Development and External Validation of a Multi-State Clinical Prediction Model for Chronic Kidney Disease Patients Progressing onto Renal Replacement Therapy and Death

MA Barrowman   GP Martin   N Peek   M Lambie   W Hulme

R Chinnadurai   J Lees   P Kalra   P Mark   J Traynor   M Sperrin

Contents

Last updated: 27 May

Abstract

Introduction

Clinical Prediction Models (CPMs) provide individualised predictions for patient outcomes. Traditionally, these models provide predictions for single outcomes, however in many circumstances, the ability to predict ~~multiple outcomes~~a multi-dimensional outcome with a single model can be advantageous. ~~Multi-State Models are a method to provide these kinds of predictions.~~ Many CPMs have been developed to predict the risk of different outcomes in individuals following chronic kidney disease (CKD) onset, but few allow the ability to predict the risk of patients transitioning onto renal replacement therapy (RRT) as well as death. For example, the risk of having a transplant within 1 year following dialysis, or the risk of remaining on dialysis until death. Multi-state models provide the vehicle to make such predictions, but have not been used within the CKD context.

Methods

We developed a Multi-State Clinical Prediction Model (MSCPM) using tertiary care data from the Salford Kidney Study as our development data set and secondary care data from the West of Scotland (SERPR) dataset as our external validation set. We developed three models of different levels of complexity; a Two-State Model (Alive and Dead), a Three-State Model (Untreated CKD, Renal Replacement Therapy and Dead) and a Five-State model (Untreated CKD, Haemodialysis, Peritoneal Dialysis, Transplant and Dead).~~ We used Royston-Parmer regression techniques to allow us to provide individualised predictions for patients.~~ Model performance was assessed for accuracy, discrimination and calibration using methods both internally and externally. The best performing model was used to produce a CPM Calculator for clinical use.

Results

Of the three models produced, Age was a strong predictor of mortality in all cases and outcomes were highly dependent on primary renal diagnosis. Models performed well in both the internal and external validation with the Three-State Model out performing overall with a Brier Score of 0.67/0.62 (internal/external, respectively), c-statistic of 0.83/0.81 and an averaged calibration intercept of 0.00/0.00 and slope diagonal of 1.34/1.53 (indicating under-prediction of all non-untreated CKD states for more extreme values). The Three-State Model was used to develop the online Calculator.

Discusssion

Our CPMs provide clinicians and patients with multi-dimensional predictions across different outcome states and any time point. This implies that users of these models can get more information about their potential future without a loss to the ~~quality of that prediction.~~model's calibration nor its discriminative ability.

1   Introduction

A clinical prediction model (CPM) is a tool which provides patients and clinicians with a measure of how likely a patient is to suffer a specific clinical event, more specifically, a prognostic model allows the prediction

of future events (Steyerberg et al. 2013). CPMs use data from previous patients to estimate the outcomes of an individual patient. Prognostic models can be used in clinical practice to influence treatment decisions.

Within Chronic Kidney Disease (CKD), prognostic models have been developed to predict mortality (Johnson et al. 2007; Landray et al. 2010; Bansal et al. 2015; Marks et al. 2015; Wick et al. 2017), End-Stage Renal Disease (Landray et al. 2010), the commencements of Renal Replacement Therapy (RRT) (Marks et al. 2015; Johnson et al. 2008; Schroeder et al. 2017; Kulkarni et al. 2017) or mortality after beginning dialysis (Floege et al. 2015; Hemke et al. 2013; Cao et al. 2015). Some previous models have used the commencement of RRT as a proxy for CKD Stage V (Tangri et al. 2011, 2017; Roy et al. 2017), while others have investigated the occurrence of cardiovascular events within CKD patients(Shlipak et al. 2005; Weiner et al. 2007; McMurray et al. 2011). Reviews by Grams & Coresh (Grams and Coresh 2013), Tangri et al (Tangri et al. 2013) and Ramspek et al (Ramspek et al. 2017), which explored the different aspects of assessing risk amongst CKD or RRT patients, found that the current landscape of CKD prediction models is lacking from both a methodological and clinical perspective (Collins et al. 2015; Bouwmeester et al. 2012).

Methodologically, the majority of existing CKD prediction models fail to account for competing events (Bansal et al. 2015; Wick et al. 2017; Perotte et al. 2015), have high risks of bias (Johnson et al. 2007, 2008; Landray et al. 2010) or are otherwise flawed compared to modern clinical prediction standards (Collins et al. 2015; Steyerberg et al. 2013)

In 2013, Begun et al (Begun et al. 2013) developed a multi-State model for assessing population-level progression through the severity stages of CKD (III-V), RRT and/or death, which can be used to provide a broad statement regarding a patient's future. In 2014, Allen et al (Allen et al. 2014) applied a similar model to liver transplant recipients and their progression through the stages of CKD with a focus on the predictions of measured vs estimated glomerular filtration rate (mGFR vs eGFR). In 2017, Kulkarni et al (Kulkarni et al. 2017) developed an MSM focusing on the categories of Calculated Panel Reactive Antibodies and kidney transplant and/or death.

Most recently, in 2018, Grams et al (Grams et al. 2018) developed a multinomial clinical prediction model for CKD patients which focused on the occurrence of RRT and/or cardiovascular events. As of the publication of this paper, this is the only currently existing CPMs of this kind for CKD patients.

However, the first three of these existing models (Begun, Allen and Kulkarni) categorise continuous variables to define their states at specific cut-offs and this has been shown to be inefficient when modelling (Royston, Altman, and Sauerbrei 2006) and none of these models have undergone any validation process, whether internal or external (Altman et al. 2009).

It is also important to note that although these models can be used to predict patient outcomes, they were not designed to produce individualised patient predictions as is a key aspect of a clinical prediction model; they were designed to assess the methodological advantages of MSMs in this medical field, to describe the prevalence of over time of different CKD stages and to produce population level predictions for patients with different levels of panel-reactive antibodies (Royston et al. 2009).

The fourth model (Grams), is presented as a Multi-State Model and the transitions involved were studied and defined, however the underlying statistical model is two multinomial logistic models analysed at 2 and 4 years, which assumes homogeneity of transition times. Two downsides to the implementation of this model are that it can only produce predictions at those predefined time points and that it is unable to estimate duration of time on dialysis.

Therefore, our aim is to develop a CPM - we do this by modelling patient pathways through a Multi-State Model by choosing transition points which can be exactly identified and include states which produce a clinical difference in patient characteristics. Our modeling techniques allow for individual predictions of ~~multiple outcomes~~multi-dimensional outcomes at any time point . The models produced by this process will then be validated, both internally and externally, to compare their results and demonstrate the transportability of the clinical prediction models. We report our work in line with the TRIPOD guidelines for development and validation of clinical prediction models (Collins et al. 2015; Moons et al. 2015).

2 Methods

2.1 Data Sources

The models were developed using data from the Salford Kidney Study (SKS) cohort of patients (previously named the CRISIS cohort), established in the Department of Renal Medicine, Salford Royal NHS Foundation Trust (SRFT). The SKS is a large longitudinal CKD cohort recruiting CKD patients since 2002. This cohort collects detailed annualised phenotypic and laboratory data, and plasma, serum and whole blood stored at -80°C for biomarker and genotypic analyses. Recruitment of patients into SKS has been described in multiple previous studies (Hoefield et al. 2010; Chinnadurai, Chrysochou, and Kalra 2019) and these have included a CKD progression prognostic factor study and to evidence the increased risk of cardiovascular events in diabetic kidney patients. In brief, any patient referred to Salford renal service (catchment population 1.5 million) who is 18 years or over and has an eGFR measurement of less than $60\text{ml/min}/1.73\text{m}^2$ (calculated using the CKD-EPI formula (Levey et al. 2009)) was approached to be consented for the study participation.

At baseline, the data, including demographics, comorbidities, physical parameters, lab results and primary renal diagnosis are recorded in the database. Patients undergo an annual study visit and any changes to these parameters are captured. All data except blood results are collected via questionnaire by a dedicated team of research nurses. Blood results (baseline and annualised), first RRT modality and mortality outcome data are directly transferred to the database from Salford's Integrated Record (New et al. 2014). eGFR, uPCR, comorbidity and blood results were measured longitudinally throughout a patient's time within the cohort.

Due to limitations in our data, we were agnostic to how long since patients were diagnosed with CKD. Therefore, we defined a patient's start date for our model as their first date after consent at which their eGFR was recorded to be below $60\text{ml/min}/1.73\text{m}^2$. Some patients consented with an eGFR that was already below 60, and some entered our study later when their eGFR was measured to be below 60. This implies that our models includes both patient who have recently been diagnosed with CKD (eGFR $\lessapprox 60$) *and* those that have been suffering with CKD for an arbitrary amount of time. This timelessness of the model means it can be applied to any patient at any time during their CKD journey prior to commencement of RRT.

All patients registered in the database between October 2002 and December 2016 with available data were included in this study. As this is a retrospective convenience sample, no sample size calculations were performed prior to recruitment. All patients were followed-up within SKS until the end-points of RRT, death or loss to follow-up or were censored at their last interaction with the healthcare system prior to December 2017. Date of death for patients who commenced RRT was also included in the SKS database.

For external validation of the model, we extracted an independent cohort from the West of Scotland Electronic Renal Patient Record (SERPR). Our extract of SERPR contains all patients known to the Glasgow and Forth Valley renal service who had an eGFR measure of less than $60\text{ml/min}/1.73m^2$ between January 2006 and January 2016. This cohort has been previously used in Chronic Kidney Disease Prognosis consortium studies investigating outcomes in patients with CKD (Matsushita et al. 2013) and a similar cohort has been used for the analysis of skin tumours amongst renal transplant patients. Use of anonymised data from this database has been approved by the West of Scotland Ethics Committee for use of NHS Greater Glasgow and Clyde 'Safe Haven' data for research.

Both the internal and external validation cohort were used as part of the multinational validation cohort used by Grams et al in their multinomial CPM discussed above (Grams et al. 2018). In SERPR, start dates were calculated to be the first time point where the following conditions were met:

- eGFR is measured at less than 60
- There is at least one prior eGFR measurement
- Patient is 18 or over

The second requirement was implemented to avoid a bias in the eGFR Rate. eGFR Rate is a measure of the change in eGFR over time and is calculated as the difference between the most recent two eGFR

measurements divided by the time between them. For patients who entered the system with an eGFR < 60, their eGFR Rate would be unavailable (i.e. missing). Otherwise, patient eGFRs would *have* to drop to below 60 and thus eGFR Rate would be negative. In addition, to avoid Acute Kidney Incident events, we also filtered measurements of eGFR Rate that were more extreme than 1.5x those found in the SKS population.

