

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

Contents

1	Introduction	2
2	Methods	3
2.1	Data Sources	3
2.2	Development	4
2.3	Validation	5
3	Results	5
3.1	Data Sources	5
3.2	Development	7
3.3	Example	9
3.4	Validation	9
4	Discussion	9
	References	9

1 Introduction

Chronic Kidney Disease (CKD) is a progressive disease that affects the ability of the kidneys to filter toxins from the blood. Patients with End-stage Renal Disease (ESRD) are treated using Renal Replacement Therapy (RRT), which collectively describes the treatments designed to emulate the processes performed by the failing kidneys. The three most common treatment modalities are haemodialysis (HD), peritoneal dialysis (PD) and kidney transplant (Tx). The more severe stages of CKD (stages 3 - 5) affects approximately 2.6 million people over the age of 16 in England [1] with around 63 thousand adult patients registered for RRT in 2015 [2], of which 8 thousand were new patients [3].

Previous prognostic models have been developed to predict mortality [4]–[8], ESRD [5], the commencements of RRT [7], [9]–[11] or mortality after beginning dialysis [12]–[14]. Some previous models have used the commencement of RRT as a proxy for ESRD [15]–[17], while others have investigated the occurrence of cardiovascular events, which are common amongst CKD patients [18]–[20]. Reviews by Grams & Coresh [21], Tangri et al [22] and Ramspek et al [23], 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 [24], [25]. Methodologically, the majority of existing CKD prediction models fail to account for competing events [6], [8], [26], have high risks of bias [4], [5], [9] or are otherwise flawed compared to modern clinical prediction standards [24], [27].

In 2013, Begun et al [28] developed a Multi-State Clinical Model for assessing patient progression through the severity stages of CKD (III-V), RRT and/or death. In 2014, Allen et al [29] focused a similar model on liver transplant patients. In 2017, Kulkarni et al [11] developed an MSM focusing on the categories of Calculated Panel Reactive Antibodies and transplant and/or death.

Most recently, in 2018, Grams et al [30] 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 [31]–[49]. 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 information when done before the data analysis stage and so these models go against the current statistical recommendations (cite). These kinds of assumptions are also subject to measurement error and interval censoring (cite), 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 [11] 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 papers 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, these models 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 et al [21]), 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 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. We also model using extensions to traditional survival analysis to incorporate heterogeneity within the population as much as possible and to allow for the prediction of patient outcomes at any future time point (within the time-scales of the study). 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.

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 [52], [53] 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 [54]) 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) [55]. eGFR, uPCR, comorbidity and blood results were measured longitudinally throughout patients time within the cohort. Patient start dates for our model was assigned as their first date after their consent date at which their eGFR was recorded to be below $60\text{ml/min}/1.73\text{m}^2$.

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 $60\text{ml/min}/1.73\text{m}^2$ 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 [56] 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 [30]. In SKS, start dates were chosen to be the first date after consent where their eGFR was recorded to be less than $60\text{ml/min}/1.73\text{m}^2$. In SERPR, start dates were calculated by removing the first recorded eGFR measurement for all patients, any eGFR measurements from before a patient turned 18 and any during an AKI episode [57], [58].

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 [59]. 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 [60].

All analysis was done in R 3.6.2 [61] using the various `tidyverse` packages [62], as well as the `mice` [63], `flexsurv` [64], `nnet` [65] and `furrr` [66] packages.

2.2 Development

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 analogous to 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).

Variables considered as covariates were demographics (sex, age, smoking status and alcohol consumption), comorbidities (congestive cardiac failure (CCF), chronic obstructive pulmonary disease (COPD), cerebrovascular accident (CVA), hypertension (HT), diabetes mellitus (DM), ischemic heart disease (IHD), chronic liver disease (LD), myocardial infarction (MI), peripheral vascular disease (PVD) and solid tumour (ST)), physical parameters (BMI, blood pressure), blood results (haemoglobin, albumin, calcium and phosphate measures), urine protein creatinine ratio (uPCR) and primary renal diagnosis (grouped as per ERA-EDTA classifications [67]). Ethnicity was assessed in the populations, but due to extreme homogeneity, it was omitted as a potential predictor from the models.

uPCR and eGFR Rate of change were also calculated [68], [69] as the difference between the two most recent measures divided by time difference in days. $\log(\text{Age})$ was considered as a covariate and then Age was centred at 60 years and squared to account for the variety of effects that Age can have on the transitions involved. To account for any time trend in overall change in treatment techniques, $\log(-\text{Calendar Time})$ were included as a covariates (cite). Calendar Time was defined as time of study entry minus 1st January 2019, ensuring CalendarTime is always negative so that patients who entered the study longer ago have a higher value for this covariate.

