

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

Abstract	2
Introduction	2
Methods	2
Results	2
Discussion	2
1 Introduction	2
2 Methods	3
2.1 Data Sources	3
2.2 Model Design	5
2.3 Example	5
2.4 Calculator	6
3 Results	6
3.1 Data Sources	6
3.2 Example	9
3.3 Calculator	9
4 Discussion	9
A Supplementary Material	11
A.1 Statistical Analysis	11
A.2 Model Results	12
References	21

Last updated: 21 Apr

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 with a single model can be advantageous. Multi-State Models are a method to provide these kinds of predictions.

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 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. The Three-State Model was used to develop the online Calculator.

Discussion

Our CPMs provide clinicians and patients with multiple outcome predictions. This implies that users of these models can get more information about their potential future without a loss to the quality of that prediction.

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 condition, more specifically, a prognostic model allows the prediction of future events [1]. CPMs use data from previous patients to estimate the outcomes of an individual patient. Prognostic models are used in clinical practice to influence treatment decisions such as the prescribing of statins for cardiovascular disease via the application of the QRISK models [2].

Within Chronic Kidney Disease (CKD), prognostic models have been developed to predict mortality [3]–[7], ESRD [4], the commencements of RRT [6], [8]–[10] or mortality after beginning dialysis [11]–[13]. Some previous models have used the commencement of RRT as a proxy for ESRD [14]–[16], while others have investigated the occurrence of cardiovascular events within CKD patients [17]–[19]. Reviews by Grams & Coresh [20], Tangri et al [21] and Ramspek et al [22], 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 [23], [24].

Methodologically, the majority of existing CKD prediction models fail to account for completing events [5], [7], [25], have high risks of bias [3], [4], [8] or are otherwise flawed compared to modern clinical prediction standards [1], [23]

In 2013, Begun et al [26] 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 [27] 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 [10] developed an MSM focusing on the categories of Calculated Panel Reactive Antibodies (CPRA) and kidney transplant and/or death.

Most recently, in 2018, Grams et al [28] 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 [29]–[47]. These kinds of cut-offs can be useful when informing patients and clinicians of a patient’s diagnosis and to coincide with policy, but inherently cause a loss of information when done before the data analysis stage and so these models go against current statistical recommendations [29]–[47]. These kinds of assumptions are also subject to measurement error [48] and interval censoring [49], i.e. we do not know when exactly when a patient moved from CKD Stage III to CKD Stage IV, or whether drop in estimated Glomerular Function Rate (eGFR) was temporary or inaccurate. For example, Kulkarni [10] assumes that a patient with an CPRA of (5%) is the same as a patient with an CPRA of (75%) and that a patient with an CPRA of (89.9%) is vastly different from a patient with an CPRA of (90%). Moreover, none of these models have undergone any validation process, whether internal or external [50].

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 [51].

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 a pair of multinomial logistic models analysed at 2 and 4 years. The major downside of this model is that it can only produce predictions at those predefined time points and it assumes homogeneity of transition times. For example, the first model assumes that a patient who began RRT 1 month after study entry is the same as one who began after 1 year & 11 months into the study and then the second model assumes these patients are the same as one who begins RRT at 3 years and 11 months.

Therefore, the aim of this study was to improve on previous efforts to model a patient’s pathways through a Multi-State Model by choosing transition points which can be exactly identified and include states which produce a drastic difference in patient characteristics. Our modeling techniques allow for individual predictions (using a proportional hazards model) of multiple outcomes (using MSMs) at any time point (using cubic splines). The models produced by this process will then be validated, both internally and externally, to compare their results and demonstrate the transportability of the (statistically robust) clinical prediction models. We report our work in line with the TRIPOD guidelines for development and validation of clinical prediction models [23], [52].

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 [53], [54] 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 60ml/min/1.73m^2 (calculated using the CKD-EPI formula [55]) 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 (SIR) [56]. 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 60ml/min/1.73m^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 ($\text{eGFR} \lesssim 60$) *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.

This allows for a wider range of baseline eGFR measurements and patients who have been suffering from CKD, translating to a model which can be applied to

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 available within SIR and so 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 60ml/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 [57] 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 [28]. 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
- Patient is not enduring an AKI [58], [59].

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 $\text{eGFR} < 60$, 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.

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). The models were developed and validated as discussed in the Supplementary

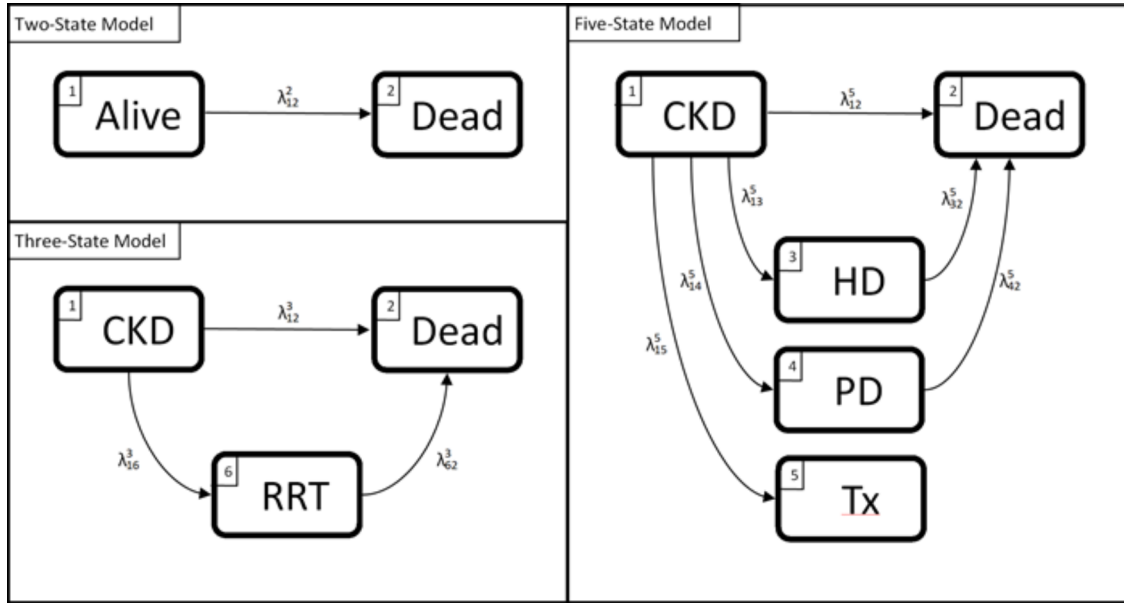


Figure 1: Diagram of the three models, the states being modelled and relevant transitions

materials.