2.2 Model Design

Three separate models were developed, so we could determine a clinically viable model while maintaining model parsimony as much as possible: a Two-State, Three-State and Five-State model, each building on the previous models' complexity (see figure 1). The Two-State model was a traditional survival analysis where a single event (death) is considered. The Three-State model expanded on this, by splitting the Alive state into transient states of (untreated) CKD and (first) RRT; patients can therefore transition from CKD to Death or CKD to RRT, and then onto RRT to Death. The Five-State model stratifies the RRT state into HD, PD and Tx and allows similar transitions into and out of the RRT states; however, the transition from Tx to Death was not considered as it was anticipated a priori that there would be insufficient patients undergoing this transition and that the process of undergoing a transplant would be medically transformative and so it would be inappropriate to assume shared parameters before and after the transition (i.e. Tx was modelled as a second absorbing state). Missing data was handled using multiple imputation (White and
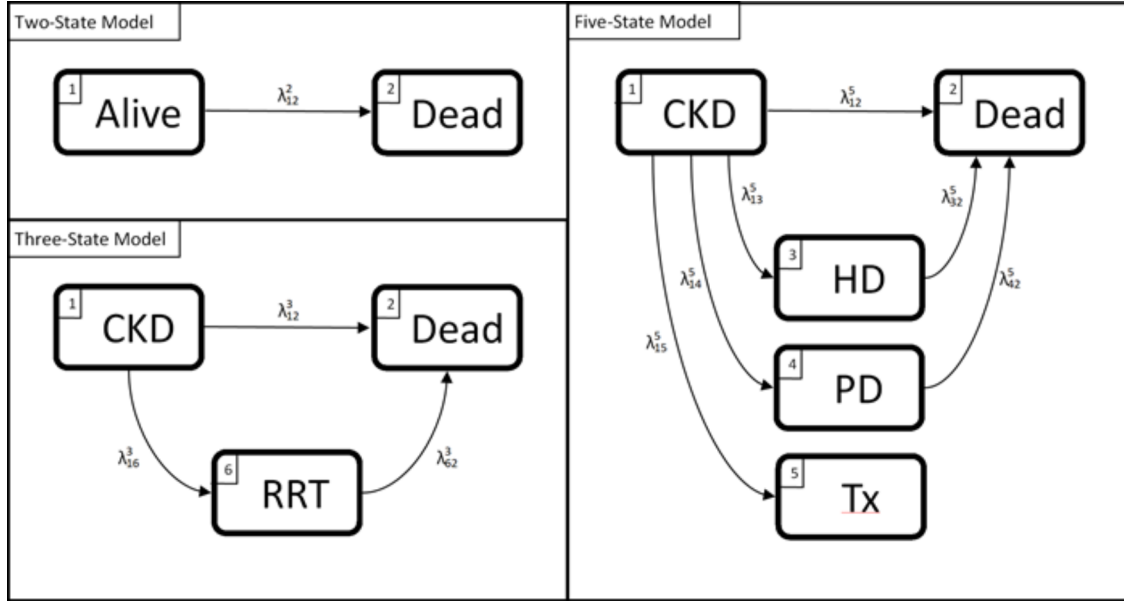

Figure 1: Diagram of the three models, the states being modelled and relevant transitions

Royston 2009) using times for all events as imputation covariates. Variables considered as covariates were demographics (sex, age, smoking status and alcohol consumption), comorbidities (congestive cardiac failure (CCF), chronic obstructive pulmonary disease (COPD), prior cerebrovascular accident (CVA), hypertension (HT), diabetes mellitus (DM), ischemic heart disease (IHD), chronic liver disease (LD), prior myocardial infarction (MI), peripheral vascular disease (PVD) and slid tumour (ST)), physical parameters (BMI, blood pressure), blood results (haemoglobin, albumin, corrected calcium and phosphate measures), urine protein creatinine ratio (uPCR) and primary renal diagnosis (grouped as per ERA-EDTA classifications (Venkat-Raman et al. 2012)). From these variables, uPCR and eGFR Rate of change were also derived (Kovesdy et al. 2016; Naimark et al. 2016) as well as their log transforms and $\log(\text{Age})$ and Age^2. For each transition, Backwards-forwards stepwise methods were used to minimise the AIC, allowing different variables to be used for each transition.

2.3 Validation

Each of the three models were internally validated in the development dataset using bootstrapping to adjust for optimism and then further externally validated in the validation dataset extracted from SERPR (Schomaker and Heumann 2018). The bootstrapping method was also used for both validations to produce confidence intervals around the performance metric estimates. To assess the performance in low eGFR patients, the models were also validated in subsets of the SKS and SERPR where patients had an eGFR $<$ $30/\mathrm{ml/min}/1.73\mathrm{m}^2$.

Model accuracy was assessed using the Brier Score, discrimination was assessed using the c-statistic and the calibration was assessed using the multi-dimensional intercept and slope-matrix (MA Barrowman, n.d.). These measures were taken at One-Year, Two-Years and Five-Years after the patient's start dates, and therefore are equivalent to validation measures at that point in time after a prediction has been applied.

Further details of how the models were developed and validated is discussed in the Supplementary materials.

2.4 Example

Once the models have been developed, we will apply them to three example patients to demonstrate their use and applicability to the general population. We will provide a direct clinical estimation of these patient outcomes based on years of nephrological experience and compare this with the results presented by our clinical prediction model.

We have chosen three (synthetic) patients to use as examples of the use of our model. Their details can be seen in table 1. Out three example patients cover a broad range of ages and other covariates. A clinically

Table 1: Details of the Example Patients

	Patient 1	Patient 2	Patient 3
Age	20	40	66
Gender	Female	Male	Female
Smoking Status	Non-Smoker	Smoker	Non-Smoker
BP	144/101	160/90	140/80
Albumin	39	40	40
Correct Calcium	2.3	3.0	2.6
Haemoglobin	150	100	14
Phosphate	0.68	2.00	0.86
eGFR	42	10	51
eGFR Previous	50 (one week ago)	30 (one year ago)	70 (one week ago)
uPCR	0.30	0.20	0.01
uPCR Previous	0.80 (one month ago)	1.20 (one year ago)	0.06 (one week ago)
Primary Diagnosis	Glomerulonephritis	Tubular Necrosis	Diabetes
Comorbities	Chronic Obstructive Pulmonary Disease Liver Disease Solid Tumour		Diabetes Chronic Obstructive Pulmonary Disease Hypertension

guided prediction for these patients would assume that Patient 1 has a high chance of proceeding as normal (with little need for RRT), Patient 2 would be recommended to start RRT soon and Patient 3 would be predicted to have a high risk of mortality with or without RRT.

2.5 Calculator

As part of this work, we have also produced an online calculator to allow patients and clinicians to easily estimate outcomes without worrying about the mathematics involved.

All analysis was done in `R 3.6.2` (Team, n.d.) using the various `tidyverse` packages (Wickham 2017), as well as the `mice` (Buuren and Groothuis-Oudshoorn 2011), `flexsurv` (Jackson, n.d.), `nnet` (Ripley and Venables 2016) and `furrr` (Vaughan and Dancho 2018) packages. The calculator was produced using the `shiny` package (Chang et al. 2020).

3 Results

3.1 Data Sources

As seen in table 2, the Age of the populations had a mean of 64.4 and 65.9 respectively with a very broad range. Due to the inclusion criteria, eGFR were capped at a maximum of 60, and was consistent across populations; however, the rate of change for eGFR was much wider in the SERPR patients than in the SKS, and it was decreasing much faster, on average (-25 vs 0) . The uPCR measures are presented in our results as g/mmol, rather than the more conventional g/mol, this is to better present results and coefficients of varying magnitudes. Levels of missingness were much higher in the SERPR dataset in most continuous variables. Table 2 also shows a breakdown of the categorical variables across the populations. In the development population, crude proportions of males were numerically higher than females whereas in the validation population the proportions are much more matched (62.6% male vs 50.4% male). Most patients were white in the SKS dataset, and ethnicity has extremely high missingness in SERPR, which also contributed to its omission from the model.

Overall, there were high levels of comorbidities within the SKS population, but these levels were much lower in the SERPR population.

The median date for the date of death was 3.9 years in the SKS population and 4.9 years in the SERPR population. The median date for transition to RRT was 2.2 years and 1.5 years (in SKS and SERPR respectively). In SKS, transitions to HD happened 6 months later than PD, and in SERPR it was 3.6 months. The Maximum followup time in SKS was 15.0 years and in SERPR it was 10.1 years. This information can be seen in table 3.

3.2 Development

Here we present the proportional hazards results for the Three-State Model. The full model description, including proportional and baseline hazards can be found in the Supplementary Materials. Older patients are predicted to be likely to transition to RRT. Increased rates of decline of eGFR were associated with the transition from CKD to RRT. The full results are shown in table 4. Female patients are predicted to be more likely to remain in the CKD state than Males, or to remain in the RRT state once there. Smokers were predicted as more likely than Non-/Former Smokers to undergo any transition, apart from CKD to Tx. Blood results had associations with all transitions in some way, and disease etiology were strongly associated with the transitions giving a wide range of predictions.

3.3 Validation

Table 5 shows the results from the internal validation in the Three-State Model. Performance was overall slightly better in patients in the <60 eGFR group than in the <30 eGFR group. All measures degraded over time, but the average scores remained strong. Table 6 shows the results from the external validation in the Three-State Model.

3.4 Example

The example patients seen in Table 1 were passed through our Three-State prediction model and the results for all time-points are shown in figure 2. The prognosis for all three patients were very different. Patient

Table 2: Patient Demographics, Continuous variables are displayed as mean (Inter-Quartile Range), and Categorical/Comorbidity data as number (percent)

Variable	SKS			SERPR	
	mean/n (IQR/p%)	missing n (p%)	range	mean/n (IQR/p%)	missing n (p%)
Age	64.378 (19.000)	0 (0.00%)	[20.000, 94.000]	66.064 (17.000)	0 (0.00%)
eGFR	30.369 (22.387)	0 (0.00%)	[3.578, 59.966]	35.647 (20.565)	0 (0.00%)
eGFR Rate	-0.016 (0.294)	1,278 (42.87%)	[-19.107, 33.782]	1.319 (21.897)	0 (0.00%)
SBP	140.193 (29.000)	50 (1.68%)	[77.000, 220.000]	145.981 (30.000)	6,345 (82.31%)
DBP	74.556 (14.000)	52 (1.74%)	[36.000, 159.000]	76.742 (17.000)	6,345 (82.31%)
BMI	28.848 (7.842)	572 (19.19%)	[13.182, 61.467]	29.020 (6.811)	7,491 (97.17%)
Albumin	42.152 (5.000)	60 (2.01%)	[12.000, 52.000]	36.475 (6.000)	3,134 (40.65%)
Calcium	2.302 (0.180)	68 (2.28%)	[1.210, 3.660]	2.407 (0.160)	4,513 (58.54%)
Haemoglobin	122.978 (23.000)	72 (2.42%)	[61.000, 195.000]	109.269 (29.000)	3,557 (46.14%)
Phosphate	1.163 (0.320)	87 (2.92%)	[0.430, 3.710]	1.194 (0.320)	4,510 (58.50%)
uPCR	0.112 (0.103)	245 (8.22%)	[0.001, 2.025]	0.178 (0.177)	7,170 (93.01%)
uPCR Rate	-0.096 (0.188)	1,777 (59.61%)	[-70.727, 28.199]	4.621 (0.905)	7,495 (97.22%)
Gender:		0 (0.00%)			0 (0.00%)
Male	1,865 (62.56%)			3,885 (50.40%)	
Female	1,116 (37.44%)			3,824 (49.60%)	
Ethnicity:		0 (0.00%)			7,009 (90.92%)
White	2,875 (96.44%)			679 (8.81%)	
Asian	75 (2.52%)			12 (0.16%)	
Black	21 (0.70%)			7 (0.09%)	
Other	10 (0.34%)			2 (0.03%)	
SmokingStatus:		42 (1.41%)			7,709 (100.00%)
Former	1,535 (51.49%)			0 (0.00%)	
Non-Smoker	979 (32.84%)			0 (0.00%)	
Smoker	379 (12.71%)			0 (0.00%)	
Former 3Y	46 (1.54%)			0 (0.00%)	
DiagGroup:		567 (19.02%)			6,640 (86.13%)
Systemic diseases affecting the kidney	1,304 (43.74%)			316 (4.10%)	
Glomerular disease	442 (14.83%)			278 (3.61%)	
Tubulointerstitial disease	268 (8.99%)			173 (2.24%)	
Miscellaneous renal disorders	227 (7.61%)			198 (2.57%)	
Familial / hereditary nephropathies	173 (5.80%)			104 (1.35%)	
Diabetes (DM)	992 (33.32%)	4 (0.13%)		1,535 (19.91%)	0 (0.00%)
Congestive Cardiac Failure (CCF)	2,414 (81.09%)	4 (0.13%)		408 (5.29%)	0 (0.00%)
Prior Myocardial Infarction (MI)	2,492 (83.71%)	4 (0.13%)		556 (7.21%)	0 (0.00%)
Ischemic Heart Disease (IHD)	2,393 (80.38%)	4 (0.13%)		863 (11.19%)	0 (0.00%)
Peripheral Vascular Disease (PVA)	2,485 (83.47%)	4 (0.13%)		376 (4.88%)	0 (0.00%)
Prior Cerebrovascular Accident (CVA)	2,727 (91.60%)	4 (0.13%)		186 (2.41%)	0 (0.00%)
Chronic Obstructive Pulmonary Disease (COPD)	2,411 (80.99%)	4 (0.13%)		0 (%)	7,709 (100.00%)
Chronic Liver Disease (LD)	2,891 (97.11%)	4 (0.13%)		0 (%)	7,709 (100.00%)
Prior Solid Tumour (ST)	2,570 (86.33%)	4 (0.13%)		0 (%)	7,709 (100.00%)
Hypertension (HT)	2,546 (91.48%)	198 (6.64%)		3,114 (40.39%)	0 (0.00%)

1 (20 year old) had a very high probability of survival, with only an 16% chance of mortality by year 10 and 0% chance of commencing RRT. Patient 2 (40 year old) was predicted almost 90% chance of starting RRT, and over 70% chance of dying overall (either with or without RR). Patient 3 (66 year old) had a fast acceleration towards high mortality, after 1 year from the recorded measurements, they had more than 50% chance of dying, and after 2 years that probability rises to over 85% with no chance of RRT.