Each transition was modelled under a proportional hazards assumption using the Royston-Parmar technique [70] 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 [71].

For variable selection, we stacked the imputed datasets together to create a larger, pseudo-population [72] and performed backwards-forwards selection based on minimising the AIC at each step ([Cite: Something]). This was repeated for each transition and for different numbers of evenly spaced knots, $K=\{0,1,2,3,4,5\}$, which 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 [73]. This gave a model fully defined by smooth cubic splines representing the cumulative cause-specific hazard and individualised proportional hazards 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. 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 1-year, 2-years, 5-years and 9-years. As well as presenting the performance measures over time.

The overall accuracy of each model was assessed using the MSM adjusted Brier Score [[**Cite: Performance Metrics**]], which is a proper score function 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 [74] [**Cite: Performance Metrics**]. 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) [75] [**Cite: Performance Metrics**] which extends the logistic regression to three or more mutually exclusive outcomes [76]. 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.

3 Results

3.1 Data Sources

As seen in table 3.1, 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 (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). 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.

Variable	SKS (Development)	SERPR (Validation)
Age	64.378 (14.573) [20.000, 94.000] < 0 (0.00%)>	65.880 (13.734) [18.000, 98.000] < 0 (0.00%)>
eGFR ^a	30.368 (14.122) [3.577, 59.965] < 0 (0.00%)>	36.132 (13.668) [1.651, 59.998] < 0 (0.00%)>
eGFR.Rate ^a	-0.015 (1.528) [-19.107, 33.781] <1,278 (42.87%)>	-25.476 (623.048) [-8,755.272, 9,260.375] < 0 (0.00%)>
SBP ^b	140.193 (21.839) [77.000, 220.000] < 50 (1.67%)>	147.746 (25.984) [82.000, 258.000] <6,880 (88.61%)>
DBP ^b	74.555 (11.754) [36.000, 159.000] < 52 (1.74%)>	76.263 (13.853) [35.000, 128.000] <6,879 (88.60%)>
BMI ^c	28.848 (6.114) [13.182, 61.466] < 572 (19.18%)>	29.331 (6.722) [15.343, 48.301] <7,681 (98.93%)>
Albumin ^d	42.152 (4.186) [12.000, 52.000] < 60 (2.01%)>	36.490 (5.647) [7.000, 53.000] <3,455 (44.50%)>
Calcium ^e	2.302 (0.147) [1.209, 3.660] < 68 (2.28%)>	2.408 (0.158) [1.419, 3.610] <5,113 (65.85%)>
Haemoglobin ^d	122.977 (17.068) [61.000, 195.000] < 72 (2.41%)>	108.588 (38.446) [6.250, 208.000] <3,968 (51.10%)>
Phosphate ^e	1.162 (0.284) [0.430, 3.710] < 87 (2.91%)>	1.203 (0.311) [0.370, 4.370] <5,127 (66.03%)>
uPCR ^f	0.112 (0.195) [0.000, 2.025] < 245 (8.21%)>	0.184 (0.477) [0.000, 6.390] <7,513 (96.76%)>
uPCR.Rate ^f	-0.096 (3.035) [-70.727, 28.198] <1,777 (59.61%)>	73.177 (465.186) [-2.255, 3,051.403] <7,721 (99.44%)>

^a(ml/min/1.73m²) or per year

^b(mmHG)

^c(kg/m²)

^d(g/l)

^e(mmol/l)

^f(gmol/l) or per year

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

3.1 shows a breakdown of the categorical variables across the populations. In the development population, there are far more males than females, which is actually against what is believed to occur in the general public (cite), whereas in the validation population the proportions are much more matched. Ethnicity was very homogeneous in the SKS dataset, and 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).

Variable	Category	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 %)
	<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 %)	
	<Missing>	42 (1.40 %)	
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 %)
	<Missing>	567 (19.02 %)	6,786 (87.40 %)

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

Overall, there were high levels of comorbidities within the SKS population as shown in 3.1, 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.

