Looking to the future: Predicting renal replacement outcomes in a large community cohort with chronic kidney disease

RRT outcome prediction in CKD

Angharad Marks PhD [1,2], Nicholas Fluck DPhil [2], Gordon J Prescott PhD [1], Lynn Robertson MSc [1], William G Simpson MBChB [2], W Cairns Smith PhD [1] & Corri Black MBChB[1,2]

[1]Aberdeen Applied Renal Research Collaboration, Division of Applied Health Sciences, University of Aberdeen, Aberdeen, Scotland

[2]NHS Grampian, Aberdeen, Scotland

Corresponding author:

Dr Angharad Marks

Division of Applied Health Sciences,

University of Aberdeen

Polwarth Building, Foresterhill,

Aberdeen. AB25 2ZD

Telephone (+44) 01224 437134

Email: a.marks@abdn.ac.uk

Abstract word count: 258
Word count : 3444
Reference count: 31
Figures: 3
Tables: 3
Equation box:1

Abstract

Background
Chronic kidney disease (CKD) is common and important due to poor outcomes. An ability to stratify CKD care based on outcome risk should improve care for all. Our objective was to develop and validate 5 year outcome prediction tools in a large population based CKD cohort. Model performance was compared to the recently reported 'Kidney Failure Risk Equation' (KFRE) models.

Methods
Those with CKD in the GLOMMS-I (3396) and –II (18687) cohorts were used to develop and validate a renal replacement therapy prediction tool. The discrimination, calibration and overall performance was assessed. The net reclassification index compared performance of the developed model and the 3- and 4-variable KFRE model to predict RRT in the validation cohort.

Results
The developed model (with measures of age, sex, excretory renal function and proteinuria) performed well with a C-statistic of 0.938 (0.918-0.957) and Hosmer-Lemeshow (HL) $\chi 2$ statistic 4.6. In the validation cohort (18687), the developed model falsely identified fewer as high risk (414 versus 3278 individuals) compared to the KFRE 3-variable model (measures of age, sex and excretory renal function), but had more false negatives (58 versus 21 individuals). The KFRE 4-variable model could only be applied to 2,274 individuals because of a lack of baseline urinary ACR data, thus limiting its use in routine clinical practice.

Conclusions
CKD outcome prediction tools have been developed by ourselves and others. These tools could be used to stratify care, but identify both false-positives and -negatives. Further refinement should optimise the balance between identifying those at increased risk with clinical utility for stratifying care.

Keywords
Chronic kidney disease, outcome, risk prediction,

Summary

An ability to appropriately stratify care for those with CKD should improve care for all. We demonstrate the development of an outcome prediction tool and compared performance in a very large cohort to the 3- and 4-variable Kidney Failure Risk Equation (KFRE) outcome prediction equations. All current models require refinements to identify those at risk without labelling all individuals as high-risk.

Introduction

In the UK, over 3.6 million adults are estimated to have chronic kidney disease (CKD)[1]; 23 million in the United States[2, 3]. While many remain undiagnosed, recognition is improving rapidly and more are coming to medical attention[4]. People with CKD are at increased risk of mortality, cardiovascular disease and progressive kidney function decline (leading to renal replacement therapy (RRT))[5, 6]. Progression to poor outcomes is highly variable and only a small proportion will require RRT[4]. Important opportunities therefore exist for improving care, maintaining function, reducing progression and minimising and managing complications. People with CKD often present to primary care, are often elderly and frequently have multiple morbidities. An ability to identify which patients would benefit most from interventions including referral to specialist services is key. Stratification of patients by predicted risk of future outcomes would potentially enable care pathways to be optimised[7].

The literature regarding prognosis prediction in CKD has been recently reviewed[8, 9] and the processes involved summarised[10]. Of the studies identified in the reviews, ten predicted progression of CKD or renal failure, three cardiovascular events and five all-cause mortality. All but two of the progression prediction models[11, 12] were developed in patients referred to nephrology services. Thus model utility in other contexts, particularly the community, is not clear[13, 14]. Some models used variables not routinely available in clinical practice e.g. cystatin C. Very few models have been externally validated. None have been applied in clinical practice. Although Tangri et al.[15] developed models using a population referred to nephrology services, these models contain commonly available variables (including measures of age, sex and excretory renal function), were externally validated by the authors in another referred population; and model performance has since been reported in 595 referred individuals[14]. Unlike many prediction model studies, model performance metrics including discrimination,

calibration and reclassification [10] were reported. Thus these 'kidney failure risk equation' (KFRE) models have the best evidence for their use to predict risk in CKD[15].

We aimed to report the development and validation of models to predict first outcome (mortality or RRT initiation) by five years in a large community based CKD cohort. We compared the performance of our RRT prediction model with the KFRE models[15], using real-life data to explore applicability to current clinical care.

Materials and methods

This work was approved by the University of Aberdeen Ethics Review Board in keeping with the principles of the declaration of Helsinki. Data-linkage of pseudoanonymised routine healthcare data provided measures of renal function, demographics, baseline comorbidity and outcome data. Data was available from the Scottish Renal Registry, Information Services Division Scotland (hospital episode data) and NHS Grampian (single Clinical Biochemistry Service, Renal management system and Health Intelligence).

Development cohort

As outlined elsewhere [16-18], the GLOMMS-I cohort (n~3,400) consists of all residents of Grampian in 2003 aged over 15 years, with a creatinine measurement between January and June 2003 of ≥150μmol/L and ≥130μmol/L for males and females respectively, who had impaired renal function (eGFR<60ml/min/1.73m^2) for at least three months. All but 1.5% of GLOMMS-I have follow-up until death or within a year of 30 June 2009.

Validation cohort

The GLOMMS-II cohort (70,780 individuals) consists of all residents of Grampian with at least one eGFR <60ml/min/1.73m^2 in 2003 (both impaired eGFR (10,857) and CKD (18,687)),

a sample (~20,000) of those with only normal eGFR values in 2003 and a sample (~20,000) of those with no measurement of eGFR in 2003 but sampling in the years pre and post 2003. Only the 18687 with sustained (for at least three months) stage 3a-5 CKD were used for validation since the aim was to predict prognosis in those with true CKD.