3.5 Calculator

The calculator is available online here:

https://michael-barrowman.shinyapps.io/MSCPM_for_CKD_Patients/.

Table 3: Event times for the two populations presented as Number of Events   Median (Inter-Quartile Range) [Max]

Transition			SKS				SERPR			
Model	From	To	n	median	IQR	max	n	median	IQR	max
Two	Alive	Dead	1,427	4.0 y	4.3 y	15.0 y	3,010	4.8 y	3.3 y	10.1 y
Three	CKD	Dead	1,125	3.5 y	4.2 y	15.0 y	2,568	4.7 y	3.3 y	10.1 y
		RRT	680	2.5 y	3.4 y	14.2 y	1,125	3.8 y	3.9 y	10.1 y
	RRT	Dead	302	2.2 y	3.3 y	13.5 y	442	1.6 y	2.5 y	9.2 y
Five	CKD	Dead	1,125	3.5 y	4.2 y	15.0 y	2,568	4.7 y	3.3 y	10.1 y
		HD	344	2.6 y	3.5 y	14.2 y	882	3.7 y	3.8 y	10.1 y
		PD	229	2.0 y	2.9 y	12.9 y	149	3.5 y	4.1 y	9.5 y
		Tx	107	3.2 y	2.7 y	12.1 y	94	5.0 y	4.4 y	9.8 y
	HD	Dead	185	2.1 y	3.3 y	11.8 y	394	1.6 y	2.5 y	9.2 y
	PD	Dead	107	2.4 y	3.2 y	11.8 y	47	2.1 y	2.3 y	8.5 y

Table 4: Proportional Hazards for each transition in the Three-State Model

Age
  (Age-60)   0.162 (-
  (Age-60)   -0.001 (-
  log(Age)   -5.725 (-
eGFR
  eGFR   -0.013 (-
  eGFR Rate
  log(eGFR Rate)   0.043 (-
uPCR
  uPCR   0.125 (-
  uPCR Rate
  log(uPCR Rate)
Measures
  SBP   -0.001 (-
  DBP   0.007 (
  BMI
  Albumin   -0.045 (-
  Corrected Calcium   0.280 (-
  Haemoglobin   -0.013 (-
  Phosphate   0.512 (
Gender
  Female   -0.236 (-
Smoking Status
  Former (3 years+)   -0.213 (-
  Non-Smoker   -0.198 (-
  Smoker   0.356 (
Primary Renal Diagnosis
  Familial / hereditary nephropathies   -0.424 (-
  Glomerular disease   -0.395 (-
  Miscellaneous renal disorders   -0.263 (-
  Tubulointerstitial disease   -0.463 (-
Comorbidity
  DM   0.122 (-
  CCF   -0.395 (-
  MI   -0.246 (-
  IHD   0.102 (-
  PVD   -0.249 (-
  CVA   -0.071 (-
  COPD   -0.290 (-
  LD   -0.170 (-
  ST   -0.274 (-
  HT

Table 5: Internal Validation of the Three-State Model, results presented as Estimate (95% CI, where possible)

Predicting	eGFR	One Year	Two Year	Five Year	Average
Brier					
Three	< 60	0.74 (0.74, 0.75)	0.68 (0.68, 0.69)	0.64 (0.64, 0.65)	0.67 (0.67, 0.68)
Three	< 30	0.75 (0.74, 0.75)	0.73 (0.73, 0.73)	0.68 (0.67, 0.68)	0.68 (0.67, 0.68)
Two	< 60	0.75 (0.75, 0.75)	0.75 (0.75, 0.76)	0.67 (0.67, 0.67)	0.67 (0.67, 0.68)
Two	< 30	0.71 (0.71, 0.72)	0.72 (0.72, 0.73)	0.65 (0.65, 0.66)	0.67 (0.67, 0.68)
c-statistic					
Three	< 60	0.87 (0.87, 0.87)	0.84 (0.84, 0.85)	0.84 (0.84, 0.84)	0.83 (0.83, 0.84)
Three	< 30	0.87 (0.86, 0.87)	0.84 (0.84, 0.84)	0.84 (0.84, 0.84)	0.83 (0.83, 0.84)
Two	< 60	0.86 (0.86, 0.86)	0.86 (0.86, 0.86)	0.83 (0.83, 0.84)	0.83 (0.83, 0.84)
Two	< 30	0.86 (0.85, 0.86)	0.86 (0.85, 0.86)	0.85 (0.85, 0.85)	0.84 (0.83, 0.84)
Intercept					
		-0.00 (-0.00, 0.00)	0.00 (-0.00, 0.00)	-0.02 (-0.02, -0.01)	0.00 (0.00, 0.00)
Three	< 60	0.00 (0.00, 0.01)	-0.01 (-0.02, -0.01)	-0.01 (-0.02, -0.01)	0.00 (-0.00, 0.00)
		0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	-0.00 (-0.01, -0.00)	0.00 (-0.00, 0.00)
		-0.01 (-0.01, -0.01)	0.00 (-0.00, 0.00)	-0.02 (-0.02, -0.01)	0.00 (0.00, 0.00)
Three	< 30	-0.00 (-0.00, -0.00)	-0.01 (-0.01, -0.00)	0.00 (-0.00, 0.00)	0.00 (-0.00, 0.00)
		0.00 (-0.00, 0.00)	0.03 (0.02, 0.03)	-0.00 (-0.01, -0.00)	0.00 (-0.00, 0.00)
Two	< 60	-0.00 (-0.01, -0.00)	0.02 (0.01, 0.02)	0.00 (-0.00, 0.00)	-0.00 (-0.00, 0.00)
Two	< 30	-0.00 (-0.00, 0.00)	-0.04 (-0.04, -0.04)	0.00 (0.00, 0.01)	-0.00 (-0.00, 0.00)
Slope					
		1.25, 0.00, 0.04	1.17, -0.06, 0.01	1.21, -0.02, -0.01	1.32, -0.01, 0.00
Three	< 60	-0.03, 1.10, -0.00	0.03, 1.25, -0.04	0.01, 1.44, -0.04	-0.00, 1.37, -0.01
		-0.00, 0.01, 1.16	-0.01, 0.01, 1.37	-0.02, 0.02, 1.27	-0.00, 0.00, 1.33
		1.21, 0.02, 0.07	1.36, -0.00, -0.01	1.35, -0.04, -0.01	1.31, 0.00, 0.00
Three	< 30	-0.04, 1.24, 0.07	0.00, 1.31, 0.03	0.04, 1.33, 0.04	-0.01, 1.33, 0.01
		0.01, 0.01, 1.16	-0.02, -0.00, 1.26	0.04, -0.02, 1.34	-0.00, 0.00, 1.35
Two	< 60	1.21	1.28	1.27	1.31
Two	< 30	1.05	1.21	1.21	1.34

Table 6: External Validation of the Three-State Model, results presented as Estimate (95% CI, where possible)

| Predicting	eGFR	One Year	Two Year	Five Year	Average
Brier					
Three	< 60	0.69 (0.68, 0.69)	0.70 (0.69, 0.70)	0.61 (0.60, 0.61)	0.62 (0.62, 0.63)
Three	< 30	0.68 (0.67, 0.68)	0.72 (0.71, 0.72)	0.65 (0.64, 0.65)	0.63 (0.62, 0.63)
Two	< 60	0.67 (0.67, 0.67)	0.70 (0.69, 0.70)	0.63 (0.63, 0.63)	0.62 (0.62, 0.63)
Two	< 30	0.66 (0.66, 0.67)	0.70 (0.70, 0.70)	0.65 (0.65, 0.66)	0.63 (0.62, 0.63)
c-statistic					
Three	< 60	0.82 (0.82, 0.82)	0.83 (0.83, 0.83)	0.79 (0.79, 0.79)	0.81 (0.80, 0.81)
Three	< 30	0.85 (0.84, 0.85)	0.84 (0.84, 0.84)	0.83 (0.83, 0.83)	0.81 (0.81, 0.81)
Two	< 60	0.85 (0.85, 0.86)	0.84 (0.84, 0.85)	0.80 (0.80, 0.80)	0.81 (0.80, 0.81)
Two	< 30	0.83 (0.83, 0.83)	0.82 (0.82, 0.82)	0.80 (0.80, 0.81)	0.81 (0.81, 0.81)
Intercept					
		0.00 (-0.00, 0.00)	0.01 (0.01, 0.02)	0.05 (0.04, 0.05)	-0.00 (-0.00, 0.00)
Three	< 60	0.00 (0.00, 0.00)	-0.01 (-0.01, -0.00)	0.01 (0.00, 0.01)	-0.00 (-0.00, 0.00)
		-0.00 (-0.00, -0.00)	0.00 (0.00, 0.01)	0.00 (0.00, 0.00)	0.00 (0.00, 0.01)
		0.04 (0.04, 0.05)	0.02 (0.01, 0.02)	0.01 (0.01, 0.01)	0.00 (-0.00, 0.00)
Three	< 30	0.01 (0.00, 0.01)	-0.00 (-0.00, 0.00)	-0.00 (-0.00, 0.00)	0.00 (-0.00, 0.00)
		0.00 (0.00, 0.01)	0.01 (0.01, 0.02)	0.00 (0.00, 0.01)	-0.00 (-0.00, 0.00)
Two	< 60	0.00 (-0.00, 0.00)	-0.05 (-0.05, -0.04)	0.01 (0.01, 0.02)	-0.00 (-0.00, 0.00)
Two	< 30	0.02 (0.01, 0.02)	-0.00 (-0.00, 0.00)	0.01 (0.00, 0.01)	-0.00 (-0.00, 0.00)
Slope					
		1.35, -0.04, 0.04	1.13, -0.00, 0.08	1.45, -0.02, 0.03	1.54, 0.00, -0.00
Three	< 60	0.06, 1.49, -0.03	-0.02, 1.31, 0.01	0.03, 1.73, 0.03	0.01, 1.52, 0.00
		0.01, 0.00, 1.25	-0.03, 0.01, 1.39	0.04, -0.00, 1.47	-0.00, 0.00, 1.54
		1.23, 0.03, 0.00	0.99, -0.01, -0.00	1.56, 0.00, 0.00	1.62, -0.01, 0.01
Three	< 30	-0.05, 1.20, -0.04	0.04, 1.34, -0.03	-0.05, 1.49, 0.03	-0.00, 1.53, 0.01
		-0.00, -0.00, 1.37	0.00, 0.01, 1.46	0.03, 0.01, 1.63	0.00, -0.00, 1.58
Two	< 60	1.28	1.26	1.64	1.51
Two	< 30	1.18	1.25	1.57	1.59
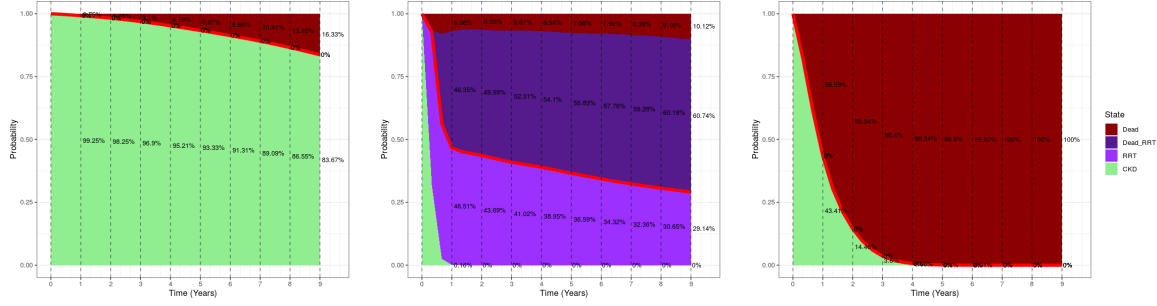
Figure 2: Results of Example Patients

4 Discussion

We have used data provided by SKS to develop a Multi-State Clinical Prediction Model and then validated this model within the SKS and SERPR datasets. Within our Models, the cause of a patient's renal disease had the widest effect on patient outcomes meaning that outcomes are highly dependent on ERA-EDTA classification of the diagnosis. Most groupings resulted in a lowered hazard of death and an increased hazard of RRT compared to the baseline of Systemic diseases.

Models performed well in model validation with the Three-State Model slightly out performing the other two models in calibration and overall predictive ability, however the Five-State model performed marginally better in terms of discriminative ability. Both Multi-State Models outperformed the Two-State (Traditional) Model.