2.3 Example

Once the models have been developed, we will apply them to two 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 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.

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
Comorbidities	Chronic Obstructive Pulmonary Disease Liver Disease Solid Tumour		Diabetes Chronic Obstructive Pulmonary Disease Hypertension

2.4 Calculator

As part of this work, we also intend to produce 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 [60] using the various `tidyverse` packages [61], as well as the `mice` [62], `flexsurv` [63], `nnet` [64] and `furrr` [65] packages. The calculator was produced using the `shiny` package [66].

3 Results

3.1 Data Sources

As seen in table 2, The Age of both populations were centred around 64-65 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) . Blood pressure was also consistent across populations (140/75 vs 148/76 for development vs validation). The blood test results (Corrected Calcium, Albumin, Haemoglobin and Phosphate) was close together, with the further difference being Haemoglobin with an average of 123 in SKS and 109 in SERPR and a much larger standard deviation in SERPR compared to SKS (38 vs 17). 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. Similar to the eGFR measures, the uPCR results were similar, but the rates of change were much broader in the validation dataset compared to the SKS and were generally increasing, whereas SKS remained stationary (73 vs 0). Levels of missingness were much higher in the SERPR dataset in most continuous variables.

Table 2: Population demographics for the continuous variables presented as: mean (IQR) [min,max] <number missing (percent missing)>

	SKS (Development)					SERPR (Validation)				
Age										
Age	64.378	(19.000)	[20.000,	94.000]	< 0 (0.00%)>	65.880	(17.000)	[18.000,	98.000]	< 0 (0.00%)>
eGFR										
eGFR ^a	30.368	(22.386)	[3.577,	59.965]	< 0 (0.00%)>	36.132	(21.444)	[1.651,	59.998]	< 0 (0.00%)>
eGFR Rate ^a	-0.015	(0.293)	[-19.107,	33.781]	<1,278 (42.87%)>	-25.476	(44.229)	[-8,755.272,	9,260.375]	< 0 (0.00%)>
uPCR										
uPCR ^f	0.112	(0.103)	[0.000,	2.025]	< 245 (8.21%)>	0.184	(0.147)	[0.000,	6.390]	<7,513 (96.76%)>
uPCR Rate ^f	-0.096	(0.188)	[-70.727,	28.198]	<1,777 (59.61%)>	73.177	(0.384)	[-2.255,	3,051.403]	<7,721 (99.44%)>
Measures										
SBP ^b	140.193	(29.000)	[77.000,	220.000]	< 50 (1.67%)>	147.746	(33.000)	[82.000,	258.000]	<6,880 (88.61%)>
DBP ^b	74.555	(14.000)	[36.000,	159.000]	< 52 (1.74%)>	76.263	(18.000)	[35.000,	128.000]	<6,879 (88.60%)>
BMI ^c	28.848	(7.842)	[13.182,	61.466]	< 572 (19.18%)>	29.331	(7.851)	[15.343,	48.301]	<7,681 (98.93%)>
Albumin ^d	42.152	(5.000)	[12.000,	52.000]	< 60 (2.01%)>	36.490	(6.000)	[7.000,	53.000]	<3,455 (44.50%)>
Corrected Calcium ^e	2.302	(0.180)	[1.209,	3.660]	< 68 (2.28%)>	2.408	(0.160)	[1.419,	3.610]	<5,113 (65.85%)>
Haemoglobin ^d	122.977	(23.000)	[61.000,	195.000]	< 72 (2.41%)>	108.588	(30.000)	[6.250,	208.000]	<3,968 (51.10%)>
Phosphate ^e	1.162	(0.320)	[0.430,	3.710]	< 87 (2.91%)>	1.203	(0.320)	[0.370,	4.370]	<5,127 (66.03%)>

^a (ml/min/1.73m²) or per year ^b (mmHG) ^c (kg/m²) ^d (g/l) ^e (mmol/l) ^f (g/mmol) or per year

Table 3 shows a breakdown of the categorical variables across the populations. In the development population, there are far more males than females, whereas in the validation population the proportions are much more matched. 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. The majority of the SKS patients were former smokers, however this information was unavailable in the SERPR dataset. Primary Renal Diagnosis suffered from very high levels of missingness in the validation dataset, but was much better recorded in the development dataset (although still far from perfect). Overall, there were high levels of comorbidities

Table 3: Population demographics for the categorical variables presented as number (percent)

	SKS (Development)	SERPR (Validation)
Gender		
Male	1,865 (62.56 %)	3,915 (50.42 %)
Female	1,116 (37.43 %)	3,849 (49.57 %)
Ethnicity		
White	2,875 (96.44 %)	683 (8.79 %)
Asian	75 (2.51 %)	12 (0.15 %)
Black	21 (0.70 %)	7 (0.09 %)
Other	10 (0.33 %)	2 (0.02 %)
<Ethnicity Missing>	0 (0.00 %)	7,060 (90.93 %)
Smoking Status		
Former	1,535 (51.49 %)	
Non-Smoker	979 (32.84 %)	
Smoker	379 (12.71 %)	
Former 3Y	46 (1.54 %)	
<Smoking Status Missing>	42 (1.40 %)	
Primary Renal Diagnosis		
Systemic diseases affecting the kidney	1,304 (43.74 %)	299 (3.85 %)
Glomerular disease	442 (14.82 %)	225 (2.89 %)
Tubulointerstitial disease	268 (8.99 %)	164 (2.11 %)
Miscellaneous renal disorders	227 (7.61 %)	188 (2.42 %)
Familial / hereditary nephropathies	173 (5.80 %)	102 (1.31 %)
<Renal Diagnosis Missing>	567 (19.02 %)	6,786 (87.40 %)

within the SKS population as shown in table 4, but these levels were much lower in the SERPR population, possibly due to the data extraction processed (where data is un-recorded, no history is assumed). In SKS, most comorbidities were at over 80% prevalence, apart from diabetes mellitus, which had a lower prevalence of 33% and over 97% (2,891) patients had a history of liver disease. In SERPR, hypertension was the highest prevalence in SERPR at 40% (3,122), followed by diabetes mellitus at 20% (1,546) and cerebrovascular accident was the lowest prevalence at 2.36% (184). Liver disease, chronic obstructive pulmonary disease and solid tumour data were unavailable in the SERPR data. The median date for the date of death was 3.9 years