Model development

Models were developed in the GLOMMS-I cohort. Only those with stage 3b CKD or worse were used for this analysis since there were only 18 with an eGFR of ≥45 ml/min/1.73m^2. Those who died on the creatinine index date were also excluded, leaving 3,396 individuals. Age, gender, stage of CKD and presence of proteinuria were pre-specified as probable predictors of outcome. The additive value of vascular comorbidity or diabetes was also explored (with forward selection) since these have been associated with outcome amongst those with CKD[19, 20]. The 4-variable IDMS-aligned MDRD equation (as used in local clinical practice) and baseline creatinine were used to calculate eGFR. The last urinary albumin (ACR) or protein creatinine ratio (PCR) measured prior to index was used, missing values were not imputed since these were considered likely to be "missing not at random" and only measured where there is clinical indication . Individuals were categorised based on whether they had either ACR, PCR or neither measured prior to baseline, to allow analysis both for all cases and for only those with a proteinuria measurement. Definitions and categorisation of the exposures available for the cohort are shown under Table 1. Outcomes included initiation of RRT and death by five years.

The utility of several model types was explored including Cox proportional hazards, logistic and multinomial regression, measures of performance as below were compared. The multinomial regression model had similar performance to the logistic regression models for both RRT and mortality, for clarity the logistic regression models are reported here. Cox

models failed to follow proportional hazards and showed worse five year predictive performance. Hence only logistic regression models for the binary outcomes RRT initiation at five years or not and being dead at five years or not, are presented further here. For each model, the performance of predictors were judged using a summary measure of fit and complexity – the Bayesian information criterion (BIC), smaller BIC values equate to better performance[21]. The models judged to offer the most parsimonious fit were then used to derive coefficients. Model 'discrimination' was assessed using C-statistics and receiver-operator-characteristics (ROC) curves[22]. Calibration plots of the predicted outcome against the actual outcome were plotted[22] and Hosmer-Lemeshow (HL) statistics calculated based on deciles of risk[22]. The 'calibration' was also assessed using the performance of the predicted outcomes compared to the actual outcomes – false positive, false negative, true predictions and overall performance.

Model validation

The chosen models were then used to predict outcome in those with CKD in the GLOMMS-II cohort. Calibration plots for model performance were plotted and the HL statistic calculated.

Comparison of model performance with 'kidney failure risk equation'

The risk of RRT initiation using the KFRE 3- and 4-variable equations (KFRE-3v and KFRE-4v) was also calculated for the GLOMMS-II cohort[15]. Net reclassification improvement (NRI)[23-25] for RRT initiation by five years using the model developed here, compared with the KFRE-3v and KFRE-4v equations[15] were described using a threshold for high-risk of initiating RRT by five years at 5%, as used by Tangri in individuals with stage 3 CKD[15]. For both the "event" and non-event" NRIs, positive values suggest the comparator model is better than the referent model at identifying "events" and "non-events" and negative values the opposite.

Results

Development

Of 3,396 individuals with stage 3b to 5 CKD, 44.0% were male, 66.8% had stage 3b CKD, 70.6% had no measure of proteinuria prior to index and the average age was 78.6 years. The outcomes at five years by baseline characteristics are shown in Table 1 and in Figure 1. Individuals who initiated RRT (4.2%, some subsequently dying) were younger, had lower eGFR, were more likely to be male and have macroalbuminuria compared to those who did not start RRT.

The best logistic regression models for the prediction of RRT initiation are shown in Table 2. Model 7 performed best (BIC 695, discrimination C-statistic 0.938 (0.918-0.957)). All models of RRT initiation had good calibration, HL χ^2 statistic 4.6 for model 7 (HL χ^2 statistic <20 is considered evidence of adequate calibration[8] and a non-significant probability observed outcome differs from predicted). Models with information on the presence of diabetes and vascular comorbidity did not improve discrimination (C-statistic 0.937 (0.918-0.957) and 0.938 (0.920-0.957) respectively), calibration (HL χ^2 statistic 6.7 and 5.8 respectively) or goodness of fit (BIC 695 and 701 respectively). For initiation of RRT, model 7, using a 5% threshold of high risk had a sensitivity of 0.82, specificity of 0.90, negative predictive value of 0.99, false negative rate of 0.18, resulting in 90% of individuals being correctly classified. Limiting analysis to those only with a measure of proteinuria (either 532 ACR or 468 PCR) again showed that a model based on model 6/7 were the best performing (results not shown here). However, only 998 individuals could be used.

Logistic regression models to predict death by five years performed poorly with little improvement with the addition of measures of excretory renal function and proteinuria (C-statistic 0.753 (0.737-0.769)) over that of age and sex alone (C-statistic 0.749 (0.733-0.765)).

There was poor calibration with statistically significant differences in the predicted and observed deaths (HL χ^2 statistic all greater than 18). No further results are presented.

Validation

The validation cohort comprised 18,687 individuals with stage 3a-5 CKD. Age distribution was similar to the development cohort (Table 1 and Figure 1), a lower proportion (1.1%) initiated RRT and there was a higher proportion of survivors (66.9%).

Model 7 (equation box 1) was applied to those with stage 3a-5 CKD in the GLOMMS-II validation cohort; using the 5% threshold of predicted risk of initiating RRT by five years as designating someone "high-risk". Only 578 individuals were "high-risk". Of the 222 individuals who initiated RRT, 58 (26.1%) were incorrectly classified as not high-risk (false-negative) (Table 3). The model had a specificity of 0.98, sensitivity of 0.74 and 94.5% of the 18,687 were correctly classified as high-risk or otherwise, if limited to the 6341 with stage 3b-5 CKD the equivalent figures were 0.93, 0.81 and 92.9% respectively.

Comparison of models

Our model 7 and the KFRE-3v model were compared for the 18,687 people with CKD in the GLOMMS-II validation cohort. The KFRE-4v model could only be applied to 2,274 individuals because of a lack of baseline urinary ACR data. Model performance measures are shown in Table 3. Model calibration are shown in Figure 2a (for all with CKD) and 2b (for all with CKD and a measure of ACR). Both the KFRE-3v and KFRE-4v models over-predicted those that would initiate RRT. For all 18,687 with CKD, our model 7 had a (discrimination) C-statistic of 0.960 (0.947-0.974) compared to 0.936 (0.918-0.954) for the KFRE-3v model. In the 2,274 where the urinary ACR was also available, the C-statistics were 0.936 (0.906-0.966), 0.881 (0.827-0.935) and 0.948 (0.922-0.974) for our model 7, and the KFRE-3v and KFRE-4v models respectively.