The application of a Multi-state clinical prediction model to this field is novel and gives a powerful tool for providing individualised predictions of different outcomes at a wide range of time points. The model can also be used to estimate an expected amount of time that a patient will be on RRT The general inclusion criteria for the development dataset, and the wide range of patient ages and measurements allows for the model to be applied to a broad spectrum of patients.

Although the inclusion criteria for SKS were broad, the demographics of the local area resulted in homogeneity of ethnicity, which may create a limitation to the applicability of our model. The Renal Department at SRFT is a tertiary care facility for CKD sufferers and is well renowned for its capabilities of care meaning that it is likely to attract less-healthy patients from a wider catchment area, making the cohort of patients in the development population in worse condition than the general population of CKD patients.

There were also high levels of missingness in the eGFR and uPCR rates of changes would also produce a bias, due to these measures likely being missing not at random. The derivation of the validation dataset ensured that all patients had an eGFR Rate measurement; this was done to avoid data missing not at random (only negative or missing data would be available as patient's eGFR dropped to less than 60), however deriving data in this way could itself induce a survivor bias in the start date used for patients.

In the Five-State Model, We omitted the analysis of the Tx to Dead state due to the anticipated low number of events within the SKS dataset. The lowest number of events for a transition was therefore PD to Dead, which had only 107. Altogether, we considered 26 covariates (with 4 categorical covariates) and so this equates to 36 predictor parameters and an events per predictor parameter (EPP) of 2.97. This is below the recommendations of Riley et al (Riley, Snell, et al. 2019), whose calculations produce a requirement of 4.54 EPP. This requirement was also not satisfied by the CKD to PD transition (EPP = 6.36,required = 10.2) or the CKD to Tx transition (EPP = 2.97, required = 17.6). Fortunately, this limitation is confined to the Five-State Model.

We have assumed a proportional hazards relationship between the predictors and probability of survival, which is considered by some to be a strong assumption to make, however we acknowledge this limitation, and the authors believe that it is mitigated by the flexibility that the assumption permits. In addition to the general PH assumption, the R-P model requires the assumption that the log cumulative hazard function

follows a cubic spline, (however this is a much weaker assumption (Royston and Parmar 2002)), which is modelled as part of the regression. We did not assess the viability of these models as it was believed this assumption to make our results more understandable.

Compared to the apparent internal validation, the model performance during the external validation was worse for all metrics. However, once adjusted for optimism, the results were much more cohesive which implies that the model is highly transportable to a new population without much alterations being required. Due to the differences in the healthcare systems of England and Scotland, it can be appreciated that despite the populations being similar, their care would be different enough to emphasise a larger difference between our populations than that shown in our (relatively homogeneous) populations.

Our paper has clearly demonstrated the accuracy of such a model. However, further research would be needed to establish the effectiveness and efficacy of its use in clinical practice (Moons et al. 2009) by comparing it to standard care and establishing whether the use of our model improves patient outcomes.

All three models produced for this work performed well in terms of accuracy, calibration and discrimination when applied internally and externally. This shows directly that the models are suitable for use in populations similar to both our development and our validation datasets. It can also be concluded that the models can be transported and applied to any population with a similar healthcare system to the UK.

A   Supplementary Material

A.1   Statistical Analysis

A.1.1   Development

Data was recorded in a time-updated manner, however all variables were measured at baseline to emulate the real-world application of the model (i.e. future prediction of states and not covariates). As well as those included in the main text, ethnicity was assessed in the populations, but as most patients were white, it was omitted as a potential predictor from the models.

log(Calendar Time) was included as a covariate to adjust for secular trends in treatment preferences (Bhatnagar et al. 2015). Calendar Time was defined as length of time between start date and 1st January 2019.

Intermediate states (RRT or modality) were considered to be medically transformative, and so a semi-markov (clock reset) method for analysis was considered to be well justified (Meira-Machado et al. 2009). Each transition was modelled under a proportional hazards assumption using the Royston-Parmar technique (Royston and Parmar 2002) to estimate coefficients for each covariate and a restricted cubic spline (on the log-time scale) for the baseline cumulative hazard. The cumulative hazards for each transition can be combined to produce estimates for the probability of a patient being in any state at any time (Putter, Fiocco, and Geskus 2007).

For variable selection, we stacked the imputed datasets together to create a larger, pseudo-population (Wood, White, and Royston 2008) and performed backwards-forwards selection based on minimising the AIC at each step. This was repeated for each transition and for different numbers of evenly spaced knots in modelling the form of the baseline hazard, K={0,1,2,3,4,5}. This allowed for different transitions to use different sets of variables and numbers of knots in the final model. Some combinations of variables resulted in models that were intractable and so these models were excluded. Once a set of variables were chosen, the R-P model was applied to each imputed dataset individually and the resulting coefficients and cubic spline parameters were aggregated across imputations using Rubin's Rules (Rubin 1984). This gave a model fully defined by smooth cubic splines representing the cumulative cause-specific hazard and individualised proportional hazards for each transition.

All missing data were assumed to be missing at random and so were multiply imputed using chained equations with the Nelson-Aalen estimators for each relevant transition as predictors (White and Royston 2009). Some

variables (smoking status and histories of COPD, LD and ST) were present in the SKS (development) dataset, but were completely missing in the SERPR extract (validation) and so these were multiply imputed from the development dataset (Janssen et al. 2009).

A.1.2 Validation

For validation purposes, we consider Death and Death after RRT/HD/PD to be distinct states meaning that for the Three-State model, we have $K = 4$ pathways a patient can take and for the Five-State model, we have $K = 7$. To compare across models, we combined states together to collapse down to simpler versions. We collapsed the Three-State model to a two-state structure by combining the CKD and RRT states into an Alive state. We collapsed the Five-State model to a three-state structure by combining the HD, PD and Tx into an RRT state and then further down to a two-state structure as with the Three-State model. We will report performance measures at 360 days (approx. 1-year), 720 days (approx. 2-years) and 1800 days (approx. 5-years). As well as presenting the performance measures over time.

The overall accuracy of each model was assessed using the MSM adjusted Brier Score (MA Barrowman, n.d.), which is a proper score function (Gneiting and Raftery 2007) assigning 0 to a non-informative model and 1 to a perfect model, with negative numbers implying the model performs worse than assuming every patient's state predictions are the same as the overall prevalence within the population.

The discrimination of each model was assessed using the MSM extension to the c-statistic (Calster et al. 2012; MA Barrowman, n.d.). The c-statistic is a score between 0 and 1 with higher scores suggesting a better model and a c-statistic of 0.5 suggesting the model performs no better than a non-informative model.

The calibration of each model was assessed using MSM multinomial logistic regression (MLR) (Hoorde et al. 2014; MA Barrowman, n.d.) which extends the logistic regression to three or more mutually exclusive outcomes (Riley, van der Windt, et al. 2019). This produces an intercept vector of length $K-1$ and a Slope-matrix of dimension $(K-1) \times (K-1)$. As with the traditional calibration intercept for a well performing model, the MLR intercept values should all be as close to 0 as possible. The traditional calibration slope should be as close to 1 as possible and so the multi-state extension of the slope, the Slope-matrix should be as close to the identity matrix (I) as possible.

A.2 Model Results

A.2.1 Two State Model

Table 7 shows the proportional hazard ratios for the transitions in the Two-State Model. Older patients have a higher hazard towards death, low adn decreasing eGFR increased hazard as did a history of diabetes. Patients with a primary renal diagnosis included in the ERA-EDTA (Venkat-Raman et al. 2012) definition of Systemic diseases affecting the kidney had the highest likelihood of death. Equation (1) below shows the baseline cumulative hazard functions for the transition from Alive to Dead in the Two-State Model.

$$\Lambda_{0,16}(t) = \begin{cases} 1.15764 \log(t) + 6.17808 & 0 \leq t < 3 \\ 0.00075 \log(t)^3 - 0.00247 \log(t)^2 + 1.16036 \log(t) + 6.17708 & 3 \leq t < 520 \\ 0.13983 \log(t)^3 - 2.61165 \log(t)^2 + 17.47602 \log(t) - 27.83122 & 520 \leq t < 984 \\ -0.19361 \log(t)^3 + 4.28227 \log(t)^2 - 30.03426 \log(t) + 81.30981 & 984 \leq t < 1454 \\ -0.15458 \log(t)^3 + 3.42966 \log(t)^2 - 23.82553 \log(t) + 66.23901 & 1454 \leq t < 2009 \\ 0.37097 \log(t)^3 - 8.56171 \log(t)^2 + 67.37553 \log(t) - 164.97265 & 2009 \leq t < 2900 \\ -0.16209 \log(t)^3 + 4.18769 \log(t)^2 - 34.26862 \log(t) + 105.14551 & 2900 \leq t < 5497 \\ 1.79561 \log(t) + 1.61764)) & 5497 \leq t \end{cases} \quad (1)$$

Table 8 shows the results from the internal validation in the Two-State Model. Calibration Intercept is close to 0, implying the model is well calibrated overall with a high c-statistic and Brier Score. Calibration Slope

Table 7: Proportional Hazards for each transition in the Two-State Model

Age	
(Age-60)	0.110 (-
(Age-60)	-0.000 (-
log(Age)	-2.853 (-:
eGFR	
eGFR	-0.013 (-
eGFR Rate	-0.007 (-
log(eGFR Rate)	0.090 (-
uPCR	
uPCR	
uPCR Rate	
log(uPCR Rate)	
Measures	
SBP	
DBP	0.004 (-
BMI	
Albumin	-0.048 (-
Corrected Calcium	0.222 (-
Haemoglobin	-0.011 (-
Phosphate	0.338 (
Gender	
Female	-0.172 (-:
Smoking Status	
Former (3 years+)	-0.403 (-
Non-Smoker	-0.226 (-
Smoker	0.376 (
Primary Renal Diagnosis	
Familial / hereditary nephropathies	-0.399 (-
Glomerular disease	-0.406 (-
Miscellaneous renal disorders	-0.220 (-
Tubulointerstitial disease	-0.452 (-
Comorbidity	
DM	0.144 (
CCF	-0.378 (-
MI	-0.165 (-
IHD	0.070 (-
PVD	-0.240 (-
CVA	-0.128 (-
COPD	-0.203 (-
LD	-0.241 (-
ST	-0.299 (-
HT	-0.080 (-

Table 8: Internal Validation of the Two-State Model, results presented as Estimate (95% CI, where possible)

Predicting	eGFR	One Year	Two Year	Five Year	Average
Brier					
Two	< 60	0.63 (0.62, 0.63)	0.69 (0.69, 0.69)	0.66 (0.66, 0.67)	0.63 (0.62, 0.63)
Two	< 30	0.71 (0.71, 0.72)	0.68 (0.68, 0.69)	0.66 (0.66, 0.66)	0.63 (0.63, 0.64)
c-statistic					
Two	< 60	0.82 (0.82, 0.82)	0.85 (0.84, 0.85)	0.81 (0.81, 0.81)	0.81 (0.81, 0.82)
Two	< 30	0.84 (0.84, 0.84)	0.83 (0.82, 0.83)	0.83 (0.82, 0.83)	0.81 (0.81, 0.81)
Intercept					
Two	< 60	0.01 (0.00, 0.01)	0.01 (0.00, 0.01)	-0.02 (-0.02, -0.01)	-0.00 (-0.01, -0.00)
Two	< 30	-0.02 (-0.02, -0.02)	0.00 (0.00, 0.01)	0.00 (0.00, 0.01)	-0.00 (-0.00, -0.00)
Slope					
Two	< 60	1.33	1.46	1.26	1.48
Two	< 30	1.23	1.25	1.30	1.51

above 1 implies that the model under-estimates outcomes. Table 9 shows the results from the external validation in the Two-State Model, which shows similar results to the internal validation with slightly impaired perfomance, which is to be expected in an external validation.