Table 4: Population comorbidity prevalence for the two populations presented as number (percent) <number missing (percent missing)>

	SKS (Development)	SERPR (Validation)
Diabetes (DM)	992 (33.32%) < 4 (0.13%)>	1,546 (19.91%) < 0 (0.00%)>
Congestive Cardiac Failure (CCF)	2,414 (81.08%) < 4 (0.13%)>	406 (5.22%) < 0 (0.00%)>
Prior Myocardial Infarction (MI)	2,492 (83.70%) < 4 (0.13%)>	556 (7.16%) < 0 (0.00%)>
Ischemic Heart Disease (IHD)	2,393 (80.38%) < 4 (0.13%)>	867 (11.16%) < 0 (0.00%)>
Peripheral Vascular Disease (PVD)	2,485 (83.47%) < 4 (0.13%)>	376 (4.84%) < 0 (0.00%)>
Prior Cerebrovascular Accident (CVA)	2,727 (91.60%) < 4 (0.13%)>	184 (2.36%) < 0 (0.00%)>
Chronic Obstructive Pulmonary Disease (COPD)	2,411 (80.98%) < 4 (0.13%)>	
Chronic Liver Disease (LD)	2,891 (97.11%) < 4 (0.13%)>	
Solid Tumour (ST)	2,570 (86.32%) < 4 (0.13%)>	
Hypertension (HT)	2,546 (91.48%) <198 (6.64%)>	3,122 (40.21%) < 0 (0.00%)>

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 5.

Table 5: Event times for the two populations presented as Number of Events Median (Inter-Quartile Range) [Max]

Transition	SKS (Development)			SERPR (Validation)		
Two						
Alive to Dead	1,427	3.9 y (4.3 y)	[15.0 y]	3,025	4.9 y (3.3 y)	[10.1 y]
Three						
CKD to Dead	1,125	3.5 y (4.2 y)	[15.0 y]	2,579	4.8 y (3.2 y)	[10.1 y]
CKD to RRT	680	2.5 y (3.3 y)	[14.1 y]	1,130	3.8 y (3.8 y)	[10.1 y]
RRT to Dead	302	2.2 y (3.2 y)	[13.5 y]	446	1.5 y (2.4 y)	[9.1 y]
CKD to Dead	1,125	3.5 y (4.2 y)	[15.0 y]	2,579	4.8 y (3.2 y)	[10.1 y]
CKD to HD	344	2.5 y (3.5 y)	[14.1 y]	887	3.8 y (3.7 y)	[10.1 y]
CKD to PD	229	2.0 y (2.9 y)	[12.9 y]	149	3.5 y (4.1 y)	[9.6 y]
CKD to Tx	107	3.2 y (2.7 y)	[12.1 y]	94	4.8 y (4.5 y)	[9.7 y]
HD to Dead	185	2.0 y (3.2 y)	[11.8 y]	398	1.5 y (2.5 y)	[9.1 y]
PD to Dead	107	2.3 y (3.2 y)	[11.7 y]	47	2.1 y (2.3 y)	[8.5 y]

3.2 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 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.

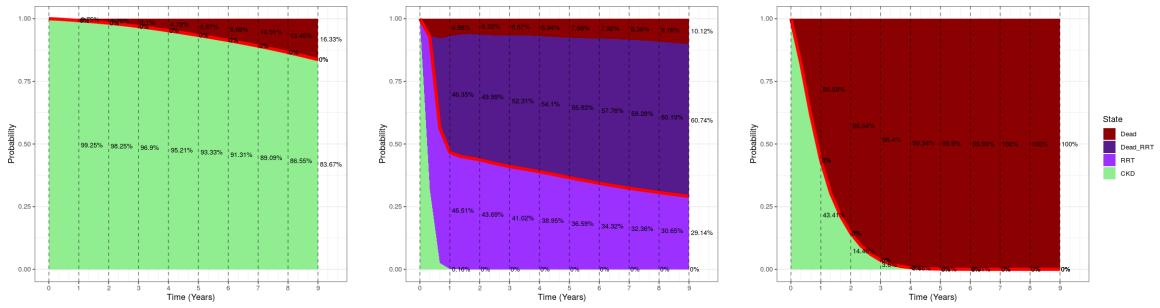


Figure 2: Results of Example Patients

3.3 Calculator

The calculator is available online here:

https://michael-barrowman.shinyapps.io/MSCPM_for_CKD_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 multiple outcomes at a wide range of time points. 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 [67], 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 [68]), 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 raw 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.

Although not directly assessing causality in regards to state-transitions, our Three-State model can be used by clinicians to either expedite or delay transition of a patient onto RRT, if it is believed that this would be beneficial. Alternatively, the Five-State Model can be interpreted to provide information regarding *which* treatment might be beneficial for a patient.

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 [69] 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). 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 solid 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 [70]). Ethnicity was assessed in the populations, but as most patients were white, it was omitted as a potential predictor from the models.

uPCR and eGFR Rate of change were also calculated [71], [72] as the difference between the two most recent measures divided by time difference in years. Age^2 , $\log(\text{Age})$, $\log(\text{eGFR Rate})$ and $\log(\text{uPCR Rate})$ were considered as transformations within the model. $\log(\text{Calendar Time})$ was included as a covariate to adjust for secular trends in treatment preferences [73]. 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 [74]. Each transition was modelled under a proportional hazards assumption using the Royston-Parmar technique [68] 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 [75].

For variable selection, we stacked the imputed datasets together to create a larger, pseudo-population [76] 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 [77]. 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 [78]. 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 [79].

A.1.2 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[80]. 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 $< 30\text{ml/min/1.73m}^2$.

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 [81], which is a proper score function [82] 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 [81], [83]. 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) [81], [84] which extends the logistic regression to three or more mutually exclusive outcomes [85]. 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 6 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 [70] 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 7 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 above 1 implies that the model under-estimates outcomes. Table 8 shows the results from the external validation in the Two-State Model, which shows similar results to the internal validation with slightly impaired performance, which is to be expected in an external validation.