For the 18,687 people with CKD, using the 5% risk threshold for identifying high or low risk, our model was more specific, 0.98 overall (vs 0.82 with KFRE-3v model) However, our model missed more cases (58 vs 21 false-negatives) who went on to initiate RRT and thus generally had a lower sensitivity. These findings were consistent for those both under and over 75 years of age. For those with stage 3a CKD the proportion predicted a false-negative by the models (predicted low risk but went on to initiate RRT) were similar (87.0% and 82.6% respectively for ours and the KFRE-3v model). For both models', performance was better in more advanced disease:- false negatives of 68.3% and 4.9% for stage 3b and 11.2% and 0% for stage 4 respectively. The better identification of high RRT risk with the KFRE-3v than our model for all CKD is reflected in the "event" NRI (Table 3) which being generally negative implies that the referent model is better at predicting events. However this better event identification came at a cost, with 2,053 and 1,246 individuals with stage 3b and 4 CKD being classified as high risk compared to 102 and 333 for our model. Overall, our model correctly reclassified 2,864 individuals from KFRE-3v 'high-risk' to 'low risk' (Table 3) – "non-event" NRI. The majority of individuals did not initiate RRT, as shown in Figure 3. Overall the $NRI^{0.05}$ was small suggesting no model was better overall than another, although for stage 4 and 5 CKD the $NRI^{0.05}$was positive thus favouring our model.

For the subset of 2,274 individuals with urinary ACR data (second part of Table 3) there was a different performance profile. The KFRE-3v model identified 512 individuals as high risk, compared to 276 with the KFRE-4v model and 120 with our model. Overall the sensitivity of the KFRE-3v model was the same as the KFRE-4v model (0.84) and better than ours (0.56), reflected in the event NRIs. However the specificity of ours was better 0.96 vs 0.79 KFRE-3v and 0.89 KFRE-4v) . In particular the KFRE-3v model identified all those with stage 4 CKD as high risk (specificity 0.00, non-event NRI positive in favour of our model). Overall the $NRI^{0.05}$s

favoured the KFRE-4v model over the KFRE-3v and our model, except for stage 4 CKD where our model was favoured over both the KFRE models (missing only one RRT initiator (false-negative 6.3%) and identifying only 68 individuals as high risk).

Discussion

We have demonstrated that it is possible to develop prediction tools for the initiation of RRT in a community CKD population, not just those referred to nephrology clinics. Using routinely available clinical biomarkers we were able to predict the five year risk of RRT using a simple prediction model. This tool could be used to stratify CKD populations by RRT risk, identifying clinically relevant sub-groups at high and low risk. The performance of our model using traditional metrics was good, and comparable to the widely cited KFRE models[15]. This is the first study to apply the KFRE 3- and 4-variable models to a non-referred population and only the second to apply them outside the Canadian health system[14]. Both KFRE models performed well on traditional metrics, but the KFRE-4v model had restricted application because of the data available from routine care. This similar performance of the KFRE models was despite differing age and gender (35% vs 56% male; (74.4 (SD 0.8) vs 70(SD 24) years), lower prevalence of diabetes (8% vs 37%) and vascular disease (24% vs 40%) in our population compared to the KFRE deriving population.

The major strength of this study lies in the availability of population-based routine clinical data with complete coverage for a large single health authority region, supporting good translation into clinical practice. The community cohort extends generalisability to beyond those already referred to nephrology care. The technical challenges of identifying those with chronic kidney disease are well-documented and have resulted in widely varying prevalence reporting in the literature[26]. Here we have defined CKD using the internationally adopted definition with an eGFR of <60 ml/min/1.73m^2 present for at least three months. We have reported a range of

model performance metrics to enable assessment of overall performance of the models. There are however some limitations to this study. The original (deriving) cohort was nested within the validating cohort and had no individuals with stage 3a disease, thus there is the potential for model over-fitting. However, this would not impact on the performance (and further external validation) of the KFRE models, and thus the issue with the over-prediction of risk for these models is appropriately highlighted. The use of data from one health board does potentially reduce generalisability, however registry data suggests that the region has similar RRT initiation rates to the rest of the UK. The majority of the population is Caucasian which limits reproducibility in more ethnically diverse populations. The use of a model that includes a group with "proteinuria not measured" is unusual. However, although there may be multiple reasons why proteinuria is not measured (we assume in the majority because the clinician in charge has not thought it relevant), the single reason it is not available in this dataset is that it was not measured. As such these individuals are an important risk group, particularly as demonstrated, this is the majority of individuals at a population level with CKD. We would expect that the use of such a model to assess risk should prompt future assessment and thus the measurement of proteinuria, and in itself is useful to consider "baseline" risk.

There is growing evidence that prediction models in CKD have the potential to stratify future risk of major health outcomes including RRT. However, reporting of performance is variable and only three studies[15, 27, 28] report external validation, the ideal in prediction model development[29]. Clinical applicability has also been limited by the variables included. For mortality prediction, we and others have found that the addition of renal function measures (including eGFR and proteinuria), added little to age and sex[27].

Our RRT prediction model performed well, with good or excellent discrimination (C-statistic 0.938 in the derivation and 0.960 in the validation cohort) and calibration (HL χ^2 statistic 4.6). Equivalent figures published for Tangri et al.[15] were C-statistics 0.89, 0.91 and 0.92 for the

KFRE 3, 4 and 8-variable models and (Nam and D'Agnostino) χ^2 statistic 37, 32 and 19 respectively.

This is one of only two studies[14] that have reported validation of the KFRE-3v and KFRE-4v models in an external population, both demonstrating excellent discrimination. However, we found that both KFRE models over-predicted risk compared to actual outcomes (calibration), likely due to the competing risk of death. This is supported by the original KFRE publication [15], reporting Nam and D'Agnostino χ^2 statistics of 37, 32 and 19 for the 3-, 4- and 8-variable models, again suggesting the observed and predicted events were rather different. Although both Tangri et al.[15] and Peeters et al.[14] report the NRI comparing the KFRE models relative performance, neither is so explicit in terms of the numbers of individuals mistakenly identified as high or low risk. To our knowledge no others have reported the numbers that would be identified as high-risk using these models in a community CKD population and thus the implications of using them, e.g. in general practice to guide referral, particularly since all individuals with stage 4 CKD were high risk according to the KFRE-3v model. We demonstrated like others[15] that overall NRI favours the use of the KFRE-4v over the KFRE-3v model, however this was only applicable to 2,274 of the 18,687 individuals with CKD.

Head-to-head comparison of our model (more specific, less sensitive) to others allows assessment of the potential clinical utility of introducing models into routine practice. Although CKD outcome prediction models have the potential to identify individuals at high- and low-risk, currently available models have limitations. Using the risk thresholds ($P \geq 0.05$=high-risk) in this study, the KFRE-3v model identified all with stage 4 CKD as high-risk and as such adds little to CKD stage. The KFRE-4v equation was unusable in 16,413 of 18,687 individuals in this population-level CKD cohort because of no measure of urinary ACR, although ACR measurement will increase given the most recent KDIGO guidelines (30). Our model

incorporating categorical information on whether urinary protein was measured (ACR or PCR), and if so, the level, does add some value to prediction estimates, facilitating use in real-world data. Although a more viable number of individuals were identified as high-risk by our model, offering potential to guide referral to nephrology care, the false negative rate limits current clinical application. This balance of false-positive to false-negative (whatever the clinical decision as a result – referral to nephrology or access surgery) is important both in terms of service (clinics, dialysis-access-surgery lists) and human (anxiety, risks of inappropriate surgery) costs against the missed opportunities to intervene earlier and thus change and improve prognosis. Other issues that would need to be considered if such models were to be introduced are the effect of using different thresholds for high-risk.

