



## OPEN ACCESS

## EDITED BY

Md. Mohaimenul Islam,  
The Ohio State University, United States

## REVIEWED BY

Danilo Lofaro,  
University of Calabria, Italy  
Luca Ongaro,  
Royal Free Hospital, United Kingdom

## \*CORRESPONDENCE

Daniel Christiadi  
✉ daniel.christiadi@anu.edu.au

RECEIVED 05 May 2024

ACCEPTED 28 November 2024

PUBLISHED 11 December 2024

## CITATION

Christiadi D, Chai K, Chuah A, Loong B,  
Andrews TD, Chakera A, Walters GD and  
Jiang SH-T (2024) Dynamic survival  
prediction of end-stage kidney disease using  
random survival forests for competing risk  
analysis.

*Front. Med.* 11:1428073.

doi: 10.3389/fmed.2024.1428073

## COPYRIGHT

© 2024 Christiadi, Chai, Chuah, Loong,  
Andrews, Chakera, Walters and Jiang. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Dynamic survival prediction of end-stage kidney disease using random survival forests for competing risk analysis

Daniel Christiadi<sup>1,2,3\*</sup>, Kevin Chai<sup>4</sup>, Aaron Chuah<sup>1</sup>,  
Bronwyn Loong<sup>5</sup>, Thomas D. Andrews<sup>1</sup>, Aron Chakera<sup>6</sup>,  
Giles Desmond Walters<sup>2,3,7</sup> and Simon Hee-Tang Jiang<sup>1,2,3</sup>

<sup>1</sup>Department of Immunology and Infectious Disease, John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia, <sup>2</sup>Department of Renal Medicine, The Canberra Hospital, Garran, ACT, Australia, <sup>3</sup>Centre of Personalised Medicine, Australian National University and Canberra Health Services, Canberra, ACT, Australia, <sup>4</sup>School of Population Health, Curtin University, Perth, WA, Australia, <sup>5</sup>Research School of Finance, Actuarial Studies & Statistics, Australian National University, Canberra, ACT, Australia, <sup>6</sup>Department of Renal Medicine, Sir Charles Gairdner Osborn Park Health Care Group, Nedlands, WA, Australia, <sup>7</sup>Australian National University Medical School, Garran, ACT, Australia

**Background and hypothesis:** A static predictive model relying solely on baseline clinicopathological data cannot capture the heterogeneity in predictor trajectories observed in the progression of chronic kidney disease (CKD). To address this, we developed and validated a dynamic survival prediction model using longitudinal clinicopathological data to predict end-stage kidney disease (ESKD), with death as a competing risk.

**Methods:** We trained a sequence of random survival forests using a landmarking approach and optimized the model with a pre-specified prediction horizon of 5 years. The predicted cumulative incidence function (CIF) values were used to generate a personalized dynamic prediction plot.

**Results:** The model was developed using baseline demographics and 13 longitudinal clinicopathological variables from 4,950 patients. Variable importance analysis for ESKD and death informed the creation of a sequence of reduced models that utilized six key variables: age, serum albumin, bicarbonate, chloride, eGFR, and hemoglobin. The models demonstrated robust predictive performance, with a median concordance index of 84.84% for ESKD and 84.1% for death. The median integrated Brier scores were 0.03 for ESKD and 0.038 for death across all landmark times. External validation with 8,729 patients confirmed these results.

**Conclusion:** We successfully developed and validated a dynamic survival prediction model using common longitudinal clinicopathological data. This model predicts ESKD with death as a competing risk and aims to assist clinicians in dialysis planning for patients with CKD.

## KEYWORDS

dynamic prediction model, end-stage kidney disease, landmarking, random survival forests, competing risk

## Introduction

Chronic kidney disease (CKD) is a global health challenge, with a reported prevalence of 13.4% (1), and is associated with significant mortality (2) and morbidity (3). As patients approach end-stage kidney disease (ESKD), clinicians and patients face complex decisions regarding the optimal timing for permanent dialysis access formation and evaluating transplantation suitability. These decisions are further complicated by the highly variable and often unpredictable rate of CKD progression. Therefore, an accurate model to predict ESKD is vital.

The Kidney Failure Risk Equation (KFRE) (4) is widely recognized as the gold-standard predictive model for ESKD. It is based on a Cox-proportional hazard model that uses a combination of baseline clinicopathological predictors. Similar to KFRE, other studies have explored various machine learning models and neural networks that rely on baseline patient characteristics and clinicopathological data as predictors (5–9).

However, these static predictive models have notable limitations. They often fail to account for the variability and non-linear relationships between predictors and ESKD over an extended follow-up period, which is critical given the chronic nature of the disease. Additionally, they may miss valuable prognostic information derived from changes in these variables over time. For example, the estimated glomerular filtration rate (eGFR) has been demonstrated to exhibit highly variable trajectories among patients with CKD (10). Despite this variability, time-varying eGFR remains a critical predictor of ESKD and all-cause mortality (11, 12). Consequently, static predictive models are primarily applicable to new patients at baseline rather than to those being monitored during follow-up (13).

To incorporate time-varying covariates, Chuah et al. and Wang et al. featured time-series data into features to predict ESKD as a binary classification problem (14, 15). While this approach demonstrated superior predictive performance, using binary classification to address a time-to-event problem introduces certain limitations.

First, this method potentially sacrifices the interpretability and flexibility of modeling event probabilities as a function of time. Additionally, it does not account for patients lost to follow-up or censored, nor does it consider the impact of competing events such as death. In populations with CKD, which are often characterized by older age and chronic comorbidities, failing to account for death can result in inaccurate ESKD risk predictions (16).

Several publications have introduced dynamic prediction models to address the limitations of static approaches (13, 17–20). In statistical literature, a clinical predictive model capable of processing longitudinal data to update predictions over time is referred to as a dynamic prediction model. Not only would this methodology mimic a nephrologist's approach to patient evaluation, but the methodology has also been shown to improve model fit and performance (19).

There are two primary methods for dynamic prediction: joint models (21, 22) and landmarking (23–25). Joint models generate predictions by linking a Cox proportional hazard model with linear mixed models to account for time-varying covariate trajectories. However, correctly specifying a joint statistical model for all covariate trajectories may be challenging or unstable, and incorporating additional time-varying covariates may be computationally expensive (26–28).

In contrast, the landmarking approach is more straightforward and better equipped to handle multiple time-varying covariates. Pickett et al. proposed combining random survival forests with the landmarking approach, effectively bypassing the limitation of Cox models, including the proportional hazards assumption (29). Simulation studies demonstrated the superior performance of this approach over joint models and Cox-based landmarking in capturing complex relationships between survival outcomes and longitudinal covariates.

Our study builds on this technique, extending it to analyze multiple longitudinal clinicopathological data points to predict ESKD while accounting for death as a competing event. The model has been externally validated and offers the potential for clinical decision support in real-world settings.

## Methods

### Study population

This observational, time-to-event study included all adult patients aged over 17 years ( $N = 10,198$ ) who attended a single tertiary nephrology unit at the Canberra Hospital, a university hospital in Australia. Clinicopathological data, encompassing both inpatient and outpatient data, along with patient demographics, were extracted from an internal database covering the period from 1st September 1996 to 14th January 2022 (Supplementary Table 1). Patients with pre-existing ESKD (defined as those undergoing chronic dialysis or with a history of kidney transplantation) were excluded from the analysis.

For external validation, a cohort of patients aged  $\geq 18$  years ( $N = 8,729$ ) was obtained from a Western Australian pathology provider, spanning data from 1st January 2006 to 21st April 2022. This external cohort was prepared using the same exclusion criteria and filtering methodology as the internal cohort (see Results). The study was conducted in accordance with the Declaration of Helsinki and was approved by the ACT Health Research Ethics and Governance Office (ETHLR.18.040). Patient consent was waived per the National Statement on Ethical Conduct in Human Research (2007, Chapter 2.3).