Table 6: Proportional Hazards for each transition in the Two-State Model

Age
(Age-60)
(Age-60)
log(Age)
eGFR
eGFR
eGFR Rate
log(eGFR Rate)
uPCR
uPCR
uPCR Rate
log(uPCR Rate)
Measures
SBP
DBP
BMI
Albumin
Corrected Calcium
Haemoglobin
Phosphate
Gender
Female
Smoking Status
Former (3 years+)
Non-Smoker
Smoker
Primary Renal Diagnosis
Familial / hereditary nephropathies
Glomerular disease
Miscellaneous renal disorders
Tubulointerstitial disease
Comorbidity
DM
CCF
MI
IHD
PVD
CVA
COPD
LD
ST
HT

Table 7: 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

Table 8: 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

In the Three-State Model, 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 9.

Table 9: Proportional Hazards for each transition in the Three-State Model

Age		
(Age-60)	0.161	(-0.000
(Age-60)	-0.000	(-0.000
log(Age)	-5.725	(-17.000
eGFR		
eGFR	-0.013	(-0.000
eGFR Rate		
log(eGFR Rate)	0.042	(-0.000
uPCR		
uPCR	0.125	(-0.000
uPCR Rate		
log(uPCR Rate)		
Measures		
SBP	-0.001	(-0.000
DBP	0.006	(0.000
BMI		
Albumin	-0.044	(-0.000
Corrected Calcium	0.280	(-0.000
Haemoglobin	-0.013	(-0.000
Phosphate	0.511	(0.000
Gender		
Female	-0.235	(-0.000
Smoking Status		
Former (3 years+)	-0.212	(-0.000
Non-Smoker	-0.198	(-0.000
Smoker	0.356	(0.000
Primary Renal Diagnosis		
Familial / hereditary nephropathies	-0.424	(-0.000
Glomerular disease	-0.394	(-0.000
Miscellaneous renal disorders	-0.263	(-0.000
Tubulointerstitial disease	-0.463	(-0.000
Comorbidity		
DM	0.122	(-0.000
CCF	-0.394	(-0.000
MI	-0.246	(-0.000
IHD	0.102	(-0.000
PVD	-0.248	(-0.000
CVA	-0.070	(-0.000
COPD	-0.289	(-0.000
LD	-0.169	(-0.000
ST	-0.274	(-0.000
HT		

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.

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.

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

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

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

Table 10 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 11 shows the results from the external validation in the Three-State Model.

A.2.3 Five State Model

Table 12 shows the proportional hazard ratios for the transitions in the Five-State Model.

Table 10: 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													
Three	< 60	-0.00	(-0.00,	0.00)	0.00	(-0.00,	0.00)	-0.02	(-0.02,	-0.01)	0.00	(0.00,	0.00)
		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)
Three	< 30	-0.01	(-0.01,	-0.01)	0.00	(-0.00,	0.00)	-0.02	(-0.02,	-0.01)	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	(-0.00,	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													
Three	< 60	1.25,	0.00,	0.04	1.17,	-0.06,	0.01	1.21,	-0.02,	-0.01	1.32,	-0.01,	0.00
		-0.03,	1.10,	-0.00	0.03,	1.25,	-0.01	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
Three	< 30	1.21,	0.02,	0.07	1.36,	-0.00,	-0.01	1.35,	-0.04,	-0.01	1.31,	0.00,	0.00
		-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 11: 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													
Three	< 60	0.00	(-0.00,	0.00)	0.01	(0.01,	0.02)	0.05	(0.04,	0.05)	-0.00	(-0.00,	0.00)
		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)
Three	< 30	0.04	(0.04,	0.05)	0.02	(0.01,	0.02)	0.01	(0.01,	0.01)	0.00	(-0.00,	0.00)
		0.01	(0.00,	0.01)	-0.00	(-0.00,	0.00)	-0.00	(-0.00,	0.00)	0.00	(-0.00,	0.00)
		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													
Three	< 60	1.35,	-0.04,	0.04	1.13,	-0.00,	0.08	1.45,	-0.02,	0.03	1.54,	0.00,	-0.00
		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
Three	< 30	1.23,	0.03,	0.00	0.99,	-0.01,	-0.00	1.56,	0.00,	0.00	1.62,	-0.01,	0.01
		-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

Table 12: Proportional Hazards for each transition in the Five-State Model

Age		
(Age-60)	0.161	(-0.000 (-0.322)
(Age-60)	-0.000	(-0.000 (-0.000)
log(Age)	-5.725	(-17.125 (5.675)
eGFR		
eGFR	-0.013	(-0.013 (-0.013)
eGFR Rate		
log(eGFR Rate)	0.042	(-0.042 (0.126)
uPCR		
uPCR	0.125	(-0.125 (0.375)
uPCR Rate		
log(uPCR Rate)		
Measures		
SBP	-0.001	(-0.001 (0.001)
DBP	0.006	(0.006 (0.006)
BMI		
Albumin	-0.044	(-0.044 (-0.044)
Corrected Calcium	0.280	(-0.280 (0.840)
Haemoglobin	-0.013	(-0.013 (-0.013)
Phosphate	0.511	(0.511 (0.511)
Gender		
Female	-0.235	(-0.235 (-0.235)
Smoking Status		
Former (3 years+)	-0.212	(-0.212 (-0.212)
Non-Smoker	-0.198	(-0.198 (-0.198)
Smoker	0.356	(0.356 (0.356)
Primary Renal Diagnosis		
Familial / hereditary nephropathies	-0.424	(-0.424 (-0.424)
Glomerular disease	-0.394	(-0.394 (-0.394)
Miscellaneous renal disorders	-0.263	(-0.263 (-0.263)
Tubulointerstitial disease	-0.463	(-0.463 (-0.463)
Comorbidity		
DM	0.122	(-0.122 (0.366)
CCF	-0.394	(-0.394 (-0.394)
MI	-0.246	(-0.246 (-0.246)
IHD	0.102	(-0.102 (0.306)
PVD	-0.248	(-0.248 (-0.248)
CVA	-0.070	(-0.070 (-0.070)
COPD	-0.289	(-0.289 (-0.289)
LD	-0.169	(-0.169 (-0.169)
ST	-0.274	(-0.274 (-0.274)
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.