This study has two major implications for current clinical practice. First it highlights the need for more timely investigation of those identified with CKD to identify risk factors such as proteinuria. This need has also been identified in the most recent KDIGO guidelines (30) by cross-classifying eGFR CKD stage by ACR. Second, with further refinement, prediction models could form part of a CKD pathway for shared care, potentially as a signal for first or re-referral.

Model refinements require further research to improve performance. In particular:- exploration of alternative methods (competing risk of death); and prior course of renal function (in clinical practice considered very important, but so far no measure of this in outcome prediction models). In terms of patient information, a model of patient "survival" i.e. being alive and not requiring RRT, would be appealing.

Conclusion

CKD is common with serious consequences for some patients. Tools that predict the initiation of RRT have been developed and perform well using traditional metrics. Prediction models

offer the potential to target and tailor clinical care within a carefully managed care-pathway. For clinical utility, further refinement is needed to optimise the balance between those labelled as high-risk and false-negatives, and clinical performance.

Transparency declaration: No conflicts of interest declared

Acknowledgements

This work was supported by the Chief Scientists Office for Scotland [CZH/4/656], NHS Grampian Endowment Research Fund and NHS Grampian Renal Endowment.

We thank Information Services Division Scotland, Scottish Renal Registry and NHS Grampian who provided data. We also thank the University of Aberdeen Data Safe Haven, who hosted the data and provided data management support and the linkage service, and also the staff of NHS Grampian Renal Unit.

References

1. Kerr M, Bray B, Medcalf J, O'Donoghue DJ, Matthews B. Estimating the financial cost of chronic kidney disease to the NHS in England. Nephrol.Dial.Transplant. 2012; 27: iii73-iii80.

2. Coresh J, Selvin E, Stevens L, et al. Prevalence of Chronic Kidney Disease in the United States. JAMA 2007; 298: 2038-2047.

3. Cukor D, Kimmel PL. Education and end of life in chronic kidney disease: disparities in black and white. Clin.J.Am.Soc.Nephrol. 2010; 5: 163-166.

4. Eckardt KU, Coresh J, Devuyst O, et al. Evolving importance of kidney disease: from subspecialty to global health burden. Lancet 2013; 382: 158-169.

5. Chronic Kidney Disease Prognosis Consortium (CKD-PC). Matsushita,K., van der Velde M, Astor BC, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet 2010; 375: 2073-2081.

6. McCullough PA, Li S, Jurkovitz CT, et al. Chronic kidney disease, prevalence of premature cardiovascular disease, and relationship to short-term mortality. Am.Heart J. 2008; 156: 277-283.

7. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ 2013; 346: e5793.

8. Tangri N, Kitsios GD, Inker LA, et al. Risk prediction models for patients with chronic kidney disease: a systematic review. Ann.Intern.Med. 2013; 158: 596-603.

9. Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review. PLoS Med. 2012; 9: e1001344.

10. Tripepi G, Heinze G, Jager KJ, Stel VS, Dekker FW, Zoccali C. Risk prediction models. Nephrol.Dial.Transplant. 2013; 28: 1975-1980.

11. Hallan SI, Ritz E, Lydersen S, Romundstad S, Kvenild K, Orth SR. Combining GFR and albuminuria to classify CKD improves prediction of ESRD. Journal of the American Society of Nephrology 2009; 20: 1069-1077.

12. Johnson ES, Thorp ML, Platt RW, Smith DH. Predicting the Risk of Dialysis and Transplant Among Patients With CKD: A Retrospective Cohort Study. Am.J.Kidney Dis. 2008; 52: 653-660.

13. Acedillo RR, Tangri N, Garg AX. The kidney failure risk equation: on the road to being clinically useful? Nephrol.Dial.Transplant. 2013; 28: 1623-1624.

14. Peeters MJ, van Zuilen AD, van den Brand JA, et al. Validation of the kidney failure risk equation in European CKD patients. Nephrol.Dial.Transplant. 2013; 28: 1773-1779.

15. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. JAMA 2011; 305: 1553-1559.

16. Marks A, Black C, Fluck N, et al. Translating chronic kidney disease epidemiology into patient care--the individual/public health risk paradox. Nephrol.Dial.Transplant. 2012; 27 Suppl 3: iii65-iii72.

17. Marks A, Macleod C, McAteer A, et al. Chronic kidney disease, a useful trigger for proactive primary care? Mortality results from a large UK cohort. Fam.Pract. 2013; 30: 282-289.

18. Marks A, Fluck N, Prescott GJ, et al. Definitions of progression in chronic kidney disease--predictors and relationship to renal replacement therapy in a population cohort with 6 years follow-up. Nephrol Dial Transplant 2014; 29: 333.

19. Bang H, Vupputuri S, Shoham DA, et al. SCreening for Occult REnal Disease (SCORED): a simple prediction model for chronic kidney disease. Arch.Intern.Med. 2007; 167: 374-381.

20. Bash LD, Coresh J, Kottgen A, et al. Defining incident chronic kidney disease in the research setting: The ARIC Study. Am.J.Epidemiol. 2009; 170: 414-424.

21. Cleves M., Gould W., Gutierrez R. G., Marchenko Y. V. An introduction to Survival Analysis Using Stata. Stata Press, Texas, USA: 2010; .

22. Steyerberg E. W. Chapter 15 - Evaluation of Performance. In: Gail M., Tsiatis A., Krickeberg K., Wong W., Sarnet J., eds. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer, New York, USA: 2009; 255-280.

23. Pencina MJ, D'Agostino RB S, D'Agostino RB,Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat.Med. 2008; 27: 157-72; discussion 207-12.

24. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. Epidemiology 2014; 25: 114-121.

25. Pencina MJ, D'Agostino RB S, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat.Med. 2011; 30: 11-21.

26. McCullough K, Sharma P, Ali T, et al. Measuring the population burden of chronic kidney disease: a systematic literature review of the estimated prevalence of impaired kidney function. Nephrol Dial Transplant. 2012; 27: 1812.