### Events and censoring definition

The primary event of the study was ESKD, with death from all causes considered a competing event. ESKD was defined as a sustained eGFR value less than or equal to  $15 \text{ mL/min/1.73m}^2$  or the commencement of chronic dialysis or kidney transplantation. Specifically, this required evidence of persistent eGFR  $< 15 \text{ mL/min/1.73m}^2$  without subsequent recovery (eGFR  $> 17 \text{ mL/min/1.73m}^2$ ) during the follow-up period or the commencement of dialysis, including peritoneal dialysis, hemodialysis, hemodiafiltration, or nocturnal hemodialysis.

The first recorded eGFR value marked the start of the observation period (designated as time 0). Age at presentation was defined as the patient's age at time 0. Observed time was measured from time 0 to the earliest event (ESKD or death) or to censoring. For patients who did not experience an event, the follow-up duration was censored at the data extraction date.

## Statistical analysis

The normality of continuous data was assessed using the Shapiro–Wilk test. For normally distributed data, differences between groups were evaluated using analysis of variance (ANOVA) and summarized as mean  $\pm$  standard deviation (SD). For non-parametric data, the Kruskal–Wallis test was used to assess group differences, and results were summarized as median  $\pm$  interquartile range (IQR). Pairwise comparisons between two groups were conducted using either the *t*-test or the Mann–Whitney *U* test, depending on the normality of the data. Values of  $p < 0.05$  were considered significant, and all tests were two-tailed.

## Dataset imputation

Within the internal cohort, only eGFR and demographic data were complete at baseline (time 0), while all other clinicopathological predictors had a median missing value rate of 18% (Supplementary Table 2). During model training and performance analysis, missing baseline values in the training set (time 0) were imputed using the R package “mice” (version 3.14.0) with the “defaultMethod” set to predictive mean matching. After imputing missing baseline values, missing values for each variable at subsequent time points were filled out using the most recent entry.

For the test dataset, missing values were filled exclusively using the most recent entry. No additional imputation was performed to preserve the reliability of the prediction performance and to reflect clinical practice, where clinicians often rely on prior available data when results are unavailable.

For the external dataset, data cleaning was conducted to address errors and inconsistencies in patient demographic data (e.g., birth dates and sex). Due to the high rate of missing test values (Supplementary Table 2), serum glucose was imputed using a combination of previous and future data entries. In contrast, all other clinicopathological test values were filled exclusively using the last available observation for each patient.

## Dataset splitting

We evaluated performance using a 5-fold event-stratified cross-validation approach. The dataset was divided into training and test sets for each fold in an 80:20 ratio. The training set was used to train the model, while model performance was evaluated on the corresponding test set.

## Model development

As proposed by Van Houwelingen, the landmarking method applies a survival model to patients who are still at risk at the time of interest, a pre-specified landmark time (23, 24). In this method, the survival model is fitted as a function of predictors measured up to the landmark time. By selecting a series of landmark time points, the method produces a sequence of survival predictions focused on events occurring in the medically relevant timeframe, known as the

prediction horizon. A detailed explanation of this approach is provided in the Supplementary material.

We implemented the landmarking approach with pre-specified landmark times at 0.5, 1, 1.5, 2, 2.5, and 3 years, with a prediction horizon of 5 years. These landmark times were chosen because stable chronic kidney disease patients are typically reviewed in the clinic every 6 months. Following the recommendation of the landmarking method’s originator, administrative censoring was applied to individuals who had not experienced an event by the end of each landmark prediction horizon.

Instead of the Cox-proportional hazard model, we used a random survival forests. The random survival forests is a non-parametric, ensemble, tree-based method (30). The algorithm trains multiple survival trees to analyze right-censored survival data.

The model uses a weighted log-rank splitting rule based on Gray’s test (31) and calculates the cumulative incidence function (CIF) to ensure accurate risk prediction in the presence of competing risks (16). Additional advantages of the model include its ability to model non-linear effects and interactions among high-dimensional predictors and its robustness in cases where the proportional hazard assumption is violated (29, 32, 33).

We implemented the built-in function in the R package “randomForestSRC” to tune the random survival forests model. The tuning process was performed to find the optimal “*mtry*,” the number of predictors that should be included as candidates for splitting, and “*nodesize*,” which represents the number of unique cases in each terminal node, determined based on the out-of-bag prediction error rate. We used the default parameters for “*mtryStart*,” which equals the number of predictors divided by 2, and “*nodesizeTry*” of (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100). Then, we used the optimal “*mtry*” and “*nodesize*” based on the out-of-bag error. We set the number of trees at 1,000.

We used baseline data—including gender, diagnosis of glomerulonephritis or vasculitis, and initial age at presentation—and longitudinal, time-varying clinicopathological data as predictors. The diagnoses of glomerulonephritis or vasculitis were verified through kidney biopsy. All other diagnoses were inputted as missing values. We then compared the performance of the standard landmark approach, consisting of adjusting the survival model on the last observed value of the biomarkers before each landmark time (LOCF) or based on the predicted value at the beginning of each landmark time obtained from the linear mixed model (34, 35). The linear mixed model is the standard approach in analyzing repeated measures and dependent data of longitudinal biomarkers. We modeled the predicted longitudinal clinicopathological data values using baseline measurements and the timing of these measurements, applying either an unconditional growth model (LME) or an unconditional quadratic growth model (LMEpoly).

The details of the R package version and the codes used in the analysis are posted at [https://github.com/daniel-christiadi/dynamic\\_prediction\\_ESKD.git](https://github.com/daniel-christiadi/dynamic_prediction_ESKD.git).

The trained models can be downloaded from <https://drive.google.com/drive/folders/1yj7McUimsEsYxZjuMG26-oELjY3oMFmf?usp=sharing>.

## Model performance

Model performance analysis in the internal cohort was performed on the test dataset obtained using a 5-fold cross-validation approach

and measured on each pre-specified landmark time described above. Model discrimination was reported as the concordance index (C-index), which measures the ratio of subjects with the worse predicted outcome that fail earlier than other subjects in a randomly selected pair of subjects to the number of all available subject pairs (32). A higher concordance index indicates superior model discrimination, with a concordance index of 50% representing a random prediction.

The parsimonious model (Top 5) performance was also analyzed using the area under the time-dependent ROC curve (td-AUC) (36). The predicted value of the random survival forests for each landmark was assessed at the pre-specified prediction horizon. Similar to the concordance index, a higher td-AUC indicates superior model discrimination.

Model discrimination and calibration were assessed using the integrated Brier score (iBS) (37), calculated from the start of each landmark time to the prediction horizon. Unlike the C-index, lower iBS values indicate better model performance, with an iBS below 0.25 considered indicative of a useful model.

Performance on the external validation cohort was performed using models trained on the entire internal dataset, which was then evaluated on all external datasets. Due to the lack of death data in the external cohort, we could only report the external ESKD error rate.

To calculate the performance of our model against the 8-variable KFRE, we filtered the internal dataset. Only patients who had three or more serum measurements of albumin, bicarbonate, chloride, eGFR, hemoglobin, phosphate, calcium, and urine albumin-creatinine ratio, which are the combined variables used by our models and the 8-variable KFRE. Then, the dataset was manipulated to have death-censored outcomes to accommodate KFRE. To allow a fairer comparison, we input the value of the covariates at the last landmark time into the KFRE formula (38) to predict ESKD at 2 and 5 years from the last landmark time, which were the same times of interest for our model.

All analyses were performed using R 4.3.1 (39).