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

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

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

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

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

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

Table 13 shows the results from the internal validation in the Five-State Model. The calibration slope results are shown in a separate table for both the internal and external validation. Table 14 shows the results from

Table 13: 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)				
Five	< 30	-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)				
		-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 15 shows the calibration slopes for the model in the internal and external datasets in both the < 60 eGFR and <30 eGFR sub-populations.

Table 14: 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)	
		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)	
Five	< 30	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)	
		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)	
		-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)	
		0.00 (0.00, 0.01)		-0.02 (-0.02, -0.02)		-0.00 (-0.00, 0.00)		0.00 (-0.00, 0.00)	
		-0.02 (-0.02, -0.01)		0.02 (0.02, 0.03)		0.01 (0.01, 0.01)		-0.00 (-0.00, 0.00)	
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	
		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	
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	
		1.24, -0.00, -0.00		1.39, 0.03, 0.00		1.28, 0.02, 0.03		1.51, -0.00, 0.00	
Three	< 30	-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

- [1] E. W. Steyerberg *et al.*, “Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research,” *PLOS Medicine*, vol. 10, no. 2, p. e1001381, Feb. 2013, doi: 10.1371/journal.pmed.1001381.
- [2] J. Hippisley-Cox, C. Coupland, and P. Brindle, “Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: Prospective cohort study,” *BMJ*, vol. 357, May 2017, doi: 10.1136/bmj.j2099.
- [3] E. S. Johnson, M. L. Thorp, X. Yang, O. L. Charansonney, and D. H. Smith, “Predicting renal replacement therapy and mortality in CKD,” *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation*, vol. 50, no. 4, pp. 559–565, Oct. 2007, doi: 10.1053/j.ajkd.2007.07.006.
- [4] M. J. Landray *et al.*, “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*, vol. 56, no. 6, pp. 1082–1094, Dec. 2010, doi:

Table 15: 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
	-0.00,	1.09,	-0.05,	-0.01,	0.01,	0.00	0.00,	1.18,	0.01,	0.00,	-0.06,	0.01
	-0.04,	0.06,	1.22,	0.02,	0.03,	0.00	0.01,	0.00,	1.03,	0.00,	-0.05,	0.00
	-0.06,	0.02,	0.01,	1.25,	0.00,	-0.02	-0.04,	-0.03,	-0.03,	1.11,	0.01,	-0.03
	-0.02,	0.03,	0.02,	-0.02,	1.13,	0.00	0.00,	-0.03,	-0.02,	-0.02,	1.15,	0.01
Two Year	0.05,	-0.04,	-0.06,	-0.02,	-0.00,	1.30	0.02,	-0.01,	0.00,	-0.00,	-0.01,	1.11
	1.42,	-0.05,	0.03,	-0.01,	-0.05,	0.00	1.14,	-0.02,	0.04,	0.02,	0.00,	-0.09
	0.02,	1.12,	0.03,	-0.01,	0.03,	-0.00	-0.01,	1.10,	0.00,	-0.07,	-0.01,	-0.00
	0.01,	-0.01,	1.22,	-0.00,	0.00,	0.04	-0.03,	0.02,	1.12,	0.03,	0.00,	-0.04
	-0.03,	0.00,	-0.05,	1.21,	-0.01,	0.00	0.00,	0.02,	-0.03,	1.11,	-0.02,	0.01
Five Year	-0.02,	-0.03,	-0.00,	-0.01,	1.39,	-0.00	-0.00,	-0.06,	-0.05,	0.02,	1.11,	-0.03
	-0.03,	0.04,	0.04,	-0.00,	0.03,	1.15	0.07,	-0.01,	-0.03,	0.02,	0.04,	1.18
	1.22,	-0.00,	0.05,	-0.00,	-0.05,	0.05	1.27,	0.02,	0.02,	-0.03,	0.04,	-0.06
	-0.04,	1.11,	-0.01,	0.03,	0.04,	0.03	-0.00,	1.20,	-0.00,	0.00,	0.04,	-0.00
	0.02,	-0.03,	1.24,	-0.03,	-0.03,	0.01	-0.03,	-0.02,	1.22,	0.02,	0.05,	0.03
Average	0.02,	-0.03,	-0.00,	1.20,	-0.05,	0.01	0.00,	-0.02,	-0.01,	1.30,	-0.00,	0.04
	0.00,	0.08,	-0.00,	0.01,	1.25,	-0.06	0.00,	-0.05,	0.00,	-0.05,	1.31,	0.01
	0.01,	-0.00,	-0.00,	0.03,	-0.05,	1.14	-0.01,	-0.04,	-0.05,	0.02,	-0.00,	1.19
	1.28,	0.00,	-0.00,	0.01,	0.00,	0.00	1.31,	0.01,	-0.00,	0.00,	0.00,	0.00
	0.00,	1.28,	0.00,	0.00,	-0.00,	0.01	-0.00,	1.29,	-0.00,	0.00,	-0.00,	0.00
	0.00,	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,	1.28,	-0.00,	0.00	0.00,	-0.00,	-0.00,	1.28,	0.00,	-0.01
	-0.00,	0.00,	-0.00,	0.01,	1.26,	-0.00	0.00,	0.00,	0.00,	0.00,	1.27,	-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

10.1053/j.ajkd.2010.07.016.

[5] N. Bansal *et al.*, “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*, vol. 10, no. 3, pp. 363–371, Mar. 2015, doi: 10.2215/CJN.04650514.

[6] A. Marks *et al.*, “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*, vol. 30, no. 9, pp. 1507–1517, Sep. 2015, doi: 10.1093/ndt/gfv089.

[7] J. P. Wick *et al.*, “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*, vol. 69, no. 5, pp. 568–575, May 2017, doi: 10.1053/j.ajkd.2016.08.035.

[8] E. S. Johnson, M. L. Thorp, R. W. Platt, and D. H. Smith, “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*, vol. 52, no. 4, pp. 653–660, Oct. 2008, doi: 10.1053/j.ajkd.2008.04.026.

[9] E. B. Schroeder *et al.*, “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*, vol. 12, no. 1, pp. 87–94, Jun. 2017, doi: 10.2215/CJN.01290216.