27. Landray MJ, Emberson JR, Blackwell L, et al. Prediction of ESRD and death among people with CKD: the Chronic Renal Impairment in Birmingham (CRIB) prospective cohort study. Am.J.Kidney Dis. 2010; 56: 1082-1094.

28. Desai AS, Toto R, Jarolim P, et al. Association between cardiac biomarkers and the development of ESRD in patients with type 2 diabetes mellitus, anemia, and CKD. Am.J.Kidney Dis. 2011; 58: 717-728.

29. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013; 10: e1001381.

30. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int. Suppl. 2013; 3: 1-150.

31. Soo M, Robertson LM, Ali T, et al. Approaches to ascertaining comorbidity information: validation of routine hospital episode data with clinician-based case note review. BMC Res.Notes 2014; 7: 253-0500-7-253.

Tables

Table 1 - Outcome at 5 years follow-up by baseline characteristics

Baseline characteristics		Outcomes at 5 years GLOMMS-I cohort					Outcomes at 5 years GLOMMS-II cohort with CKD 3a-5				
			RRT Alive	RRT then died	Died no RRT	Survive		RRT Alive	RRT then died	Died no RRT	Survive
All	N (%)	3396	(2.4)	(1.8)	(53.0)	(42.8)	18687	(0.7)	(0.4)	(31.9)	(66.9)
Sex	Male	1493	(3.2)	(2.7)	(52.7)	(41.4)	6580	(1.3)	(0.8)	(34.5)	(63.5)
	Female	1903	(1.7)	(1.1)	(53.2)	(43.9)	12107	(0.4)	(0.3)	(30.5)	(68.8)
Age	Median (25-75%)		58 (46-71)	70 (63-76)	82 (76-87)	75 (67-80)		58 (44-70)	70 (59-76)	81 (75-87)	73 (66-79)
years	15-44	77	(24.7)	(6.5)	(3.9)	(64.9)	305	(11.8)	(2.3)	(5.2)	(80.7)
	45-54	106	(12.3)	(6.6)	(11.3)	(69.8)	660	(3.2)	(1.2)	(6.8)	(88.8)
	55-64	280	(6.8)	(2.1)	(22.9)	(68.2)	2201	(1.5)	(0.5)	(12.2)	(85.7)
	65-74	758	(2.2)	(3.3)	(37.7)	(56.7)	5630	(0.6)	(0.6)	(19.2)	(79.6)
	75-84	1418	(0.9)	(1.3)	(58.4)	(39.4)	7119	(0.2)	(0.3)	(38.2)	(61.2)
	85+	757	(0.0)	(0.0)	(80.2)	(19.8)	2772	(0.0)	(0.0)	(65.8)	(34.2)
eGFR	Median (25-75%)		19 (13-28)	19 (14-28)	33 (27-36)	35 (30-37)		19 (13-32)	22 (15-34)	46 (37-53)	51 (44-56)
Index eGFR	45-59	0					12346	(0.1)	(0.1)	(26.1)	(73.7)
ml/min/1.73m²	30-44	2268	(0.8)	(0.7)	(50.2)	(48.3)	4951	(0.4)	(0.4)	(40.2)	(59.0)
	15-29	1036	(3.5)	(2.7)	(59.7)	(34.2)	1246	(4.5)	(2.6)	(55.1)	(37.7)
	0-14	92	(29.3)	(19.6)	(46.7)	(4.3)	144	(32.6)	(15.3)	(36.1)	(16.0)
ACR	Median (25-75%)		129 (32-215)	135 (6-319)	4 (1-16)	2 (0.9-9)		143 (35-240)	135 (9-319)	3 (0.9-10)	1 (0.9-4)
PCR	Median (25-75%)		261 (125-415)	228 (73-441)	39 (17-83)	31 (13-79)		216 (85-414)	228 (73-382)	26 (13-59)	21 (9-53)
mg/mmol	Normoalbuminuria	504	(1.2)	(1.4)	(45.6)	(51.8)	2125	(0.6)	(0.3)	(29.1)	(69.9)
	Microalbuminuria	181	(1.7)	(1.7)	(54.1)	(42.5)	602	(0.5)	(0.5)	(39.7)	(59.3)
	Macroalbuminuria	313	(15.0)	(10.2)	(42.2)	(32.6)	548	(13.1)	(7.3)	(35.2)	(44.3)
	Not measured	2398	(1.0)	(0.8)	(55.9)	(42.3)	15412	(0.3)	(0.2)	(31.8)	(67.6)

Microalbuminuria= ≥2.5mg/mmol albumin creatinine ratio (ACR) for men, ≥3.5mg/mmol ACR for women; macroalbuminuria = ≥30mg/mmol ACR, or ≥50mg/mmol protein creatinine ratio (PCR); normoalbuminuria = ACR or PCR below thresholds for micro/macroalbuminuria; not measured= no measure of ACR or PCR prior to index. In GLOMMS-I, the presence of vascular disease was defined by a case-note review record of:- ischaemic heart disease (angina, myocardial infarction, abnormal coronary angiogram, coronary angioplasty or coronary artery bypass grafting); congestive cardiac failure (record of New York Heart Association criteria symptoms); peripheral vascular disease (claudication, suggestive angiogram or tissue loss due to vascular disease); cerebrovascular disease (transient ischaemic attack, stroke (with or without full recovery) or an abnormal CT scan in keeping with cerebrovascular disease), up to the time of "index". Diabetes was defined as present in GLOMMS-I if case-note review up to the time of index had this recorded. In GLOMMS-II, the presence of any hospital episode record with ICD-10 (World Health Organisation, international classification of disease) coding in the five years prior to index consistent with vascular disease (I11.0, I13.0, I13.2, I20.x-I25.x, I42.0, I42.5–I42.9, I43.x, I50.x, I60.x-I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, G45.x-G46.x, H34.0) or diabetes (E10.x-E14.x, O24.0-O24.1) was taken to indicate these comorbidities were present. Coding performance compared to case-note review has been reported previously[31].