Variable	SKS (Development)	SERPR (Validation)
DM	992 (33.32%) < 4 (0.13%)>	1,546 (19.91%) < 0 (0.00%)>
CCF	2,414 (81.08%) < 4 (0.13%)>	406 (5.22%) < 0 (0.00%)>
MI	2,492 (83.70%) < 4 (0.13%)>	556 (7.16%) < 0 (0.00%)>
IHD	2,393 (80.38%) < 4 (0.13%)>	867 (11.16%) < 0 (0.00%)>
PVD	2,485 (83.47%) < 4 (0.13%)>	376 (4.84%) < 0 (0.00%)>
CVA	2,727 (91.60%) < 4 (0.13%)>	184 (2.36%) < 0 (0.00%)>
COPD	2,411 (80.98%) < 4 (0.13%)>	
LD	2,891 (97.11%) < 4 (0.13%)>	
ST	2,570 (86.32%) < 4 (0.13%)>	
HT	2,546 (91.48%) <198 (6.64%)>	3,122 (40.21%) < 0 (0.00%)>

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

3.2 Development

Table 4 shows the full results from the Three-State Models, the results for the Two-State and Five-State Models can be seen in Supplementary Material. Older patients are more likely to transition to the Dead state, regardless of whether from CKD or an RRT state and Older patients were less likely to transition into the RRT states. Increased rates of decline of eGFR had large effects on the transition from CKD into RRT, due to an increase in rates to HD and the high prevalence of transitions to HD since the effect on transitions into PD were relatively low, and eGFR decline actually reduced the rate of transition to Tx.

Var	CKD -> Dead	CKD -> RRT	RRT -> Dead
DiagGroup: Familial / hereditary nephropathies	-0.424 (-0.854, 0.006)	1.029 (0.720, 1.338)	-0.562 (-1.084, -0.040)
DiagGroup: Glomerular disease	-0.394 (-0.635, -0.154)	-0.165 (-0.465, 0.134)	-0.488 (-0.883, -0.094)
DiagGroup: Miscellaneous renal disorders	-0.263 (-0.505, -0.021)	-0.649 (-1.143, -0.155)	0.033 (-0.553, 0.620)
DiagGroup: Tubulointerstitial disease	-0.463 (-0.741, -0.184)	-0.265 (-0.577, 0.046)	-0.310 (-0.803, 0.181)
Gender: Female	-0.235 (-0.371, -0.099)	-0.277 (-0.455, -0.099)	
SmokingStatus: Former_3Y	-0.212 (-0.879, 0.453)	-0.133 (-0.757, 0.490)	-0.282 (-1.082, 0.518)
SmokingStatus: Non_Smoker	-0.198 (-0.345, -0.051)	-0.162 (-0.364, 0.039)	-0.294 (-0.598, 0.009)
SmokingStatus: Smoker	0.356 (0.160, 0.551)	0.175 (-0.076, 0.428)	0.387 (0.068, 0.706)
(Age-60)	0.161 (-0.051, 0.374)	-0.041 (-0.051, -0.031)	0.063 (0.050, 0.076)
(Age-60)^2	-0.000 (-0.002, 0.000)	-0.000 (-0.000, -0.000)	
Albumin	-0.044 (-0.064, -0.024)	-0.032 (-0.059, -0.004)	-0.044 (-0.079, -0.009)
Calcium	0.280 (-0.192, 0.752)	-0.515 (-1.207, 0.177)	
CCF	-0.394 (-0.535, -0.253)		-0.299 (-0.597, -0.002)
COPD	-0.289 (-0.433, -0.145)		
CVA	-0.070 (-0.252, 0.111)		-0.168 (-0.577, 0.240)
DBP	0.006 (0.000, 0.013)	0.006 (-0.001, 0.015)	
DM	0.122 (-0.011, 0.255)	0.141 (-0.074, 0.358)	0.200 (-0.096, 0.496)
eGFR	-0.013 (-0.019, -0.006)	-0.095 (-0.108, -0.082)	0.011 (-0.001, 0.025)
eGFR.Rate		0.055 (-0.021, 0.131)	-0.056 (-0.363, 0.250)
Haemoglobin	-0.013 (-0.017, -0.008)	-0.005 (-0.012, 0.001)	-0.005 (-0.014, 0.003)
HT		0.274 (-0.176, 0.726)	-0.416 (-1.104, 0.271)
IHD	0.102 (-0.041, 0.245)	-0.077 (-0.334, 0.179)	-0.097 (-0.424, 0.228)
LD	-0.169 (-0.578, 0.239)	-0.316 (-0.731, 0.097)	-0.270 (-0.858, 0.318)
log(Age)	-5.725 (-17.969, 6.518)		
log.Calendar	-0.144 (-0.331, 0.041)	-0.416 (-0.682, -0.150)	0.254 (-0.083, 0.592)
log.eGFR.Rate	0.042 (-0.125, 0.210)		0.227 (-0.770, 1.225)
MI	-0.246 (-0.397, -0.094)	0.234 (-0.061, 0.530)	0.186 (-0.199, 0.572)
Phosphate	0.511 (0.132, 0.890)	0.869 (-0.059, 1.799)	
PVD	-0.248 (-0.394, -0.103)	-0.168 (-0.405, 0.068)	-0.183 (-0.492, 0.126)
ST	-0.274 (-0.431, -0.117)	-0.181 (-0.516, 0.153)	-0.278 (-0.611, 0.055)
SBP	-0.001 (-0.004, 0.002)	0.005 (-0.000, 0.011)	
uPCR	0.125 (-0.318, 0.569)	0.700 (0.112, 1.288)	-0.108 (-0.736, 0.519)
log.uPCR.Rate		0.218 (-0.310, 0.747)	-0.198 (-0.534, 0.137)
uPCR.Rate		-0.019 (-0.045, 0.005)	0.036 (-0.062, 0.136)
BMI			