Table 9: External Validation of the Two-State Model, results presented as Estimate (95% CI, where possible)

Predicting	eGFR	One Year	Two Year	Five Year	Average
Brier					
Two	< 60	0.64 (0.63, 0.64)	0.57 (0.56, 0.57)	0.57 (0.56, 0.58)	0.56 (0.56, 0.57)
Two	< 30	0.67 (0.66, 0.67)	0.64 (0.63, 0.64)	0.57 (0.56, 0.57)	0.57 (0.56, 0.57)
c-statistic					
Two	< 60	0.81 (0.81, 0.82)	0.81 (0.80, 0.81)	0.80 (0.79, 0.80)	0.78 (0.78, 0.78)
Two	< 30	0.81 (0.81, 0.81)	0.80 (0.80, 0.81)	0.78 (0.78, 0.79)	0.78 (0.78, 0.78)
Intercept					
Two	< 60	-0.00 (-0.00, 0.00)	0.02 (0.01, 0.02)	0.00 (0.00, 0.01)	-0.00 (-0.00, 0.00)
Two	< 30	0.02 (0.01, 0.02)	-0.05 (-0.05, -0.04)	0.01 (0.01, 0.02)	-0.00 (-0.00, 0.00)
Slope					
Two	< 60	1.29	1.25	1.72	2.21
Two	< 30	1.37	1.37	2.05	1.88

A.2.2 Three State Model

The equations (2), (3) and (4) shows the baseline cumulative hazard functions for the transition from CKD to Dead, CKD to RRT and RRT to Dead, respectively in the Three-State Model.

$$\Lambda_{0,16}(t) = \begin{cases} 1.19795\log(t) + 17.68798 & 0 \le t < 3 \\ -9e-05\log(t)^3 + 3e-04\log(t)^2 + 1.19761\log(t) + 17.68811 & 3 \le t < 443 \\ 0.15869\log(t)^3 - 2.9019\log(t)^2 + 18.88018\log(t) - 18.22403 & 443 \le t < 873 \\ -0.30096\log(t)^3 + 6.43659\log(t)^2 - 44.36299\log(t) + 124.54338 & 873 \le t < 1295 \\ -0.04158\log(t)^3 + 0.86028\log(t)^2 - 4.40166\log(t) + 29.08554 & 1295 \le t < 1876 \\ 0.51263\log(t)^3 - 11.67048\log(t)^2 + 90.03919\log(t) - 208.17245 & 1876 \le t < 2738 \\ -0.23992\log(t)^3 + 6.19863\log(t)^2 - 51.3924\log(t) + 164.96467 & 2738 \le t < 5497 \\ 1.98997\log(t) + 11.72244)) & 5497 \le t \end{cases} \tag{2}$$

$$\Lambda_{0,15}(t) = \begin{cases} 1.55753\log(t) - 5.44635 & 0 \le t < 18 \\ -0.00279\log(t)^3 + 0.0242\log(t)^2 + 1.48757\log(t) - 5.37894 & 18 \le t < 270 \\ -0.13576\log(t)^3 + 2.25776\log(t)^2 - 11.01817\log(t) + 17.96111 & 270 \le t < 538 \\ 0.49133\log(t)^3 - 9.57146\log(t)^2 + 63.36209\log(t) - 137.93609 & 538 \le t < 919 \\ -0.76978\log(t)^3 + 16.24541\log(t)^2 - 112.80782\log(t) + 262.78176 & 919 \le t < 1316 \\ 0.30039\log(t)^3 - 6.81452\log(t)^2 + 52.82254\log(t) - 133.77073 & 1316 \le t < 2000 \\ -0.0123\log(t)^3 + 0.31552\log(t)^2 - 1.37092\log(t) + 3.53243 & 2000 \le t < 5173 \\ 1.32717\log(t) - 4.15818)) & 5173 \le t \end{cases} \tag{3}$$

$$\Lambda_{0,56}(t) = \begin{cases} 1.35522\log(t) - 7.7618 & 0 \le t < 8 \\ -0.01704\log(t)^3 + 0.1063\log(t)^2 + 1.13417\log(t) - 7.60859 & 8 \le t < 196 \\ 0.21761\log(t)^3 - 3.6103\log(t)^2 + 20.75671\log(t) - 42.14228 & 196 \le t < 506 \\ -0.67558\log(t)^3 + 13.07415\log(t)^2 - 83.12956\log(t) + 173.47479 & 506 \le t < 816 \\ 0.8043\log(t)^3 - 16.69103\log(t)^2 + 116.42771\log(t) - 272.49495 & 816 \le t < 1388 \\ -1.26732\log(t)^3 + 28.27738\log(t)^2 - 208.94656\log(t) + 512.26643 & 1388 \le t < 1927 \\ 0.17019\log(t)^3 - 4.34244\log(t)^2 + 37.78861\log(t) - 109.83234 & 1927 \le t < 4940 \\ 0.85568\log(t) - 5.12598)) & 4940 \le t \end{cases} \tag{4}$$

15

Validation results for the Three-State Model can be found in the main text.

A.2.3   Five State Model

Table 10 shows the proportional hazard ratios for the transitions in the Five-State Model.

Table 10: Proportional Hazards for each transition in the Five-State Model

	CKD to Dead
Age	
(Age-60)	0.161 (-0.051, 0.374)
(Age-60)	-0.000 (-0.002, 0.000)
log(Age)	-5.725 (-17.969, 6.518)
eGFR	
eGFR	-0.013 (-0.019, -0.006)
eGFR Rate	
log(eGFR Rate)	0.042 (-0.125, 0.210)
uPCR	
uPCR	0.125 (-0.318, 0.569)
uPCR Rate	
log(uPCR Rate)	
Measures	
SBP	-0.001 (-0.004, 0.002)
DBP	0.006 (0.000, 0.013)
BMI	
Albumin	-0.044 (-0.064, -0.024)
Corrected Calcium	0.280 (-0.192, 0.752)
Haemoglobin	-0.013 (-0.017, -0.008)
Phosphate	0.511 (0.132, 0.890)
Gender	
Female	-0.235 (-0.371, -0.099)
Smoking Status	
Former (3 years+)	-0.212 (-0.879, 0.453)
Non-Smoker	-0.198 (-0.345, -0.051)
Smoker	0.356 (0.160, 0.551)
Primary Renal Diagnosis	
Familial / hereditary nephropathies	-0.424 (-0.854, 0.006)
Glomerular disease	-0.394 (-0.635, -0.154)
Miscellaneous renal disorders	-0.263 (-0.505, -0.021)
Tubulointerstitial disease	-0.463 (-0.741, -0.184)
Comorbidity	
DM	0.122 (-0.011, 0.255)
CCF	-0.394 (-0.535, -0.253)
MI	-0.246 (-0.397, -0.094)
IHD	0.102 (-0.041, 0.245)
PVD	-0.248 (-0.394, -0.103)
CVA	-0.070 (-0.252, 0.111)
COPD	-0.289 (-0.433, -0.145)
LD	-0.169 (-0.578, 0.239)
ST	-0.274 (-0.431, -0.117)
HT	

The equations (6), (7), (8) and (5) show the baseline cumulative hazard functions from the CKD state to HD, PD, Tx and Dead, respectively. Equation (9) shows the baseline cumulative hazard function from HD to Dead and Equation (10) shows the baseline cumulative hazard function from PD to Dead.

$$\Lambda_{0,16}(t) = \begin{cases} 1.19795\log(t) + 17.68798 & 0 \le t < 3 \\ -9e-05\log(t)^3 + 3e-04\log(t)^2 + 1.19761\log(t) + 17.68811 & 3 \le t < 443 \\ 0.15869\log(t)^3 - 2.9019\log(t)^2 + 18.88018\log(t) - 18.22403 & 443 \le t < 873 \\ -0.30096\log(t)^3 + 6.43659\log(t)^2 - 44.36299\log(t) + 124.54338 & 873 \le t < 1295 \\ -0.04158\log(t)^3 + 0.86028\log(t)^2 - 4.40166\log(t) + 29.08554 & 1295 \le t < 1876 \\ 0.51263\log(t)^3 - 11.67048\log(t)^2 + 90.03919\log(t) - 208.17245 & 1876 \le t < 2738 \\ -0.23992\log(t)^3 + 6.19863\log(t)^2 - 51.3924\log(t) + 164.96467 & 2738 \le t < 5497 \\ 1.98997\log(t) + 11.72244)) & 5497 \le t \end{cases} \quad (5)$$

$$\Lambda_{0,12}(t) = \begin{cases} 2.10248\log(t) - 8.40415 & 0 \le t < 22 \\ -0.02708\log(t)^3 + 0.25114\log(t)^2 + 1.3262\log(t) - 7.60432 & 22 \le t < 398 \\ 0.1382\log(t)^3 - 2.71684\log(t)^2 + 19.09196\log(t) - 43.05188 & 398 \le t < 939 \\ -0.08093\log(t)^3 + 1.78288\log(t)^2 - 11.70775\log(t) + 27.2208 & 939 \le t < 1680 \\ 0.00603\log(t)^3 - 0.15477\log(t)^2 + 2.68263\log(t) - 8.40353 & 1680 \le t < 5173 \\ 1.35914\log(t) - 4.63106)) & 5173 \le t \end{cases} \quad (6)$$

$$\Lambda_{0,13}(t) = \begin{cases} 1.54717\log(t) - 2.95572 & 0 \le t < 18 \\ -0.0099\log(t)^3 + 0.08586\log(t)^2 + 1.29901\log(t) - 2.71663 & 18 \le t < 399 \\ 0.06206\log(t)^3 - 1.20706\log(t)^2 + 9.04226\log(t) - 18.17463 & 399 \le t < 1143 \\ -0.02442\log(t)^3 + 0.61979\log(t)^2 - 3.82137\log(t) + 12.01805 & 1143 \le t < 4720 \\ 1.42179\log(t) - 2.7669)) & 4720 \le t \end{cases} \quad (7)$$

$$\Lambda_{0,14}(t) = \begin{cases} 1.2532\log(t) - 26.76925 & 0 \le t < 57 \\ 0.05585\log(t)^3 - 0.67736\log(t)^2 + 3.99179\log(t) - 30.46001 & 57 \le t < 673 \\ -1.53678\log(t)^3 + 30.432\log(t)^2 - 198.5656\log(t) + 409.16648 & 673 \le t < 968 \\ 1.6017\log(t)^3 - 34.30313\log(t)^2 + 246.5168\log(t) - 610.87894 & 968 \le t < 1408 \\ -1.63794\log(t)^3 + 36.16075\log(t)^2 - 264.35866\log(t) + 623.7665 & 1408 \le t < 1874 \\ 0.33702\log(t)^3 - 8.48952\log(t)^2 + 72.12758\log(t) - 221.49141 & 1874 \le t < 4432 \\ 0.84439\log(t) - 21.97914)) & 4432 \le t \end{cases} \quad (8)$$

$$\Lambda_{0,26}(t) = \begin{cases} 1.29244\log(t) - 8.98979 & 0 \le t < 8 \\ -0.01397\log(t)^3 + 0.08716\log(t)^2 + 1.1112\log(t) - 8.86417 & 8 \le t < 204 \\ 0.26151\log(t)^3 - 4.30937\log(t)^2 + 24.49956\log(t) - 50.33752 & 204 \le t < 479 \\ -0.93294\log(t)^3 + 17.80365\log(t)^2 - 111.9599\log(t) + 230.35972 & 479 \le t < 764 \\ 1.30576\log(t)^3 - 26.7818\log(t)^2 + 184.02359\log(t) - 424.6091 & 764 \le t < 1217 \\ -1.24908\log(t)^3 + 27.66815\log(t)^2 - 202.79666\log(t) + 491.39983 & 1217 \le t < 1828 \\ 0.18579\log(t)^3 - 4.66437\log(t)^2 + 40.05781\log(t) - 116.63908 & 1828 \le t < 4310 \\ 1.02311\log(t) - 7.74927)) & 4310 \le t \end{cases} \quad (9)$$

$$\Lambda_{0,36}(t) = \begin{cases} 1.55188\log(t) - 8.40785 & 0 \le t < 9 \\ -0.02867\log(t)^3 + 0.18901\log(t)^2 + 1.13657\log(t) - 8.10367 & 9 \le t < 167 \\ 0.24473\log(t)^3 - 4.00711\log(t)^2 + 22.60308\log(t) - 44.70978 & 167 \le t < 553 \\ -0.99362\log(t)^3 + 19.45322\log(t)^2 - 125.54725\log(t) + 267.14313 & 553 \le t < 875 \\ 1.22972\log(t)^3 - 25.73085\log(t)^2 + 180.53981\log(t) - 424.02428 & 875 \le t < 1410 \\ -1.86859\log(t)^3 + 41.67205\log(t)^2 - 308.23775\log(t) + 757.44556 & 1410 \le t < 1937 \\ 0.31596\log(t)^3 - 7.93077\log(t)^2 + 67.19201\log(t) - 189.72839 & 1937 \le t < 4302 \\ 0.83657\log(t) - 4.66673)) & 4302 \le t \end{cases} \tag{10}$$

Table 11 shows the results from the internal validation in the Five-State Model. The calibration slope results are shown in a seperate table for both the internal and external validation. Table 12 shows the results from