## Dynamic prediction plot

As the output of the trained model, we created a dynamic prediction plot. The predicted CIF values from each model for the events of interest were aggregated by each time point (from year three to year eight) using the arithmetic mean to draw the CIF curves. Higher percentages indicate a greater likelihood of the event occurring. To create the dynamic prediction plot, readers are recommended to follow the steps described at [https://github.com/daniel-christiadi/dynamic\\_prediction\\_ESKD.git](https://github.com/daniel-christiadi/dynamic_prediction_ESKD.git).

## Sensitivity analysis

We performed two sensitivity analyses. First, we created a dataset (named “sens”) with a different censoring definition. Instead of using the definition mentioned previously, for patients with no ESKD or death event, we censored the subjects on the last recorded clinicopathological test date. Furthermore, we incorporate only the tests within the previous 10 years from events or the final clinicopathological tests (the alternative censoring definition). The date of the earliest eGFR within the 10 years became time 0, and the initial age at presentation was calculated from this time point.

Additionally, we created another dataset with an additional three or more urine albumin/creatinine ratio (uACR) filter criteria (named “acr3”). To reduce the missing value, we converted the urine protein/creatinine ratio to uACR according to the published formula (40).

## Results

### Dataset creation and baseline characteristics

Of the 10,198 patients, 15 of them were excluded due to incomplete baseline demographic information (absence of gender and birth year). Another 495 patients were removed due to pre-existing ESKD, and two patients were excluded due to incorrect dates (the initial clinicopathology test occurred after the documented death date), leaving 9,686 patients with 4,458,014 individual clinicopathological values for further analysis.

Some patients exhibited skewed test frequencies, likely due to recurrent hospitalisations, prolonged stays, or frequent clinic visits (Supplementary Figure 1A). Therefore, we aggregated (resampled) the daily test results into mean monthly values to prevent this cohort from skewing the results (Supplementary Figure 1B). Due to the uneven data density of the available clinicopathological tests and to ensure longitudinal data availability, we filtered patients on three or more of the following criteria: hemoglobin, total white cell count, platelet count, CKD-EPI-based eGFR, and serum measurement of sodium, potassium, chloride, bicarbonate, calcium, phosphate, albumin, alkaline phosphates, and glucose, leaving 4,950 patients (Supplementary Figure 1C). This dataset will now be referred to as dense3.

The median initial age of the study population was 62 (IQR 47 to 73), with a median baseline eGFR of 50 mL/min/1.73m<sup>2</sup>. The patients had a median follow-up time of 7 years (IQR 3 to 13). Only 500 of 4,950 patients were documented to have any diagnosis of glomerulonephritis or vasculitis. ESKD events occurred in 1,270 (25%) and death in 733 (15%) patients. In the external cohort, the total cohort size was 8,729 subjects with a median age of 57 (IQR 47 to 66). The patients had relatively similar follow-up durations. Notably, patients in the external cohort had better baseline kidney function (median eGFR 89 mL/min/1.73m<sup>2</sup>). Due to the absence of death information, only an ESKD event was reported. A total of 246 patients of 8,729 (3%) experienced ESKD. The baseline demographic and clinicopathological data from the internal and external cohorts are displayed in Table 1.

### Models comparison

To incorporate longitudinal data, the landmarking approach can utilize the last observed value of the longitudinal biomarkers, known as the last observed carried forward (LOCF), or incorporate all the previous longitudinal biomarkers to model the value of the respective biomarkers at the beginning of each landmark time. The linear mixed model is the standard approach in analyzing repeated measures and dependent data of longitudinal biomarkers. Two linear mixed model methods were used for comparison: the unconditional growth model (LME) and the unconditional quadratic growth model (LME poly). At each landmark time, there was no significant performance difference

TABLE 1 Baseline summary of internal and external cohorts.

Variables	Dense3 (N = 4,950)	Acr3 (N = 2,916)	External cohort (N = 8,729)
<b>Demographic data</b>			
Gender - Female	2,217 (45%)	1,280 (44%)	4,865 (56%)
Initial age of presentation (median, IQR)	62 (47–73)	59 (45–70)	57 (47–66)
Glomerulonephritis or vasculitis diagnosis	500 (10%)	445 (15%)	No information
Follow-up time in years (median, IQR)	7 (3–13)	8 (4–14)	8 (5–13)
<b>Event</b>			
ESKD	1,270 (25%)	639 (22%)	246 (3%)
Death	733 (15%)	278 (9.5%)	No information
Censored	2,947 (60%)	1,999 (69%)	8,483 (97%)
<b>Baseline clinicopathological data</b>			
Serum albumin in g/L (median, IQR)	42 (39–44)	42 (39–44)	43 (41–45)
Serum alkaline phosphatase in U/L (median, IQR)	78 (63–97)	77 (63–95)	73 (60–89)
Serum bicarbonate in mmol/L (median, IQR)	23 (21–26)	24 (22–26)	28 (26–29)
Serum calcium in mmol/L (median, IQR)	2.36 (2.27–2.44)	2.36 (2.27–2.43)	2.37 (2.31–2.44)
Serum chloride in mmol/L (median, IQR)	105 (102–107)	105 (102–107)	103 (101–105)
eGFR in mL/min/1.73m <sup>2</sup> (median, IQR)	50 (30–73)	57 (36–77)	89 (74–90)
Serum glucose in mmol/L (median, IQR)	5.7 (5–7.20)	5.6 (5–7.3)	5.3 (4.8–5.9)
Hemoglobin in g/L (median, IQR)	133 (118–146)	136 (122–148)	139 (129–149)
Serum phosphate in mmol/L (median, IQR)	1.13 (0.99–1.28)	1.12 (0.99–1.26)	1.10 (1–1.22)
Platelet count in x10 <sup>9</sup> /L (median, IQR)	240 (198–290)	245 (203–294)	255 (212–305)
Serum potassium in mmol/L (median, IQR)	4.2 (3.9–4.6)	4.3 (4–4.6)	4.3 (4.1–4.6)
Serum sodium in mmol/L (median, IQR)	139 (137.5–141)	140 (138–141)	140 (139–141)
White cell count in x10 <sup>9</sup> /L (median, IQR)	7.3 (6–8.9)	7.3 (6–8.8)	6.6 (5.4–8.2)
Albumin creatinine ratio in mg/mmol (median, IQR)	NA	13 (2–92)	NA

between models that used time-varying clinicopathological data value using the LOCF, LME, and LME poly approach in predicting ESKD or death (Figure 1 and Supplementary Table 3).

## Important predictors

To examine the importance of each covariate in the model prediction, we explored the variable importance (VIMP, Figure 2). A large VIMP indicates the model relies on the respective predictors to achieve prediction accuracy and is thus likely a potentially predictive variable. We calculated the median of variable importance for individual predictors using 5-fold cross-validation and the interquartile range. eGFR is the most important predictor for ESKD prediction in all landmark times, followed by serum albumin and bicarbonate. In contrast, initial age at presentation, eGFR, and serum chloride are crucial for death prediction. To test whether we could reduce the number of predictors to allow visualization of the dynamic prediction plot, we combined the VIMP value for ESKD and death for all the landmark times and calculated their median (Supplementary Table 4).

