astra Zeneca outside the submitted work. Dr Tangri reported receiving grants and personal fees from Astra Zeneca and Tricida and receiving personal fees from Boehringer Ingelheim/Eli Lilly and Co, Janssen Pharmaceuticals, and Otsuka outside the submitted work. Dr Quinn reported having a patent for Dialysis Measurement Analysis and Reporting data system issued for Oliver Medical Management. Dr Tonelli reported receiving grants from Canadian Institutes of Health Research during the conduct of the study. No other disclosures were reported.

Funding/Support: Dr Al-Wahsh is supported by a Cumming School of Medicine of the University of Calgary grant. Drs Ravani and Liu hold Canadian Institutes for Health Research funding (grant FRN 173359 to Dr Ravani; grant FRN-MFE-152465 to Dr Liu) to support studies in chronic kidney disease. Dr Ravani is supported by the Roy and Vi Baay Chair in Kidney Disease. Drs Quinn, Lam, Tonelli, and Tangri are supported by operating grants or foundation awards from the Canadian Institutes for Health Research, by Alberta Innovates-Health Solutions, and by Canada Foundation for Innovation.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: This study is based in part on data provided by Alberta Health and Alberta Health Services. The interpretation and conclusions contained herein are those of the researchers and do not represent the views of the government of Alberta or Alberta Health Services. Neither the government of Alberta, Alberta Health, nor Alberta Health Services express any opinion in relation to this study. The results and conclusions are those of the authors, and no official endorsement by the Manitoba Centre for Health Policy, Manitoba Health, or other data providers is intended or should be inferred. Data used in this study are from the Manitoba Population Research Data Repository housed at the Manitoba Centre for Health Policy, University of Manitoba and were derived from data provided by Manitoba Health and Shared Health.

Additional Contributions: The authors also acknowledge the Manitoba Centre for Health Policy for use of data contained in the Manitoba Population Research Data Repository under project #2018-039 (HIPC#2018/2019-12).

Additional Information: The authors are not able to make their data set available to other researchers because of their contractual arrangements with the provincial health ministries of Alberta Health and Manitoba Health, who are the data custodians. Researchers may make requests to obtain a similar data set from each custodian.

REFERENCES

1. Stevens PE, Levin A, Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members. Evaluation and management of chronic kidney disease: synopsis of the kidney disease: improving global outcomes 2012 clinical practice guideline. *Ann Intern Med.* 2013;158(11):825-830. doi:10.7326/0003-4819-158-11-201306040-00007
2. Ravani P, Quinn R, Fiocco M, et al. Association of age with risk of kidney failure in adults with stage IV chronic kidney disease in Canada. *JAMA Netw Open.* 2020;3(9):e2017150. doi:10.1001/jamanetworkopen.2020.17150
3. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA.* 2011;305(15):1553-1559. doi:10.1001/jama.2011.451

4. Tangri N, Grams ME, Levey AS, et al; CKD Prognosis Consortium. Multinational assessment of accuracy of equations for predicting risk of kidney failure: a meta-analysis. *JAMA*. 2016;315(2):164-174. doi:10.1001/jama.2015.18202
5. Ravani P, Fiocco M, Liu P, et al. Influence of mortality on estimating the risk of kidney failure in people with stage 4 CKD. *J Am Soc Nephrol*. 2019;30(11):2219-2227. doi:10.1681/ASN.2019060640
6. Grams ME, Sang Y, Ballew SH, et al. Predicting timing of clinical outcomes in patients with chronic kidney disease and severely decreased glomerular filtration rate. *Kidney Int*. 2018;93(6):1442-1451. doi:10.1016/j.kint.2018.01.009
7. Benchimol EI, Smeeth L, Guttman A, et al; RECORD Working Committee. The Reporting of Studies Conducted Using Observational Routinely Collected Health Data (RECORD) statement. *PLoS Med*. 2015;12(10):e1001885. doi:10.1371/journal.pmed.1001885
8. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63. doi:10.7326/M14-0697
9. Levey AS, Stevens LA, Schmid CH, et al; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150(9):604-612. doi:10.7326/0003-4819-150-9-200905050-00006
10. Statistics Canada. Census Profile, 2016 Census—Alberta [Province] and Canada [Country]. Accessed October 2, 2020. <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/Page.cfm?Lang=E&Geo1=PR&Code1=48&Geo2=&Code2=&Data=Count&SearchText=Alberta&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=48>
11. Statistics Canada. Census Profile, 2016 Census—Manitoba [Province] and Canada [Country]. Accessed October 2, 2020. <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=PR&Code1=46&Geo2=PR&Code2=01&Data=Count&SearchText=46&SearchType=Begins&SearchPR=01&B1=All&Custom=&TABID=3>
12. Ishani A, Nelson D, Clothier B, et al. The magnitude of acute serum creatinine increase after cardiac surgery and the risk of chronic kidney disease, progression of kidney disease, and death. *Arch Intern Med*. 2011;171(3):226-233. doi:10.1001/archinternmed.2010.514
13. Tonelli M, Wiebe N, Fortin M, et al. Methods for identifying 30 chronic conditions: application to administrative data healthcare information systems. *BMC Med Inform Decis Mak*. 2015;15:31. doi:10.1186/s12911-015-0155-5
14. Cox D, Oakes D. *Analyses of Survival Data*. Chapman and Hall; 1984.
15. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*. 2007;26(11):2389-2430. doi:10.1002/sim.2712
16. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med*. 2012;31(11-12):1089-1097. doi:10.1002/sim.4384
17. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol*. 2012;41(3):861-870. doi:10.1093/ije/dyr213
18. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496-509. doi:10.1080/01621459.1999.10474144
19. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*. 2016;133(6):601-609. doi:10.1161/CIRCULATIONAHA.115.017719
20. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81(3):515-526. doi:10.1093/biomet/81.3.515
21. Zhou B, Fine J, Laird G. Goodness-of-fit test for proportional subdistribution hazards model. *Stat Med*. 2013;32(22):3804-3811. doi:10.1002/sim.5815
22. Ontario Renal Network. Ontario 2016 CKD system atlas: trends in kidney disease and care. Published 2016. Accessed March 14, 2021. <https://www.renalnetwork.on.ca/common/pages/UserFile.aspx?fileId=362165>
23. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiol*. 2010;21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2
24. Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiol*. 2009;20(4):555-561. doi:10.1097/EDE.0b013e3181a39056
25. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem*. 2008;54(1):17-23. doi:10.1373/clinchem.2007.096529