Table 11: Internal Validation of the Five-State Model, results presented as Estimate (95% CI, where possible)

Predicting	eGFR	One Year	Two Year	Five Year	Average
Brier					
Five	< 60	0.74 (0.74, 0.75)	0.72 (0.72, 0.72)	0.67 (0.66, 0.67)	0.69 (0.69, 0.69)
Five	< 30	0.76 (0.75, 0.76)	0.71 (0.71, 0.72)	0.65 (0.65, 0.66)	0.68 (0.68, 0.69)
Three	< 60	0.72 (0.72, 0.73)	0.72 (0.72, 0.72)	0.66 (0.66, 0.67)	0.68 (0.68, 0.69)
Three	< 30	0.72 (0.71, 0.72)	0.74 (0.73, 0.74)	0.67 (0.67, 0.68)	0.69 (0.68, 0.69)
Two	< 60	0.74 (0.74, 0.74)	0.72 (0.72, 0.73)	0.67 (0.67, 0.68)	0.69 (0.68, 0.69)
Two	< 30	0.74 (0.73, 0.74)	0.72 (0.72, 0.73)	0.72 (0.71, 0.72)	0.69 (0.69, 0.69)
c-statistic					
Five	< 60	0.88 (0.88, 0.88)	0.86 (0.85, 0.86)	0.83 (0.83, 0.84)	0.84 (0.84, 0.84)
Five	< 30	0.88 (0.87, 0.88)	0.87 (0.87, 0.87)	0.86 (0.86, 0.86)	0.84 (0.84, 0.85)
Three	< 60	0.87 (0.87, 0.87)	0.87 (0.86, 0.87)	0.84 (0.84, 0.84)	0.84 (0.84, 0.84)
Three	< 30	0.86 (0.86, 0.86)	0.87 (0.87, 0.87)	0.85 (0.85, 0.85)	0.84 (0.84, 0.85)
Two	< 60	0.86 (0.86, 0.86)	0.86 (0.86, 0.87)	0.84 (0.84, 0.84)	0.84 (0.84, 0.84)
Two	< 30	0.86 (0.86, 0.86)	0.87 (0.87, 0.88)	0.81 (0.81, 0.81)	0.84 (0.84, 0.84)
Intercept					
Five	< 60	0.00 (-0.00, 0.00)	0.01 (0.00, 0.01)	-0.01 (-0.01, -0.00)	0.00 (-0.00, 0.00)
		-0.01 (-0.01, -0.01)	0.01 (0.01, 0.01)	-0.00 (-0.01, -0.00)	0.00 (-0.00, 0.00)
		-0.01 (-0.01, -0.01)	-0.00 (-0.00, 0.00)	-0.01 (-0.01, -0.00)	-0.00 (-0.00, 0.00)
		0.01 (0.00, 0.01)	-0.00 (-0.00, -0.00)	-0.04 (-0.05, -0.04)	-0.00 (-0.00, 0.00)
		-0.00 (-0.00, 0.00)	0.00 (0.00, 0.01)	0.01 (0.00, 0.01)	-0.00 (-0.00, 0.00)
		-0.01 (-0.01, -0.00)	0.00 (0.00, 0.00)	-0.00 (-0.00, -0.00)	0.00 (0.00, 0.01)
		-0.02 (-0.02, -0.02)	0.00 (0.00, 0.00)	-0.00 (-0.00, 0.00)	-0.00 (-0.01, -0.00)
		0.00 (0.00, 0.01)	0.00 (0.00, 0.00)	0.00 (0.00, 0.01)	0.00 (0.00, 0.00)
Five	< 30	-0.00 (-0.01, -0.00)	-0.01 (-0.02, -0.01)	-0.00 (-0.00, 0.00)	0.00 (-0.00, 0.00)
		-0.00 (-0.01, -0.00)	-0.02 (-0.02, -0.01)	-0.01 (-0.01, -0.01)	-0.00 (-0.00, 0.00)
		0.01 (0.01, 0.01)	-0.00 (-0.00, -0.00)	-0.01 (-0.01, -0.01)	0.00 (-0.00, 0.00)
		0.00 (0.00, 0.00)	-0.00 (-0.00, -0.00)	-0.00 (-0.00, -0.00)	-0.00 (-0.00, 0.00)
Three	< 60	-0.00 (-0.01, -0.00)	0.01 (0.01, 0.02)	0.02 (0.02, 0.02)	0.00 (-0.00, 0.00)
		0.00 (-0.00, 0.00)	0.00 (-0.00, 0.00)	-0.02 (-0.02, -0.02)	-0.00 (-0.00, 0.00)
		-0.01 (-0.01, -0.00)	0.00 (0.00, 0.00)	0.02 (0.02, 0.02)	-0.00 (-0.00, -0.00)
Three	< 30	-0.00 (-0.01, -0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	-0.00 (-0.00, 0.00)
		0.00 (-0.00, 0.00)	0.01 (0.00, 0.01)	-0.01 (-0.02, -0.01)	0.00 (-0.00, 0.00)
		0.01 (0.00, 0.01)	0.01 (0.01, 0.01)	-0.01 (-0.01, -0.00)	0.00 (-0.00, 0.00)
Two	< 60	-0.01 (-0.01, -0.00)	0.01 (0.00, 0.01)	0.00 (-0.00, 0.00)	-0.00 (-0.01, -0.00)
Two	< 30	-0.03 (-0.03, -0.03)	0.02 (0.01, 0.02)	-0.00 (-0.00, 0.00)	0.00 (-0.00, 0.00)
Slope					
Three	< 60	1.19, 0.01, 0.01	1.16, -0.01, -0.01	1.24, -0.00, 0.00	1.30, 0.01, -0.00
		-0.01, 1.07, -0.02	-0.00, 1.11, -0.01	-0.07, 1.28, 0.00	0.00, 1.29, -0.00
		-0.07, 0.00, 1.11	0.01, -0.00, 1.35	-0.03, 0.01, 1.18	-0.00, 0.00, 1.32
Three	< 30	1.21, -0.01, 0.01	1.12, 0.01, 0.00	1.34, 0.12, -0.02	1.27, 0.00, -0.00
		-0.02, 1.18, -0.01	-0.02, 1.27, 0.00	0.02, 1.43, -0.05	-0.00, 1.28, 0.00
		0.02, -0.01, 1.15	0.03, 0.04, 1.18	0.00, 0.09, 1.19	-0.00, -0.00, 1.29
Two	< 60	1.21	1.03	1.25	1.25
Two	< 30	1.05	1.16	1.38	1.28

the external validation in the Five-State Model. Table 13 shows the calibration slopes for the model in the internal and external datasets in both the < 60 eGFR and <30 eGFR sub-populations.

Table 12: External Validation of the Five-State Model, results presented as Estimate (95% CI, where possible)

Predicting	eGFR	One Year	Two Year	Five Year	Average
Brier					
Five	< 60	0.70 (0.70, 0.71)	0.72 (0.71, 0.72)	0.64 (0.64, 0.65)	0.63 (0.63, 0.64)
Five	< 30	0.73 (0.72, 0.73)	0.71 (0.70, 0.71)	0.67 (0.67, 0.68)	0.64 (0.64, 0.65)
Three	< 60	0.69 (0.68, 0.69)	0.71 (0.71, 0.71)	0.67 (0.66, 0.67)	0.64 (0.63, 0.64)
Three	< 30	0.69 (0.68, 0.69)	0.68 (0.68, 0.69)	0.63 (0.63, 0.64)	0.63 (0.62, 0.63)
Two	< 60	0.69 (0.68, 0.69)	0.68 (0.68, 0.69)	0.60 (0.60, 0.60)	0.64 (0.63, 0.64)
Two	< 30	0.73 (0.72, 0.73)	0.67 (0.67, 0.67)	0.64 (0.63, 0.64)	0.63 (0.63, 0.64)
c-statistic					
Five	< 60	0.85 (0.85, 0.85)	0.85 (0.84, 0.85)	0.82 (0.82, 0.82)	0.82 (0.81, 0.82)
Five	< 30	0.85 (0.85, 0.85)	0.83 (0.82, 0.83)	0.82 (0.82, 0.82)	0.81 (0.81, 0.82)
Three	< 60	0.83 (0.83, 0.84)	0.83 (0.82, 0.83)	0.81 (0.81, 0.82)	0.81 (0.81, 0.82)
Three	< 30	0.87 (0.87, 0.87)	0.84 (0.84, 0.84)	0.82 (0.82, 0.82)	0.82 (0.81, 0.82)
Two	< 60	0.84 (0.84, 0.85)	0.85 (0.84, 0.85)	0.81 (0.81, 0.81)	0.82 (0.82, 0.82)
Two	< 30	0.84 (0.83, 0.84)	0.84 (0.84, 0.84)	0.81 (0.81, 0.82)	0.82 (0.81, 0.82)
Intercept					
Five	< 60	-0.01 (-0.01, -0.01)	0.01 (0.00, 0.01)	0.00 (0.00, 0.01)	-0.00 (-0.00, 0.00)
		-0.02 (-0.02, -0.01)	-0.01 (-0.02, -0.01)	-0.00 (-0.01, -0.00)	-0.00 (-0.01, -0.00)
		-0.02 (-0.02, -0.01)	-0.01 (-0.02, -0.01)	-0.02 (-0.03, -0.02)	0.00 (-0.00, 0.00)
		0.04 (0.03, 0.04)	-0.02 (-0.03, -0.02)	0.01 (0.01, 0.02)	-0.00 (-0.00, 0.00)
		0.00 (0.00, 0.00)	-0.01 (-0.01, -0.00)	0.00 (-0.00, 0.00)	-0.00 (-0.01, -0.00)
		-0.01 (-0.01, -0.00)	0.01 (0.00, 0.01)	0.00 (0.00, 0.00)	0.00 (-0.00, 0.00)
Five	< 30	0.02 (0.01, 0.02)	0.02 (0.02, 0.03)	0.03 (0.02, 0.03)	0.00 (-0.00, 0.00)
		-0.01 (-0.01, -0.00)	-0.02 (-0.02, -0.01)	0.00 (0.00, 0.01)	-0.00 (-0.00, -0.00)
		0.02 (0.01, 0.02)	-0.02 (-0.02, -0.01)	-0.02 (-0.02, -0.01)	0.00 (-0.00, 0.00)
		-0.02 (-0.03, -0.02)	-0.00 (-0.00, -0.00)	-0.00 (-0.01, -0.00)	0.00 (-0.00, 0.00)
		0.03 (0.02, 0.03)	0.02 (0.02, 0.02)	-0.00 (-0.00, 0.00)	-0.00 (-0.00, -0.00)
		0.01 (0.01, 0.01)	-0.01 (-0.01, -0.00)	0.02 (0.01, 0.02)	-0.00 (-0.00, 0.00)
Three	< 60	0.00 (0.00, 0.00)	-0.01 (-0.01, -0.00)	0.01 (0.00, 0.01)	0.00 (0.00, 0.00)
		-0.01 (-0.02, -0.01)	0.02 (0.02, 0.03)	-0.03 (-0.04, -0.03)	-0.00 (-0.00, 0.00)
		-0.02 (-0.02, -0.02)	0.01 (0.01, 0.02)	-0.02 (-0.03, -0.02)	0.00 (-0.00, 0.00)
Three	< 30	-0.00 (-0.01, -0.00)	0.01 (0.01, 0.02)	0.02 (0.02, 0.02)	-0.00 (-0.00, 0.00)
		-0.02 (-0.02, -0.02)	-0.00 (-0.01, -0.00)	-0.05 (-0.05, -0.04)	-0.00 (-0.00, 0.00)
		-0.03 (-0.04, -0.03)	-0.00 (-0.01, -0.00)	0.01 (0.01, 0.01)	-0.00 (-0.00, 0.00)
Two	< 60	0.00 (0.00, 0.01)	-0.02 (-0.02, -0.02)	-0.00 (-0.00, 0.00)	0.00 (-0.00, 0.00)
Two	< 30	-0.02 (-0.02, -0.01)	0.02 (0.02, 0.03)	0.01 (0.01, 0.01)	-0.00 (-0.00, 0.00)
Slope					
Three	< 60	*1.01*, -0.00, -0.00	*1.40*, 0.02, -0.02	*1.61*, -0.03, 0.04	*1.46*, -0.00, -0.00
		0.01, *1.21*, -0.05	-0.03, *1.28*, -0.03	0.03, *1.76*, 0.00	0.00, *1.50*, -0.01
		0.08, -0.05, *1.22*	-0.07, 0.06, *1.12*	0.04, -0.00, *1.45*	-0.00, -0.00, *1.45*
Three	< 30	*1.24*, -0.00, -0.00	*1.39*, 0.03, 0.00	*1.28*, 0.02, 0.03	*1.51*, -0.00, 0.00
		-0.02, *1.20*, -0.05	-0.04, *1.24*, 0.04	-0.09, *1.38*, 0.02	0.00, *1.41*, -0.00
		-0.01, -0.07, *1.34*	-0.03, -0.01, *1.18*	-0.07, -0.00, *1.44*	-0.00, -0.00, *1.47*
Two	< 60	*1.38*	*1.12*	*1.25*	*1.46*
Two	< 30	*1.26*	*1.28*	*1.35*	*1.55*

References

Allen, Alina M., W. Ray Kim, Terry M. Therneau, Joseph J. Larson, Julie K. Heimbach, and Andrew D. Rule. 2014. "Chronic Kidney Disease and Associated Mortality After Liver Transplantation–a Time-Dependent Analysis Using Measured Glomerular Filtration Rate." *Journal of Hepatology* 61 (2): 286–92. https://doi.org/10.1016/j.jhep.2014.03.034.