Table 2 – Variables, odds ratios and performance of logistic regression models for predicting having initiated RRT by 5 years follow-up

	Models						
Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Sex (female vs male)	**0.55 †**	**0.38 †**	**0.35 †**	**0.27 †**	**0.35 †**	**0.35 †**	**0.36 †**
Age/10 (years)	**0.47 †**	**0.44 †**	**0.42 †**		**0.46 †**	**0.47 †**	**0.47 †**
(Mean age)^2				**1.00 †**			
CKD stage 4 vs 3		**6.55 †**					
CKD stage 5 vs 3		**93.37 †**					
Index eGFR			**0.83 †**	**0.83 †**	**0.85 †**	**0.84 †**	**0.84 †**
Age.eGFR interaction					**1.00 †**		
A **= Baseline proteinuria not measured**							
Normoalbuminuria vs A						**0.84**	
Microalbuminuria vs A						**2.08**	
Macroalbuminuria vs A						**5.09 †**	
B **= Baseline proteinuria not measured / Normoalbuminuria**							
Microalbuminuria vs B							**2.16**
Macroalbuminuria vs B							**5.31 †**
C statistic	**0.817**	**0.9022**	**0.917**	**0.865**	**0.918**	**0.938**	**0.938**
(C statistic 95% confidence interval)	**(0.784-0.849)**	**(0.876-0.928)**	**(0.893-0.941)**	**(0.827-0.903)**	**(0.894-0.941)**	**(0.918-0.957)**	**(0.918-0.957)**
P value*		**<0.0001**	**0.0077**	**<0.0001**	**0.2164**	**0.0004**	**0.8706**
versus model		**1**	**2**	**3**	**3**	**3**	**6**
Bayesian information criterion	**1011**	**801**	**729**	**830**	**737**	**703**	**695**
Hosmer-Lemeshow Chi sq statistic	**10.6**	**6.3**	**8.9**	**11.7**	**10.3**	**4.6**	**4.6**
Hosmer-Lemeshow Chi sq statistic probability	**0.229**	**0.609**	**0.353**	**0.165**	**0.244**	**0.801**	**0.800**

*Probability that C-statistic differs significantly from previous model (specified below)

† = statistically significant odds ratios

Addition of diabetes or vascular comorbidity to model 7 yielded C-statistics of 0.937 and 0.938, BICs of 695 and 701 and HL statistics of 6.7 and 5.8 respectively.

C-statistic higher indicates better discrimination, BIC lower indicates better goodness of fit, HL statistic lower value indicates that observed and predicted instances of RRT initiation are similar, P value examines the probability that the distribution of observed and predicted instances are significantly different or otherwise

Table 3 – Comparison of the performance of the KFRE 3- and 4-variable and our model to predict RRT initiation at 5 years in those with CKD in the GLOMMS-II cohort

	Comparison			Referent							Comparator							Event NRI$^{0.05}$		Non-event NRI$^{0.05}$		NRI$^{0.05}$	(95%CI)
	Referent	Comparator	Total number of patients at risk	Start RRT		Do not start RRT		Sens	Spec	Proportion correctly classified	Start RRT		Do not start RRT		Sens	Spec	Proportion correctly classified	Number of events correctly given higher risk	Difference in proportions of events reclassified	Number of non-events correctly given lower risk	Difference in proportions of non-events reclassified		
				TP	FN	TN	FP				TP	FN	TN	FP				ΔFN		ΔFP			
All																							
Stage 3a-5	KFRE-3v	Ours	18687	201	21	15187	3278	0.91	0.82	0.823	164	58	18051	414	0.74	0.98	0.975	-37	-0.167	2864	0.155	-0.012	(-0.061 , 0.038)
Stage 3a-5 (under 75years)	KFRE-3v	Ours	8796	163	18	7281	1334	0.90	0.85	0.846	137	44	8340	275	0.76	0.97	0.964	-26	-0.144	1059	0.123	-0.021	(-0.072 , 0.031)
Stage 3a-5 (over 75years)	KFRE-3v	Ours	9891	38	3	7906	1944	0.93	0.80	0.803	27	14	9711	139	0.66	0.99	0.985	-11	-0.268	1805	0.183	-0.085	(-0.221 , 0.051)
Stage 3a	KFRE-3v	Ours	12346	4	19	12291	32	0.17	1.00	0.996	3	20	12320	3	0.13	1.00	0.998	-1	-0.043	29	0.002	-0.041	(-0.124 , 0.042)
Stage 3b	KFRE-3v	Ours	4951	39	2	2896	2014	0.95	0.59	0.593	13	28	4821	89	0.32	0.98	0.976	-26	-0.634	1925	0.392	-0.242	(-0.390 , -0.094)
Stage 4	KFRE-3v	Ours	1246	89	0	0	1157	1.00	0.00	0.071	79	10	903	254	0.89	0.78	0.788	-10	-0.112	903	0.780	0.668	(0.598 , 0.738)
Stage 5	KFRE-3v	Ours	144	69	0	0	75	1.00	0.00	0.479	69	0	7	68	1.00	0.09	0.528	0	0.000	7	0.093	0.093	(0.027 , 0.159)
Those with urinary ACR data at baseline																							
All stage 3a - 5	KFRE-4v	KFRE-3v	2274	36	7	1991	240	0.84	0.89	0.891	36	7	1755	476	0.84	0.79	0.788	0	0.000	-236	-0.106	-0.106	(-0.172 , -0.040)
	KFRE-4v	Ours	2274	36	7	1991	240	0.84	0.89	0.891	24	19	2135	96	0.56	0.96	0.949	-12	-0.279	144	0.065	-0.215	(-0.349 , -0.080)
	KFRE-3v	Ours	2274	36	7	1755	476	0.84	0.79	0.788	24	19	2135	96	0.56	0.96	0.949	-12	-0.279	380	0.170	-0.109	(-0.244 , 0.026)
Stage 3a	KFRE-4v	KFRE-3v	1456	2	5	1436	13	0.29	0.99	0.988	1	6	1443	6	0.14	1.00	0.992	-1	-0.143	7	0.005	-0.138	(-0.397 , 0.121)
	KFRE-4v	Ours	1456	2	5	1436	13	0.29	0.99	0.988	1	6	1447	2	0.14	1.00	0.995	-1	-0.143	11	0.008	-0.135	(-0.395 , 0.124)
	KFRE-3v	Ours	1456	1	6	1443	6	0.14	1.00	0.992	1	6	1447	2	0.14	1.00	0.995	0	0.000	4	0.003	0.003	(0.000 , 0.005)
Stage 3b	KFRE-4v	KFRE-3v	650	12	2	527	109	0.86	0.83	0.829	13	1	312	324	0.93	0.49	0.500	1	0.071	-215	-0.338	-0.267	(-0.407 , -0.126)
	KFRE-4v	Ours	650	12	2	527	109	0.86	0.83	0.829	2	12	604	32	0.14	0.95	0.932	-10	-0.714	77	0.121	-0.593	(-0.831 , -0.355)
	KFRE-3v	Ours	650	13	1	312	324	0.93	0.49	0.500	2	12	604	32	0.14	0.95	0.932	-11	-0.786	292	0.459	-0.327	(-0.545 , -0.108)
Stage 4	KFRE-4v	KFRE-3v	153	16	0	28	109	1.00	0.20	0.288	16	0	0	137	1.00	0.00	0.105	0	0.000	-28	-0.204	-0.204	(-0.272 , -0.137)
	KFRE-4v	Ours	153	16	0	28	109	1.00	0.20	0.288	15	1	84	53	0.94	0.61	0.647	-1	-0.063	56	0.409	0.346	(0.202 , 0.491)
	KFRE-3v	Ours	153	16	0	0	137	1.00	0.00	0.105	15	1	84	53	0.94	0.61	0.647	-1	-0.063	84	0.613	0.551	(0.407 , 0.695)
Stage 5	KFRE-4v	KFRE-3v	15	6	0	0	9	1.00	0.00	0.400	6	0	0	9	1.00	0.00	0.400	0	0.000	0	0.000	0.000	‡
	KFRE-4v	Ours	15	6	0	0	9	1.00	0.00	0.400	6	0	0	9	1.00	0.00	0.400	0	0.000	0	0.000	0.000	‡
	KFRE-3v	Ours	15	6	0	0	9	1.00	0.00	0.400	6	0	0	9	1.00	0.00	0.400	0	0.000	0	0.000	0.000	‡