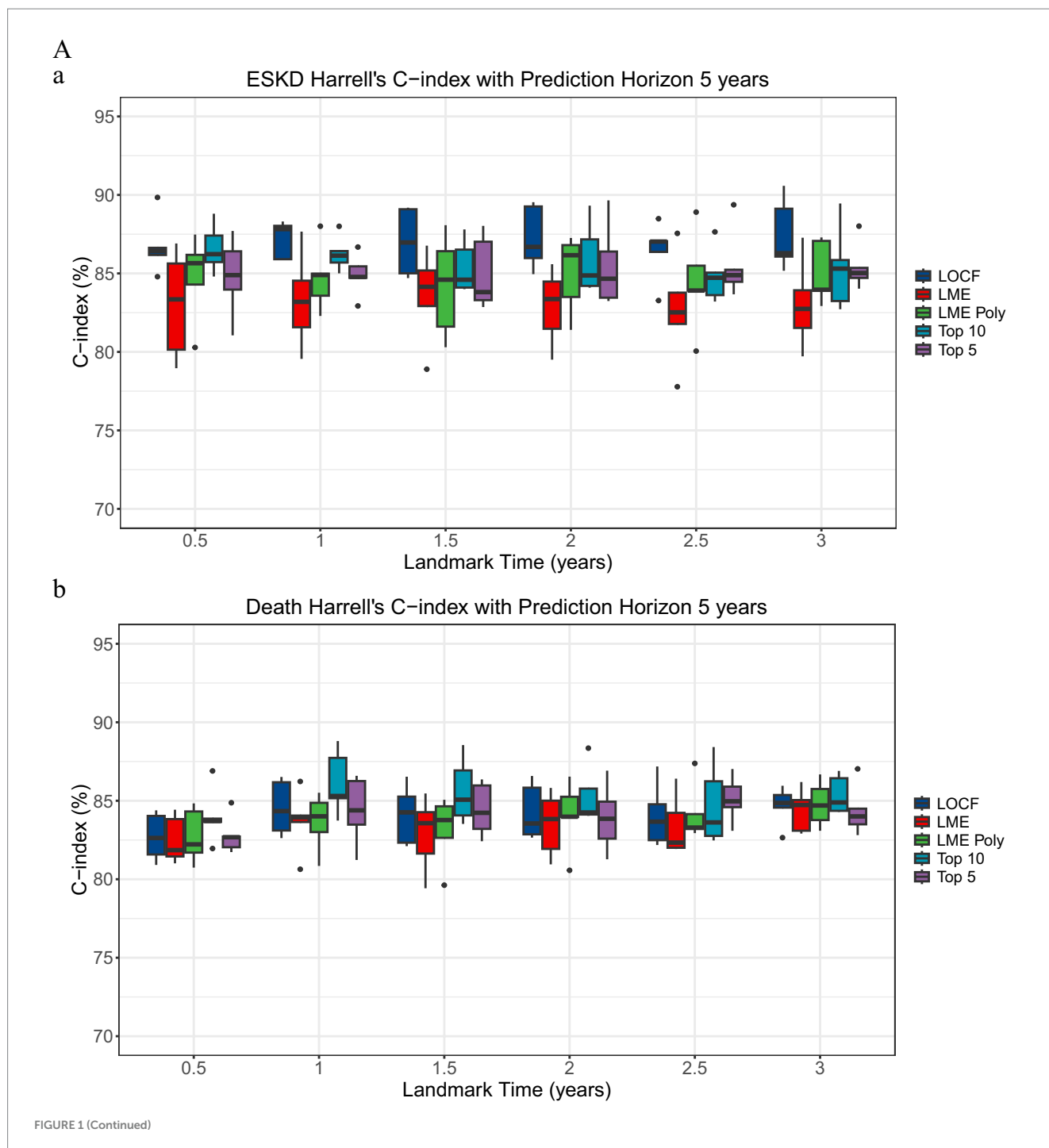
We then compared the performance of the full model (LOCF) against a sequence of models using the top 10 (Top 10) longitudinal

predictors (serum measurement of albumin, alkaline phosphatase, bicarbonate, chloride, potassium, and sodium, eGFR, hemoglobin, platelet count, and white cell count) and a sequence of models using the top five longitudinal predictors (serum measurement of albumin, bicarbonate, chloride, eGFR, and hemoglobin). For the top 10 and top 5 models, we removed the diagnosis of glomerulonephritis or vasculitis and gender due to low median VIMP values.

The ESKD concordance index for the top 5 models was close to 85%, except at the 1.5-year landmark, where the median was 83.81% (IQR: 83.3 to 87.02%, Supplementary Table 3A), with a median td-AUC for ESKD of approximately 84.9% (Supplementary Table 3C). Additionally, the ESKD iBS for the top 5 was approximately 0.03 (Supplementary Table 3B). No significant performance differences were found when comparing the top 5 models to the full model or the top 10 models (Figure 1), indicating the top 5 variables accounted for most of the model's predictive capability.

## External validation performance

In external validation, the ESKD median concordance index across all landmark times ranged from 83.1 to 88.3%



(Supplementary Table 3). The Wilcoxon rank-sum test comparing the ESKD concordance index at each landmark between the top 5 models and external validation showed no significant difference.

## Comparison to 8-variable KFRE

Using the clinicopathological data of death-censored patients in the test datasets on the last landmark time, the 8-variable KFRE model (38) was used to calculate ESKD prediction at 2 years and 5 years. There is no performance difference between our models and KFRE at

5 years, but the 8-variable KFRE concordance index at 2 years was superior (Supplementary Table 6).

## Dynamic prediction plot

We used three case studies to demonstrate the utility of the dynamic prediction plot (Figure 3). For case 1, an 80-year-old patient had the last clinicopathological test at 6.91 years and died at 7.2 years. The predicted death CIF curve peaked at 7.5 years. Case 2, a 62-year-old subject, was followed for 11.66 years with an ESKD event at