Altman, Douglas G., Yvonne Vergouwe, Patrick Royston, and Karel G. M. Moons. 2009. "Prognosis and Prognostic Research: Validating a Prognostic Model." *BMJ* 338 (May): b605. https://doi.org/10.1136/bmj.b605.

Bansal, Nisha, Ronit Katz, Ian H. De Boer, Carmen A. Peralta, Linda F. Fried, David S. Siscovick, Dena E. Rifkin, et al. 2015. "Development and Validation of a Model to Predict 5-Year Risk of Death Without ESRD Among Older Adults with CKD." *Clinical Journal of the American Society of Nephrology: CJASN* 10 (3): 363–71. https://doi.org/10.2215/CJN.04650514.

Table 13: Calibration Slope results for both the External and Internal Validation for the Five-State Model

	Internal < 60	Internal < 30	
One Year	*1.14*, -0.01, 0.00, 0.02, 0.00, 0.01	*1.07*, 0.03, 0.02, 0.02, 0.01, -0.00	*1.16*
	-0.00, *1.09*, -0.05, -0.01, 0.01, 0.00	0.00, *1.18*, 0.01, 0.00, -0.06, 0.01	-0.03
	-0.04, 0.06, *1.22*, 0.02, 0.03, 0.00	0.01, 0.00, *1.03*, 0.00, -0.05, 0.00	-0.05
	-0.06, 0.02, 0.01, *1.25*, 0.00, -0.02	-0.04, -0.03, -0.03, *1.11*, 0.01, -0.03	-0.01
	-0.02, 0.03, 0.02, -0.02, *1.13*, 0.00	0.00, -0.03, -0.02, -0.02, *1.15*, 0.01	-0.01
	0.05, -0.04, -0.06, -0.02, -0.00, *1.30*	0.02, -0.01, 0.00, -0.00, -0.01, *1.11*	0.
Two Year	*1.42*, -0.05, 0.03, -0.01, -0.05, 0.00	*1.14*, -0.02, 0.04, 0.02, 0.00, -0.09	*1.25*
	0.02, *1.12*, 0.03, -0.01, 0.03, -0.00	-0.01, *1.10*, 0.00, -0.07, -0.01, -0.00	0.00
	0.01, -0.01, *1.22*, -0.00, 0.00, 0.04	-0.03, 0.02, *1.12*, 0.03, 0.00, -0.04	-0.01
	-0.03, 0.00, -0.05, *1.21*, -0.01, 0.00	0.00, 0.02, -0.03, *1.11*, -0.02, 0.01	-0.00
	-0.02, -0.03, -0.00, -0.01, *1.39*, -0.00	-0.00, -0.06, -0.05, 0.02, *1.11*, -0.03	-0.02
	-0.03, 0.04, 0.04, -0.00, 0.03, *1.15*	0.07, -0.01, -0.03, 0.02, 0.04, *1.18*	-0.
Five Year	*1.22*, -0.00, 0.05, -0.00, -0.05, 0.05	*1.27*, 0.02, 0.02, -0.03, 0.04, -0.06	*1.45*
	-0.04, *1.11*, -0.01, 0.03, 0.04, 0.03	-0.00, *1.20*, -0.00, 0.00, 0.04, -0.00	-0.01
	0.02, -0.03, *1.24*, -0.03, -0.03, 0.01	-0.03, -0.02, *1.22*, 0.02, 0.05, 0.03	0.01
	0.02, -0.03, -0.00, *1.20*, -0.05, 0.00	0.00, -0.02, -0.01, *1.30*, -0.00, 0.04	-0.01
	0.00, 0.08, -0.00, 0.01, *1.25*, -0.06	0.00, -0.05, 0.00, -0.05, *1.31*, 0.01	-0.09
	0.01, -0.00, -0.00, 0.03, -0.05, *1.14*	-0.01, -0.04, -0.05, 0.02, -0.00, *1.19*	-0.
Average	*1.28*, 0.00, -0.00, 0.01, 0.00, 0.00	*1.31*, 0.01, -0.00, 0.00, 0.00, 0.00	*1.46*
	0.00, *1.28*, 0.00, 0.00, -0.00, 0.01	-0.00, *1.29*, -0.00, 0.00, -0.00, 0.00	0.01
	0.00, 0.01, *1.25*, 0.00, -0.00, 0.00	-0.00, 0.00, *1.30*, -0.00, -0.00, -0.00	-0.00
	-0.00, 0.00, 0.00, *1.28*, -0.00, 0.00	0.00, -0.00, -0.00, *1.28*, 0.00, -0.01	0.00
	-0.00, 0.00, -0.00, 0.01, *1.26*, -0.00	0.00, 0.00, 0.00, 0.00, *1.27*, -0.01	-0.01
	-0.00, 0.01, 0.00, -0.00, -0.00, *1.31*	0.01, -0.00, -0.00, -0.00, 0.00, *1.28*	0.

Begun, A., A. Icks, R. Waldeyer, S. Landwehr, M. Koch, and G. Giani. 2013. "Identification of a Multistate Continuous-Time Nonhomogeneous Markov Chain Model for Patients with Decreased Renal Function." *Medical Decision Making : An International Journal of the Society for Medical Decision Making* 33 (2): 298–306. https://doi.org/10.1177/0272989X12466731.

Bhatnagar, Prachi, Kremlin Wickramasinghe, Julianne Williams, Mike Rayner, and Nick Townsend. 2015. "The Epidemiology of Cardiovascular Disease in the UK 2014." *Heart* 101 (15): 1182–9. https://doi.org/10.1136/heartjnl-2015-307516.

Bouwmeester, Walter, Nicolaas P. A. Zuithoff, Susan Mallett, Mirjam I. Geerlings, Yvonne Vergouwe, Ewout W. Steyerberg, Douglas G. Altman, and Karel G. M. Moons. 2012. "Reporting and Methods in Clinical Prediction Research: A Systematic Review." *PLOS Medicine* 9 (5): e1001221. https://doi.org/10.1371/journal.pmed.1001221.

Buuren, Stef van, and Karin Groothuis-Oudshoorn. 2011. "Mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software* 45 (1): 1–67. https://doi.org/10.18637/jss.v045.i03.

Calster, Ben Van, Vanya Van Belle, Yvonne Vergouwe, Dirk Timmerman, Sabine Van Huffel, and Ewout W. Steyerberg. 2012. "Extending the c-Statistic to Nominal Polytomous Outcomes: The Polytomous Discrimination Index." *Statistics in Medicine* 31 (23): 2610–26. https://doi.org/10.1002/sim.5321.

Cao, Xue-Ying, Jian-Hui Zhou, Guang-Yan Cai, Ni-Na Tan, Jing Huang, Xiang-Cheng Xie, Li Tang, and Xiang-Mei Chen. 2015. "Predicting One-Year Mortality in Peritoneal Dialysis Patients: An Analysis of the China Peritoneal Dialysis Registry." *International Journal of Medical Sciences* 12 (4): 354–61. https://doi.org/10.7150/ijms.11694.

Chang, Winston, Joe Cheng, J. J. Allaire, Yihui Xie, Jonathan McPherson, RStudio, jQuery Foundation (jQuery library and jQuery UI library), et al. 2020. "Shiny: Web Application Framework for R."

Chinnadurai, Rajkumar, Constantina Chrysochou, and Philip A. Kalra. 2019. "Increased Risk for Cardiovascular Events in Patients with Diabetic Kidney Disease and Non-Alcoholic Fatty Liver Disease." *Nephron* 141 (1): 24–30. https://doi.org/10.1159/000493472.

Collins, Gary S., Johannes B. Reitsma, Douglas G. Altman, and Karel GM Moons. 2015. "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The

TRIPOD Statement." *BMC Medicine* 13 (1): 1. https://doi.org/10.1186/s12916-014-0241-z.

Floege, Jürgen, Iain A. Gillespie, Florian Kronenberg, Stefan D. Anker, Ioanna Gioni, Sharon Richards, Ronald L. Pisoni, et al. 2015. "Development and Validation of a Predictive Mortality Risk Score from a European Hemodialysis Cohort." *Kidney International* 87 (5): 996–1008. https://doi.org/10.1038/ki.2014.419.

Gneiting, Tilmann, and Adrian E. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102 (477): 359–78. https://doi.org/10.1198/016214506000001437.

Grams, Morgan E., and Josef Coresh. 2013. "Assessing Risk in Chronic Kidney Disease: A Methodological Review." *Nature Reviews. Nephrology* 9 (1): 18–25. https://doi.org/10.1038/nrneph.2012.248.

Grams, Morgan E., Yingying Sang, Shoshana H. Ballew, Juan Jesus Carrero, Ognjenka Djurdjev, Hiddo J. L. Heerspink, Kevin Ho, et al. 2018. "Predicting Timing of Clinical Outcomes in Patients with Chronic Kidney Disease and Severely Decreased Glomerular Filtration Rate." *Kidney International* 93 (6): 1442–51. https://doi.org/10.1016/j.kint.2018.01.009.

Hemke, Aline C, Martin BA Heemskerk, Merel van Diepen, Willem Weimar, Friedo W Dekker, and Andries J Hoitsma. 2013. "Survival Prognosis After the Start of a Renal Replacement Therapy in the Netherlands: A Retrospective Cohort Study." *BMC Nephrology* 14 (November): 258. https://doi.org/10.1186/1471-2369-14-258.

Hoefield, Richard A., Philip A. Kalra, Patricia Baker, Beverley Lane, John P. New, Donal J. O'Donoghue, Robert N. Foley, and Rachel J. Middleton. 2010. "Factors Associated with Kidney Disease Progression and Mortality in a Referred CKD Population." *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation* 56 (6): 1072–81. https://doi.org/10.1053/j.ajkd.2010.06.010.

Hoorde, Kirsten Van, Yvonne Vergouwe, Dirk Timmerman, Sabine Van Huffel, Ewout W. Steyerberg, and Ben Van Calster. 2014. "Assessing Calibration of Multinomial Risk Prediction Models." *Statistics in Medicine* 33 (15): 2585–96. https://doi.org/10.1002/sim.6114.

Jackson, Christopher. n.d. "Flexsurv: A Platform for Parametric Survival Modelling in R," 33.

Janssen, Kristel J. M., Yvonne Vergouwe, A. Rogier T. Donders, Frank E. Harrell, Qingxia Chen, Diederick E. Grobbee, and Karel G. M. Moons. 2009. "Dealing with Missing Predictor Values When Applying Clinical Prediction Models." *Clinical Chemistry* 55 (5): 994–1001. https://doi.org/10.1373/clinchem.2008.115345.

Johnson, Eric S., Micah L. Thorp, Robert W. Platt, and David H. Smith. 2008. "Predicting the Risk of Dialysis and Transplant Among Patients with CKD: A Retrospective Cohort Study." *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation* 52 (4): 653–60. https://doi.org/10.1053/j.ajkd.2008.04.026.

Johnson, Eric S., Micah L. Thorp, Xiuhai Yang, Olivier L. Charansonney, and David H. Smith. 2007. "Predicting Renal Replacement Therapy and Mortality in CKD." *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation* 50 (4): 559–65. https://doi.org/10.1053/j.ajkd.2007.07.006.