‡All individuals predicted to initiate RRT by all models therefore NRI probability is inappropriate.
TP=true positive; FN=false negative; TN=true negative; FP=false positive; Sens=sensitivity; Spec=specificity; NRI0.05=net reclassification index, p<0.05 is low risk; ΔFN=difference in false negatives between referent and comparator model; ΔFP=difference in false positives between referent and comparator model.

Figures and legends

Equation box 1

P (y =1, initiate RRT by 5 years) $= e^{lp}/(1 + e^{lp})$

$= 1 / (1 + e^{-lp})$

Lp = 8.090-1.031 x (sex) - 0.761 x (age/10) - 0.175 x (eGFR) if normoalbuminuric / not known

Lp = 8.090-1.031 x (sex) - 0.761 x (age/10) - 0.175 x (eGFR) + 0.772 if microalbuminuric

Lp = 8.090-1.031 x (sex) - 0.761 x (age/10) - 0.175 x (eGFR) + 1.670 if macroalbuminuric

Sex = 1 if male, 2 if female

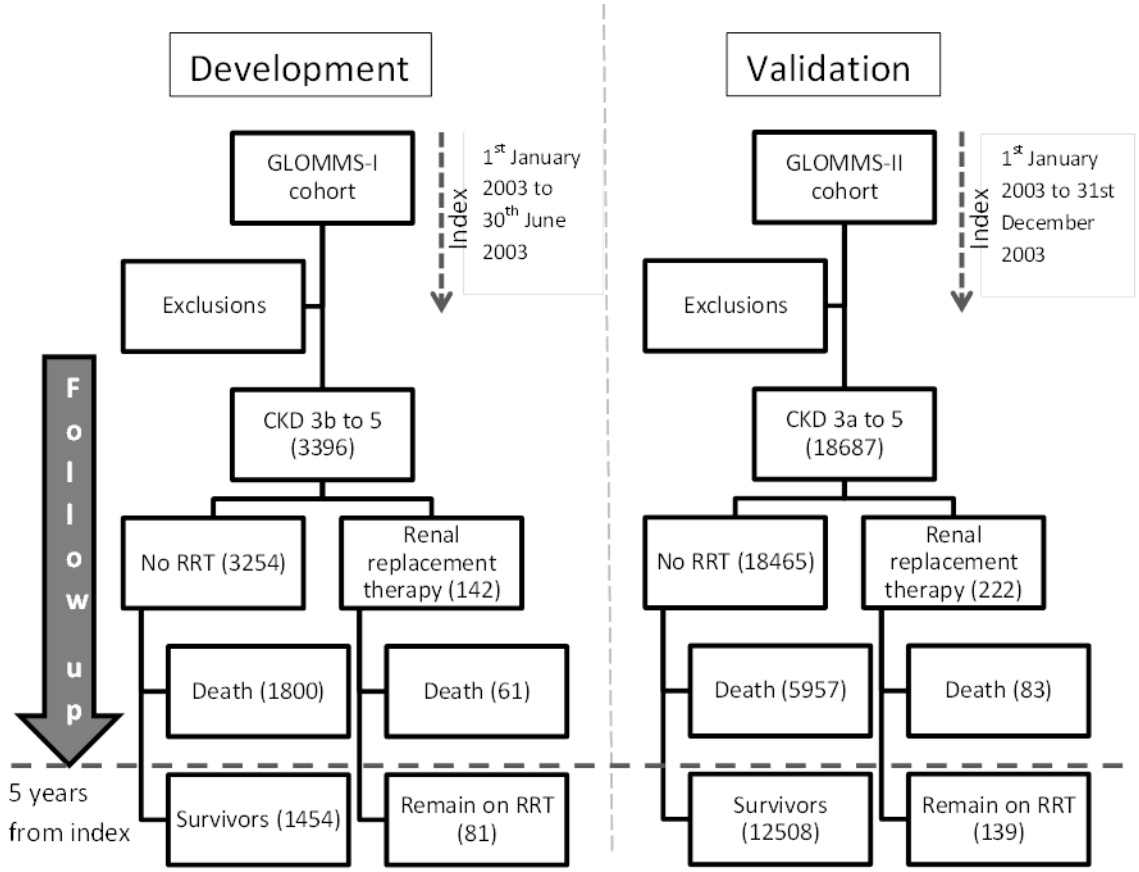


Figure 1 – Timeline and outcomes of development and validation cohorts

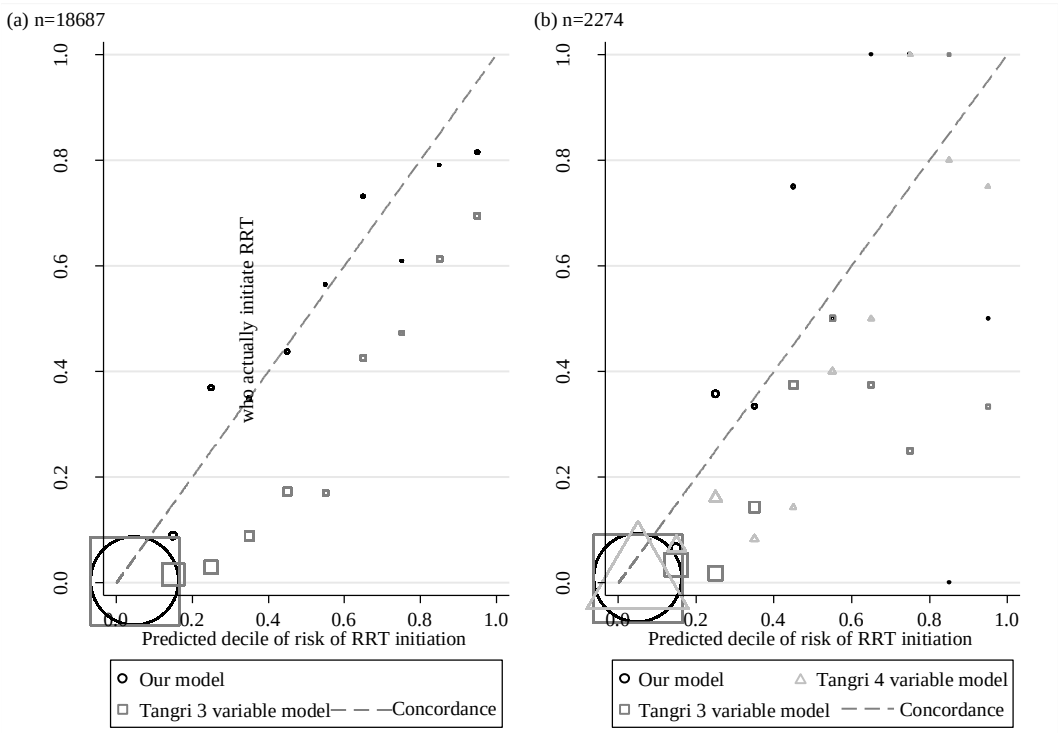
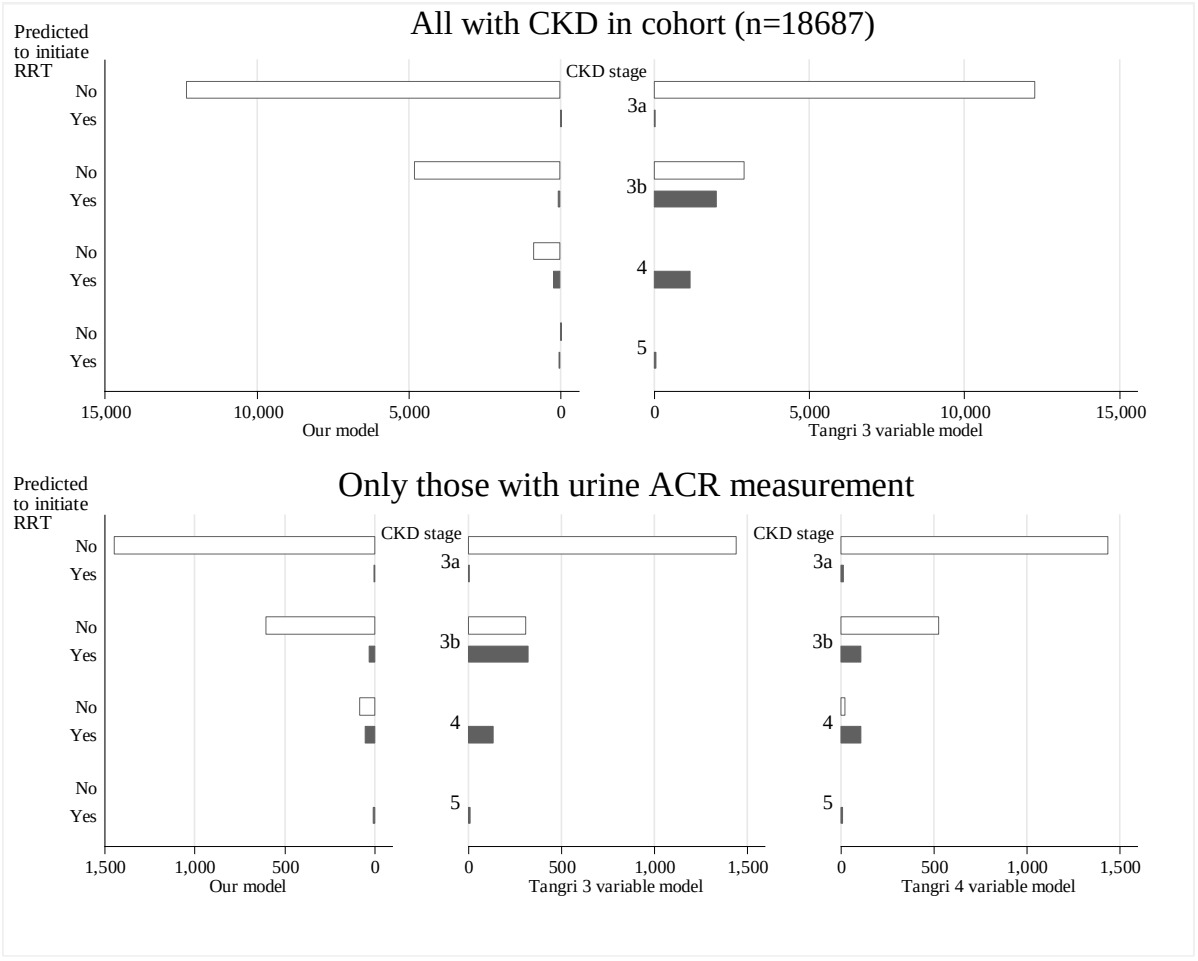


Figure 2 – Calibration plots for proportion of individuals in each predicted decile of risk who actually initiate RRT (a) with CKD in GLOMMS-II (b) with CKD in GLOMMS-II who had a measure of ACR available at baseline