[10] S. Kulkarni *et al.*, “Transition probabilities between changing sensitization levels, waitlist activity status and competing-risk kidney transplant outcomes using multi-state modeling,” *PLOS ONE*, vol. 12, no. 12, p. e0190277, Dec. 2017, doi: 10.1371/journal.pone.0190277.

[11] J. Floege *et al.*, “Development and validation of a predictive mortality risk score from a European hemodialysis cohort,” *Kidney International*, vol. 87, no. 5, pp. 996–1008, May 2015, doi: 10.1038/ki.2014.419.

[12] A. C. Hemke, M. B. Heemskerk, M. van Diepen, W. Weimar, F. W. Dekker, and A. J. Hoitsma, “Survival prognosis after the start of a renal replacement therapy in the Netherlands: A retrospective cohort study,” *BMC Nephrology*, vol. 14, p. 258, Nov. 2013, doi: 10.1186/1471-2369-14-258.

- [13] X.-Y. Cao *et al.*, “Predicting one-year mortality in peritoneal dialysis patients: An analysis of the China Peritoneal Dialysis Registry,” *International Journal of Medical Sciences*, vol. 12, no. 4, pp. 354–361, 2015, doi: 10.7150/ijms.11694.
- [14] N. Tangri *et al.*, “A predictive model for progression of chronic kidney disease to kidney failure,” *JAMA*, vol. 305, no. 15, pp. 1553–1559, Apr. 2011, doi: 10.1001/jama.2011.451.
- [15] J. Roy *et al.*, “Statistical Methods for Cohort Studies of CKD: Prediction Modeling,” *Clinical journal of the American Society of Nephrology: CJASN*, vol. 12, no. 6, pp. 1010–1017, Jun. 2017, doi: 10.2215/CJN.06210616.
- [16] N. Tangri *et al.*, “A Dynamic Predictive Model for Progression of CKD,” *American Journal of Kidney Diseases: The Official Journal of the National Kidney Foundation*, vol. 69, no. 4, pp. 514–520, Apr. 2017, doi: 10.1053/j.ajkd.2016.07.030.
- [17] M. G. Shlipak *et al.*, “Cardiovascular mortality risk in chronic kidney disease: Comparison of traditional and novel risk factors,” *JAMA*, vol. 293, no. 14, pp. 1737–1745, Apr. 2005, doi: 10.1001/jama.293.14.1737.
- [18] D. E. Weiner *et al.*, “The Framingham predictive instrument in chronic kidney disease,” *Journal of the American College of Cardiology*, vol. 50, no. 3, pp. 217–224, Jul. 2007, doi: 10.1016/j.jacc.2007.03.037.
- [19] J. J. V. McMurray *et al.*, “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*, vol. 162, no. 4, pp. 748–755.e3, Oct. 2011, doi: 10.1016/j.ahj.2011.07.016.
- [20] M. E. Grams and J. Coresh, “Assessing risk in chronic kidney disease: A methodological review,” *Nature Reviews. Nephrology*, vol. 9, no. 1, pp. 18–25, Jan. 2013, doi: 10.1038/nrneph.2012.248.
- [21] N. Tangri *et al.*, “Risk prediction models for patients with chronic kidney disease: A systematic review,” *Annals of Internal Medicine*, vol. 158, no. 8, pp. 596–603, Apr. 2013, doi: 10.7326/0003-4819-158-8-201304160-00004.
- [22] C. L. Ramspek, P. W. Voskamp, F. J. van Ittersum, R. T. Krediet, F. W. Dekker, and M. van Diepen, “Prediction models for the mortality risk in chronic dialysis patients: A systematic review and independent external validation study,” *Clinical Epidemiology*, vol. 9, pp. 451–464, 2017, doi: 10.2147/CLEP.S139748.
- [23] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. Moons, “Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement,” *BMC Medicine*, vol. 13, no. 1, p. 1, Jan. 2015, doi: 10.1186/s12916-014-0241-z.
- [24] W. Bouwmeester *et al.*, “Reporting and Methods in Clinical Prediction Research: A Systematic Review,” *PLOS Medicine*, vol. 9, no. 5, p. e1001221, May 2012, doi: 10.1371/journal.pmed.1001221.
- [25] A. Perotte, R. Ranganath, J. S. Hirsch, D. Blei, and N. Elhadad, “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*, vol. 22, no. 4, pp. 872–880, Jul. 2015, doi: 10.1093/jamia/ocv024.
- [26] A. Begun, A. Icks, R. Waldeyer, S. Landwehr, M. Koch, and G. Giani, “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*, vol. 33, no. 2, pp. 298–306, Feb. 2013, doi: 10.1177/0272989X12466731.
- [27] A. M. Allen, W. R. Kim, T. M. Therneau, J. J. Larson, J. K. Heimbach, and A. D. Rule, “Chronic kidney disease and associated mortality after liver transplantation—a time-dependent analysis using measured glomerular filtration rate,” *Journal of Hepatology*, vol. 61, no. 2, pp. 286–292, Aug. 2014, doi: 10.1016/j.jhep.2014.03.034.
- [28] M. E. Grams *et al.*, “Predicting timing of clinical outcomes in patients with chronic kidney disease and severely decreased glomerular filtration rate,” *Kidney International*, vol. 93, no. 6, pp. 1442–1451, Jun. 2018, doi: 10.1016/j.kint.2018.01.009.