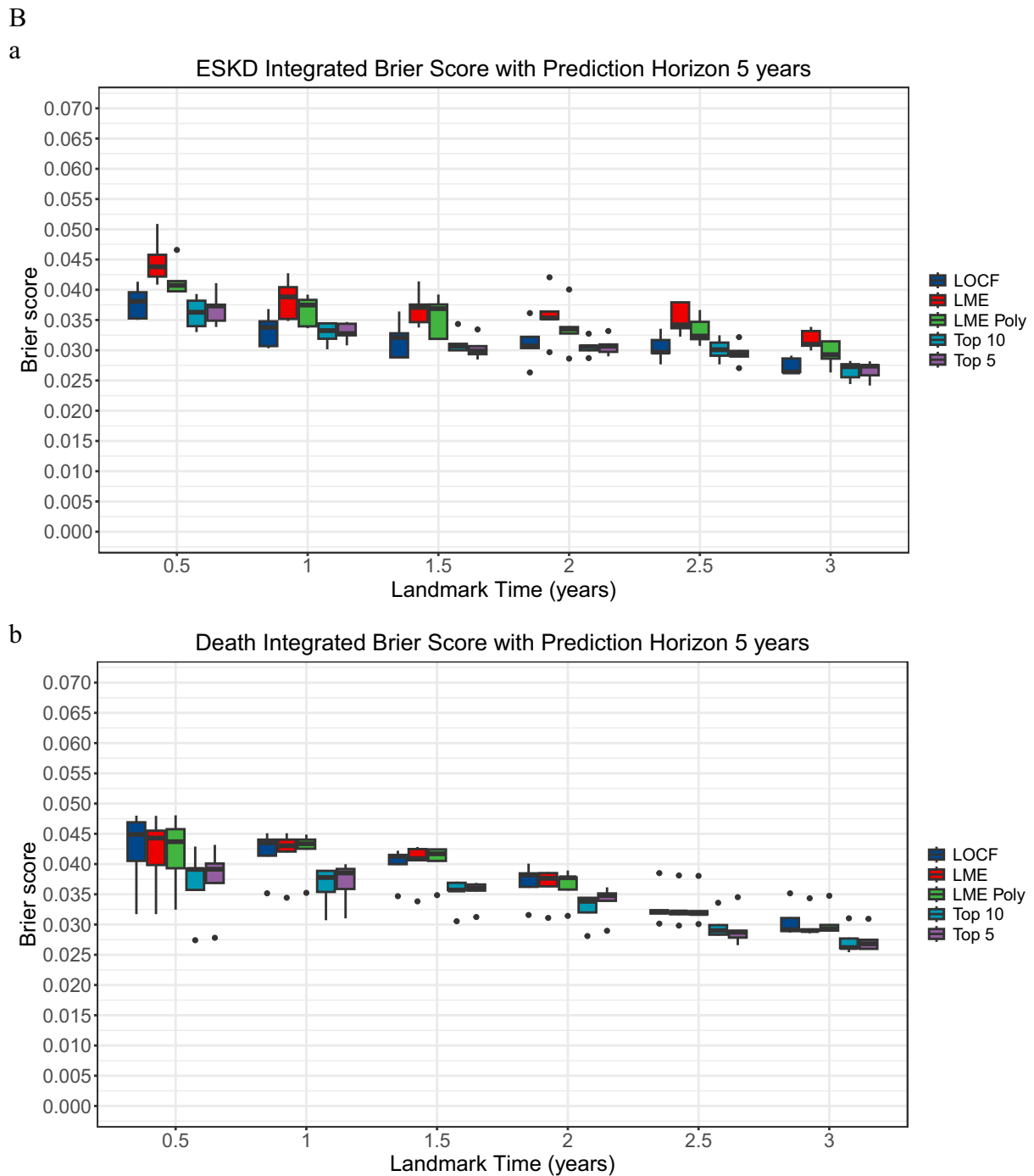


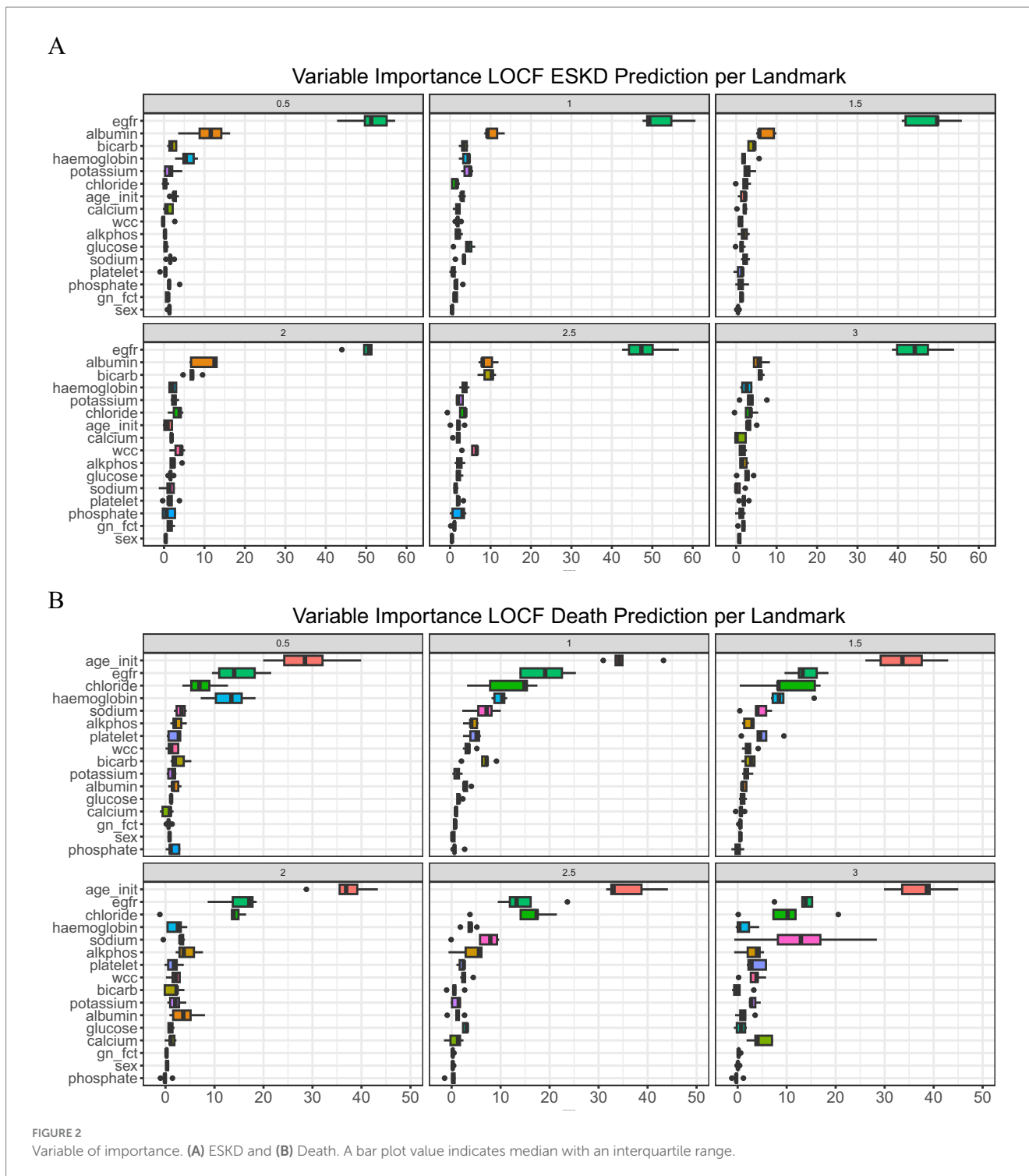
FIGURE 1

Comparison of landmarking approach. **(A)** The Concordance index. **(a)** ESKD: The Kruskal–Wallis test comparing the ESKD concordance index of the models for each landmark time was non-significant. Each boxplot was created using the model’s performance on the five-fold cross-validation test dataset. **(b)** Death: The Kruskal–Wallis test comparing the death concordance index of the models for each landmark time was non-significant. Each boxplot was created using the model’s performance on the test dataset of the five-fold cross-validation. **(B)** Integrated Brier score. **(a)** ESKD: The Kruskal–Wallis test comparing the ESKD integrated Brier score of the models for each landmark time was non-significant. Each boxplot was created using the model’s performance on the test dataset of the five-fold cross-validation. **(b)** Death: The Kruskal–Wallis test comparing the death-integrated Brier score of the models for each landmark time was non-significant. Each boxplot was created using the model’s performance on the test dataset of the five-fold cross-validation.

5.08 years. The ESKD CIF curve steadily rose and overtook the death curve from 4 years onwards. Finally, a 67-year-old patient was followed until 6.69 years without any event. ESKD and death CIF curves peaked at approximately 10%.