Size of the shape (circle, square or triangle) represents the number of individuals out of the 18687 or 2274 individuals who had a predicted risk within the given decile. For example in (a) our model had 18313 with predicted risk between 0.0 and 0.1, 124 with predicted risk between 0.1 and 0.2, 57 between 0.2 and 0.3, 43 between 0.3 and 0.4, 32 between 0.4 and 0.5, 23 between 0.5 and 0.6, 26 between 0.6 and 0.7, 23 between 0.7 and 0.8, 19 between 0.8 and 0.9, 27 between 0.9 and 1.0Similarly in (a) the KFRE 3-variable model had 16814 with predicted risk between 0.0 and 0.1, 1001 with predicted risk between 0.1 and 0.2, 346 between 0.2 and 0.3, 170 between 0.3 and 0.4, 110 between 0.4 and 0.5, 53 between 0.5 and 0.6, 54 between 0.6 and 0.7, 36 between 0.7 and 0.8, 44 between 0.8 and 0.9, 59 between 0.9 and 1.0.

Figure 3 – Predicted need for RRT at 5 years follow up, among those who did not initiate RRT: a comparison of models in the GLOMMS II validation cohort.



For each model and CKD stage, those who did not initiate RRT are shown, they are represented in white if correctly predicted not to initiate and black if incorrectly predicted to initiate RRT.