- [29] D. G. Altman, "Problems in dichotomizing continuous variables," *American Journal of Epidemiology*, vol. 139, no. 4, pp. 442–445, Feb. 1994, doi: 10.1093/oxfordjournals.aje.a117020.
- [30] D. G. Altman, B. Lausen, W. Sauerbrei, and M. Schumacher, "Dangers of Using 'Optimal' Cutpoints in the Evaluation of Prognostic Factors," *JNCI: Journal of the National Cancer Institute*, vol. 86, no. 11, pp. 829–835, Jun. 1994, doi: 10.1093/jnci/86.11.829.
- [31] D. G. Altman and P. Royston, "The cost of dichotomising continuous variables," *BMJ*, vol. 332, no. 7549, p. 1080, May 2006, doi: 10.1136/bmj.332.7549.1080.
- [32] C. Bennette and A. Vickers, "Against quantiles: Categorization of continuous variables in epidemiologic research, and its discontents," *BMC Medical Research Methodology*, vol. 12, no. 1, p. 21, Feb. 2012, doi: 10.1186/1471-2288-12-21.
- [33] M. M. Butts and T. W. H. Ng, "Chopped liver? OK. Chopped data? Not OK," in *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*, New York, NY, US: Routledge/Taylor & Francis Group, 2009, pp. 361–386.
- [34] P. M. Cumberland *et al.*, "Ophthalmic statistics note: The perils of dichotomising continuous variables," *British Journal of Ophthalmology*, vol. 98, no. 6, pp. 841–843, Jun. 2014, doi: 10.1136/bjophthalmol-2014-304930.
- [35] N. V. Dawson and R. Weiss, "Dichotomizing continuous variables in statistical analysis: A practice to avoid," *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, vol. 32, no. 2, pp. 225–226, doi: 10.1177/0272989X12437605.
- [36] T. E. Dinero, "Seven reasons why you should not categorize continuous data," *Journal of Health & Social Policy*, vol. 8, no. 1, pp. 63–72, 1996, doi: 10.1300/J045v08n01_06.
- [37] J. R. Irwin and G. H. McClelland, "Negative Consequences of Dichotomizing Continuous Predictor Variables," *Journal of Marketing Research*, vol. 40, no. 3, pp. 366–371, Aug. 2003, doi: 10.1509/jmkr.40.3.366.19237.
- [38] O. Kuss, "The danger of dichotomizing continuous variables: A visualization," *Teaching Statistics*, vol. 35, no. 2, pp. 78–79, 2013, doi: 10.1111/test.12006.
- [39] K. Metze, "Dichotomization of continuous data—a pitfall in prognostic factor studies," *Pathology, Research and Practice*, vol. 204, no. 3, pp. 213–214, 2008, doi: 10.1016/j.prp.2007.12.002.
- [40] O. Naggara, J. Raymond, F. Guilbert, D. Roy, A. Weill, and D. G. Altman, "Analysis by categorizing or dichotomizing continuous variables is inadvisable: An example from the natural history of unruptured aneurysms," *AJNR. American journal of neuroradiology*, vol. 32, no. 3, pp. 437–440, Mar. 2011, doi: 10.3174/ajnr.A2425.
- [41] S. V. Owen and R. D. Froman, "Why carve up your continuous data?" *Research in Nursing & Health*, vol. 28, no. 6, pp. 496–503, 2005, doi: 10.1002/nur.20107.
- [42] P. Royston, D. G. Altman, and W. Sauerbrei, "Dichotomizing continuous predictors in multiple regression: A bad idea," *Statistics in Medicine*, vol. 25, no. 1, pp. 127–141, Jan. 2006, doi: 10.1002/sim.2331.
- [43] J. M. Schellingerhout, M. W. Heymans, H. C. W. de Vet, B. W. Koes, and A. P. Verhagen, "Categorizing continuous variables resulted in different predictors in a prognostic model for nonspecific neck pain," *Journal of Clinical Epidemiology*, vol. 62, no. 8, pp. 868–874, Aug. 2009, doi: 10.1016/j.jclinepi.2008.10.010.
- [44] D. L. Streiner, "Breaking up is hard to do: The heartbreak of dichotomizing continuous data," *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie*, vol. 47, no. 3, pp. 262–266, Apr. 2002, doi: 10.1177/070674370204700307.
- [45] C. van Walraven and R. G. Hart, "Leave 'em alone - why continuous variables should be analyzed as such," *Neuroepidemiology*, vol. 30, no. 3, pp. 138–139, 2008, doi: 10.1159/000126908.
- [46] A. M. Vintzileos, Y. Oyelese, and C. V. Ananth, "The 'anathema' of arbitrary categorization of continuous predictors," *American Journal of Obstetrics and Gynecology*, vol. 210, no. 3, pp. 200–203, Mar. 2014, doi: 10.1016/j.ajog.2013.09.042.

- [47] C. R. Weinberg, “How Bad Is Categorization?” *Epidemiology*, vol. 6, no. 4, pp. 345–347, 1995.
- [48] M. van Smeden, T. L. Lash, and R. H. H. Groenwold, “Reflection on modern methods: Five myths about measurement error in epidemiological research,” *International Journal of Epidemiology*, Oct. 19AD, doi: 10.1093/ije/dyz251.
- [49] J. Sun, “Interval Censoring,” in *Encyclopedia of Biostatistics*, American Cancer Society, 2005.
- [50] D. G. Altman, Y. Vergouwe, P. Royston, and K. G. M. Moons, “Prognosis and prognostic research: Validating a prognostic model,” *BMJ*, vol. 338, p. b605, May 2009, doi: 10.1136/bmj.b605.
- [51] P. Royston, K. G. M. Moons, D. G. Altman, and Y. Vergouwe, “Prognosis and prognostic research: Developing a prognostic model,” *BMJ*, vol. 338, p. b604, Mar. 2009, doi: 10.1136/bmj.b604.
- [52] K. G. M. Moons *et al.*, “Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration,” *Annals of Internal Medicine*, vol. 162, no. 1, p. W1, Jan. 2015, doi: 10.7326/M14-0698.
- [53] R. A. Hoefield *et al.*, “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*, vol. 56, no. 6, pp. 1072–1081, Dec. 2010, doi: 10.1053/j.ajkd.2010.06.010.
- [54] R. Chinnadurai, C. Chrysochou, and P. A. Kalra, “Increased Risk for Cardiovascular Events in Patients with Diabetic Kidney Disease and Non-Alcoholic Fatty Liver Disease,” *Nephron*, vol. 141, no. 1, pp. 24–30, 2019, doi: 10.1159/000493472.
- [55] A. S. Levey *et al.*, “A new equation to estimate glomerular filtration rate,” *Annals of Internal Medicine*, vol. 150, no. 9, pp. 604–612, May 2009, doi: 10.7326/0003-4819-150-9-200905050-00006.
- [56] J. P. New, N. D. Bakerly, D. Leather, and A. Woodcock, “Obtaining real-world evidence: The Salford Lung Study,” *Thorax*, vol. 69, pp. 1152–1154, 2014, doi: <http://dx.doi.org/10.1136/thoraxjnl-2014-205259>.
- [57] K. Matsushita *et al.*, “Cohort Profile: The Chronic Kidney Disease Prognosis Consortium,” *International Journal of Epidemiology*, vol. 42, no. 6, pp. 1660–1668, Dec. 2013, doi: 10.1093/ije/dys173.
- [58] L. G. Forni *et al.*, “Renal recovery after acute kidney injury,” *Intensive Care Medicine*, vol. 43, no. 6, pp. 855–866, 2017, doi: 10.1007/s00134-017-4809-x.
- [59] “KDIGO Clinical Practice Guideline for Acute Kidney Injury,” *OFFICIAL JOURNAL OF THE INTERNATIONAL SOCIETY OF NEPHROLOGY*, p. 141, 2012.
- [60] R. C. Team, “R: A Language and Environment for Statistical Computing.” R Foundation for Statistical Computing, Vienna, Austria, Vienna,
- [61] H. Wickham, “The tidy tools manifesto.” <https://cran.r-project.org/web/packages/tidyverse/vignettes/manifesto.html>, Nov-2017.
- [62] S. van Buuren and K. Groothuis-Oudshoorn, “Mice: Multivariate Imputation by Chained Equations in R,” *Journal of Statistical Software*, vol. 45, no. 1, pp. 1–67, Dec. 2011, doi: 10.18637/jss.v045.i03.
- [63] C. Jackson, “Flexsurv: A Platform for Parametric Survival Modelling in R,” p. 33.
- [64] B. Ripley and W. Venables, “Package ‘nnet’,” Feb-2016.
- [65] D. Vaughan and M. Dancho, “Furrr: Apply Mapping Functions in Parallel using Futures.” May-2018.
- [66] W. Chang *et al.*, “Shiny: Web Application Framework for R.” Mar-2020.
- [67] R. D. Riley *et al.*, “Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes,” *Statistics in Medicine*, vol. 38, no. 7, pp. 1276–1296, 2019, doi: 10.1002/sim.7992.
- [68] P. Royston and M. K. B. Parmar, “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*, vol. 21, no. 15, pp. 2175–2197, Aug. 2002, doi: 10.1002/sim.1203.