26. Blanche P, Dartigues J-F, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013;32(30):5381-5397. doi:10.1002/sim.5958
27. Hosmer DW, Lemeshow S. Assessing the fit of the model. In: *Applied Logistic Regression*. 2nd Ed. John Wiley and Sons; 2000: 160-164
28. Gerds TA, Andersen PK, Kattan MW. Calibration plots for risk prediction models in the presence of competing risks. *Stat Med*. 2014;33(18):3191-3203. doi:10.1002/sim.6152
29. Wojciechowski P, Tangri N, Rigatto C, Komenda P. Risk prediction in CKD: the rational alignment of health care resources in CKD 4/5 care. *Adv Chronic Kidney Dis*. 2016;23(4):227-230. doi:10.1053/j.ackd.2016.04.002
30. Major RW, Shepherd D, Medcalf JF, Xu G, Gray LJ, Brunskill NJ. Correction: the kidney failure risk equation for prediction of end stage renal disease in UK primary care: an external validation and clinical impact projection cohort study. *PLoS Med*. 2020;17(7):e1003313. doi:10.1371/journal.pmed.1003313
31. Al-Wahsh H, Lam NN, Liu P, et al. Investigating the relationship between age and kidney failure in adults with category 4 chronic kidney disease. *Can J Kidney Health Dis*. 2020;7:2054358120966819. doi:10.1177/2054358120966819
32. Van Der Pas S, Nelissen R, Fiocco M. Different competing risks models for different questions may give similar results in arthroplasty registers in the presence of few events. *Acta Orthop*. 2018;89(2):145-151. doi:10.1080/17453674.2018.1427314

SUPPLEMENT.

eTable 1. Coding Definitions for Baseline Characteristics

eTable 2. Coding Definitions for Identifying Dialysis or Transplantation

eTable 3. Baseline Characteristics by Cohort Entry on or Before vs After the Median Index Date (Development Cohort)

eTable 4. Model Coefficients

eTable 5. Difference in C Statistics and Brier Scores

eTable 6. Kidney Failure Risk Reclassification at Year 2 Considering Different Risk Cutoffs: Subgroup Analyses (Development Cohort)

eTable 7. Kidney Failure Risk Reclassification at Year 5 Considering Different Risk Cutoffs: Subgroup Analyses (Development Cohort)

eFigure 1. Derivation of Development and Validation Cohorts

eFigure 2. Calibration Plots at 1, 3, 5, and 7 Years in the Development Cohort

eFigure 3. Calibration Plots at 2, 4, 6, and 8 Years in the Development Cohort

eFigure 4. Calibration Plots at 1, 3, 5, and 7 Years in the Validation Cohort

eFigure 5. Reclassification Table With Cutoffs at 2 Years of 10% and 20% and at 5 Years of 15% and 30% (Cause-Specific Cox)

eFigure 6. Calibration Plots at 2 and 5 Years by Age (Development Cohort)

eFigure 7. Calibration Plots at 2 and 5 Years by Sex (Development Cohort)

eFigure 8. Calibration Plots at 2 and 5 Years by Albumin-Creatinine Ratio (ACR; Development Cohort)

eFigure 9. Calibration Plots at 2 and 5 Years by Median Estimated Glomerular Filtration Rate (Development Cohort)

eFigure 10. Calibration Plots at 2 and 5 Years by Diabetes (Development Cohort)

eFigure 11. Calibration Plots at 2 and 5 Years by Cardiovascular Disease (Development Cohort)

eFigure 12. Reclassification Table With Cutoffs at 2 Years of 5% and 15% and at 5 Years of 10% and 20% (Fine and Gray)

eFigure 13. Reclassification Table With Cutoffs at 2 Years of 5% and 15% and at 5 Years 10% and 20% (Cause-Specific Cox)

eFigure 14. Reclassification Table With Cutoffs at 2 Years of 10% and 40% and at 5 Years of 5% and 50% (Fine and Gray)

eFigure 15. Reclassification Table With Cutoffs at 2 Years of 10% and 40% and at 5 Years of 5% and 50% (Cause-Specific Cox)

eFigure 16. Calibration Plots at 2 and 5 Years by Cohort Entry on or Before vs After the Median Index Date (Development Cohort)

eFigure 17. Calibration Plots at 1, 3, 5, and 7 Years (Sensitivity Analysis)

eFigure 18. Calibration Plots at 2, 4, 6, and 8 Years (Sensitivity Analysis)

eReferences.