### Sensitivity analysis

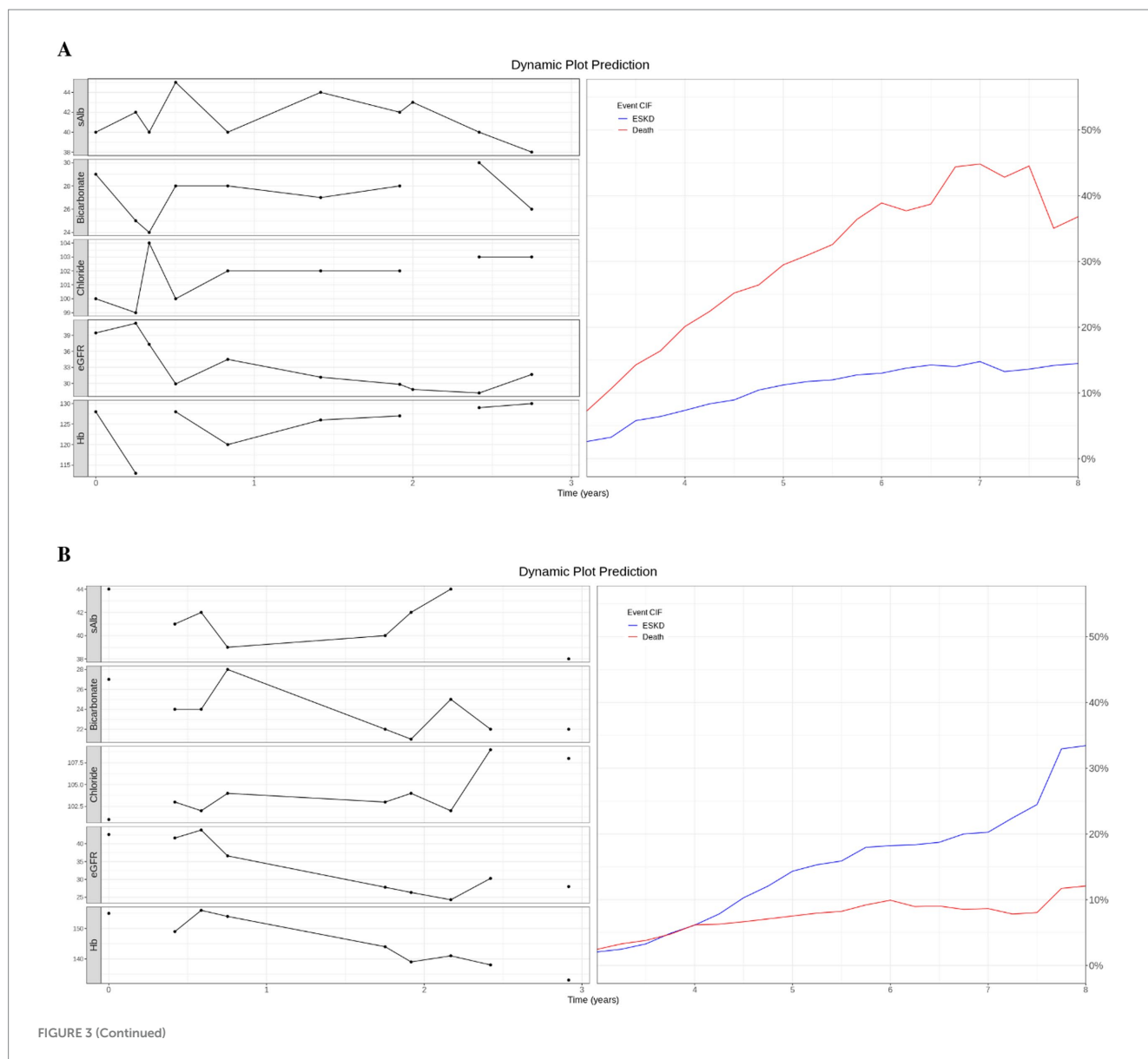
To create a “sens” dataset, 10,198 patients were extracted. A total of 15 patients were excluded due to incomplete baseline demographic



information (absence of gender and birth year). Another 495 patients were removed due to pre-existing ESKD. A total of 31 patients did not have any clinicopathological tests within 10 years of the event, resulting in 9,657 patients for further analysis. Applying the same filter criteria as for 'dense3', the 'sens' dataset included 4,225 patients. To create "acr3," we applied additional filter criteria (three or more uACR) to the dense3 dataset, resulting in 2,916 patients.

There was no significant difference in discrimination or calibration performance between the original, conventional censoring definition (LOCF, censoring date based on the data extraction date) and the alternative censoring definition (sens, censoring date based on the last clinicopathological data). There was no improvement in performance when the urine albumin/creatinine ratio was included in the acr3 model training (Supplementary Figure 2 and Supplementary Table 5).





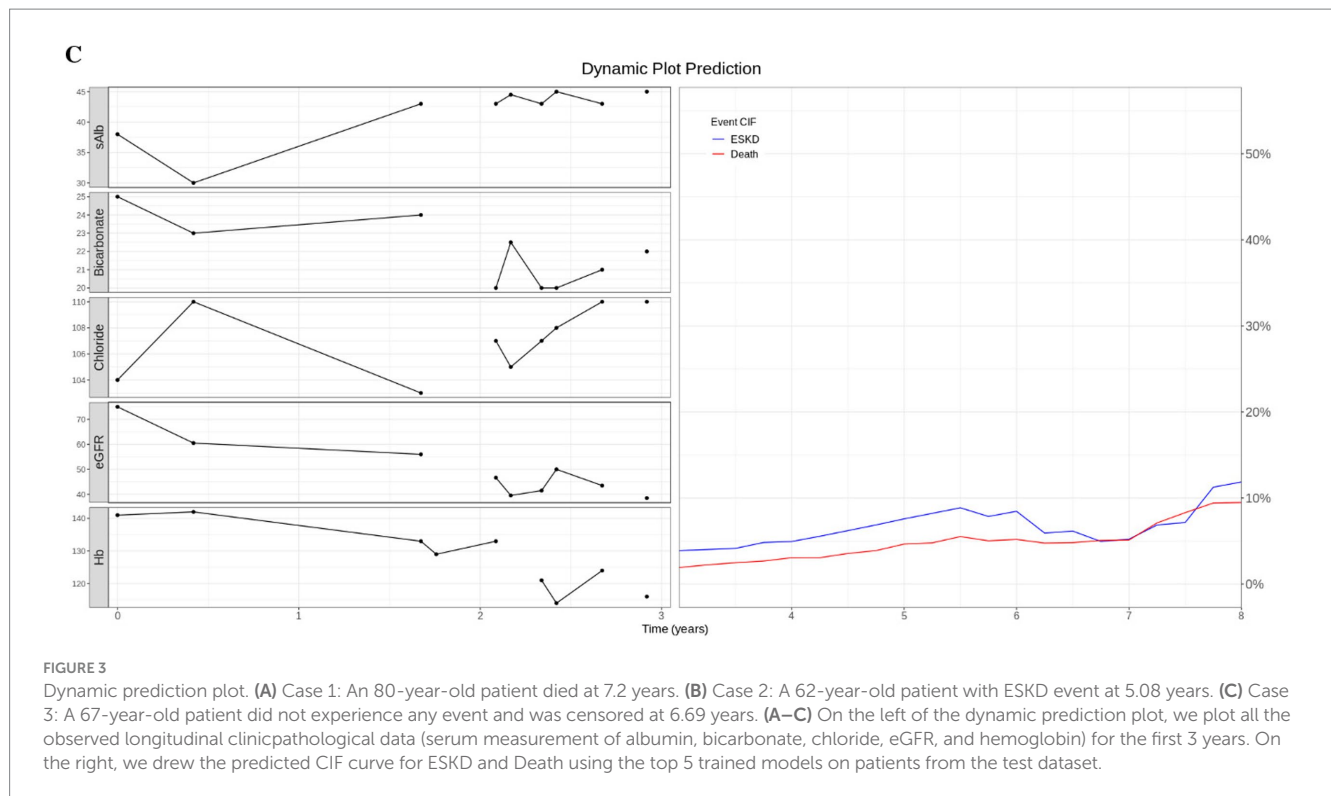
## Discussion

We demonstrated how longitudinal clinicopathological data can be incorporated to create a dynamic prediction plot that estimates the risks of death and ESKD. The graph, designed as an end-user interface, resembles the workflow of specialist clinical practice. CKD patients are typically monitored quarterly to annually, generating dense clinicopathological data points over extended follow-up periods. In contrast, static predictive models require nephrologists to disregard valuable data points or previous predictions, treating each visit as if the patients were new.

Instead of relying on cross-sectional values, clinicians could use all accumulated and updated data points to generate personalized dynamic predictions. These predictions can be used during patient education, decision-making regarding dialysis preparation, and resource planning, especially for older patients with underlying poor kidney function who are at risk for both events. By integrating clinical expertise with the predicted risks provided by dynamic predictions,

clinicians can better guide discussions about whether patients at high risk of death should undergo dialysis preparation, such as fistula creation.