- [69] K. G. M. Moons, D. G. Altman, Y. Vergouwe, and P. Royston, “Prognosis and prognostic research: Application and impact of prognostic models in clinical practice,” *BMJ*, vol. 338, p. b606, Jun. 2009, doi: 10.1136/bmj.b606.
- [70] G. Venkat-Raman *et al.*, “New primary renal diagnosis codes for the ERA-EDTA,” *Nephrology Dialysis Transplantation*, vol. 27, no. 12, pp. 4414–4419, Dec. 2012, doi: 10.1093/ndt/gfs461.
- [71] C. P. Kovesdy *et al.*, “Past Decline Versus Current eGFR and Subsequent ESRD Risk,” *Journal of the American Society of Nephrology*, vol. 27, no. 8, pp. 2447–2455, Aug. 2016, doi: 10.1681/ASN.2015060687.
- [72] D. M. J. Naimark *et al.*, “Past Decline Versus Current eGFR and Subsequent Mortality Risk,” *Journal of the American Society of Nephrology*, vol. 27, no. 8, pp. 2456–2466, Aug. 2016, doi: 10.1681/ASN.2015060688.
- [73] P. Bhatnagar, K. Wickramasinghe, J. Williams, M. Rayner, and N. Townsend, “The epidemiology of cardiovascular disease in the UK 2014,” *Heart*, vol. 101, no. 15, pp. 1182–1189, Aug. 2015, doi: 10.1136/heartjnl-2015-307516.
- [74] L. Meira-Machado, J. de Uña-Álvarez, C. Cadarso-Suárez, and P. K. Andersen, “Multi-state models for the analysis of time-to-event data,” *Statistical methods in medical research*, vol. 18, no. 2, pp. 195–222, Apr. 2009, doi: 10.1177/0962280208092301.
- [75] H. Putter, M. Fiocco, and R. B. Geskus, “Tutorial in biostatistics: Competing risks and multi-state models,” *Statistics in Medicine*, vol. 26, no. 11, pp. 2389–2430, May 2007, doi: 10.1002/sim.2712.
- [76] A. M. Wood, I. R. White, and P. Royston, “How should variable selection be performed with multiply imputed data?” *Statistics in Medicine*, vol. 27, no. 17, pp. 3227–3246, Jul. 2008, doi: 10.1002/sim.3177.
- [77] D. B. Rubin, *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons, Inc, 1984.
- [78] I. R. White and P. Royston, “Imputing missing covariate values for the Cox model,” *Statistics in Medicine*, vol. 28, no. 15, pp. 1982–1998, Jul. 2009, doi: 10.1002/sim.3618.
- [79] K. J. M. Janssen *et al.*, “Dealing with missing predictor values when applying clinical prediction models,” *Clinical Chemistry*, vol. 55, no. 5, pp. 994–1001, May 2009, doi: 10.1373/clinchem.2008.115345.
- [80] M. Schomaker and C. Heumann, “Bootstrap inference when using multiple imputation,” *Statistics in Medicine*, vol. 37, no. 14, pp. 2252–2266, 2018, doi: 10.1002/sim.7654.
- [81] N. P. MA Barrowman GP Martin, “Prediction model performance metrics for the validation of multi-state clinical prediction models,” *Unpublished Draft*.
- [82] T. Gneiting and A. E. Raftery, “Strictly Proper Scoring Rules, Prediction, and Estimation,” *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, Mar. 2007, doi: 10.1198/016214506000001437.
- [83] B. V. Calster, V. V. Belle, Y. Vergouwe, D. Timmerman, S. V. Huffel, and E. W. Steyerberg, “Extending the c-statistic to nominal polytomous outcomes: The Polytomous Discrimination Index,” *Statistics in Medicine*, vol. 31, no. 23, pp. 2610–2626, 2012, doi: 10.1002/sim.5321.
- [84] K. V. Hoorde, Y. Vergouwe, D. Timmerman, S. V. Huffel, E. W. Steyerberg, and B. V. Calster, “Assessing calibration of multinomial risk prediction models,” *Statistics in Medicine*, vol. 33, no. 15, pp. 2585–2596, Jul. 2014, doi: 10.1002/sim.6114.
- [85] R. D. Riley, D. van der Windt, P. Croft, and K. G. M. Moons, *Prognosis Research in Healthcare: Concepts, Methods, and Impact*, First. Oxford University Press, 2019.