Hazard Ratios for the Three-State Model

Female patients were more likely to remain in the CKD state than Males, or to remain in the RRT state once there. Smokers were more likely than Non-/Former Smokers to undergo any transition, apart from CKD to Tx. Blood results had effects on all transitions in some way, and Primary Renal Diagnosis were strongly related to predictions. Patients with Liver Disease were much more likely to undergo a Transplant than those without Liver Disease ($\beta = 12.4$).

3.3 Example

3.4 Validation

4 Discussion

Within our Multi-State Clinical Prediction Model, [covariate], [covariate] and [covariate] had the largest effect on the predictions for patient outcomes. This was caused by the strong relationship between the time to [RRT(?)] and the level of [covariate], which also had a strong link to [mortality(?)], after adjusting for the other covariates within our model.

[Describe the baseline hazards, seen in suppl. material]

[Describe the validation results]

Paragraph 2

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, 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.

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, 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.

Paragraph 4

Paragraph 5

Paragraph 6

Paragraph 7

References

- [1] E. Barron, "Chronic Kidney Disease (CKD) prevalence model," Public Health England, Oct. 2014.
- [2] S. J. MacNeill, D. Ford, K. Evans, and J. F. Medcalf, "Chapter 2 UK Renal Replacement Therapy Adult Prevalence in 2016: National and Centre-specific Analyses," *Nephron*, vol. 139 Suppl 1, pp. 47–74, 2018, doi: 10.1159/000490960.

- [3] B. Hole, J. Gilg, A. Casula, S. Methven, and C. Castledine, "Chapter 1 UK Renal Replacement Therapy Adult Incidence in 2016: National and Centre-specific Analyses," *Nephron*, vol. 139 Suppl 1, pp. 13–46, 2018, doi: 10.1159/000490959.
- [4] 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.
- [5] 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: 10.1053/j.ajkd.2010.07.016.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.

- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.

- [34] 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.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] 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.
- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] 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.
- [49] C. R. Weinberg, "How Bad Is Categorization?" *Epidemiology*, vol. 6, no. 4, pp. 345–347, 1995.
- [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] 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.
- [53] 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.
- [54] 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.
- [55] 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>.
- [56] 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.
- [57] 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.
- [58] “KDIGO Clinical Practice Guideline for Acute Kidney Injury,” *OFFICIAL JOURNAL OF THE INTERNATIONAL SOCIETY OF NEPHROLOGY*, p. 141, 2012.
- [59] 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.
- [60] 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.
- [61] R. C. Team, “R: A Language and Environment for Statistical Computing.” R Foundation for Statistical Computing, Vienna, Austria, Vienna,
- [62] H. Wickham, “The tidy tools manifesto.” <https://cran.r-project.org/web/packages/tidyverse/vignettes/manifesto.html>, Nov-2017.
- [63] 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.
- [64] C. Jackson, “Flexsurv: A Platform for Parametric Survival Modelling in R,” p. 33.
- [65] B. Ripley and W. Venables, “Package ‘nnet’,” Feb-2016.
- [66] D. Vaughan and M. Dancho, “Furrr: Apply Mapping Functions in Parallel using Futures.” May-2018.
- [67] 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.
- [68] 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.
- [69] 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.
- [70] 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.
- [71] 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.
- [72] 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.

- [73] D. B. Rubin, *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons, Inc, 1984.
- [74] 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.
- [75] 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.
- [76] 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.