Our limited comparison showed that our model performed similarly to KFRE at predicting ESKD at 5 years and was inferior to KFRE at 2 years. However, this is not a valid comparison because the KFRE was designed using baseline data and provided a static prediction at a single time point. The KFRE model was not designed to be applied to subsequent clinicopathological data. Moreover, KFRE was created from a cohort of patients with eGFR less than 60 mL/min/1.73m<sup>2</sup> with a development cohort comprised of patients with a mean age of 70 years (71% of patients were more than 65 years old) to predict ESKD without taking death into account. Of all patients aged 70 years or older in our cohort (“dense3”), 495 patients of 1,612 (31%) died before developing ESKD, and 423 (26.2%) experienced ESKD as the first event. Therefore, analyzing the performance of KFRE in our cohort will likely produce biased results. Our model seeks to move to an



analysis of continuous data and outputs estimates of both ESKD and death to better inform clinical decision-making.

We also observed a higher prevalence of ESKD (25%) during the study period in the internal cohort. To compare the event, we reviewed datasets from another state in Australia (Tasmania). Irish et al. reported 2.2% ESKD and 6.8% death within 2 years of observation (41). The difference is likely due to the data source. Our cohort came from patients referred to nephrologists with CKD. TASlink was a cohort created through data linkage. It is also worth pointing out that our external validation cohort (from Western Australia) had a relatively similar ESKD prevalence of 3%, and the parallel performance we demonstrated to the external cohort implied the external validity of our models.

The difference in ESKD prevalence potentially affected the comparison of internal versus external performance. The median external ESKD concordance index at the landmark time of 0.5 years was better than the internal ESKD concordance index. However, it had a wider interquartile range and did not achieve statistical significance in formal comparison ( $p = 0.056$ ). Reviewing the respective landmarks, 489 patients in the internal cohort had ESKD events, while only 50 ESKD events occurred in the external cohort. In addition, unlike the internal cohort used to train the algorithm, our validation cohort did not have mortality information. As a result, we could not estimate the C-index for death and the integrated Brier score.

Our experiment showed no performance difference between the standard landmarking approach that used the LOCF approach versus modeling the clinicopathological values using linear mixed models. Our findings further emphasized the difficulties of modeling real-life biomarker trajectories using linear mixed models observed in the literature (27, 28). This finding might extrapolate to the joint model's likely suboptimal performance, as the approach combines a Cox model with linear mixed models that estimate the trajectories of longitudinal biomarkers (35).

Furthermore, random survival forests with landmarking can tackle the limitation of restrictive proportional hazard assumptions, outperform joint modeling or Cox landmarking in the setting of a complex relationship between the survival and longitudinal biomarkers (29), and are also less sensitive to misspecification of longitudinal biomarker trajectories (34).

The strength of our study lies in the incorporation of high-dimensional, longitudinal clinicopathological data with a landmarking approach and a random survival forests capable of predicting ESKD with death as a competing risk. The models were built with a relatively large dataset and externally validated in a larger dataset in another Australian center. Another advantage of our approach is that it utilizes common clinicopathological data and does not require a urine albumin/creatinine ratio.

However, this approach has several limitations. We had to train six different models as we pre-specified six landmark times, requiring significant computing resources. Extending the landmark times would increase the entire model's size and limit the model's applicability. To ensure the models provide a continuous prediction for patients with accumulated data points over 3 years, we could apply the models as a rolling window. For example, patient Mr. X had clinicopathological data repeated for 4 years. Our model cannot process the clinicopathological data after 3 years and can only provide a prediction of ESKD or death up to 8 years when, theoretically, we have the information to predict Mr. X's outcome at 9 years. Therefore, we disregard the longitudinal data of the preceding 1 year and use the clinicopathological data at year 1 as input for the landmark 0.5-year model. As a result, this also means longitudinal data starting from 1.5 years will be used for the landmark 1-year model, longitudinal data from 2 years for the landmark 1.5-year model, and data from 2.5 years for the landmark 2-year model. Applying the approach will allow the model to provide an extended prediction when the patients have repeated tests after 3 years of

follow-up. To support this approach, our sensitivity analysis showed similar performance of models trained using the conventional time 0 definition (first kidney function test) versus using the earliest kidney function test within 10 years of events (“sens,” [Supplementary Figure 2](#) and [Supplementary Table 5](#)).

Another limitation is that we could only categorize the etiology as glomerulonephritis or vasculitis, while other ESKD aetiologies—such as diabetes, hypertension, post-nephrectomy, or other causes—had to be grouped into an “others” category. A kidney biopsy is required to definitively diagnose glomerulonephritis or vasculitis. However, biopsies are not always mandatory and can sometimes be relatively contraindicated, as in the case of single kidneys.

For patients with underlying conditions such as diabetes and hypertension, which are highly prevalent globally, determining the exact cause of ESKD can be particularly challenging.

Furthermore, the diagnosis category was identified as having low importance for model performance ([Figure 2](#)). We suspect that the varying kidney or mortality outcomes associated with different underlying conditions could be better predicted by the model using longitudinal data, such as eGFR or other biomarker measurements, rather than relying on baseline ESKD etiology.

Finally, it is crucial that we test the model performance in patients with reduced kidney function following an intervention, such as post-nephrectomy ([42](#)) or cryoablation ([43](#)), in future studies. We expect the longitudinal model, similar to ours, will have the advantage when predicting the outcomes, as the model can incorporate multiple sequential biomarker values before and after the procedure, and the cohort might experience some improvement in kidney function due to hypertrophy of the remaining kidney tissue.

In conclusion, we developed and validated a dynamic survival prediction model to predict ESKD, incorporating death as a competing risk. The model utilized common longitudinal clinicopathological data and demonstrated performance in satisfactory discrimination and calibration.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by ACT Health Research Ethics and Governance Office (ETHLR.18.040). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants’ legal guardians/next of kin because as per the conditions of the National Statement on Ethical Conduct in Human Research 2007 chapter 2.3.

## Author contributions

DC: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft,

Writing – review & editing. KC: Formal analysis, Validation, Writing – review & editing. AChu: Methodology, Writing – review & editing. BL: Methodology, Writing – review & editing. TA: Methodology, Writing – review & editing. ACha: Writing – review & editing. GW: Data curation, Methodology, Supervision, Writing – review & editing. SJ: Conceptualization, Supervision, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. DC was supported by the Australian Government Research Training Program, TALO Innovator Grant, and Jacquot Research Entry Scholarship. SJ is supported by the NHMRC (MRF2025693 and MRF20223294), Medical Research Future Fund, Jacquot Research Establishment and Career Development Awards, Canberra Hospital Private Practice Fund, The Pryor Bequest (from Jenny and Bruce Pryor), and the Hindmarsh Family and McCusker Charitable Foundation.

## Acknowledgments

The collection and validation of the external Western Australian dataset was supported by the Digital Health Cooperative Research Centre (“DHCRC”) and is part of a larger 4-year collaborative partnership between Curtin University, La Trobe University, Deakin University, WA Department of Health, WA Country Health Service, WA Primary Health Alliance, and the DHCRC. DHCRC is funded under the Commonwealth’s Cooperative Research Centres (CRC) Program.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2024.1428073/full#supplementary-material>