Kovesdy, Csaba P., Josef Coresh, Shoshana H. Ballew, Mark Woodward, Adeera Levin, David M. J. Naimark, Joseph Nally, et al. 2016. "Past Decline Versus Current eGFR and Subsequent ESRD Risk." *Journal of the American Society of Nephrology* 27 (8): 2447–55. https://doi.org/10.1681/ASN.2015060687.

Kulkarni, Sanjay, Isaac Hall, Richard Formica, Carrie Thiessen, Darren Stewart, Geliang Gan, Erich Greene, and Yanhong Deng. 2017. "Transition Probabilities Between Changing Sensitization Levels, Waitlist Activity Status and Competing-Risk Kidney Transplant Outcomes Using Multi-State Modeling." *PLOS ONE* 12 (12): e0190277. https://doi.org/10.1371/journal.pone.0190277.

Landray, Martin J., Jonathan R. Emberson, Lisa Blackwell, Tanaji Dasgupta, Rosita Zakeri, Matthew D. Morgan, Charlie J. Ferro, et al. 2010. "Prediction of ESRD and Death Among People with CKD:

The Chronic Renal Impairment in Birmingham (CRIB) Prospective Cohort Study." *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation* 56 (6): 1082–94. https://doi.org/10.1053/j.ajkd.2010.07.016.

Levey, Andrew S., Lesley A. Stevens, Christopher H. Schmid, Yaping Lucy Zhang, Alejandro F. Castro, Harold I. Feldman, John W. Kusek, et al. 2009. "A New Equation to Estimate Glomerular Filtration Rate." *Annals of Internal Medicine* 150 (9): 604–12. https://doi.org/10.7326/0003-4819-150-9-200905050-00006.

MA Barrowman, N Peek, GP Martin. n.d. "Prediction Model Performance Metrics for the Validation of Multi-State Clinical Prediction Models." *Unpublished Draft.*

Marks, Angharad, Nicholas Fluck, Gordon J. Prescott, Lynn Robertson, William G. Simpson, William Cairns Smith, and Corri Black. 2015. "Looking to the Future: Predicting Renal Replacement Outcomes in a Large Community Cohort with Chronic Kidney Disease." *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association* 30 (9): 1507–17. https://doi.org/10.1093/ndt/gfv089.

Matsushita, Kunihiro, Shoshana H. Ballew, Brad C. Astor, Paul E. de Jong, Ron T. Gansevoort, Brenda R. Hemmelgarn, Andrew S. Levey, et al. 2013. "Cohort Profile: The Chronic Kidney Disease Prognosis Consortium." *International Journal of Epidemiology* 42 (6): 1660–8. https://doi.org/10.1093/ije/dys173.

McMurray, John J. V., Hajime Uno, Petr Jarolim, Akshay S. Desai, Dick de Zeeuw, Kai-Uwe Eckardt, Peter Ivanovich, et al. 2011. "Predictors of Fatal and Nonfatal Cardiovascular Events in Patients with Type 2 Diabetes Mellitus, Chronic Kidney Disease, and Anemia: An Analysis of the Trial to Reduce Cardiovascular Events with Aranesp (Darbepoetin-Alfa) Therapy (TREAT)." *American Heart Journal* 162 (4): 748–755.e3. https://doi.org/10.1016/j.ahj.2011.07.016.

Meira-Machado, Luís, Jacobo de Una-Alvarez, Carmen Cadarso-Suarez, and Per K Andersen. 2009. "Multi-State Models for the Analysis of Time-to-Event Data." *Statistical Methods in Medical Research* 18 (2): 195–222. https://doi.org/10.1177/0962280208092301.

Moons, Karel G. M., Douglas G. Altman, Johannes B. Reitsma, John P. A. Ioannidis, Petra Macaskill, Ewout W. Steyerberg, Andrew J. Vickers, David F. Ransohoff, and Gary S. Collins. 2015. "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and Elaboration." *Annals of Internal Medicine* 162 (1): W1. https://doi.org/10.7326/M14-0698.

Moons, Karel G. M., Douglas G. Altman, Yvonne Vergouwe, and Patrick Royston. 2009. "Prognosis and Prognostic Research: Application and Impact of Prognostic Models in Clinical Practice." *BMJ* 338 (June): b606. https://doi.org/10.1136/bmj.b606.

Naimark, David M. J., Morgan E. Grams, Kunihiro Matsushita, Corri Black, Iefke Drion, Caroline S. Fox, Lesley A. Inker, et al. 2016. "Past Decline Versus Current eGFR and Subsequent Mortality Risk." *Journal of the American Society of Nephrology* 27 (8): 2456–66. https://doi.org/10.1681/ASN.2015060688.

New, John P, Nawar Diar Bakerly, David Leather, and Ashley Woodcock. 2014. "Obtaining Real-World Evidence: The Salford Lung Study." *Thorax* 69: 1152–4. https://doi.org/http://dx.doi.org/10.1136/thoraxjnl-2014-205259.

Perotte, Adler, Rajesh Ranganath, Jamie S. Hirsch, David Blei, and Noémie Elhadad. 2015. "Risk Prediction for Chronic Kidney Disease Progression Using Heterogeneous Electronic Health Record Data and Time Series Analysis." *Journal of the American Medical Informatics Association: JAMIA* 22 (4): 872–80. https://doi.org/10.1093/jamia/ocv024.

Putter, H., M. Fiocco, and R. B. Geskus. 2007. "Tutorial in Biostatistics: Competing Risks and Multi-State Models." *Statistics in Medicine* 26 (11): 2389–2430. https://doi.org/10.1002/sim.2712.

Ramspek, Chava L., Pauline Wm Voskamp, Frans J. van Ittersum, Raymond T. Krediet, Friedo W. Dekker, and Merel van Diepen. 2017. "Prediction Models for the Mortality Risk in Chronic Dialysis Patients: A Systematic Review and Independent External Validation Study." *Clinical Epidemiology* 9: 451–64. https://doi.org/10.2147/CLEP.S139748.

Riley, Richard D., Kym IE Snell, Joie Ensor, Danielle L. Burke, Frank E. Harrell Jr, Karel GM Moons, and Gary S. Collins. 2019. "Minimum Sample Size for Developing a Multivariable Prediction Model: PART II - Binary and Time-to-Event Outcomes." *Statistics in Medicine* 38 (7): 1276–96. https://doi.org/10.1002/sim.7992.

Riley, Richard D., Danielle van der Windt, Peter Croft, and Karel G. M. Moons. 2019. *Prognosis Research in Healthcare: Concepts, Methods, and Impact.* First. Oxford University Press.

Ripley, Brian, and William Venables. 2016. "Package 'Nnet'."

Roy, Jason, Haochang Shou, Dawei Xie, Jesse Y. Hsu, Wei Yang, Amanda H. Anderson, J. Richard Landis, et al. 2017. "Statistical Methods for Cohort Studies of CKD: Prediction Modeling." *Clinical Journal of the American Society of Nephrology: CJASN* 12 (6): 1010–7. https://doi.org/10.2215/CJN.06210616.

Royston, Patrick, Douglas G. Altman, and Willi Sauerbrei. 2006. "Dichotomizing Continuous Predictors in Multiple Regression: A Bad Idea." *Statistics in Medicine* 25 (1): 127–41. https://doi.org/10.1002/sim.2331.

Royston, Patrick, Karel G. M. Moons, Douglas G. Altman, and Yvonne Vergouwe. 2009. "Prognosis and Prognostic Research: Developing a Prognostic Model." *BMJ* 338 (March): b604. https://doi.org/10.1136/bmj.b604.

Royston, Patrick, and Mahesh K. B. Parmar. 2002. "Flexible Parametric Proportional-Hazards and Proportional-Odds Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects." *Statistics in Medicine* 21 (15): 2175–97. https://doi.org/10.1002/sim.1203.

Rubin, Donald B. 1984. *Multiple Imputation for Nonresponse in Surveys.* New York, NY: John Wiley & Sons, Inc.

Schomaker, Michael, and Christian Heumann. 2018. "Bootstrap Inference When Using Multiple Imputation." *Statistics in Medicine* 37 (14): 2252–66. https://doi.org/10.1002/sim.7654.

Schroeder, Emily B., Xiuhai Yang, Micah L. Thorp, Brent M. Arnold, David C. Tabano, Amanda F. Petrik, David H. Smith, Robert W. Platt, and Eric S. Johnson. 2017. "Predicting 5-Year Risk of RRT in Stage 3 or 4 CKD: Development and External Validation." *Clinical Journal of the American Society of Nephrology: CJASN* 12 (1): 87–94. https://doi.org/10.2215/CJN.01290216.

Shlipak, Michael G., Linda F. Fried, Mary Cushman, Teri A. Manolio, Do Peterson, Catherine Stehman-Breen, Anthony Bleyer, Anne Newman, David Siscovick, and Bruce Psaty. 2005. "Cardiovascular Mortality Risk in Chronic Kidney Disease: Comparison of Traditional and Novel Risk Factors." *JAMA* 293 (14): 1737–45. https://doi.org/10.1001/jama.293.14.1737.

Steyerberg, Ewout W., Karel G. M. Moons, Danielle A. van der Windt, Jill A. Hayden, Pablo Perel, Sara Schroter, Richard D. Riley, Harry Hemingway, Douglas G. Altman, and for the PROGRESS Group. 2013. "Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research." *PLOS Medicine* 10 (2): e1001381. https://doi.org/10.1371/journal.pmed.1001381.

Tangri, Navdeep, Lesley A. Inker, Brett Hiebert, Jenna Wong, David Naimark, David Kent, and Andrew S. Levey. 2017. "A Dynamic Predictive Model for Progression of CKD." *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation* 69 (4): 514–20. https://doi.org/10.1053/j.ajkd.2016.07.030.

Tangri, Navdeep, Georgios D. Kitsios, Lesley Ann Inker, John Griffith, David M. Naimark, Simon Walker, Claudio Rigatto, Katrin Uhlig, David M. Kent, and Andrew S. Levey. 2013. "Risk Prediction Models for Patients with Chronic Kidney Disease: A Systematic Review." *Annals of Internal Medicine* 158 (8): 596–603. https://doi.org/10.7326/0003-4819-158-8-201304160-00004.

Tangri, Navdeep, Lesley A. Stevens, John Griffith, Hocine Tighiouart, Ognjenka Djurdjev, David Naimark, Adeera Levin, and Andrew S. Levey. 2011. "A Predictive Model for Progression of Chronic Kidney Disease to Kidney Failure." *JAMA* 305 (15): 1553–9. https://doi.org/10.1001/jama.2011.451.

Team, R Core. n.d. "R: A Language and Environment for Statistical Computing." Vienna, R Foundation for Statistical Computing, Vienna, Austria.

Vaughan, Davis, and Matt Dancho. 2018. "Furrr: Apply Mapping Functions in Parallel Using Futures."

Venkat-Raman, Gopalakrishnan, Charles R. V. Tomson, Yongsheng Gao, Ronald Cornet, Benedicte Stengel, Carola Gronhagen-Riska, Chris Reid, et al. 2012. "New Primary Renal Diagnosis Codes for the ERA-EDTA." *Nephrology Dialysis Transplantation* 27 (12): 4414–9. https://doi.org/10.1093/ndt/gfs461.

Weiner, Daniel E., Hocine Tighiouart, Essam F. Elsayed, John L. Griffith, Deeb N. Salem, Andrew S. Levey, and Mark J. Sarnak. 2007. "The Framingham Predictive Instrument in Chronic Kidney Disease." *Journal of the American College of Cardiology* 50 (3): 217–24. https://doi.org/10.1016/j.jacc.2007.03.037.

White, Ian R, and Patrick Royston. 2009. "Imputing Missing Covariate Values for the Cox Model." *Statistics in Medicine* 28 (15): 1982–98. https://doi.org/10.1002/sim.3618.

Wick, James P., Tanvir C. Turin, Peter D. Faris, Jennifer M. MacRae, Robert G. Weaver, Marcello Tonelli, Braden J. Manns, and Brenda R. Hemmelgarn. 2017. "A Clinical Risk Prediction Tool for 6-Month Mortality After Dialysis Initiation Among Older Adults." *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation* 69 (5): 568–75. https://doi.org/10.1053/j.ajkd.2016.08.035.

Wickham, Hadley. 2017. "The Tidy Tools Manifesto." https://cran.r-project.org/web/packages/tidyverse/vignettes/manifest

Wood, Angela M., Ian R. White, and Patrick Royston. 2008. "How Should Variable Selection Be Performed with Multiply Imputed Data?" *Statistics in Medicine* 27 (17): 3227–46. https://doi.org/10.1002/sim.3177.