## References

- Hill NR, Fatoba ST, Oke JL, et al. Global prevalence of chronic kidney disease—a systematic review and Meta-analysis. *PLoS One*. (2016) 11:e0158765. doi: 10.1371/journal.pone.0158765
- Wang H, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the global burden of disease study 2015. *Lancet*. (2016) 388:1459–544. doi: 10.1016/S0140-6736(16)31012-1
- Bowe B, Xie Y, Li T, Mokdad AH, Xian H, Yan Y, et al. Changes in the US burden of chronic kidney disease from 2002 to 2016: an analysis of the global burden of disease study. *JAMA Netw Open*. (2018) 1:e184412. doi: 10.1001/jamanetworkopen.2018.4412
- Tangri N. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA*. (2011) 305:1553. doi: 10.1001/jama.2011.451
- Inoguchi T, Okui T, Nojiri C, Eto E, Hasuzawa N, Inoguchi Y, et al. A simplified prediction model for end-stage kidney disease in patients with diabetes. *Sci Rep*. (2022) 12:12482. doi: 10.1038/s41598-022-16451-5
- Bai Q, Su C, Tang W, Li Y. Machine learning to predict end-stage kidney disease in chronic kidney disease. *Sci Rep*. (2022) 12:8377. doi: 10.1038/s41598-022-12316-z
- Hui M, Ma J, Yang H, Gao B, Wang F, Wang J, et al. ESKD risk prediction model in a multicenter chronic kidney disease cohort in China: a derivation, validation, and comparison study. *J Clin Med*. (2023) 12:1504. doi: 10.3390/jcm12041504
- Schena FP, Anelli VW, Trotta J, di Noia T, Manno C, Tripepi G, et al. Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin a nephropathy. *Kidney Int*. (2021) 99:1179–88. doi: 10.1016/j.kint.2020.07.046
- 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. doi: 10.1016/j.ekir.2022.05.004
- Li L, Astor BC, Lewis J, Hu B, Appel LJ, Lipkowitz MS, et al. Longitudinal progression trajectory of GFR among patients with CKD. *Am J Kidney Dis*. (2012) 59:504–12. doi: 10.1053/j.ajkd.2011.12.009
- Matsushita K, Selvin E, Bash LD, Franceschini N, Astor BC, Coresh J. Change in estimated GFR associates with coronary heart disease and mortality. *J Am Soc Nephrol*. (2009) 20:2617–24. doi: 10.1681/ASN.2009010025
- Turin TC, Coresh J, Tonelli M, Stevens PE, de Jong PE, Farmer CKT, et al. One-year change in kidney function is associated with an increased mortality risk. *Am J Nephrol*. (2012) 36:41–9. doi: 10.1159/000339289
- Li L, Luo S, Hu B, Greene T. Dynamic prediction of renal failure using longitudinal biomarkers in a cohort study of chronic kidney disease. *Stat Biosci*. (2017) 9:357–78. doi: 10.1007/s12561-016-9183-7
- Wang S, Han J, Jung SY, Oh TJ, Yao S, Lim S, et al. Development and implementation of patient-level prediction models of end-stage renal disease for type 2 diabetes patients using fast healthcare interoperability resources. *Sci Rep*. (2022) 12:12322. doi: 10.1038/s41598-022-15036-6
- Chuah A, Walters G, Christiadi D, Karpe K, Kennard A, Singer R, et al. Machine learning improves upon clinicians' prediction of end stage kidney disease. *Front Med*. (2022) 9:837232. doi: 10.3389/fmed.2022.837232
- Noordzij M, Leffondre K, van Stralen KJ, Zoccali C, Dekker FW, Jager KJ. When do we need competing risks methods for survival analysis in nephrology? *Nephrol Dial Transplant*. (2013) 28:2670–7. doi: 10.1093/ndt/gft355
- Raynaud M, Aubert O, Divard G, Reese PP, Kamar N, Yoo D, et al. Dynamic prediction of renal survival among deeply phenotyped kidney transplant recipients using artificial intelligence: an observational, international, multicohort study. *Lancet Digit Health*. (2021) 3:e795–805. doi: 10.1016/S2589-7500(21)00209-0
- van den Brand JA, Dijkstra TMH, Wetzels J, Stengel B, Metzger M, Blankestijn PJ, et al. Predicting kidney failure from longitudinal kidney function trajectory: a comparison of models. *PLoS One*. (2019) 14:e0216559. doi: 10.1371/journal.pone.0216559
- 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. doi: 10.1053/j.ajkd.2016.07.030
- Wu C, Li L, Li R. Dynamic prediction of competing risk events using landmark sub-distribution hazard model with multiple longitudinal biomarkers. *Stat Methods Med Res*. (2020) 29:3179–91. doi: 10.1177/0962280220921553
- Rizopoulos D. Joint models for longitudinal and time-to-event data: with applications in R. New York: CRC press (2012).
- Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*. (2011) 67:819–29. doi: 10.1111/j.1541-0420.2010.01546.x
- Van Houwelingen HC. Dynamic prediction by Landmarking in event history analysis. *Scand J Stat*. (2007) 34:70–85. doi: 10.1111/j.1467-9469.2006.00529.x
- van Houwelingen H, Putter H. Dynamic prediction in clinical survival analysis. New York: CRC Press (2011).
- Zheng Y, Heagerty PJ. Partly conditional survival models for longitudinal data. *Biometrics*. (2005) 61:379–91. doi: 10.1111/j.1541-0420.2005.00323.x
- Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Med Res Methodol*. (2016) 16:117. doi: 10.1186/s12874-016-0212-5
- Hsieh F, Tseng Y, Wang J. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*. (2006) 62:1037–43. doi: 10.1111/j.1541-0420.2006.00570.x
- Wu L, Liu W, Yi GY, Huang Y. Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *J Probab Stat*. (2012) 2012:1–17. doi: 10.1155/2012/640153
- Pickett KL, Suresh K, Campbell KR, Davis S, Juarez-Colunga E. Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker. *BMC Med Res Methodol*. (2021) 21:216. doi: 10.1186/s12874-021-01375-x
- Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. (2008) 2:3. doi: 10.1214/08-AOAS169
- Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. (1999) 94:496–509. doi: 10.1080/01621459.1999.10474144
- Ishwaran H, Gerds TA, Kogalur UB, Moore RD, Gange SJ, Lau BM. Random survival forests for competing risks. *Biostatistics*. (2014) 15:757–73. doi: 10.1093/biostatistics/kxu010
- Wang H, Li G. A selective review on random survival forests for high dimensional data. *Quant Biosci*. (2017) 36:85–96. doi: 10.22283/qbs.2017.36.2.85
- Ferrer L, Putter H, Proust-Lima C. Individual dynamic predictions using landmarking and joint modelling: validation of estimators and robustness assessment. *Stat Methods Med Res*. (2019) 28:3649–66. doi: 10.1177/0962280218811837
- Rizopoulos D, Molenberghs G, Lesaffre EMEH. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biom J*. (2017) 59:1261–76. doi: 10.1002/bimj.201600238
- Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. (2013) 32:5381–97. doi: 10.1002/sim.5958
- Mogensen UB, Ishwaran H, Gerds TA. Evaluating random forests for survival analysis using prediction error curves. *J Stat Softw*. (2012) 50:1–23. doi: 10.18637/jss.v050.i11
- Tangri N, Grams ME, Levey AS, Coresh J, Appel LJ, Astor BC, et al. Multinational assessment of accuracy of equations for predicting risk of kidney failure: a Meta-analysis. *JAMA*. (2016) 315:164–74. doi: 10.1001/jama.2015.18202
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/> (Accessed June, 2023).
- Weaver RG, James MT, Ravani P, Weaver CGW, Lamb EJ, Tonelli M, et al. Estimating urine albumin-to-creatinine ratio from protein-to-creatinine ratio: development of equations using same-day measurements. *J Am Soc Nephrol*. (2020) 31:591–601. doi: 10.1681/ASN.2019060605
- Irish GL, Cuthbertson L, Kitsos A, Saunder T, Clayton PA, Jose MD. The kidney failure risk equation predicts kidney failure: validation in an Australian cohort. *Nephrology*. (2023) 28:328–35. doi: 10.1111/nep.14160
- Boltri M, Traunero F, Ongaro L, Migliozzi F, Vianello F, Lenardon O, et al. The added value of controlling nutritional status (Conut) score for preoperative counselling on significant early loss of renal function after radical nephrectomy for renal cell carcinoma. *Cancers*. (2024) 16:3519. doi: 10.3390/cancers16203519
- Piasentin A, Claps F, Silvestri T, Rebez G, Traunero F, Mir MC, et al. Assessing trifecta achievement after percutaneous Cryoablation of small renal masses: results from a multi-institutional collaboration. *Medicina*. (2022) 58:1041. doi: 10.3390/medicina58081041